CMS Books in Mathematics

Canadian
Mathematical Society
Société mathématique
du Canada

Heinz H. Bauschke
Patrick L. Combettes

Convex Analysis and Monotone Operator Theory in Hilbert Spaces

Second Edition

Springer

More information about this series at http://www.springer.com/series/4318

Heinz H. Bauschke • Patrick L. Combettes

Convex Analysis and Monotone Operator Theory in Hilbert Spaces

Second Edition

Springer

Heinz H. Bauschke
Department of Mathematics
University of British Columbia
Kelowna, BC, Canada

Patrick L. Combettes
Department of Mathematics
North Carolina State University
Raleigh, NC, USA

ISSN 1613-5237 ISSN 2197-4152 (electronic)
CMS Books in Mathematics
ISBN 978-3-319-48310-8 ISBN 978-3-319-48311-5 (eBook)
DOI 10.1007/978-3-319-48311-5

Library of Congress Control Number: 2016954949

Mathematics Subject Classification (2010): 47H05, 49M27, 65K05, 90C25, 47H10, 46C05

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Für Steffi, Andrea & Kati

À ma famille

Foreword

This self-contained book offers a modern unifying presentation of three basic areas of nonlinear analysis, namely, convex analysis, monotone operator theory, and the fixed point theory of nonexpansive mappings.

This turns out to be a judicious choice. Showing the rich connections and interplay between these topics gives a strong coherence to the book. Moreover, these particular topics are at the core of modern optimization and its applications.

Choosing to work in Hilbert spaces offers a wide range of applications while keeping the mathematics accessible to a large audience. Each topic is developed in a self-contained fashion, and the presentation often draws on recent advances.

The organization of the book makes it accessible to a large audience. Each chapter is illustrated by several exercises, which makes the monograph an excellent textbook. In addition, it offers deep insights into algorithmic aspects of optimization, especially splitting algorithms, which are important in theory and applications.

Let us point out the high quality of the writing and presentation. The authors combine an uncompromising demand for rigorous mathematical statements and a deep concern for applications, which makes this book remarkably accomplished.

Montpellier, France
October 2010 — Hédy Attouch

Preface to the Second Edition

This second edition contains over 140 pages of new material, over 270 new results, and more than 100 new exercises. It features a new chapter on proximity operators (Chapter 24) including two sections on proximity operators of matrix functions, as well as new sections on cocoercive operators (Section 4.2), quasi-Fejér monotone sequences (Section 5.4), nonlinear ergodic theorems (Section 5.5), recession functions (Section 9.4), the subdifferential of a composition (Section 16.5), directional derivatives and convexity (Section 17.5), the partial inverse (Sections 20.3, 26.2, and 28.2), local maximal monotonicity (Section 25.2), parallel compositions of monotone operators (Section 25.6), the Peaceman-Rachford algorithm (Sections 26.4 and 28.4), algorithms for solving composite monotone inclusion problems (Section 26.8) and composite minimization problems (Section 28.8), subgradient projections (Section 29.6), and Halpern's algorithm (Section 30.1). Furthermore, many existing results have been improved. Finally, the list of references has been updated.

We are again grateful to Isao Yamada for his careful reading and his constructive suggestions. We thank Radu Boţ, Luis Briceño-Arias, Minh Bùi, Minh Đào, Lilian Glaudin, Sarah Moffat, Walaa Moursi, Quang Văn Nguyễn, Audrey Repetti, Saverio Salzo, Bằng Công Vũ, Xianfu Wang, and Liangjin Yao for various pertinent and helpful comments.

We acknowledge support of our work by France's Centre National de la Recherche Scientifique, the Canada Research Chair Program, and the Natural Sciences and Engineering Research Council of Canada.

We grieve the passing of Jonathan Borwein (1951–2016) and Jean Jacques Moreau (1923–2014), who have deeply influenced and inspired us in our work.

Kelowna, BC, Canada — Heinz H. Bauschke
Paris, France and Raleigh, NC, USA — Patrick L. Combettes
August 2016

Preface to the First Edition

Three important areas of nonlinear analysis emerged in the early 1960s: convex analysis, monotone operator theory, and the theory of nonexpansive mappings. Over the past four decades, these areas have reached a high level of maturity, and an increasing number of connections have been identified between them. At the same time, they have found applications in a wide array of disciplines, including mechanics, economics, partial differential equations, information theory, approximation theory, signal and image processing, game theory, optimal transport theory, probability and statistics, and machine learning.

The purpose of this book is to present a largely self-contained account of the main results of convex analysis, monotone operator theory, and the theory of nonexpansive operators in the context of Hilbert spaces. Authoritative monographs are already available on each of these topics individually. A novelty of this book, and indeed, its central theme, is the tight interplay among the key notions of convexity, monotonicity, and nonexpansiveness. We aim at making the presentation accessible to a broad audience and to reach out in particular to the applied sciences and engineering communities, where these tools have become indispensable. We chose to cast our exposition in the Hilbert space setting. This allows us to cover many applications of interest to practitioners in infinite-dimensional spaces and yet to avoid the technical difficulties pertaining to general Banach space theory that would exclude a large portion of our intended audience. We have also made an attempt to draw on recent developments and modern tools to simplify the proofs of key results, exploiting heavily, for instance, the concept of a Fitzpatrick function in our exposition of monotone operators, the notion of Fejér monotonicity to unify the convergence proofs of several algorithms, and the notion of a proximity operator throughout the second half of the book.

The book is organized in 29 chapters. Chapters 1 and 2 provide background material. Chapters 3–7 cover set convexity and nonexpansive operators. Various aspects of the theory of convex functions are discussed in

Chapters 8–19. Chapters 20–25 are dedicated to monotone operator theory. In addition to these basic building blocks, we also address certain themes from different angles in several places. Thus, optimization theory is discussed in Chapters 11, 19, 26, and 27. Best approximation problems are discussed in Chapters 3, 19, 27, 28, and 29. Algorithms are also present in various parts of the book: fixed point and convex feasibility algorithms in Chapter 5, proximal-point algorithms in Chapter 23, monotone operator splitting algorithms in Chapter 25, optimization algorithms in Chapter 27, and best approximation algorithms in Chapters 27 and 29. More than 400 exercises are distributed throughout the book, at the end of each chapter.

Preliminary drafts of the first edition of this book have been used in courses in our institutions, and we have benefited from the input of postdoctoral fellows and many students. To all of them, many thanks. In particular, HHB thanks Liangjin Yao for his helpful comments. We are grateful to Hédy Attouch, Jon Borwein, Stephen Simons, Jon Vanderwerff, Shawn Wang, and Isao Yamada for helpful discussions and pertinent comments. PLC also thanks Oscar Wesler. Finally, we thank the Natural Sciences and Engineering Research Council of Canada, the Canada Research Chair Program, and France's Agence Nationale de la Recherche for their support.

Kelowna, BC, Canada — Heinz H. Bauschke
Paris, France — Patrick L. Combettes
October 2010

Contents

The original version of this book was revised. The correction to this book is available at https://doi.org/10.1007/978-3-319-48311-5_31

Chapter 1
Background

This chapter reviews basic definitions, facts, and notation from topology and set-valued analysis that will be used throughout the book.

1.1 Sets in Vector Spaces

Let $\mathcal{X}$ be a real vector space, let C and D be subsets of $\mathcal{X}$, and let $z \in \mathcal{X}$. Then $C + D = \{x + y \mid x \in C,\ y \in D\}$, $C - D = \{x - y \mid x \in C,\ y \in D\}$, $z + C = \{z\} + C$, $C - z = C - \{z\}$, and, for every $\lambda \in \mathbb{R}$, $\lambda C = \{\lambda x \mid x \in C\}$. If Λ is a nonempty subset of $\mathbb{R}$, then $\Lambda C = \bigcup_{\lambda \in \Lambda} \lambda C$ and $\Lambda z = \Lambda\{z\} = \{\lambda z \mid \lambda \in \Lambda\}$. In particular, C is a *cone* if

$$C = \mathbb{R}_{++} C, \tag{1.1}$$

where $\mathbb{R}_{++} = \{\lambda \in \mathbb{R} \mid \lambda > 0\}$. Now suppose that $u \in \mathcal{X} \smallsetminus \{0\}$. Then C is a *ray* if $C = \mathbb{R}_+ u$, where $\mathbb{R}_+ = \{\lambda \in \mathbb{R} \mid \lambda \geqslant 0\}$, and C is a *line* if $C = \mathbb{R}\, u$. Furthermore, C is an *affine subspace* if

$$C \neq \varnothing \quad \text{and} \quad (\forall \lambda \in \mathbb{R}) \quad C = \lambda C + (1 - \lambda) C. \tag{1.2}$$

Suppose that $C \neq \varnothing$. The intersection of all the linear subspaces of $\mathcal{X}$ containing C, i.e., the smallest linear subspace of $\mathcal{X}$ containing C, is called the *span* of C and is denoted by $\operatorname{span} C$; its closure is the smallest closed linear subspace of $\mathcal{X}$ containing C and it is denoted by $\overline{\operatorname{span}}\, C$. Likewise, the intersection of all the affine subspaces of $\mathcal{X}$ containing C, i.e., the smallest affine subspace of $\mathcal{X}$ containing C, is denoted by $\operatorname{aff} C$ and called the *affine hull* of C. If C is an affine subspace, then $V = C - C$ is the *linear subspace parallel to C* and $(\forall x \in C)\ C = x + V$.

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-48311-5_31

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_1

The four types of *line segments* between two points x and y in $\mathcal{X}$ are

$$[x,y] = \big\{(1-\alpha)x + \alpha y \;\big|\; 0 \leqslant \alpha \leqslant 1\big\}, \tag{1.3}$$

$]x,y[\, = \big\{(1-\alpha)x + \alpha y \;\big|\; 0 < \alpha < 1\big\}$, $[x,y[\, = \big\{(1-\alpha)x + \alpha y \;\big|\; 0 \leqslant \alpha < 1\big\}$, and $]x,y] = [y,x[$.

1.2 Operators

Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be nonempty sets, and let $2^{\mathcal{Y}}$ be the *power set* of $\mathcal{Y}$, i.e., the family of all subsets of $\mathcal{Y}$. The notation $T\colon \mathcal{X} \to \mathcal{Y}$ means that the operator (also called mapping) T maps every point x in $\mathcal{X}$ to a point Tx in $\mathcal{Y}$. Thus, the notation $A\colon \mathcal{X} \to 2^{\mathcal{Y}}$ means that A is a *set-valued operator* from $\mathcal{X}$ to $\mathcal{Y}$, i.e., A maps every point $x \in \mathcal{X}$ to a set $Ax \subset \mathcal{Y}$. Let $A\colon \mathcal{X} \to 2^{\mathcal{Y}}$. Then A is characterized by its *graph*

$$\operatorname{gra} A = \big\{(x,u) \in \mathcal{X} \times \mathcal{Y} \;\big|\; u \in Ax\big\}. \tag{1.4}$$

If C is a subset of $\mathcal{X}$, then $A(C) = \bigcup_{x \in C} Ax$. Given $B\colon \mathcal{Y} \to 2^{\mathcal{Z}}$, the *composition* $B \circ A$ is

$$B \circ A\colon \mathcal{X} \to 2^{\mathcal{Z}} \colon x \mapsto B(Ax) = \bigcup_{y \in Ax} By. \tag{1.5}$$

The *domain* and the *range* of A are

$$\operatorname{dom} A = \big\{x \in \mathcal{X} \;\big|\; Ax \neq \varnothing\big\} \quad \text{and} \quad \operatorname{ran} A = A(\mathcal{X}), \tag{1.6}$$

respectively. If $\mathcal{X}$ is a topological space, the closure of $\operatorname{dom} A$ is denoted by $\overline{\operatorname{dom}}\, A$; likewise, if $\mathcal{Y}$ is a topological space, the closure of $\operatorname{ran} A$ is denoted by $\overline{\operatorname{ran}}\, A$. The *inverse* of A, denoted by A^{-1}, is defined through its graph

$$\operatorname{gra} A^{-1} = \big\{(u,x) \in \mathcal{Y} \times \mathcal{X} \;\big|\; (x,u) \in \operatorname{gra} A\big\}. \tag{1.7}$$

Thus, for every $(x,u) \in \mathcal{X} \times \mathcal{Y}$, $u \in Ax \Leftrightarrow x \in A^{-1}u$. Moreover, $\operatorname{dom} A^{-1} = \operatorname{ran} A$ and $\operatorname{ran} A^{-1} = \operatorname{dom} A$. If $\mathcal{Y}$ is a vector space, the set of zeros of A is

$$\operatorname{zer} A = A^{-1}0 = \big\{x \in \mathcal{X} \;\big|\; 0 \in Ax\big\}. \tag{1.8}$$

Suppose that $\operatorname{dom} A \neq \varnothing$. If, for every $x \in \operatorname{dom} A$, Ax is a singleton, say $Ax = \{Tx\}$, then A is said to be *at most single-valued* from $\mathcal{X}$ to $\mathcal{Y}$ and it can be identified with the operator $T\colon \operatorname{dom} A \to \mathcal{Y}$. Conversely, if D is a nonempty subset of $\mathcal{X}$, an operator $T\colon D \to \mathcal{Y}$ can be identified with an at most single-valued operator from $\mathcal{X}$ to $\mathcal{Y}$, namely

$$A\colon \mathcal{X} \to 2^{\mathcal{Y}} \colon x \mapsto \begin{cases} \{Tx\}, & \text{if } x \in D; \\ \varnothing, & \text{otherwise.} \end{cases} \tag{1.9}$$

A *selection* of a set-valued operator $A\colon \mathcal{X} \to 2^{\mathcal{Y}}$ is an operator $T\colon \operatorname{dom} A \to \mathcal{Y}$ such that $(\forall x \in \operatorname{dom} A)\ Tx \in Ax$. Now let $T\colon \mathcal{X} \to \mathcal{Y}$, let $C \subset \mathcal{X}$, and let $D \subset \mathcal{Y}$. Then $T(C) = \big\{Tx \;\big|\; x \in C\big\}$ and $T^{-1}(D) = \big\{x \in \mathcal{X} \;\big|\; Tx \in D\big\}$.

Suppose that $\mathcal{Y}$ is a real vector space, let $A\colon \mathcal{X} \to 2^{\mathcal{Y}}$, let $B\colon \mathcal{X} \to 2^{\mathcal{Y}}$, and let $\lambda \in \mathbb{R}$. Then

$$A + \lambda B\colon \mathcal{X} \to 2^{\mathcal{Y}}\colon x \mapsto Ax + \lambda Bx. \tag{1.10}$$

Thus, $\operatorname{gra}\big(A + \lambda B\big) = \big\{(x, u + \lambda v) \;\big|\; (x,u) \in \operatorname{gra} A,\ (x,v) \in \operatorname{gra} B\big\}$ and $\operatorname{dom}\big(A + \lambda B\big) = \operatorname{dom} A \cap \operatorname{dom} B$. Now suppose that $\mathcal{X}$ is a real vector space and let $T\colon \mathcal{X} \to \mathcal{Y}$. Then T is *positively homogeneous* if

$$(\forall x \in \mathcal{X})(\forall \lambda \in \mathbb{R}_{++}) \quad T\big(\lambda x\big) = \lambda Tx, \tag{1.11}$$

and T is *affine* if

$$(\forall x \in \mathcal{X})(\forall y \in \mathcal{X})(\forall \lambda \in \mathbb{R}) \quad T\big(\lambda x + (1-\lambda)y\big) = \lambda Tx + (1-\lambda)Ty. \tag{1.12}$$

Note that T is affine if and only if $x \mapsto Tx - T0$ is linear.

Finally, suppose that $\mathcal{X}$ is a real vector space and let $A\colon \mathcal{X} \to 2^{\mathcal{Y}}$. Then the *translation* of A by $y \in \mathcal{X}$ is $\tau_y A\colon x \mapsto A(x-y)$ and the *reversal* of A is $A^{\vee}\colon x \mapsto A(-x)$.

1.3 Order

Let A be a nonempty set and let $\preccurlyeq$ be a binary relation on $A \times A$. Consider the following statements:

① $(\forall a \in A) \quad a \preccurlyeq a$.
② $(\forall a \in A)(\forall b \in A)(\forall c \in A) \quad [\,a \preccurlyeq b \text{ and } b \preccurlyeq c\,] \;\Rightarrow\; a \preccurlyeq c$.
③ $(\forall a \in A)(\forall b \in A)(\exists\, c \in A) \quad a \preccurlyeq c \text{ and } b \preccurlyeq c$.
④ $(\forall a \in A)(\forall b \in A) \quad [\,a \preccurlyeq b \text{ and } b \preccurlyeq a\,] \;\Rightarrow\; a = b$.
⑤ $(\forall a \in A)(\forall b \in A) \quad a \preccurlyeq b \text{ or } b \preccurlyeq a$.

We shall also write $b \succcurlyeq a$ if $a \preccurlyeq b$. If ①, ②, and ③ are satisfied, then $(A, \preccurlyeq)$ is a *directed set*. If ①, ②, and ④ hold, then $(A, \preccurlyeq)$ is a *partially ordered set*; in this case, we define a *strict order* $\prec$ on A by

$$(\forall a \in A)(\forall b \in A) \quad a \prec b \quad \Leftrightarrow \quad \big[a \preccurlyeq b \quad \text{and} \quad a \neq b\big]. \tag{1.13}$$

We shall also write $b \succ a$ if $a \prec b$. If $(A, \preccurlyeq)$ is a partially ordered set such that ⑤ holds, then $(A, \preccurlyeq)$ is a *totally ordered set*. Unless mentioned otherwise, nonempty subsets of $\mathbb{R}$ will be totally ordered and directed by $\leqslant$. A totally ordered subset of a partially ordered set is often called a *chain*. Let $(A, \preccurlyeq)$ be a partially ordered set and let B be a subset of A. Then $a \in A$ is an *upper* *bound* of B if $(\forall b \in B)\ b \preccurlyeq a$, and a *lower bound* of B if $(\forall b \in B)\ a \preccurlyeq b$. Furthermore, $b \in B$ is the *least element* of B if $(\forall c \in B)\ b \preccurlyeq c$. Finally, $a \in A$ is a *maximal element* of A if $(\forall c \in A)\ a \preccurlyeq c \Rightarrow c = a$.

Fact 1.1 (Zorn's lemma) *Let A be a partially ordered set such that every chain in A has an upper bound. Then A contains a maximal element.*

1.4 Nets

Let $(A, \preccurlyeq)$ be a directed set and let $\mathcal{X}$ be a nonempty set. A *net* (or *generalized sequence*) in $\mathcal{X}$ indexed by A is an operator from A to $\mathcal{X}$ and it is denoted by $(x_a)_{a\in A}$. Let $\mathbb{N} = \{0, 1, \ldots\}$. Since $(\mathbb{N}, \leqslant)$ is a directed set, every sequence is a net; $(a)_{a\in]0,1[}$ is an example of a net that is not a sequence.

Let $(x_a)_{a\in A}$ be a net in $\mathcal{X}$ and let $\mathcal{Y}$ be a subset of $\mathcal{X}$. Then $(x_a)_{a\in A}$ is *eventually* in $\mathcal{Y}$ if

$$(\exists\, c \in A)(\forall a \in A) \quad a \succcurlyeq c \quad \Rightarrow \quad x_a \in \mathcal{Y}, \tag{1.14}$$

and it is *frequently* in $\mathcal{Y}$ if

$$(\forall c \in A)(\exists\, a \in A) \quad a \succcurlyeq c \quad \text{and} \quad x_a \in \mathcal{Y}. \tag{1.15}$$

A net $(y_b)_{b\in B}$ is a *subnet* of $(x_a)_{a\in A}$ via $k\colon B \to A$ if

$$(\forall b \in B) \quad y_b = x_{k(b)} \tag{1.16}$$

and

$$(\forall a \in A)(\exists\, d \in B)(\forall b \in B) \quad b \succcurlyeq d \ \Rightarrow\ k(b) \succcurlyeq a. \tag{1.17}$$

The notations $(x_{k(b)})_{b\in B}$ and $(x_{k_b})_{b\in B}$ will also be used for subnets. Thus, $(y_b)_{b\in B}$ is a subsequence of $(x_a)_{a\in A}$ when $A = B = \mathbb{N}$ and $(y_b)_{b\in B}$ is a subnet of $(x_a)_{a\in A}$ via some strictly increasing function $k\colon \mathbb{N} \to \mathbb{N}$.

Remark 1.2 A subnet of a sequence $(x_n)_{n\in\mathbb{N}}$ need not be a subsequence. For instance, let B be a nonempty subset of $\mathbb{R}$ that is unbounded above and suppose that the function $k\colon B \to \mathbb{N}$ satisfies $k(b) \to +\infty$ as $b \to +\infty$. Then $(x_{k(b)})_{b\in B}$ is a subnet of $(x_n)_{n\in\mathbb{N}}$. However, if B is uncountable, then $(x_{k(b)})_{b\in B}$ is not a subsequence. Likewise, if $B = \mathbb{N}$ and k is not strictly increasing, e.g., $k\colon b \mapsto b(2 + (-1)^b)$, then $(x_{k(b)})_{b\in B}$ is not a subsequence.

1.5 The Extended Real Line

One obtains the *extended real line* $[-\infty, +\infty] = \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$ by adjoining the elements $-\infty$ and $+\infty$ to the real line $\mathbb{R}$ and extending the order via $(\forall \xi \in \mathbb{R})\ -\infty < \xi < +\infty$. Arithmetic rules are extended to elements of $[-\infty,+\infty]$ in the usual fashion, leaving expressions such as $+\infty+(-\infty)$, $0\cdot(+\infty)$, and $+\infty/+\infty$ undefined, unless mentioned otherwise. Given $\xi\in\mathbb{R}$, $]\xi,+\infty] =]\xi,+\infty[\cup\{+\infty\}$; the other extended intervals are defined similarly.

Remark 1.3 Throughout this book, we use the following terminology.

(i) For extended real numbers, *positive* means $\geqslant 0$, *strictly positive* means >0, *negative* means $\leqslant 0$, and *strictly negative* means <0. Moreover, $\mathbb{R}_+ = [0,+\infty[= \{\xi\in\mathbb{R} \mid \xi\geqslant 0\}$ and $\mathbb{R}_{++} =]0,+\infty[= \{\xi\in\mathbb{R} \mid \xi>0\}$. Likewise, if N is a strictly positive integer, the *positive orthant* is $\mathbb{R}^N_+ = [0,+\infty[^N$ and the *strictly positive orthant* is $\mathbb{R}^N_{++} =]0,+\infty[^N$. The sets $\mathbb{R}_-$ and $\mathbb{R}_{--}$, as well as the *negative orthants*, $\mathbb{R}^N_-$ and $\mathbb{R}^N_{--}$, are defined similarly.

(ii) Let D be a subset of $\mathbb{R}$. Then a function $f\colon D\to[-\infty,+\infty]$ is *increasing* if, for every ξ and η in D such that $\xi<\eta$, we have $f(\xi)\leqslant f(\eta)$ (*strictly increasing* if $f(\xi)<f(\eta)$). Applying these definitions to $-f$ yields the notions of a *decreasing* and of a *strictly decreasing* function, respectively.

Let S be a subset of $[-\infty,+\infty]$. A number $\gamma\in[-\infty,+\infty]$ is the (necessarily unique) *infimum* (or the greatest lower bound) of S if it is a lower bound of S and if, for every lower bound δ of S, we have $\delta\leqslant\gamma$. This number is denoted by $\inf S$, and by $\min S$ when $\inf S\in S$. The *supremum* (least upper bound) of S is $\sup S = -\inf\{-\alpha \mid \alpha\in S\}$. This number is denoted by $\max S$ when $\sup S\in S$. The set S always admits an infimum and a supremum. Note that $\inf\varnothing = +\infty$ and $\sup\varnothing=-\infty$.

The *limit inferior* of a net $(\xi_a)_{a\in A}$ in $[-\infty,+\infty]$ is

$$\varliminf \xi_a = \sup_{a\in A}\ \inf_{\substack{b\in A\\ a\preccurlyeq b}} \xi_b \tag{1.18}$$

and its *limit superior* is

$$\varlimsup \xi_a = \inf_{a\in A}\ \sup_{\substack{b\in A\\ a\preccurlyeq b}} \xi_b. \tag{1.19}$$

It is clear that $\varliminf \xi_a \leqslant \varlimsup \xi_a$.

1.6 Functions

Let $\mathcal{X}$ be a nonempty set.

Definition 1.4 Let $f\colon\mathcal{X}\to[-\infty,+\infty]$. The *domain* of f is

$$\operatorname{dom} f = \{x\in\mathcal{X} \mid f(x)<+\infty\}, \tag{1.20}$$

the *graph* of f is

$$\operatorname{gra} f = \big\{(x,\xi) \in \mathcal{X} \times \mathbb{R} \;\big|\; f(x) = \xi\big\}, \tag{1.21}$$

the *epigraph* of f is

$$\operatorname{epi} f = \big\{(x,\xi) \in \mathcal{X} \times \mathbb{R} \;\big|\; f(x) \leqslant \xi\big\}, \tag{1.22}$$

the *lower level set* of f at height $\xi \in \mathbb{R}$ is

$$\operatorname{lev}_{\leqslant \xi} f = \big\{x \in \mathcal{X} \mid f(x) \leqslant \xi\big\}, \tag{1.23}$$

and the *strict lower level set* of f at height $\xi \in \mathbb{R}$ is

$$\operatorname{lev}_{< \xi} f = \big\{x \in \mathcal{X} \mid f(x) < \xi\big\}. \tag{1.24}$$

The function f is *proper* if $-\infty \notin f(\mathcal{X})$ and $\operatorname{dom} f \neq \varnothing$. In addition, the closures of $\operatorname{dom} f$ and $\operatorname{epi} f$ are respectively denoted by $\overline{\operatorname{dom}}\, f$ and $\overline{\operatorname{epi}}\, f$.

Remark 1.5

(i) Strictly speaking, $\operatorname{dom} f$ in (1.20) does not correspond to the domain of $f\colon \mathcal{X} \to [-\infty,+\infty]$ viewed as an operator (which is $\mathcal{X}$ in this case in light of our conventions) and it is sometimes called the *effective domain*. However, it is customary in convex analysis to call it simply the domain of f and to denote it still by $\operatorname{dom} f$.

(ii) Let $f\colon \mathcal{X} \to [-\infty,+\infty]$ and $g\colon \mathcal{X} \to [-\infty,+\infty]$. Then the sum $f + g\colon \mathcal{X} \to [-\infty,+\infty]$ is defined pointwise using the convention $+\infty + (-\infty) = +\infty$, which yields $\operatorname{dom}(f+g) = \operatorname{dom} f \cap \operatorname{dom} g$.

Lemma 1.6 *Let $(f_i)_{i\in I}$ be a family of functions from $\mathcal{X}$ to $[-\infty,+\infty]$. Then the following hold:*

(i) $\operatorname{epi}\big(\sup_{i\in I} f_i\big) = \bigcap_{i\in I} \operatorname{epi} f_i$.

(ii) *If I is finite, then* $\operatorname{epi}\big(\min_{i\in I} f_i\big) = \bigcup_{i\in I} \operatorname{epi} f_i$.

Proof. Take $(x,\xi) \in \mathcal{X} \times \mathbb{R}$.

(i): $(x,\xi) \in \operatorname{epi}\big(\sup_{i\in I} f_i\big) \Leftrightarrow \sup_{i\in I} f_i(x) \leqslant \xi \Leftrightarrow (\forall i \in I)\ f_i(x) \leqslant \xi \Leftrightarrow (\forall i \in I)\ (x,\xi) \in \operatorname{epi} f_i \Leftrightarrow (x,\xi) \in \bigcap_{i\in I} \operatorname{epi} f_i$.

(ii): $(x,\xi) \in \operatorname{epi}\big(\min_{i\in I} f_i\big) \Leftrightarrow \min_{i\in I} f_i(x) \leqslant \xi \Leftrightarrow (\exists\, i \in I)\ f_i(x) \leqslant \xi \Leftrightarrow (\exists\, i \in I)\ (x,\xi) \in \operatorname{epi} f_i \Leftrightarrow (x,\xi) \in \bigcup_{i\in I} \operatorname{epi} f_i$. ☐

Definition 1.7 Let $f\colon \mathcal{X} \to [-\infty,+\infty]$ and let C be a subset of $\mathcal{X}$. The *infimum* of f over C is $\inf f(C)$; it is also denoted by $\inf_{x\in C} f(x)$. Moreover, f achieves its infimum over C if there exists $y \in C$ such that $f(y) = \inf f(C)$. In this case, we write $f(y) = \min f(C)$ or $f(y) = \min_{x\in C} f(x)$ and call $\min f(C)$ the minimum of f over C. Likewise, the *supremum* of f over C is $\sup f(C)$; it is also denoted by $\sup_{y\in C} f(y)$. Moreover, f achieves its supremum over

C if there exists $x \in C$ such that $f(x) = \sup f(C)$. In this case, we write $f(x) = \max f(C)$ or $f(x) = \max_{y \in C} f(y)$ and call $\max f(C)$ the maximum of f over C.

Definition 1.8 Let $f\colon \mathcal{X} \to [-\infty, +\infty]$ and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\operatorname{dom} f$. Then $(x_n)_{n\in\mathbb{N}}$ is a *minimizing sequence of* f if $f(x_n) \to \inf f(\mathcal{X})$.

Finally, suppose that $\mathcal{X}$ is a real vector space and let $f\colon \mathcal{X} \to [-\infty, +\infty]$. The *translation* of f by $y \in \mathcal{X}$ is $\tau_y f\colon x \mapsto f(x-y)$ and the *reversal* of f is $f^\vee\colon x \mapsto f(-x)$.

1.7 Topological Spaces

Let $\mathcal{X}$ be a set and let T be a family of subsets of $\mathcal{X}$ that contains $\mathcal{X}$, $\varnothing$, as well as all arbitrary unions and finite intersections of its elements. Then T is a *topology* and $(\mathcal{X}, \mathsf{T})$—or simply $\mathcal{X}$ if a topology is assumed—is a *topological space*. The elements of T are called *open* sets and their complements in $\mathcal{X}$ *closed* sets. A *neighborhood* of $x \in \mathcal{X}$ is a subset V of $\mathcal{X}$ such that $x \in U \subset V$ for some $U \in \mathsf{T}$. The family of all neighborhoods of x is denoted by $\mathcal{V}(x)$. A subfamily B of T is a *base* of T if, for every $x \in \mathcal{X}$ and every $V \in \mathcal{V}(x)$, there exists a set $B \in \mathsf{B}$ such that $x \in B \subset V$. If B is a base of T, then every set in T can be written as a union of elements in B. Now let C be a subset of $\mathcal{X}$. Then the *interior* of C is the largest open set that is contained in C; it is denoted by $\operatorname{int} C$. A point $x \in \mathcal{X}$ belongs to $\operatorname{int} C$ if and only if $(\exists\, V \in \mathcal{V}(x))\ V \subset C$. The *closure* of C is the smallest closed set that contains C; it is denoted by $\overline{C}$. If $\overline{C} = \mathcal{X}$, then C is *dense* in $\mathcal{X}$. A point $x \in \mathcal{X}$ belongs to $\overline{C}$ if and only if $(\forall\, V \in \mathcal{V}(x))\ V \cap C \neq \varnothing$. The *boundary* of C is $\operatorname{bdry} C = \overline{C} \smallsetminus (\operatorname{int} C)$. If $\mathcal{X}_1$ and $\mathcal{X}_2$ are topological spaces with respective bases B_1 and B_2, the Cartesian product $\mathcal{X}_1 \times \mathcal{X}_2$ will be considered as a topological space equipped with the *product topology*, i.e., the topology that has

$$\mathsf{B} = \big\{ B_1 \times B_2 \;\big|\; B_1 \in \mathsf{B}_1 \text{ and } B_2 \in \mathsf{B}_2 \big\} \tag{1.25}$$

as a base.

The topological space $\mathcal{X}$ is a *Hausdorff space* if, for any two distinct points x_1 and x_2 in $\mathcal{X}$, there exist $V_1 \in \mathcal{V}(x_1)$ and $V_2 \in \mathcal{V}(x_2)$ such that $V_1 \cap V_2 = \varnothing$. Let $\mathcal{X}$ be a Hausdorff space. A subset C of $\mathcal{X}$ is *compact* if, whenever C is contained in the union of a family of open sets, it is also contained in the union of a finite subfamily from that family. Now let $(x_a)_{a\in A}$ be a net in $\mathcal{X}$. Then $(x_a)_{a\in A}$ *converges* to a (necessarily unique) limit point $x \in \mathcal{X}$, in symbols, $x_a \to x$ or $\lim x_a = x$, if $(x_a)_{a\in A}$ lies eventually in every neighborhood of x, i.e.,

$$(\forall\, V \in \mathcal{V}(x))(\exists\, b \in A)(\forall a \in A) \quad a \succcurlyeq b \quad \Rightarrow \quad x_a \in V. \tag{1.26}$$

Fact 1.9 *Let $(x_a)_{a\in A}$ be a net in a Hausdorff space $\mathcal{X}$ that converges to a point $x \in \mathcal{X}$ and let $(x_{k(b)})_{b\in B}$ be a subnet of $(x_a)_{a\in A}$. Then $x_{k(b)} \to x$.*

A point $x \in \mathcal{X}$ is a *cluster point* (or *accumulation point*) of $(x_a)_{a\in A}$ if $(x_a)_{a\in A}$ lies frequently in every neighborhood of x, i.e.,

$$(\forall\, V \in \mathcal{V}(x))(\forall b \in A)(\exists\, a \in A) \quad a \succcurlyeq b \quad \text{and} \quad x_a \in V. \tag{1.27}$$

Alternatively, x is a cluster point of $(x_a)_{a\in A}$ if $(x_a)_{a\in A}$ possesses a subnet that converges to x. If a sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{X}$ possesses a subsequence that converges to a point $x \in \mathcal{X}$, then x is a *sequential cluster point* of $(x_n)_{n\in\mathbb{N}}$.

Topological notions can be conveniently characterized using nets.

Lemma 1.10 *Let C be a subset of a Hausdorff space $\mathcal{X}$ and let $x \in \mathcal{X}$. Then $x \in \overline{C}$ if and only if there exists a net in C that converges to x.*

Proof. First, suppose that $x \in \overline{C}$ and direct $A = \mathcal{V}(x)$ via $(\forall a \in A)(\forall b \in A)$ $a \preccurlyeq b \Leftrightarrow a \supset b$. For every $a \in A$, there exists $x_a \in C \cap a$. The net $(x_a)_{a\in A}$ lies in C and, by (1.26), it converges to x. Conversely, let $(x_a)_{a\in A}$ be a net in C such that $x_a \to x$ and let $V \in \mathcal{V}(x)$. Then $(x_a)_{a\in A}$ is eventually in V and therefore $C \cap V \neq \varnothing$. Thus, $x \in \overline{C}$. ☐

Let C be a subset of $\mathcal{X}$. It follows from Lemma 1.10 that C is closed if and only if the limit of every convergent net that lies in C belongs to C. Likewise, C is open if and only if, for every point $x \in C$, every net in $\mathcal{X}$ that converges to x is eventually in C.

Fact 1.11 *Let C be a subset of a Hausdorff space $\mathcal{X}$. Then the following are equivalent:*

(i) *C is compact.*
(ii) *$C \cap \bigcap_{j\in J} C_j \neq \varnothing$ for every family $(C_j)_{j\in J}$ of closed subsets of $\mathcal{X}$ such that, for every finite subset I of J, $C \cap \bigcap_{i\in I} C_i \neq \varnothing$.*
(iii) *Every net in C has a cluster point in C.*
(iv) *Every net in C has a subnet that converges to a point in C.*

Lemma 1.12 *Let C be a compact subset of a Hausdorff space $\mathcal{X}$. Then C is closed, and every closed subset of C is compact.*

Proof. Let $(x_a)_{a\in A}$ be a net in C that converges to a point $x \in \mathcal{X}$. By Fact 1.11, there exists a subnet $(x_{k(b)})_{b\in B}$ of $(x_a)_{a\in A}$ that converges to a point $y \in C$. Therefore $x = y \in C$ and C is closed. Now let D be a closed subset of C and let $(x_a)_{a\in A}$ be a net in D. Then $(x_a)_{a\in A}$ lies in C and, as above, there exists a subnet $(x_{k(b)})_{b\in B}$ of $(x_a)_{a\in A}$ that converges to a point $y \in C$. Since D is closed, $y \in D$. We conclude that D is compact. ☐

Remark 1.13 Let C be a compact subset of a Hausdorff space $\mathcal{X}$ and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in C. By Fact 1.11, $(x_n)_{n\in\mathbb{N}}$ possesses a convergent subnet. However, it may happen that no subsequence of $(x_n)_{n\in\mathbb{N}}$ converges (see [148, Chapter 13] for an example).

Lemma 1.14 *Let C be a compact subset of a Hausdorff space $\mathcal{X}$ and suppose that $(x_a)_{a\in A}$ is a net in C that has a unique cluster point $x \in \mathcal{X}$. Then $x_a \to x$.*

Proof. Suppose that $x_a \not\to x$. Then it follows from (1.26) that there exist a subnet $(x_{k(b)})_{b\in B}$ of $(x_a)_{a\in A}$ and an open set $V \in \mathcal{V}(x)$ such that $(x_{k(b)})_{b\in B}$ lies in $C \smallsetminus V$. Since $C \smallsetminus V$ is compact by Lemma 1.12, $(x_{k(b)})_{b\in B}$ possesses a cluster point $y \in C \smallsetminus V$. Hence, $y \neq x \in V$ and y is a cluster point of $(x_a)_{a\in A}$, which contradicts the assumption that x is the unique cluster point of $(x_a)_{a\in A}$. $\square$

1.8 Two-Point Compactification of the Real Line

Unless stated otherwise, the real line $\mathbb{R}$ will always be equipped with the usual topology, a base of which is the family of open intervals. With this topology, $\mathbb{R}$ is a Hausdorff space that is not compact. The extended real line $[-\infty,+\infty]$ equipped with the topology that has as a base the open real intervals and the intervals of the form $[-\infty,\xi[$ and $]\xi,+\infty]$, where $\xi \in \mathbb{R}$ is a compact space.

Fact 1.15 *Let $(\xi_a)_{a\in A}$ be a net in $[-\infty,+\infty]$. Then the following hold:*

(i) *The nets*

$$\Bigg(\inf_{\substack{b\in A\\ a\preccurlyeq b}} \xi_b\Bigg)_{a\in A} \quad \text{and} \quad \Bigg(\sup_{\substack{b\in A\\ a\preccurlyeq b}} \xi_b\Bigg)_{a\in A} \tag{1.28}$$

converge to $\underline{\lim}\,\xi_a$ and $\overline{\lim}\,\xi_a$, respectively.

(ii) *The net $(\xi_a)_{a\in A}$ possesses subnets that converge to $\underline{\lim}\,\xi_a$ and $\overline{\lim}\,\xi_a$, respectively.*

(iii) *The net $(\xi_a)_{a\in A}$ converges if and only if $\underline{\lim}\,\xi_a = \overline{\lim}\,\xi_a$, in which case $\lim \xi_a = \underline{\lim}\,\xi_a = \overline{\lim}\,\xi_a$.*

Moreover, if $(\xi_a)_{a\in A}$ is a sequence, then (ii) remains true if subnets are replaced by subsequences.

Lemma 1.16 *Let $(\xi_a)_{a\in A}$ and $(\eta_a)_{a\in A}$ be nets in $[-\infty,+\infty]$ such that $\underline{\lim}\,\xi_a > -\infty$ and $\underline{\lim}\,\eta_a > -\infty$. Then $\underline{\lim}\,\xi_a + \underline{\lim}\,\eta_a \leqslant \underline{\lim}(\xi_a + \eta_a)$.*

Proof. Let $a \in A$. Clearly,

$$(\forall c \in A) \quad a \preccurlyeq c \quad \Rightarrow \quad \inf_{\substack{b\in A\\ a\preccurlyeq b}} \xi_b + \inf_{\substack{b\in A\\ a\preccurlyeq b}} \eta_b \leqslant \xi_c + \eta_c. \tag{1.29}$$

Hence

$$\inf_{\substack{b\in A\\ a\preccurlyeq b}} \xi_b + \inf_{\substack{b\in A\\ a\preccurlyeq b}} \eta_b \leqslant \inf_{\substack{c\in A\\ a\preccurlyeq c}} (\xi_c + \eta_c). \tag{1.30}$$

In view of Fact 1.15(i), the result follows by taking limits over a in (1.30). $\square$

1.9 Continuity

Definition 1.17 Let $(\mathcal{X}, \mathsf{T}_{\mathcal{X}})$ and $(\mathcal{Y}, \mathsf{T}_{\mathcal{Y}})$ be topological spaces and let $T\colon \mathcal{X} \to \mathcal{Y}$. Then T is *continuous* at $x \in \mathcal{X}$ if

$$\big(\forall\, W \in \mathcal{V}(Tx)\big)\big(\exists\, V \in \mathcal{V}(x)\big) \quad T(V) \subset W. \tag{1.31}$$

Moreover, T is continuous if it is continuous at every point in $\mathcal{X}$.

Fact 1.18 *Let $(\mathcal{X}, \mathsf{T}_{\mathcal{X}})$ and $(\mathcal{Y}, \mathsf{T}_{\mathcal{Y}})$ be topological spaces, let $T\colon \mathcal{X} \to \mathcal{Y}$, and suppose that $\mathsf{B}_{\mathcal{Y}}$ is a base of $\mathsf{T}_{\mathcal{Y}}$. Then T is continuous if and only if $(\forall B \in \mathsf{B}_{\mathcal{Y}})\ T^{-1}(B) \in \mathsf{T}_{\mathcal{X}}$.*

Fact 1.19 *Let $\mathcal{X}$ and $\mathcal{Y}$ be Hausdorff spaces, let $T\colon \mathcal{X} \to \mathcal{Y}$, and let $x \in \mathcal{X}$. Then T is continuous at x if and only if, for every net $(x_a)_{a\in A}$ in $\mathcal{X}$ that converges to x, $Tx_a \to Tx$.*

Lemma 1.20 *Let $\mathcal{X}$ and $\mathcal{Y}$ be Hausdorff spaces, let $T\colon \mathcal{X} \to \mathcal{Y}$ be continuous, and let $C \subset \mathcal{X}$ be compact. Then $T(C)$ is compact.*

Proof. Let $(y_a)_{a\in A}$ be a net in $T(C)$. Then there exists a net $(x_a)_{a\in A}$ in C such that $(\forall a \in A)\ y_a = Tx_a$. By Fact 1.11, we can find a subnet $(x_{k(b)})_{b\in B}$ of $(x_a)_{a\in A}$ that converges to a point $x \in C$. Fact 1.19 implies that $Tx_{k(b)} \to Tx \in T(C)$. Therefore, the claim follows from Fact 1.11. ☐

1.10 Lower Semicontinuity

Definition 1.21 Let $\mathcal{X}$ be a Hausdorff space, let $f\colon \mathcal{X} \to [-\infty, +\infty]$, and let $x \in \mathcal{X}$. Then f is *lower semicontinuous* at x if, for every net $(x_a)_{a\in A}$ in $\mathcal{X}$,

$$x_a \to x \quad \Rightarrow \quad f(x) \leqslant \underline{\lim}\, f(x_a) \tag{1.32}$$

or, equivalently, if (see Figure 1.1)

$$(\forall\, \xi \in \left]-\infty, f(x)\right[)(\exists\, V \in \mathcal{V}(x)) \quad f(V) \subset \left]\xi, +\infty\right]. \tag{1.33}$$

Moreover, f is lower semicontinuous if it is lower semicontinuous at every point in $\mathcal{X}$. *Upper semicontinuity* of f at $x \in \mathcal{X}$ holds if $-f$ is lower semicontinuous at x, i.e., if, for every net $(x_a)_{a\in A}$ in $\mathcal{X}$, $x_a \to x \Rightarrow \overline{\lim}\, f(x_a) \leqslant f(x)$; if f is lower and upper semicontinuous at x, it is *continuous* at x, i.e., for every net $(x_a)_{a\in A}$ in $\mathcal{X}$, $x_a \to x \Rightarrow f(x_a) \to f(x)$. Finally, the *domain of continuity* of f is

$$\operatorname{cont} f = \big\{x \in \mathcal{X} \;\big|\; f(x) \in \mathbb{R} \text{ and } f \text{ is continuous at } x\big\}. \tag{1.34}$$

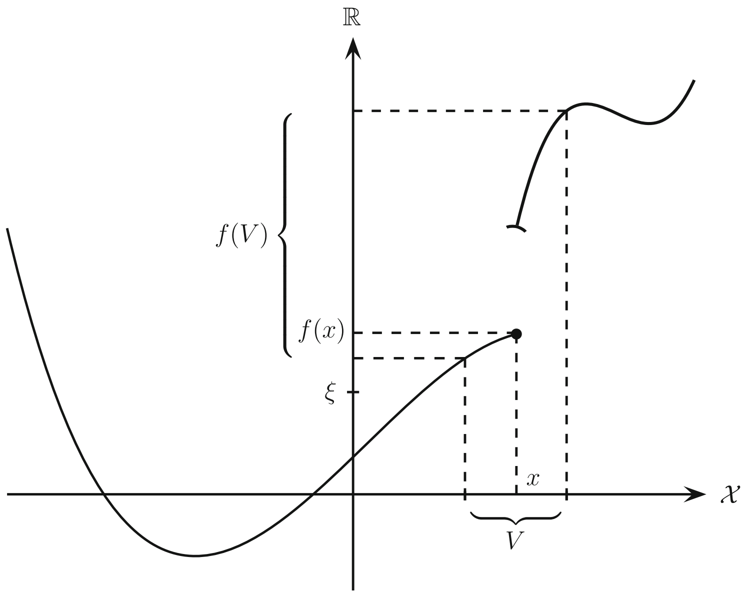

Fig. 1.1 The function f is lower semicontinuous at x: for every $\xi \in \left]-\infty, f(x)\right[$, we can find a neighborhood V of x such that $f(V) \subset \left]\xi, +\infty\right]$.

Note that $\operatorname{cont} f \subset \operatorname{int} \operatorname{dom} f$, since f cannot be both continuous and real-valued on the boundary of its domain. Let us also observe that a continuous function may take on the values $-\infty$ and $+\infty$.

Example 1.22 Set

$$f \colon \mathbb{R} \to [-\infty, +\infty] : x \mapsto \begin{cases} 1/x, & \text{if } x > 0; \\ +\infty, & \text{otherwise.} \end{cases} \tag{1.35}$$

Then f is continuous on $\mathbb{R}$ and $\operatorname{cont} f = \mathbb{R}_{++}$ is a proper open subset of $\mathbb{R}$.

Let $\mathcal{X}$ be a Hausdorff space, let $f \colon \mathcal{X} \to [-\infty, +\infty]$, and let $x \in \mathcal{X}$. Then

$$\varliminf_{y \to x} f(y) = \sup_{V \in \mathcal{V}(x)} \inf f(V). \tag{1.36}$$

Lemma 1.23 *Let $\mathcal{X}$ be a Hausdorff space, let $f \colon \mathcal{X} \to [-\infty, +\infty]$, let $x \in \mathcal{X}$, and let $\mathrm{N}(x)$ be the set of all nets in $\mathcal{X}$ that converge to x. Then*

$$\varliminf_{y \to x} f(y) = \min_{(x_a)_{a \in A} \in \mathrm{N}(x)} \varliminf f(x_a). \tag{1.37}$$

Proof. Let $(x_a)_{a \in A} \in \mathrm{N}(x)$ and set $(\forall a \in A)$ $\mu_a = \inf \left\{ f(x_b) \;\middle|\; a \preccurlyeq b \right\}$. By Fact 1.15(i), $\lim \mu_a = \varliminf f(x_a)$. For every $V \in \mathcal{V}(x)$, there exists $a_V \in A$

such that $(\forall a \in A)\ a \succcurlyeq a_V \Rightarrow x_a \in V$. Thus

$$(\forall a \in A) \quad a \succcurlyeq a_V \;\Rightarrow\; \mu_a \geqslant \inf f(V). \tag{1.38}$$

Hence, for every $V \in \mathcal{V}(x)$, $\underline{\lim}\, f(x_a) = \lim \mu_a = \lim_{a \succcurlyeq a_V} \mu_a \geqslant \inf f(V)$, which implies that

$$\inf \big\{ \underline{\lim}\, f(x_a) \;\big|\; x_a \to x \big\} \geqslant \sup_{V \in \mathcal{V}(x)} \inf f(V). \tag{1.39}$$

Set $B = \big\{ (y, V) \;\big|\; y \in V \in \mathcal{V}(x) \big\}$ and direct B via $(y, V) \preccurlyeq (z, W) \Leftrightarrow W \subset V$. For every $b = (y, V) \in B$, set $x_b = y$. Then $x_b \to x$ and

$$\begin{aligned} \underline{\lim}\, f(x_b) &= \sup_{b \in B} \inf \big\{ f(x_c) \;\big|\; b \preccurlyeq c \big\} \\ &= \sup_{y \in V \in \mathcal{V}(x)} \inf \big\{ f(z) \;\big|\; z \in V \big\} \\ &= \sup_{V \in \mathcal{V}(x)} \inf f(V). \end{aligned} \tag{1.40}$$

Altogether, (1.36), (1.39), and (1.40) yield (1.37). ☐

Lemma 1.24 *Let $\mathcal{X}$ be a Hausdorff space and let $f \colon \mathcal{X} \to [-\infty, +\infty]$. Then the following are equivalent:*

(i) *f is lower semicontinuous.*
(ii) $\operatorname{epi} f$ *is closed in $\mathcal{X} \times \mathbb{R}$.*
(iii) *For every $\xi \in \mathbb{R}$, $\operatorname{lev}_{\leqslant \xi} f$ is closed in $\mathcal{X}$.*

Proof. (i)⇒(ii): Let $(x_a, \xi_a)_{a \in A}$ be a net in $\operatorname{epi} f$ that converges to (x, ξ) in $\mathcal{X} \times \mathbb{R}$. Then $f(x) \leqslant \underline{\lim}\, f(x_a) \leqslant \underline{\lim}\, \xi_a = \xi$ and, hence, $(x, \xi) \in \operatorname{epi} f$.

(ii)⇒(iii): Fix $\xi \in \mathbb{R}$ and assume that $(x_a)_{a \in A}$ is a net in $\operatorname{lev}_{\leqslant \xi} f$ that converges to x. Then the net $(x_a, \xi)_{a \in A}$ lies in $\operatorname{epi} f$ and converges to (x, ξ). Since $\operatorname{epi} f$ is closed, we deduce that $(x, \xi) \in \operatorname{epi} f$ and, hence, that $x \in \operatorname{lev}_{\leqslant \xi} f$.

(iii)⇒(i): Fix $x \in \mathcal{X}$, let $(x_a)_{a \in A}$ be a net in $\mathcal{X}$ that converges to x, and set $\mu = \underline{\lim}\, f(x_a)$. Then it suffices to show that $f(x) \leqslant \mu$. When $\mu = +\infty$, the inequality is clear and we therefore assume that $\mu < +\infty$. By Fact 1.15(ii), there exists a subnet $(x_{k(b)})_{b \in B}$ of $(x_a)_{a \in A}$ such that $f(x_{k(b)}) \to \mu$. Now fix $\xi \in \,]\mu, +\infty[$. Then $(f(x_{k(b)}))_{b \in B}$ is eventually in $[-\infty, \xi]$ and, therefore, there exists $c \in B$ such that $\big\{ x_{k(b)} \;\big|\; c \preccurlyeq b \in B \big\} \subset \operatorname{lev}_{\leqslant \xi} f$. Since $x_{k(b)} \to x$ and since $\operatorname{lev}_{\leqslant \xi} f$ is closed, we deduce that $x \in \operatorname{lev}_{\leqslant \xi} f$, i.e., $f(x) \leqslant \xi$. Letting $\xi \downarrow \mu$, we conclude that $f(x) \leqslant \mu$. ☐

Example 1.25 Let $\mathcal{X}$ be a Hausdorff space. The *indicator function* of a subset C of $\mathcal{X}$, i.e., the function

$$\iota_C \colon \mathcal{X} \to [-\infty, +\infty] : x \mapsto \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{otherwise}, \end{cases} \tag{1.41}$$

is lower semicontinuous if and only if C is closed.

Proof. Take $\xi \in \mathbb{R}$. Then $\operatorname{lev}_{\leqslant \xi} \iota_C = \varnothing$ if $\xi < 0$, and $\operatorname{lev}_{\leqslant \xi} \iota_C = C$ otherwise. Hence, the result follows from Lemma 1.24. ☐

Lemma 1.26 *Let $\mathcal{X}$ be a Hausdorff space and let $(f_i)_{i \in I}$ be a family of lower semicontinuous functions from $\mathcal{X}$ to $[-\infty, +\infty]$. Then $\sup_{i \in I} f_i$ is lower semicontinuous. If I is finite, then $\min_{i \in I} f_i$ is lower semicontinuous.*

Proof. A direct consequence of Lemma 1.6 and Lemma 1.24. ☐

Lemma 1.27 *Let $\mathcal{X}$ be a Hausdorff space, let $(f_i)_{i \in I}$ be a finite family of lower semicontinuous functions from $\mathcal{X}$ to $]-\infty, +\infty]$, and let $(\alpha_i)_{i \in I}$ be in $\mathbb{R}_{++}$. Then $\sum_{i \in I} \alpha_i f_i$ is lower semicontinuous.*

Proof. It is clear that, for every $i \in I$, $\alpha_i f_i$ is lower semicontinuous. Let f and g be lower semicontinuous functions from $\mathcal{X}$ to $]-\infty, +\infty]$, and let $(x_a)_{a \in A}$ be a net in $\mathcal{X}$ that converges to some point $x \in \mathcal{X}$. By Lemma 1.16,

$$
\begin{aligned}
(f+g)(x) &= f(x) + g(x) \\
&\leqslant \underline{\lim}\, f(x_a) + \underline{\lim}\, g(x_a) \\
&\leqslant \underline{\lim}\, \big(f(x_a) + g(x_a)\big) \\
&= \underline{\lim}(f+g)(x_a), \qquad (1.42)
\end{aligned}
$$

which yields the result when I contains two elements. The general case follows by induction on the number of elements in I. ☐

Lemma 1.28 *Let $\mathcal{X}$ and $\mathcal{Y}$ be Hausdorff spaces, let $T\colon \mathcal{X} \to \mathcal{Y}$ be continuous, let $g\colon \mathcal{Y} \to [-\infty, +\infty]$ be lower semicontinuous, and set $f = g \circ T$. Then f is lower semicontinuous.*

Proof. Let $x \in \mathcal{X}$ and let $(x_a)_{a \in A}$ be a net in $\mathcal{X}$ that converges to x. Then Fact 1.19 asserts that $Tx_a \to Tx$. In turn, (1.32) yields $f(x) = g(Tx) \leqslant \underline{\lim}\, g(Tx_a) = \underline{\lim}\, f(x_a)$. ☐

The classical Weierstrass theorem states that a continuous function defined on a compact set achieves its infimum and its supremum on that set. The following refinement is a fundamental tool in proving the existence of solutions to minimization problems.

Theorem 1.29 (Weierstrass) *Let $\mathcal{X}$ be a Hausdorff space, let $f\colon \mathcal{X} \to [-\infty, +\infty]$ be lower semicontinuous, and let C be a compact subset of $\mathcal{X}$. Suppose that $C \cap \operatorname{dom} f \neq \varnothing$. Then f achieves its infimum over C.*

Proof. By definition of $\inf f(C)$, there exists a minimizing sequence $(x_n)_{n \in \mathbb{N}}$ of $f + \iota_C$. By Fact 1.11 and the compactness of C, we can extract a subnet $(x_{k(b)})_{b \in B}$ that converges to a point $x \in C$. Therefore $f(x_{k(b)}) \to \inf f(C) \leqslant f(x)$. On the other hand, by lower semicontinuity, we get $f(x) \leqslant \underline{\lim}\, f(x_{k(b)}) = \lim f(x_{k(b)}) = \inf f(C)$. Thus, $f(x) = \inf f(C)$. ☐

Lemma 1.30 *Let $\mathcal{X}$ be a Hausdorff space, let $\mathcal{Y}$ be a compact Hausdorff space, and let $\varphi\colon \mathcal{X} \times \mathcal{Y} \to [-\infty, +\infty]$ be lower semicontinuous. Then the* marginal function

$$f\colon \mathcal{X} \to [-\infty, +\infty] : x \mapsto \inf \varphi(x, \mathcal{Y}) \tag{1.43}$$

is lower semicontinuous and $(\forall x \in \mathcal{X})\ f(x) = \min \varphi(x, \mathcal{Y})$.

Proof. First let us note that, for every $x \in \mathcal{X}$, $\{x\} \times \mathcal{Y}$ is compact and hence Theorem 1.29 implies that $f(x) = \min \varphi(x, \mathcal{Y})$. Now fix $\xi \in \mathbb{R}$ and suppose that $(x_a)_{a \in A}$ is a net in $\operatorname{lev}_{\leqslant \xi} f$ that converges to some point $x \in \mathcal{X}$. Then there exists a net $(y_a)_{a \in A}$ in $\mathcal{Y}$ such that $(\forall a \in A)\ f(x_a) = \varphi(x_a, y_a)$. Since $\mathcal{Y}$ is compact, Fact 1.11 yields the existence of a subnet $(y_{k(b)})_{b \in B}$ that converges to a point $y \in \mathcal{Y}$. It follows that $(x_{k(b)}, y_{k(b)}) \to (x, y)$ and, by lower semicontinuity of φ, that $f(x) \leqslant \varphi(x, y) \leqslant \underline{\lim}\, \varphi(x_{k(b)}, y_{k(b)}) = \underline{\lim}\, f(x_{k(b)}) \leqslant \xi$. Thus, $x \in \operatorname{lev}_{\leqslant \xi} f$ and f is therefore lower semicontinuous by Lemma 1.24. ☐

Definition 1.31 Let $\mathcal{X}$ be a Hausdorff space. The *lower semicontinuous envelope* of $f\colon \mathcal{X} \to [-\infty, +\infty]$ is

$$\bar{f} = \sup \big\{ g\colon \mathcal{X} \to [-\infty, +\infty] \;\big|\; g \leqslant f \text{ and } g \text{ is lower semicontinuous} \big\}. \tag{1.44}$$

Lemma 1.32 *Let $\mathcal{X}$ be a Hausdorff space and let $f\colon \mathcal{X} \to [-\infty, +\infty]$. Then the following hold:*

(i) *$\bar{f}$ is the largest lower semicontinuous function majorized by f.*
(ii) $\operatorname{epi} \bar{f}$ *is closed.*
(iii) $\operatorname{dom} f \subset \operatorname{dom} \bar{f} \subset \overline{\operatorname{dom}}\, f$.
(iv) $(\forall x \in \mathcal{X})\ \bar{f}(x) = \underline{\lim}_{y \to x} f(y)$.
(v) *Let $x \in \mathcal{X}$. Then f is lower semicontinuous at x if and only if* $\bar{f}(x) = f(x)$.
(vi) $\operatorname{epi} \bar{f} = \overline{\operatorname{epi}}\, f$.

Proof. (i): This follows from (1.44) and Lemma 1.26.

(ii): Since $\bar{f}$ is lower semicontinuous by (i), the closedness of $\operatorname{epi} \bar{f}$ follows from Lemma 1.24.

(iii): Since $\bar{f} \leqslant f$, we have $\operatorname{dom} f \subset \operatorname{dom} \bar{f}$. Now set

$$g\colon \mathcal{X} \to [-\infty, +\infty] : x \mapsto \begin{cases} \bar{f}(x), & \text{if } x \in \overline{\operatorname{dom}}\, f; \\ +\infty, & \text{if } x \notin \overline{\operatorname{dom}}\, f. \end{cases} \tag{1.45}$$

It follows from (ii) that $\operatorname{epi} g = \operatorname{epi} \bar{f} \cap (\overline{\operatorname{dom}}\, f \times \mathbb{R})$ is closed and hence from Lemma 1.24 that g is lower semicontinuous. On the other hand, for every $x \in \mathcal{X}$, $g(x) = \bar{f}(x) \leqslant f(x)$ if $x \in \overline{\operatorname{dom}}\, f$, and $g(x) = f(x) = +\infty$ if $x \notin \overline{\operatorname{dom}}\, f$. Hence, $g \leqslant f$ and thus $g = \bar{g} \leqslant \bar{f}$. We conclude that $\operatorname{dom} \bar{f} \subset \operatorname{dom} g \subset \overline{\operatorname{dom}}\, f$.

(iv): Set $\widetilde{f}\colon x \mapsto \underline{\lim}_{y\to x} f(y)$ and let $x \in \mathcal{X}$. We first show that $\widetilde{f}$ is lower semicontinuous. To this end, suppose that $\widetilde{f}(x) > -\infty$ and fix $\xi \in]-\infty, \widetilde{f}(x)[$. In view of (1.36), there exists $V \in \mathcal{V}(x)$ such that $\xi < \inf f(V)$. Now let U be an open set such that $x \in U \subset V$. Then $\xi < \inf f(U)$ and $(\forall y \in U)$ $U \in \mathcal{V}(y)$. Hence, $(\forall y \in U)$ $\xi < \sup_{W\in\mathcal{V}(y)} \inf f(W) = \widetilde{f}(y)$ and therefore $\widetilde{f}(U) \subset]\xi, +\infty]$. It follows from (1.33) that $\widetilde{f}$ is lower semicontinuous at x. Thus, since (1.36) yields $\widetilde{f} \leqslant f$, we derive from (i) that $\widetilde{f} \leqslant \bar{f}$. Next, let $g\colon \mathcal{X} \to [-\infty, +\infty]$ be a lower semicontinuous function such that $g \leqslant f$. Then, in view of (1.44), it remains to show that $g \leqslant \widetilde{f}$ to prove that $\bar{f} \leqslant \widetilde{f}$, and hence conclude that $\bar{f} = \widetilde{f}$. To this end, suppose that $g(x) > -\infty$ and let $\eta \in]-\infty, g(x)[$. By (1.33), there exists $V \in \mathcal{V}(x)$ such that $g(V) \subset]\eta, +\infty]$. Hence, (1.36) yields $\eta \leqslant \inf g(V) \leqslant \underline{\lim}_{y\to x} g(y)$. Letting $\eta \uparrow g(x)$, we obtain $g(x) \leqslant \underline{\lim}_{y\to x} g(y)$. Thus,

$$g(x) \leqslant \underline{\lim}_{y\to x} g(y) = \sup_{W\in\mathcal{V}(x)} \inf g(W) \leqslant \sup_{W\in\mathcal{V}(x)} \inf f(W) = \widetilde{f}(x). \tag{1.46}$$

(v): Suppose that f is lower semicontinuous at x. Then it follows from (iv) that $\bar{f}(x) \leqslant f(x) \leqslant \underline{\lim}_{y\to x} f(y) = \bar{f}(x)$. Therefore, $\bar{f}(x) = f(x)$. Conversely, suppose that $\bar{f}(x) = f(x)$. Then $f(x) = \underline{\lim}_{y\to x} f(y)$ by (iv) and therefore f is lower semicontinuous at x.

(vi): Since $\bar{f} \leqslant f$, $\operatorname{epi} f \subset \operatorname{epi} \bar{f}$, and therefore (ii) yields $\overline{\operatorname{epi} f} \subset \overline{\operatorname{epi} \bar{f}} = \operatorname{epi} \bar{f}$. Conversely, let $(x, \xi) \in \operatorname{epi} \bar{f}$ and fix a neighborhood of (x, ξ) of the form $W = V \times [\xi - \varepsilon, \xi + \varepsilon]$, where $V \in \mathcal{V}(x)$ and $\varepsilon \in \mathbb{R}_{++}$. To show that $(x, \xi) \in \overline{\operatorname{epi} f}$, it is enough to show that $W \cap \operatorname{epi} f \neq \varnothing$. To this end, note that (1.36) and (iv) imply that $\bar{f}(x) \geqslant \inf f(V)$. Hence, there exists $y \in V$ such that $f(y) \leqslant \bar{f}(x) + \varepsilon \leqslant \xi + \varepsilon$ and therefore $(y, \xi + \varepsilon) \in W \cap \operatorname{epi} f$. $\square$

1.11 Sequential Topological Notions

Let $\mathcal{X}$ be a Hausdorff space and let C be a subset of $\mathcal{X}$. Then C is *sequentially closed* if the limit of every convergent sequence $(x_n)_{n\in\mathbb{N}}$ that lies in C is also in C. A closed set is sequentially closed, but the converse is false (see Example 3.33), which shows that, in general, sequences are not adequate to describe topological notions.

Definition 1.33 A subset C of a Hausdorff space $\mathcal{X}$ is *sequentially compact* if every sequence in C has a sequential cluster point in C, i.e., if every sequence in C has a subsequence that converges to a point in C.

The notions of *sequential continuity* and *sequential lower (upper) semicontinuity* are obtained by replacing nets by sequences in Fact 1.19 and Definition 1.21, respectively. Thus, by replacing nets by sequences and subnets by subsequences in the proofs of Lemma 1.12, Lemma 1.14, and Lemma 1.24, we obtain the following sequential versions.

Lemma 1.34 *Let C be a sequentially compact subset of a Hausdorff space $\mathcal{X}$. Then C is sequentially closed, and every sequentially closed subset of C is sequentially compact.*

Lemma 1.35 *Let C be a sequentially compact subset of a Hausdorff space $\mathcal{X}$ and suppose that $(x_n)_{n\in\mathbb{N}}$ is a sequence in C that has a unique sequential cluster point x. Then $x_n \to x$.*

Lemma 1.36 *Let $\mathcal{X}$ be a Hausdorff space and let $f\colon \mathcal{X} \to [-\infty,+\infty]$. Then the following are equivalent:*

(i) *f is sequentially lower semicontinuous.*
(ii) $\operatorname{epi} f$ *is sequentially closed in $\mathcal{X} \times \mathbb{R}$.*
(iii) *For every $\xi \in \mathbb{R}$, $\operatorname{lev}_{\leqslant\xi} f$ is sequentially closed in $\mathcal{X}$.*

Remark 1.37 Let $\mathcal{X}$ be a topological space. Then $\mathcal{X}$ is called *sequential* if every sequentially closed subset of $\mathcal{X}$ is closed, i.e., if the notions of closedness and sequential closedness coincide. It follows that in sequential spaces the notions of lower semicontinuity and sequential lower semicontinuity are equivalent. Alternatively, $\mathcal{X}$ is sequential if, for every topological space $\mathcal{Y}$ and every operator $T\colon \mathcal{X} \to \mathcal{Y}$, the notions of continuity and sequential continuity coincide. Note, however, that in Hausdorff sequential spaces, the notions of compactness and sequential compactness need not coincide (see [340, Counterexample 43]).

1.12 Metric Spaces

Let $\mathcal{X}$ be a metric space with *distance* (or *metric*) d, and let C be a subset of $\mathcal{X}$. The *diameter* of C is $\operatorname{diam} C = \sup_{(x,y)\in C\times C} d(x,y)$. The *distance to* C is the function

$$d_C\colon \mathcal{X} \to [0,+\infty]\colon x \mapsto \inf d(x,C). \tag{1.47}$$

Note that if $C = \varnothing$ then $d_C \equiv +\infty$. The *closed* and *open balls* with center $x \in \mathcal{X}$ and radius $\rho \in \mathbb{R}_{++}$ in $\mathcal{X}$ are defined as $B(x;\rho) = \big\{y \in \mathcal{X} \bigm| d(x,y) \leqslant \rho\big\}$ and $\big\{y \in \mathcal{X} \bigm| d(x,y) < \rho\big\}$, respectively. The *metric topology* of $\mathcal{X}$ is the topology that has the family of all open balls as a base. A topological space is *metrizable* if its topology coincides with a metric topology.

A sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{X}$ converges to a point $x \in \mathcal{X}$ if $d(x_n,x) \to 0$. Moreover, $\mathcal{X}$ is a sequential Hausdorff space and thus, as seen in Remark 1.37, the notions of closedness, continuity, and lower semicontinuity are equivalent to their sequential counterparts.

Fact 1.38 *Let $\mathcal{X}$ be a metric space, let $\mathcal{Y}$ be a Hausdorff space, and let $T\colon \mathcal{X} \to \mathcal{Y}$. Then T is continuous if and only if it is sequentially continuous.*

In addition, in metric spaces, the notions of compactness and sequential compactness are equivalent.

Fact 1.39 *Let C be a subset of a metric space $\mathcal{X}$. Then C is compact if and only if it is sequentially compact.*

Lemma 1.40 *Let C be a subset of a metric space $\mathcal{X}$ and let $z \in \mathcal{X}$ be such that, for every strictly positive integer n, $C \cap B(z;n)$ is closed. Then C is closed.*

Proof. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in C that converges to a point x. Since $(x_n)_{n\in\mathbb{N}}$ is bounded, there exists $m \in \mathbb{N} \smallsetminus \{0\}$ such that $(x_n)_{n\in\mathbb{N}}$ lies in $C \cap B(z;m)$. The hypothesis implies that $x \in C \cap B(z;m)$ and hence that $x \in C$. $\square$

Lemma 1.41 *Let C be a compact subset of a metric space $\mathcal{X}$. Then C is closed and bounded.*

Proof. Closedness follows from Lemma 1.12. Now suppose that C is not bounded. Then it contains a sequence $(x_n)_{n\in\mathbb{N}}$ such that $d(x_0, x_n) \to +\infty$. Clearly, $(x_n)_{n\in\mathbb{N}}$ has no convergent subsequence and C is therefore not sequentially compact. In view of Fact 1.39, C is not compact. $\square$

Lemma 1.42 *Let $\mathcal{X}$ be a metric space, let $f\colon \mathcal{X} \to [-\infty, +\infty]$, let $x \in \mathcal{X}$, and let $\mathrm{S}(x)$ be the set of all sequences in $\mathcal{X}$ that converge to x. Then*

$$\varliminf_{y\to x} f(y) = \min_{(x_n)_{n\in\mathbb{N}}\in \mathrm{S}(x)} \varliminf f(x_n). \tag{1.48}$$

Proof. Let $(\varepsilon_n)_{n\in\mathbb{N}}$ be a strictly decreasing sequence in $\mathbb{R}_{++}$ such that $\varepsilon_n \downarrow 0$, and set $(\forall n \in \mathbb{N})\ V_n = \big\{ y \in \mathcal{X} \bigm| d(x,y) < \varepsilon_n \big\}$ and $\sigma = \sup_{n\in\mathbb{N}} \inf f(V_n)$. By Lemma 1.23,

$$\inf \Big\{ \varliminf_{n\in\mathbb{N}} f(x_n) \Bigm| x_n \to x \Big\} \geqslant \varliminf_{y\to x} f(y) = \sup_{V\in\mathcal{V}(x)} \inf f(V) = \sigma. \tag{1.49}$$

Since $\inf f(V_n) \uparrow \sigma$, it suffices to provide a sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{X}$ such that $x_n \to x$ and $\varliminf f(x_n) \leqslant \sigma$. If $\sigma = +\infty$, then (1.49) implies that $f(x_n) \to \sigma$ for every sequence $(x_n)_{n\in\mathbb{N}}$ that converges to x. Now assume that $\sigma \in \mathbb{R}$ and that $(x_n)_{n\in\mathbb{N}}$ is a sequence such that $(\forall n \in \mathbb{N})\ x_n \in V_n$ and $f(x_n) \leqslant \varepsilon_n + \inf f(V_n)$. Then $\varliminf f(x_n) \leqslant \varlimsup f(x_n) \leqslant \varlimsup \varepsilon_n + \varlimsup \inf f(V_n) = \sigma$. Finally, assume that $\sigma = -\infty$. Then, for every $n \in \mathbb{N}$, $\inf f(V_n) = -\infty$, and we take $x_n \in V_n$ such that $f(x_n) \leqslant -n$. Then $\varliminf f(x_n) = -\infty = \sigma$, and the proof is complete. $\square$

A sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{X}$ is a *Cauchy sequence* if $d(x_m, x_n) \to 0$ as $\min\{m,n\} \to +\infty$. The metric space $\mathcal{X}$ is *complete* if every Cauchy sequence in $\mathcal{X}$ converges to a point in $\mathcal{X}$.

Lemma 1.43 (Cantor) *Let $\mathcal{X}$ be a complete metric space and let $(C_n)_{n\in\mathbb{N}}$ be a sequence of nonempty closed sets in $\mathcal{X}$ such that $(\forall n \in \mathbb{N})\ C_n \supset C_{n+1}$ and $\operatorname{diam} C_n \to 0$. Then $\bigcap_{n\in\mathbb{N}} C_n$ is a singleton.*

Proof. Set $C = \bigcap_{n\in\mathbb{N}} C_n$. For every $n \in \mathbb{N}$, fix $x_n \in C_n$ and set $A_n = \{x_m\}_{m\geqslant n} \subset C_n$. Then $\operatorname{diam} A_n \to 0$, i.e., $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. By completeness, there exists $x \in \mathcal{X}$ such that $x_n \to x$. Hence, for every $n \in \mathbb{N}$, $C_n \ni x_{n+p} \to x$ as $p \to +\infty$ and, by closedness of C_n, we get $x \in C_n$. Thus, $x \in C$ and $\operatorname{diam} C \leqslant \operatorname{diam} C_n \to 0$. Altogether, $C = \{x\}$. ☐

Lemma 1.44 (Ursescu) *Let $\mathcal{X}$ be a complete metric space. Then the following hold:*

(i) *Suppose that $(C_n)_{n\in\mathbb{N}}$ is a sequence of closed subsets of $\mathcal{X}$. Then $\overline{\bigcup_{n\in\mathbb{N}} \operatorname{int} C_n} = \overline{\operatorname{int} \bigcup_{n\in\mathbb{N}} C_n}$.*
(ii) *Suppose that $(C_n)_{n\in\mathbb{N}}$ is a sequence of open subsets of $\mathcal{X}$. Then $\operatorname{int} \bigcap_{n\in\mathbb{N}} \overline{C_n} = \operatorname{int} \overline{\bigcap_{n\in\mathbb{N}} C_n}$.*

Proof. For every subset C of $\mathcal{X}$, one has $\mathcal{X} \smallsetminus \overline{C} = \operatorname{int}(\mathcal{X} \smallsetminus C)$ and $\mathcal{X} \smallsetminus \operatorname{int} C = \overline{\mathcal{X} \smallsetminus C}$. This and De Morgan's laws imply that (i) and (ii) are equivalent.

(ii): The inclusion $\operatorname{int} \overline{\bigcap_{n\in\mathbb{N}} C_n} \subset \operatorname{int} \bigcap_{n\in\mathbb{N}} \overline{C_n}$ is clear. To show the reverse inclusion, let us fix

$$z \in \operatorname{int} \bigcap_{n\in\mathbb{N}} \overline{C_n} \quad \text{and} \quad \varepsilon \in \mathbb{R}_{++}. \tag{1.50}$$

Using strong (also known as complete) induction, we shall construct sequences $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{X}$ and $(\varepsilon_n)_{n\in\mathbb{N}}$ in $\mathbb{R}_{++}$ such that $x_0 = z$ and, for every $n \in \mathbb{N}$,

$$B(x_{n+1}; \varepsilon_{n+1}) \subset B(x_n; \varepsilon_n) \cap C_n \quad \text{with} \quad \varepsilon_n \in \left]0, \varepsilon/2^n\right[. \tag{1.51}$$

For every $n \in \mathbb{N}$, let us denote by U_n the open ball with center x_n and radius ε_n. First, set $x_0 = z$ and let $\varepsilon_0 \in \left]0, \varepsilon\right[$ be such that

$$B(x_0; \varepsilon_0) \subset \bigcap_{n\in\mathbb{N}} \overline{C_n}. \tag{1.52}$$

Since $x_0 \in \overline{C_0}$, the set $U_0 \cap C_0$ is nonempty and open. Thus, there exist $x_1 \in \mathcal{X}$ and $\varepsilon_1 \in \left]0, \varepsilon/2\right[$ such that

$$B(x_1; \varepsilon_1) \subset U_0 \cap C_0 \subset B(x_0; \varepsilon_0) \cap C_0. \tag{1.53}$$

Now assume that $(x_k)_{0\leqslant k\leqslant n}$ and $(\varepsilon_k)_{0\leqslant k\leqslant n}$ are already constructed. Then, using (1.52), we obtain

$$x_n \in U_n \subset B(x_n; \varepsilon_n) \subset B(x_0; \varepsilon_0) \subset \overline{C_n}. \tag{1.54}$$

Hence, there exists $x_{n+1} \in C_n$ such that $d(x_n, x_{n+1}) < \varepsilon_n/2$. Moreover, x_{n+1} belongs to the open set $U_n \cap C_n$. As required, there exists $\varepsilon_{n+1} \in \left]0, \varepsilon_n/2\right[\subset \left]0, \varepsilon/2^{n+1}\right[$ such that

$$B(x_{n+1}; \varepsilon_{n+1}) \subset U_n \cap C_n \subset B(x_n; \varepsilon_n) \cap C_n. \tag{1.55}$$

Since the sequence $(B(x_n; \varepsilon_n))_{n\in\mathbb{N}}$ is decreasing and $\varepsilon_n \to 0$, we have $\operatorname{diam} B(x_n; \varepsilon_n) \leqslant 2\varepsilon_n \to 0$. Therefore, Lemma 1.43 yields a point $z_\varepsilon \in \mathcal{X}$ such that

$$\bigcap_{n\in\mathbb{N}} B(x_n; \varepsilon_n) = \{z_\varepsilon\}. \tag{1.56}$$

Combining (1.51) and (1.56), we deduce that $z_\varepsilon \in B(z; \varepsilon) \cap \bigcap_{n\in\mathbb{N}} C_n$. Letting $\varepsilon \downarrow 0$, we conclude that $z \in \overline{\bigcap_{n\in\mathbb{N}} C_n}$. ☐

A countable intersection of open sets in a Hausdorff space is called a G_δ set.

Corollary 1.45 *Let $\mathcal{X}$ be a complete metric space, and let $(C_n)_{n\in\mathbb{N}}$ be a sequence of dense open subsets of $\mathcal{X}$. Then $\bigcap_{n\in\mathbb{N}} C_n$ is a dense G_δ subset of $\mathcal{X}$.*

Proof. It is clear that $C = \bigcap_{n\in\mathbb{N}} C_n$ is a G_δ set. Using Lemma 1.44(ii), we obtain $\mathcal{X} = \operatorname{int} \bigcap_{n\in\mathbb{N}} \overline{C_n} = \operatorname{int} \overline{\bigcap_{n\in\mathbb{N}} C_n} = \operatorname{int} \overline{C} \subset \overline{C} \subset \mathcal{X}$. Hence $\overline{C} = \mathcal{X}$. ☐

Theorem 1.46 (Ekeland) *Let $(\mathcal{X}, d)$ be a complete metric space, let $f\colon \mathcal{X} \to \left]-\infty, +\infty\right]$ be proper, lower semicontinuous, and bounded below, let $\alpha \in \mathbb{R}_{++}$, let $\beta \in \mathbb{R}_{++}$, and suppose that $y \in \operatorname{dom} f$ satisfies $f(y) \leqslant \alpha + \inf f(\mathcal{X})$. Then there exists $z \in \mathcal{X}$ such that the following hold:*

(i) $f(z) + (\alpha/\beta)d(y, z) \leqslant f(y)$.
(ii) $d(y, z) \leqslant \beta$.
(iii) $(\forall x \in \mathcal{X} \smallsetminus \{z\})\ f(z) < f(x) + (\alpha/\beta)d(x, z)$.

Proof. We fix $x_0 \in \mathcal{X}$ and define inductively sequences $(x_n)_{n\in\mathbb{N}}$ and $(C_n)_{n\in\mathbb{N}}$ as follows. Given $x_n \in \mathcal{X}$, where $n \in \mathbb{N}$, set

$$C_n = \big\{x \in \mathcal{X} \;\big|\; f(x) + (\alpha/\beta)d(x_n, x) \leqslant f(x_n)\big\} \tag{1.57}$$

and take $x_{n+1} \in \mathcal{X}$ such that

$$x_{n+1} \in C_n \quad \text{and} \quad f(x_{n+1}) \leqslant \tfrac{1}{2} f(x_n) + \tfrac{1}{2} \inf f(C_n). \tag{1.58}$$

Since $x_{n+1} \in C_n$, we have $(\alpha/\beta)d(x_n, x_{n+1}) \leqslant f(x_n) - f(x_{n+1})$. Thus,

$$\big(f(x_n)\big)_{n\in\mathbb{N}} \quad \text{is decreasing and bounded below, hence convergent,} \tag{1.59}$$

and

$$(\forall n \in \mathbb{N})(\forall m \in \mathbb{N}) \quad n \leqslant m \;\Rightarrow\; (\alpha/\beta)d(x_n, x_m) \leqslant f(x_n) - f(x_m). \tag{1.60}$$

Combining (1.59) and (1.60), we see that $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. Set $z = \lim x_n$. Since f is lower semicontinuous at z, it follows from (1.59) that

$$f(z) \leqslant \lim f(x_n). \tag{1.61}$$

Letting $m \to +\infty$ in (1.60), we deduce that

$$(\forall n \in \mathbb{N}) \quad (\alpha/\beta)d(x_n, z) \leqslant f(x_n) - f(z). \tag{1.62}$$

Recalling that $x_0 = y$ and setting $n = 0$ in (1.62), we obtain (i). In turn, (i) implies that $f(z) + (\alpha/\beta)d(y,z) \leqslant f(y) \leqslant \alpha + \inf f(\mathcal{X}) \leqslant \alpha + f(z)$. Thus (ii) holds. Now assume that (iii) is false. Then there exists $x \in \mathcal{X} \smallsetminus \{z\}$ such that

$$f(x) \leqslant f(z) - (\alpha/\beta)d(x,z) < f(z). \tag{1.63}$$

In view of (1.62), we get

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad f(x) &\leqslant f(z) - (\alpha/\beta)d(x,z)\\ &\leqslant f(x_n) - (\alpha/\beta)\big(d(x,z) + d(x_n,z)\big)\\ &\leqslant f(x_n) - (\alpha/\beta)d(x,x_n). \end{aligned}\tag{1.64}$$

Thus $x \in \bigcap_{n\in\mathbb{N}} C_n$. Using (1.58), we deduce that $(\forall n \in \mathbb{N})$ $2f(x_{n+1}) - f(x_n) \leqslant f(x)$. Hence $\lim f(x_n) \leqslant f(x)$. This, (1.63), and (1.61) imply that $\lim f(x_n) \leqslant f(x) < f(z) \leqslant \lim f(x_n)$, which is impossible. Therefore, (iii) holds. ☐

Definition 1.47 Let $(\mathcal{X}_1, d_1)$ and $(\mathcal{X}_2, d_2)$ be metric spaces, let $T\colon \mathcal{X}_1 \to \mathcal{X}_2$, and let C be a subset of $\mathcal{X}_1$. Then T is *Lipschitz continuous* with constant $\beta \in \mathbb{R}_+$ if

$$(\forall x \in \mathcal{X}_1)(\forall y \in \mathcal{X}_1) \quad d_2(Tx, Ty) \leqslant \beta d_1(x,y), \tag{1.65}$$

locally Lipschitz continuous near a point $x \in \mathcal{X}_1$ if there exists $\rho \in \mathbb{R}_{++}$ such that the operator $T|_{B(x;\rho)}$ is Lipschitz continuous, and locally Lipschitz continuous on C if it is locally Lipschitz continuous near every point in C. Finally, T is *Lipschitz continuous relative to* C with constant $\beta \in \mathbb{R}_+$ if

$$(\forall x \in C)(\forall y \in C) \quad d_2(Tx, Ty) \leqslant \beta d_1(x,y). \tag{1.66}$$

Example 1.48 Let C be a nonempty subset of a metric space $(\mathcal{X}, d)$. Then

$$(\forall x \in \mathcal{X})(\forall y \in \mathcal{X}) \quad |d_C(x) - d_C(y)| \leqslant d(x,y). \tag{1.67}$$

Proof. Take x and y in $\mathcal{X}$. Then $(\forall z \in \mathcal{X})$ $d(x,z) \leqslant d(x,y) + d(y,z)$. Taking the infimum over $z \in C$ yields $d_C(x) \leqslant d(x,y) + d_C(y)$, hence $d_C(x) - d_C(y) \leqslant d(x,y)$. Interchanging x and y, we obtain $d_C(y) - d_C(x) \leqslant d(x,y)$. Altogether, $|d_C(x) - d_C(y)| \leqslant d(x,y)$. ☐

Theorem 1.49 (Ostrowski) *Let C be a nonempty compact subset of a metric space $(\mathcal{X}, d)$, let $(x_n)_{n\in\mathbb{N}}$ be a sequence in C such that $d(x_n, x_{n+1}) \to 0$, and denote the set of cluster points of $(x_n)_{n\in\mathbb{N}}$ by S. Then S is connected.*

Proof. Suppose to the contrary that S is not connected. Then there exist two nonempty closed subsets A and B of C such that $A \cup B = S$ and $A \cap B = \varnothing$. Since $A \times B$ is compact, $A \cap B = \varnothing$, and d is continuous, it follows from Theorem 1.29 that there exists $\delta \in \mathbb{R}_{++}$ such that $\min d(A, B) = \delta$. Now let $(k_n)_{n\in\mathbb{N}}$ and $(l_n)_{n\in\mathbb{N}}$ be strictly increasing sequences in $\mathbb{N}$ such that, for every $n \in \mathbb{N}$, $k_n < l_n < k_{n+1}$, $d_A(x_{k_n}) < \delta/3$, $d(x_{l_n-1}, x_{l_n}) < \delta/3$, $d_B(x_{l_n-1}) \geqslant 2\delta/3$, and $d_B(x_{l_n}) < 2\delta/3$. Then, using Example 1.48, $(\forall n \in \mathbb{N})$ $d_B(x_{l_n}) \geqslant d_B(x_{l_n-1}) - d(x_{l_n-1}, x_{l_n}) > \delta/3$. Thus, $(\forall n \in \mathbb{N})$ $\delta/3 < d_B(x_{l_n}) < 2\delta/3$. Therefore, there exists $z \in S$ such that $\delta/3 \leqslant d_B(z) \leqslant 2\delta/3$. Clearly, $z \notin B$. Hence $z \in A$, which implies the contradiction $d_B(z) \leqslant 2\delta/3 < \delta \leqslant d_B(z)$. ☐

The following result is known as the Banach–Picard fixed point theorem. The set of fixed points of an operator $T\colon \mathcal{X} \to \mathcal{X}$ is denoted by $\operatorname{Fix} T$, i.e.,

$$\operatorname{Fix} T = \big\{x \in \mathcal{X} \;\big|\; Tx = x\big\}. \tag{1.68}$$

Theorem 1.50 (Banach–Picard) *Let $(\mathcal{X}, d)$ be a complete metric space and let $T\colon \mathcal{X} \to \mathcal{X}$ be Lipschitz continuous with constant $\beta \in [0, 1[$. Given $x_0 \in \mathcal{X}$, set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = Tx_n. \tag{1.69}$$

Then there exists $x \in \mathcal{X}$ such that the following hold:

(i) *x is the unique fixed point of T.*
(ii) $(\forall n \in \mathbb{N})$ $d(x_{n+1}, x) \leqslant \beta d(x_n, x)$.
(iii) $(\forall n \in \mathbb{N})$ $d(x_n, x) \leqslant \beta^n d(x_0, x)$ *(hence $(x_n)_{n\in\mathbb{N}}$ converges linearly to x).*
(iv) *A priori error estimate:* $(\forall n \in \mathbb{N})$ $d(x_n, x) \leqslant \beta^n d(x_0, x_1)/(1-\beta)$.
(v) *A posteriori error estimate:* $(\forall n \in \mathbb{N})$ $d(x_n, x) \leqslant d(x_n, x_{n+1})/(1-\beta)$.
(vi) $d(x_0, x_1)/(1+\beta) \leqslant d(x_0, x) \leqslant d(x_0, x_1)/(1-\beta)$.

Proof. The triangle inequality and (1.69) yield

$$\begin{aligned}
(\forall m \in \mathbb{N})(\forall n \in \mathbb{N}) \quad d(x_n, x_{n+m}) &\leqslant d(x_n, x_{n+1}) + \cdots + d(x_{n+m-1}, x_{n+m}) \\
&\leqslant (1 + \beta + \cdots + \beta^{m-1}) d(x_n, x_{n+1}) \\
&= \frac{1-\beta^m}{1-\beta} d(x_n, x_{n+1}) && (1.70) \\
&\leqslant \frac{\beta^n}{1-\beta} d(x_0, x_1). && (1.71)
\end{aligned}$$

It follows that $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{X}$ and therefore that it converges to some point $x \in \mathcal{X}$. In turn, since T is continuous, $Tx = T(\lim x_n) = \lim Tx_n = \lim x_{n+1} = x$ and thus $x \in \operatorname{Fix} T$. Now let $y \in \operatorname{Fix} T$. Then $d(x, y) = d(Tx, Ty) \leqslant \beta d(x, y)$ and hence $y = x$. This establishes (i).

(ii): Observe that $(\forall n \in \mathbb{N})\ d(x_{n+1}, x) = d(Tx_n, Tx) \leqslant \beta d(x_n, x)$.

(iii): This follows from (ii).

(iv)&(v): Let $m \to +\infty$ in (1.71) and (1.70), respectively.

(vi): Since $d(x_0, x_1) \leqslant d(x_0, x) + d(x, x_1) \leqslant (1 + \beta)d(x_0, x)$, the first inequality holds. The second inequality follows from (iv) or (v). ☐

We close this chapter with a variant of the Banach–Picard theorem.

Theorem 1.51 *Let $(\mathcal{X}, d)$ be a complete metric space and let $T\colon \mathcal{X} \to \mathcal{X}$ be such that there exists a summable sequence $(\beta_n)_{n\in\mathbb{N}}$ in $\mathbb{R}_+$ such that*

$$(\forall x \in \mathcal{X})(\forall y \in \mathcal{X})(\forall n \in \mathbb{N}) \quad d(T^n x, T^n y) \leqslant \beta_n d(x, y), \tag{1.72}$$

where T^n denotes the n-fold composition of T if $n > 0$, and $T^0 = \mathrm{Id}$. Let $x_0 \in \mathcal{X}$ and set

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = Tx_n \quad \text{and} \quad \alpha_n = \sum_{k=n}^{+\infty} \beta_k. \tag{1.73}$$

Then there exists $x \in \mathcal{X}$ such that the following hold:

(i) *x is the unique fixed point of T.*
(ii) *$x_n \to x$.*
(iii) *$(\forall n \in \mathbb{N})\ \ d(x_n, x) \leqslant \alpha_n d(x_0, x_1)$.*

Proof. We deduce from (1.73) that

$$\begin{aligned}(\forall m \in \mathbb{N})(\forall n \in \mathbb{N}) \quad d(x_n, x_{n+m}) &\leqslant \sum_{k=n}^{n+m-1} d(x_k, x_{k+1}) \\ &= \sum_{k=n}^{n+m-1} d(T^k x_0, T^k x_1) \\ &\leqslant \sum_{k=n}^{n+m-1} \beta_k d(x_0, x_1) \\ &\leqslant \alpha_n d(x_0, x_1).\end{aligned} \tag{1.74}$$

(i)&(ii): Since $(\beta_n)_{n\in\mathbb{N}}$ is summable, we have $\alpha_n \to 0$. Thus, (1.74) implies that $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. Thus, it converges to some point $x \in \mathcal{X}$. It follows from the continuity of T that $Tx = T(\lim x_n) = \lim Tx_n = \lim x_{n+1} = x$ and therefore that $x \in \operatorname{Fix} T$. Now let $y \in \operatorname{Fix} T$. Then (1.72) yields $(\forall n \in \mathbb{N})\ d(x, y) = d(T^n x, T^n y) \leqslant \beta_n d(x, y)$. Since $\beta_n \to 0$, we conclude that $y = x$.

(iii): Let $m \to +\infty$ in (1.74). ☐

Exercises

Exercise 1.1 Let $\mathcal{X}$ be a real vector space and let $A\colon \mathcal{X} \to 2^{\mathcal{X}}$. Show that $\operatorname{gra} A + \operatorname{gra}(-\mathrm{Id}) = \mathcal{X} \times \mathcal{X}$ if and only if $\operatorname{ran}(A + \mathrm{Id}) = \mathcal{X}$.

Exercise 1.2 Let $\mathcal{X}$ and $\mathcal{Y}$ be nonempty sets and let $T\colon \mathcal{X} \to \mathcal{Y}$.

(i) Let $C \subset \mathcal{X}$ and $D \subset \mathcal{Y}$. Show that $C \subset T^{-1}(T(C))$ and $T(T^{-1}(D)) \subset D$. Provide an example of strict inclusion in both cases.

(ii) Let $D \subset \mathcal{Y}$ and let $(D_i)_{i\in I}$ be a family of subsets of $\mathcal{Y}$. Show the following:

(a) $T^{-1}(\mathcal{Y} \smallsetminus D) = \mathcal{X} \smallsetminus T^{-1}(D)$.
(b) $T^{-1}(\bigcap_{i\in I} D_i) = \bigcap_{i\in I} T^{-1}(D_i)$.
(c) $T^{-1}(\bigcup_{i\in I} D_i) = \bigcup_{i\in I} T^{-1}(D_i)$.

(iii) Prove Fact 1.18.

Exercise 1.3 Let C and D be subsets of a topological space $\mathcal{X}$. Show the following:

(i) $\mathcal{X} \smallsetminus \operatorname{int} C = \overline{\mathcal{X} \smallsetminus C}$ and $\mathcal{X} \smallsetminus \overline{C} = \operatorname{int}(\mathcal{X} \smallsetminus C)$.
(ii) $\operatorname{int}(C \cap D) = (\operatorname{int} C) \cap (\operatorname{int} D)$ and $\overline{C \cup D} = \overline{C} \cup \overline{D}$.
(iii) $\operatorname{int}(C \cup D) \neq (\operatorname{int} C) \cup (\operatorname{int} D)$ and $\overline{C \cap D} \neq \overline{C} \cap \overline{D}$.

Exercise 1.4 Let $A = \mathbb{Z}$ be directed by $\leqslant$ and define a net $(x_a)_{a\in A}$ in $\mathbb{R}$ by

$$(\forall a \in A) \quad x_a = \begin{cases} a, & \text{if } a \leqslant 0; \\ 1/a, & \text{if } a > 0. \end{cases} \tag{1.75}$$

Show that $(x_a)_{a\in A}$ is unbounded and that it converges.

Exercise 1.5 In this exercise, $\mathbb{N}$ designates the unordered and undirected set of positive integers. Let $A = \mathbb{N}$ be directed by $\leqslant$, and let $B = \mathbb{N}$ be directed by the relation $\preccurlyeq$ that coincides with $\leqslant$ on $\{2n\}_{n\in\mathbb{N}}$ and on $\{2n+1\}_{n\in\mathbb{N}}$, and satisfies $(\forall m \in \mathbb{N})(\forall n \in \mathbb{N})\ 2m+1 \preccurlyeq 2n$. Now let $(x_a)_{a\in A}$ be a sequence in $\mathbb{R}$.

(i) Show that $(\forall a \in A)(\forall b \in B)\ b \succcurlyeq 2a \Rightarrow 3b \geqslant a$.
(ii) Show that $(x_{3b})_{b\in B}$ is a subnet, but not a subsequence, of $(x_a)_{a\in A}$.
(iii) Set $(\forall a \in A)\ x_a = (1 - (-1)^a)a/2$. Show that the subnet $(x_{3b})_{b\in B}$ converges to 0 but that the subsequence $(x_{3a})_{a\in A}$ does not converge.

Exercise 1.6 Let $\mathcal{X}$ be a nonempty set and let $f\colon \mathcal{X} \to \mathbb{R}$. Equip $\mathcal{X}$ with a topology T and $\mathbb{R}$ with the usual topology. Provide a general condition for f to be continuous in the following cases:

(i) $\mathsf{T} = \{\varnothing, \mathcal{X}\}$.
(ii) $\mathsf{T} = 2^{\mathcal{X}}$.

Exercise 1.7 Let $f\colon \mathbb{R} \to [-\infty,+\infty]$ and set

$$(\forall z \in \mathbb{R}) \quad f_z\colon \mathbb{R} \to [-\infty,+\infty] : x \mapsto \begin{cases} f(z), & \text{if } x = z; \\ +\infty, & \text{otherwise.} \end{cases} \tag{1.76}$$

Show that the functions $(f_z)_{z\in\mathbb{R}}$ are lower semicontinuous and conclude that the infimum of an infinite family of lower semicontinuous functions need not be lower semicontinuous (compare with Lemma 1.26).

Exercise 1.8 Construct two lower semicontinuous functions from $\mathbb{R}$ to $\mathbb{R}$ such that their composition is not lower semicontinuous.

Exercise 1.9 Let $\mathcal{X}$ be a Hausdorff space, let $f\colon \mathcal{X} \to [-\infty,+\infty]$, and let $g\colon \mathcal{X} \to \mathbb{R}$ be continuous. Show that $\overline{f+g} = \bar{f} + g$.

Exercise 1.10 Let $\mathcal{X}$ be a topological space. Then $\mathcal{X}$ is *first countable* if, for every $x \in \mathcal{X}$, there exists a countable base of neighborhoods, i.e., a family $(V_n)_{n\in\mathbb{N}}$ in $\mathcal{V}(x)$ such that $(\forall\, V \in \mathcal{V}(x))(\exists\, n \in \mathbb{N})\ V_n \subset V$.

(i) Show that, without loss of generality, the sequence of sets $(V_n)_{n\in\mathbb{N}}$ can be taken to be decreasing in the above definition.
(ii) Show that if $\mathcal{X}$ is metrizable, then it is first countable.
(iii) The space $\mathcal{X}$ is a *Fréchet space* if, for every subset C of $\mathcal{X}$ and every $x \in \overline{C}$, there exists a sequence $(x_n)_{n\in\mathbb{N}}$ in C such that $x_n \to x$. Show the following:

   (a) If $\mathcal{X}$ is first countable, then it is Fréchet.
   (b) If $\mathcal{X}$ is Fréchet, then it is sequential.

(iv) Conclude that the following implications hold for $\mathcal{X}$: metrizable $\Rightarrow$ first countable $\Rightarrow$ Fréchet $\Rightarrow$ sequential.

Exercise 1.11 Let C be a subset of a metric space $\mathcal{X}$. Show the following:

(i) $\overline{C} = \big\{x \in \mathcal{X} \;\big|\; d_C(x) = 0\big\}$.
(ii) $\overline{C}$ is a G_δ set.

Exercise 1.12 A subset of a Hausdorff space $\mathcal{X}$ is an F_σ set if it is a countable union of closed sets. Show the following:

(i) The complement of an F_σ set is a G_δ set.
(ii) If $\mathcal{X}$ is a metric space, every open set in $\mathcal{X}$ is an F_σ set.

Exercise 1.13 Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in a Hausdorff space $\mathcal{X}$ that converges to some point $x \in \mathcal{X}$. Show that the set $\{x_n\}_{n\in\mathbb{N}} \cup \{x\}$ is compact.

Exercise 1.14 In the compact metric space $\mathcal{X} = [0,1]$, direct the set

$$A = \big\{a \in \left]0,1\right[\;\big|\; a \text{ is a rational number}\big\} \tag{1.77}$$

by $\leqslant$ and define $(\forall a \in A)\ x_a = a$. Show that the net $(x_a)_{a\in A}$ converges to 1, while the set $\{x_a\}_{a\in A} \cup \{1\}$ is not closed. Compare with Exercise 1.13.

Exercise 1.15 Find a metric space $(\mathcal{X}, d)$ such that $(\forall x \in \mathcal{X})$ $\operatorname{int} B(x;1) \neq \big\{y \in \mathcal{X} \;\big|\; d(x,y) < 1\big\}$.

Exercise 1.16 Let $(x_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in a metric space $\mathcal{X}$. Show that $(x_n)_{n\in\mathbb{N}}$ is bounded and that, if it possesses a sequential cluster point, it converges to that point.

Exercise 1.17 Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in a complete metric space $(\mathcal{X}, d)$ such that $\sum_{n\in\mathbb{N}} d(x_n, x_{n+1}) < +\infty$. Show that $(x_n)_{n\in\mathbb{N}}$ converges and that this is no longer true if we merely assume that $\sum_{n\in\mathbb{N}} d^2(x_n, x_{n+1}) < +\infty$.

Exercise 1.18 Show that if in Lemma 1.44 the sequence $(C_n)_{n\in\mathbb{N}}$ is replaced by an arbitrary family $(C_a)_{a\in A}$, then the conclusion fails.

Exercise 1.19 Provide an example of a function f from $\mathbb{R}$ to $\mathbb{R}$ such that f is locally Lipschitz continuous on $\mathbb{R}$ but f is not Lipschitz continuous.

Exercise 1.20 Let $(\mathcal{X}, d)$ be a metric space and let $(x_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathcal{X}$. Show that for every $x \in \mathcal{X}$, $(d(x, x_n))_{n\in\mathbb{N}}$ converges, that $f\colon \mathcal{X} \to \mathbb{R}\colon x \mapsto \lim d(x, x_n)$ is Lipschitz continuous with constant 1, and that $f(x_n) \to 0$.

Exercise 1.21 Let $(\mathcal{X}, d)$ be a metric space. Use Theorem 1.46 and Exercise 1.20 to show that $\mathcal{X}$ is complete if and only if, for every Lipschitz continuous function $f\colon \mathcal{X} \to \mathbb{R}_+$ and for every $\varepsilon \in \left]0,1\right[$, there exists $z \in \mathcal{X}$ such that $(\forall x \in \mathcal{X})$ $f(z) \leqslant f(x) + \varepsilon d(x,z)$.

Exercise 1.22 Let C be a nonempty subset of a metric space $(\mathcal{X}, d)$. Show that

$$(\forall x \in \mathcal{X})(\forall y \in \mathcal{X}) \quad d_C^2(y) \leqslant 2d^2(x,y) + 2d_C^2(x). \tag{1.78}$$

Exercise 1.23 Let $(\mathcal{X}, d)$ be a metric space and let $\mathcal{C}$ be the class of nonempty bounded closed subsets of $\mathcal{X}$. Define the *Hausdorff distance* by

$$H\colon \mathcal{C} \times \mathcal{C} \to \mathbb{R}_+ \colon (C, D) \mapsto \max\left\{\sup_{x\in C} d_D(x), \sup_{x\in D} d_C(x)\right\}. \tag{1.79}$$

The purpose of this exercise is to verify that H is indeed a distance on $\mathcal{C}$. Show the following:

(i) $(\forall x \in \mathcal{X})(\forall y \in \mathcal{X})$ $H(\{x\}, \{y\}) = d(x,y)$.
(ii) $(\forall C \in \mathcal{C})(\forall D \in \mathcal{C})$ $H(C,D) = H(D,C)$.
(iii) $(\forall C \in \mathcal{C})(\forall D \in \mathcal{C})$ $H(C,D) = 0 \;\Leftrightarrow\; C = D$.
(iv) $(\forall C \in \mathcal{C})(\forall D \in \mathcal{C})(\forall x \in \mathcal{X})$ $d_D(x) \leqslant d_C(x) + H(C,D)$.
(v) $(\mathcal{C}, H)$ is a metric space.

Exercise 1.24 Use Theorem 1.46 to prove Theorem 1.50(i).

Exercise 1.25 Let $\mathcal{X}$ be a complete metric space, suppose that $(C_i)_{i\in I}$ is a family of subsets of $\mathcal{X}$ that are open and dense in $\mathcal{X}$, and set $C = \bigcap_{i\in I} C_i$. Show the following:

(i) If I is finite, then C is open and dense in $\mathcal{X}$.
(ii) If I is countably infinite, then C is dense (hence nonempty), but C may fail to be open.
(iii) If I is uncountable, then C may be empty.

Chapter 2
Hilbert Spaces

Throughout this book, $\mathcal{H}$ is a real Hilbert space with scalar (or inner) product $\langle\cdot \mid \cdot\rangle$. The associated norm is denoted by $\|\cdot\|$ and the associated distance by d, i.e.,

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|x\| = \sqrt{\langle x \mid x\rangle} \quad \text{and} \quad d(x,y) = \|x-y\|. \tag{2.1}$$

The identity operator on $\mathcal{H}$ is denoted by Id.

In this chapter, we derive useful identities and inequalities, and we review examples and basic results from linear and nonlinear analysis in a Hilbert space setting.

2.1 Notation and Examples

The *orthogonal complement* of a subset C of $\mathcal{H}$ is denoted by $C^\perp$, i.e.,

$$C^\perp = \big\{u \in \mathcal{H} \;\big|\; (\forall x \in C) \;\; \langle x \mid u\rangle = 0\big\}. \tag{2.2}$$

An orthonormal subset C of $\mathcal{H}$ is an *orthonormal basis* of $\mathcal{H}$ if $\overline{\operatorname{span}}\, C = \mathcal{H}$. The space $\mathcal{H}$ is *separable* if it possesses a countable orthonormal basis. Now let $(x_i)_{i\in I}$ be a family of vectors in $\mathcal{H}$ and let $\mathcal{I}$ be the class of nonempty finite subsets of I, directed by $\subset$. Then $(x_i)_{i\in I}$ is *summable* if there exists $x \in \mathcal{H}$ such that the net $(\sum_{i\in J} x_i)_{J\in\mathcal{I}}$ converges to x, i.e., by (1.26),

$$(\forall \varepsilon \in \mathbb{R}_{++})(\exists\, K \in \mathcal{I})(\forall J \in \mathcal{I}) \quad J \supset K \quad \Rightarrow \quad \Big\|x - \sum_{i\in J} x_i\Big\| \leqslant \varepsilon. \tag{2.3}$$

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-48311-5_31

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_2

In this case we write $x = \sum_{i\in I} x_i$. For every family $(\alpha_i)_{i\in I}$ in $[0,+\infty]$, we have

$$\sum_{i\in I} \alpha_i = \sup_{J\in\mathcal{I}} \sum_{i\in J} \alpha_i. \tag{2.4}$$

Here are specific real Hilbert spaces that will be used in this book.

Example 2.1 Let I be a nonempty set. The *Hilbert direct sum* of a family of real Hilbert spaces $(\mathcal{H}_i, \|\cdot\|_i)_{i\in I}$ is the real Hilbert space

$$\bigoplus_{i\in I} \mathcal{H}_i = \left\{ \boldsymbol{x} = (x_i)_{i\in I} \in \bigtimes_{i\in I} \mathcal{H}_i \;\middle|\; \sum_{i\in I} \|x_i\|_i^2 < +\infty \right\} \tag{2.5}$$

equipped with the addition $(\boldsymbol{x}, \boldsymbol{y}) \mapsto (x_i + y_i)_{i\in I}$, the scalar multiplication $(\alpha, \boldsymbol{x}) \mapsto (\alpha x_i)_{i\in I}$, and the scalar product

$$(\boldsymbol{x}, \boldsymbol{y}) \mapsto \sum_{i\in I} \langle x_i \mid y_i \rangle_i\,, \tag{2.6}$$

where $\langle \cdot \mid \cdot \rangle_i$ denotes the scalar product of $\mathcal{H}_i$ (when I is finite, we shall sometimes adopt a common practice and write $\bigtimes_{i\in I} \mathcal{H}_i$ instead of $\bigoplus_{i\in I} \mathcal{H}_i$). Now suppose that, for every $i \in I$, $f_i \colon \mathcal{H}_i \to \left]-\infty, +\infty\right]$ and that, if I is infinite, $\inf_{i\in I} f_i \geqslant 0$. Then

$$\bigoplus_{i\in I} f_i \colon \bigoplus_{i\in I} \mathcal{H}_i \to \left]-\infty, +\infty\right] : (x_i)_{i\in I} \mapsto \sum_{i\in I} f_i(x_i). \tag{2.7}$$

Example 2.2 If each $\mathcal{H}_i$ is the Euclidean line $\mathbb{R}$ in Example 2.1, then we obtain $\ell^2(I) = \bigoplus_{i\in I} \mathbb{R}$, which is equipped with the scalar product $(x, y) = ((\xi_i)_{i\in I}, (\eta_i)_{i\in I}) \mapsto \sum_{i\in I} \xi_i \eta_i$. The standard unit vectors $(e_i)_{i\in I}$ of $\ell^2(I)$ are defined by

$$(\forall i \in I) \quad e_i \colon I \to \mathbb{R} \colon j \mapsto \begin{cases} 1, & \text{if } j = i; \\ 0, & \text{otherwise.} \end{cases} \tag{2.8}$$

Example 2.3 If $I = \{1, \ldots, N\}$ in Example 2.2, then we obtain the standard Euclidean space $\mathbb{R}^N$.

Example 2.4 Let M and N be strictly positive integers. Then $\mathbb{R}^{M\times N}$ denotes the Hilbert space of real $M \times N$ matrices equipped with the scalar product $(A, B) \mapsto \operatorname{tra}(A^{\mathsf{T}} B)$, where tra is the trace function. The associated norm is the *Frobenius norm* $\|\cdot\|_{\mathsf{F}}$.

Example 2.5 Let N be a strictly positive integer. The Hilbert space $\mathbb{S}^N$ is the subspace of $\mathbb{R}^{N\times N}$ that consists of all the symmetric matrices.

Example 2.6 Let $(\Omega, \mathcal{F}, \mu)$ be a (positive) measure space. A property is said to hold μ-*almost everywhere* (μ-a.e.) on Ω if there exists a set $C \in \mathcal{F}$ such that $\mu(C) = 0$ and the property holds on $\Omega \smallsetminus C$. Let $(\mathsf{H}, \langle \cdot \mid \cdot \rangle_{\mathsf{H}})$ be a separable real Hilbert space, and let $p \in \left[1, +\infty\right[$. Denote by $L^p((\Omega, \mathcal{F}, \mu); \mathsf{H})$ the space of (equivalence classes of) Borel measurable functions $x \colon \Omega \to \mathsf{H}$ such that

$\int_\Omega \|x(\omega)\|_{\mathsf{H}}^p \mu(d\omega) < +\infty$. Then $L^2((\Omega, \mathcal{F}, \mu); \mathsf{H})$ is a real Hilbert space with scalar product $(x, y) \mapsto \int_\Omega \langle x(\omega) \mid y(\omega)\rangle_{\mathsf{H}}\, \mu(d\omega)$.

Example 2.7 In Example 2.6, let $\mathsf{H} = \mathbb{R}$. Then we obtain the real Banach space $L^p(\Omega, \mathcal{F}, \mu) = L^p((\Omega, \mathcal{F}, \mu); \mathbb{R})$ and, for $p = 2$, the real Hilbert space $L^2(\Omega, \mathcal{F}, \mu)$, which is equipped with the scalar product $(x, y) \mapsto \int_\Omega x(\omega)y(\omega)\mu(d\omega)$.

Example 2.8 In Example 2.6, let $T \in \mathbb{R}_{++}$, set $\Omega = [0, T]$, and let μ be the Lebesgue measure. Then we obtain the Hilbert space $L^2([0, T]; \mathsf{H})$, which is equipped with the scalar product $(x, y) \mapsto \int_0^T \langle x(t) \mid y(t)\rangle_{\mathsf{H}}\, dt$. In particular, when $\mathsf{H} = \mathbb{R}$, we obtain the classical Lebesgue space $L^2([0, T]) = L^2([0, T]; \mathbb{R})$.

Example 2.9 Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a *probability space*, i.e., a measure space such that $\mathsf{P}(\Omega) = 1$. A property that holds P-almost everywhere on Ω is said to hold *almost surely* (a.s.). A *random variable* (r.v.) is a measurable function $X \colon \Omega \to \mathbb{R}$, and its expected value is $\mathsf{E}X = \int_\Omega X(\omega)\mathsf{P}(d\omega)$, provided that the integral exists. In this context, Example 2.7 yields the Hilbert space

$$L^2(\Omega, \mathcal{F}, \mathsf{P}) = \big\{X \text{ r.v. on } (\Omega, \mathcal{F}, \mathsf{P}) \bigm| \mathsf{E}|X|^2 < +\infty\big\} \tag{2.9}$$

of random variables with finite second absolute moment, which is equipped with the scalar product $(X, Y) \mapsto \mathsf{E}(XY)$.

Example 2.10 Let $T \in \mathbb{R}_{++}$ and let $(\mathsf{H}, \langle \cdot \mid \cdot \rangle_{\mathsf{H}})$ be a separable real Hilbert space. For every $y \in L^2([0, T]; \mathsf{H})$, the function $x \colon [0, T] \to \mathsf{H} \colon t \mapsto \int_0^t y(s)ds$ is differentiable almost everywhere (a.e.) on $]0, T[$ with $x'(t) = y(t)$ a.e. on $]0, T[$. We say that $x \colon [0, T] \to \mathsf{H}$ belongs to $W^{1,2}([0, T]; \mathsf{H})$ if there exists $y \in L^2([0, T]; \mathsf{H})$ such that

$$(\forall t \in [0, T]) \quad x(t) = x(0) + \int_0^t y(s)ds. \tag{2.10}$$

Alternatively,

$$W^{1,2}([0, T]; \mathsf{H}) = \big\{x \in L^2([0, T]; \mathsf{H}) \bigm| x' \in L^2([0, T]; \mathsf{H})\big\}. \tag{2.11}$$

The scalar product of this real Hilbert space is $(x, y) \mapsto \int_0^T \langle x(t) \mid y(t)\rangle_{\mathsf{H}}\, dt + \int_0^T \langle x'(t) \mid y'(t)\rangle_{\mathsf{H}}\, dt$.

2.2 Basic Identities and Inequalities

Fact 2.11 (Cauchy–Schwarz) *Let x and y be in $\mathcal{H}$. Then*

$$|\langle x \mid y\rangle| \leqslant \|x\|\,\|y\|. \tag{2.12}$$

Moreover, $\langle x \mid y\rangle = \|x\|\,\|y\| \Leftrightarrow (\exists\, \alpha \in \mathbb{R}_+)\ x = \alpha y$ or $y = \alpha x$.

Lemma 2.12 *Let x, y, and z be in $\mathcal{H}$. Then the following hold:*

(i) $\|x+y\|^2 = \|x\|^2 + 2\langle x \mid y\rangle + \|y\|^2$.
(ii) *Parallelogram identity:* $\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$.
(iii) *Polarization identity:* $4\langle x \mid y\rangle = \|x+y\|^2 - \|x-y\|^2$.
(iv) *Apollonius's identity:* $\|x-y\|^2 = 2\|z-x\|^2 + 2\|z-y\|^2 - 4\|z-(x+y)/2\|^2$.

Proof. (i): A simple expansion.

(ii)&(iii): It follows from (i) that

$$\|x-y\|^2 = \|x\|^2 - 2\langle x \mid y\rangle + \|y\|^2. \tag{2.13}$$

Adding this identity to (i) yields (ii), and subtracting it from (i) yields (iii).

(iv): Apply (ii) to the points $(z-x)/2$ and $(z-y)/2$. $\square$

Lemma 2.13 *Let x and y be in $\mathcal{H}$. Then the following hold:*

(i) $\langle x \mid y\rangle \leqslant 0 \Leftrightarrow (\forall \alpha \in \mathbb{R}_+)\,\|x\| \leqslant \|x-\alpha y\| \Leftrightarrow (\forall \alpha \in [0,1])\,\|x\| \leqslant \|x-\alpha y\|$.
(ii) $x \perp y \Leftrightarrow (\forall \alpha \in \mathbb{R})\ \|x\| \leqslant \|x-\alpha y\| \Leftrightarrow (\forall \alpha \in [-1,1])\ \|x\| \leqslant \|x-\alpha y\|$.

Proof. (i): Observe that

$$(\forall \alpha \in \mathbb{R}) \quad \|x-\alpha y\|^2 - \|x\|^2 = \alpha\big(\alpha\|y\|^2 - 2\langle x \mid y\rangle\big). \tag{2.14}$$

Hence, the forward implications follow immediately. Conversely, if for every $\alpha \in \left]0,1\right]$, $\|x\| \leqslant \|x-\alpha y\|$, then (2.14) implies that $\langle x \mid y\rangle \leqslant \alpha\|y\|^2/2$. As $\alpha \downarrow 0$, we obtain $\langle x \mid y\rangle \leqslant 0$.

(ii): A consequence of (i), since $x \perp y \Leftrightarrow [\langle x \mid y\rangle \leqslant 0$ and $\langle x \mid -y\rangle \leqslant 0]$. $\square$

Lemma 2.14 *Let $(x_i)_{i\in I}$ and $(u_i)_{i\in I}$ be finite families in $\mathcal{H}$ and let $(\alpha_i)_{i\in I}$ be a family in $\mathbb{R}$ such that $\sum_{i\in I}\alpha_i = 1$. Then the following hold:*

(i) $\big\langle \sum_{i\in I}\alpha_i x_i \mid \sum_{j\in I}\alpha_j u_j\big\rangle + \sum_{i\in I}\sum_{j\in I}\alpha_i\alpha_j\langle x_i - x_j \mid u_i - u_j\rangle/2 = \sum_{i\in I}\alpha_i\langle x_i \mid u_i\rangle$.
(ii) $\big\|\sum_{i\in I}\alpha_i x_i\big\|^2 + \sum_{i\in I}\sum_{j\in I}\alpha_i\alpha_j\|x_i - x_j\|^2/2 = \sum_{i\in I}\alpha_i\|x_i\|^2$.

Proof. (i): We have

$$\begin{aligned}
2\Big\langle \sum_{i\in I}\alpha_i x_i \,\Big|\, \sum_{j\in I}\alpha_j u_j\Big\rangle
&= \sum_{i\in I}\sum_{j\in I}\alpha_i\alpha_j\big(\langle x_i \mid u_j\rangle + \langle x_j \mid u_i\rangle\big)\\
&= \sum_{i\in I}\sum_{j\in I}\alpha_i\alpha_j\big(\langle x_i \mid u_i\rangle + \langle x_j \mid u_j\rangle - \langle x_i - x_j \mid u_i - u_j\rangle\big)\\
&= 2\sum_{i\in I}\alpha_i\langle x_i \mid u_i\rangle - \sum_{i\in I}\sum_{j\in I}\alpha_i\alpha_j\langle x_i - x_j \mid u_i - u_j\rangle.
\end{aligned} \tag{2.15}$$

(ii): This follows from (i) when $(u_i)_{i\in I} = (x_i)_{i\in I}$. $\square$

The following two results imply that Hilbert spaces are uniformly convex and strictly convex Banach spaces, respectively.

Corollary 2.15 *Let $x \in \mathcal{H}$, let $y \in \mathcal{H}$, and let $\alpha \in \mathbb{R}$. Then*

$$\|\alpha x + (1-\alpha)y\|^2 + \alpha(1-\alpha)\|x-y\|^2 = \alpha\|x\|^2 + (1-\alpha)\|y\|^2. \tag{2.16}$$

Corollary 2.16 *Suppose that x and y are distinct points in $\mathcal{H}$ such that $\|x\| = \|y\|$, and let $\alpha \in \,]0,1[$. Then $\|\alpha x + (1-\alpha)y\| < \|x\|$.*

Proof. An immediate consequence of Corollary 2.15. ☐

Lemma 2.17 *Let $(x,y) \in \mathcal{H} \times \mathcal{H}$. Then the following hold:*

(i) *Let $\alpha \in \,]0,1[$. Then*

$$\begin{aligned}\alpha^2\big(\|x\|^2 - \|(1-\alpha^{-1})x + \alpha^{-1}y\|^2\big)&\\ &= (2\alpha-1)\|x\|^2 + 2(1-\alpha)\langle x \mid y\rangle - \|y\|^2\\ &= 2(1-\alpha)\langle x \mid y\rangle - \big(\|y\|^2 + (1-2\alpha)\|x\|^2\big)\\ &= \alpha\big(\|x\|^2 - \alpha^{-1}(1-\alpha)\|x-y\|^2 - \|y\|^2\big).\end{aligned}$$

(ii) *We have*

$$\begin{aligned}\|x\|^2 - \|2y-x\|^2 &= 4\big(\langle x \mid y\rangle - \|y\|^2\big)\\ &= 4\langle x-y \mid y\rangle\\ &= 2\big(\|x\|^2 - \|x-y\|^2 - \|y\|^2\big).\end{aligned}$$

Proof. (i): These identities follow from Lemma 2.12(i).
(ii): Divide by α^2 in (i) and set $\alpha = 1/2$. ☐

The following inequality is classical.

Fact 2.18 (Hardy–Littlewood–Pólya) (See [196, Theorems 368 and 369]) Let x and y be in $\mathbb{R}^N$, and let $x_\downarrow$ and $y_\downarrow$ be, respectively, their rearrangement vectors with entries ordered decreasingly. Then

$$\langle x \mid y\rangle \leqslant \langle x_\downarrow \mid y_\downarrow\rangle, \tag{2.17}$$

and equality holds if and only if there exists a permutation matrix P of size $N \times N$ such that $Px = x_\downarrow$ and $Py = y_\downarrow$.

2.3 Linear Operators and Functionals

Let $\mathcal{X}$ and $\mathcal{Y}$ be real normed vector spaces. We set

$$\mathcal{B}(\mathcal{X},\mathcal{Y}) = \big\{T\colon \mathcal{X} \to \mathcal{Y} \;\big|\; T \text{ is linear and continuous}\big\} \tag{2.18}$$

and $\mathcal{B}(\mathcal{X}) = \mathcal{B}(\mathcal{X}, \mathcal{X})$. Equipped with the norm

$$(\forall T \in \mathcal{B}(\mathcal{X}, \mathcal{Y})) \quad \|T\| = \sup \big\|T(B(0;1))\big\| = \sup_{\substack{x \in \mathcal{X},\\ \|x\| \leqslant 1}} \|Tx\|, \tag{2.19}$$

$\mathcal{B}(\mathcal{X}, \mathcal{Y})$ is a normed vector space, and it is a Banach space if $\mathcal{Y}$ is a Banach space.

Example 2.19 Let $A \in \mathbb{R}^{M \times N}$. Then $A \in \mathcal{B}(\mathbb{R}^N, \mathbb{R}^M)$ and the operator norm of A given by (2.19) is the spectral norm of A, i.e., the largest singular value of A, and it is denoted by $\|A\|_2$. We have $\|A\|_2 \leqslant \|A\|_{\mathsf{F}}$.

Fact 2.20 *Let $\mathcal{X}$ and $\mathcal{Y}$ be real normed vector spaces and let $T\colon \mathcal{X} \to \mathcal{Y}$ be linear. Then T is continuous at a point in $\mathcal{X}$ if and only if it is Lipschitz continuous.*

Fact 2.21 (See [116, Proposition III.6.1]) Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be real normed vector spaces and let $T\colon \mathcal{X} \oplus \mathcal{Y} \to \mathcal{Z}$ be a bilinear operator. Then T is continuous if and only if

$$(\exists\, \beta \in \mathbb{R}_+)(\forall x \in \mathcal{X})(\forall y \in \mathcal{Y}) \quad \|T(x,y)\| \leqslant \beta \|x\|\,\|y\|. \tag{2.20}$$

The following result is also known as the *Banach–Steinhaus theorem*.

Lemma 2.22 (Uniform boundedness principle) *Let $\mathcal{X}$ be a real Banach space, let $\mathcal{Y}$ be a real normed vector space, and let $(T_i)_{i \in I}$ be a family of operators in $\mathcal{B}(\mathcal{X}, \mathcal{Y})$ that is* pointwise bounded, *i.e.,* $(\forall x \in \mathcal{X})\ \sup_{i \in I} \|T_i x\| < +\infty$. *Then $(T_i)_{i \in I}$ is* uniformly bounded, *i.e.,* $\sup_{i \in I} \|T_i\| < +\infty$.

Proof. Apply Lemma 1.44(i) to $\big(\big\{x \in \mathcal{X} \;\big|\; \sup_{i \in I} \|T_i x\| \leqslant n\big\}\big)_{n \in \mathbb{N}}$. ☐

Definition 2.23 Let $T\colon \mathcal{H} \to \mathcal{H}$ be linear and let $\alpha \in \mathbb{R}_{++}$. Then T is:

(i) *monotone* if $(\forall x \in \mathcal{H})\ \langle Tx \mid x \rangle \geqslant 0$;
(ii) *strictly monotone* if $(\forall x \in \mathcal{H} \smallsetminus \{0\})\ \langle Tx \mid x \rangle > 0$;
(iii) *α-strongly monotone* if $(\forall x \in \mathcal{H})\ \langle Tx \mid x \rangle \geqslant \alpha \|x\|^2$.

The Riesz–Fréchet representation theorem states that every continuous linear functional on the real Hilbert space $\mathcal{H}$ can be identified with a vector in $\mathcal{H}$.

Fact 2.24 (Riesz–Fréchet representation) *Let $f \in \mathcal{B}(\mathcal{H}, \mathbb{R})$. Then there exists a unique vector $u \in \mathcal{H}$ such that $(\forall x \in \mathcal{H})\ f(x) = \langle x \mid u \rangle$. Moreover, $\|f\| = \|u\|$.*

If $\mathcal{K}$ is a real Hilbert space and $T \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, then the *adjoint* of T is the unique operator $T^* \in \mathcal{B}(\mathcal{K}, \mathcal{H})$ that satisfies

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{K}) \quad \langle Tx \mid y \rangle = \langle x \mid T^* y \rangle\,. \tag{2.21}$$

Fact 2.25 *Let $\mathcal{K}$ be a real Hilbert space, let $T \in \mathcal{B}(\mathcal{H},\mathcal{K})$, and let* $\ker T = \big\{x \in \mathcal{H} \;\big|\; Tx = 0\big\}$ *be the kernel of T. Then the following hold:*

(i) $T^{**} = T$.
(ii) $\|T^*\| = \|T\| = \sqrt{\|T^*T\|}$.
(iii) *Suppose that $\mathcal{K} = \mathcal{H}$ and that $T = T^*$. Then*

$$\|T\| = \sup\big\{|\langle Tx \mid x\rangle| \;\big|\; x \in B(0;1)\big\}. \tag{2.22}$$

(iv) $(\ker T)^{\perp} = \overline{\operatorname{ran}}\, T^*$.
(v) $(\operatorname{ran} T)^{\perp} = \ker T^*$.
(vi) $\ker T^*T = \ker T$ *and* $\overline{\operatorname{ran}}\, TT^* = \overline{\operatorname{ran}}\, T$.
(vii) $(\operatorname{gra} T)^{\perp} = \big\{(u,v) \in \mathcal{H} \oplus \mathcal{K} \;\big|\; u = -T^*v\big\}$.

Fact 2.26 *Let $\mathcal{K}$ be a real Hilbert space and let $T \in \mathcal{B}(\mathcal{H},\mathcal{K})$. Then* $\operatorname{ran} T$ *is closed* $\Leftrightarrow$ $\operatorname{ran} T^*$ *is closed* $\Leftrightarrow$ $\operatorname{ran} TT^*$ *is closed* $\Leftrightarrow$ $\operatorname{ran} T^*T$ *is closed* $\Leftrightarrow$ $(\exists\, \alpha \in \mathbb{R}_{++})(\forall x \in (\ker T)^{\perp})\ \|Tx\| \geqslant \alpha\|x\|$.

Suppose that $\mathcal{H} \neq \{0\}$. Let $f\colon \mathcal{H} \to \mathbb{R}$ be nonzero and linear, and let $\eta \in \mathbb{R}$. A *hyperplane* in $\mathcal{H}$ is a set of the form

$$\big\{x \in \mathcal{H} \;\big|\; f(x) = \eta\big\}, \tag{2.23}$$

and it is closed if and only if f is continuous; if it is not closed, it is dense in $\mathcal{H}$. Alternatively, let $u \in \mathcal{H} \smallsetminus \{0\}$. Then it follows from Fact 2.24 that a *closed hyperplane* in $\mathcal{H}$ is a set of the form

$$\big\{x \in \mathcal{H} \;\big|\; \langle x \mid u\rangle = \eta\big\}. \tag{2.24}$$

Moreover, a *closed half-space* with *outer normal* u is a set of the form

$$\big\{x \in \mathcal{H} \;\big|\; \langle x \mid u\rangle \leqslant \eta\big\}, \tag{2.25}$$

and an *open half-space* with *outer normal* u is a set of the form

$$\big\{x \in \mathcal{H} \;\big|\; \langle x \mid u\rangle < \eta\big\}. \tag{2.26}$$

The distance function to $C = \big\{x \in \mathcal{H} \;\big|\; \langle x \mid u\rangle = \eta\big\}$ is (see (1.47))

$$d_C\colon \mathcal{H} \to \mathbb{R}_+ : x \mapsto \frac{|\langle x \mid u\rangle - \eta|}{\|u\|}. \tag{2.27}$$

We conclude this section with an example of a discontinuous linear functional.

Example 2.27 Assume that $\mathcal{H}$ is infinite-dimensional and let H be a Hamel basis of $\mathcal{H}$, i.e., a maximally linearly independent subset. Then H is uncountable. Indeed, if $\mathcal{H} = \bigcup_{n\in\mathbb{N}} \operatorname{span}\{h_k\}_{0\leqslant k\leqslant n}$ for some Hamel basis

$H = \{h_n\}_{n\in\mathbb{N}}$, then Lemma 1.44(i) implies that some finite-dimensional linear subspace $\operatorname{span}\{h_k\}_{0\leqslant k\leqslant n}$ has nonempty interior, which is absurd. The Gram–Schmidt orthonormalization procedure thus guarantees the existence of an orthonormal set $B = \{e_n\}_{n\in\mathbb{N}}$ and an uncountable set $C = \{c_a\}_{a\in A}$ such that $B \cup C$ is a Hamel basis of $\mathcal{H}$. Thus, every point in $\mathcal{H}$ is a (finite) linear combination of elements in $B \cup C$ and, therefore, the function

$$f\colon \mathcal{H} \to \mathbb{R}\colon x = \sum_{n\in\mathbb{N}} \xi_n e_n + \sum_{a\in A} \eta_a c_a \mapsto \sum_{n\in\mathbb{N}} \xi_n \tag{2.28}$$

is well defined and linear. Now take $(\alpha_n)_{n\in\mathbb{N}} \in \ell^2(\mathbb{N}) \smallsetminus \ell^1(\mathbb{N})$ (e.g., $(\forall n \in \mathbb{N})$ $\alpha_n = 1/(n+1)$) and set

$$(\forall n \in \mathbb{N}) \quad x_n = \sum_{k=0}^{n} \alpha_k e_k. \tag{2.29}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges strongly to some point $z \in \mathcal{H}$ and $f(x_n) \to +\infty$. This proves that f is discontinuous at z and hence discontinuous everywhere by Fact 2.20. Now set $(\forall n \in \mathbb{N})$ $y_n = (x_n - f(x_n)e_0)/\max\{f(x_n), 1\}$. Then $(y_n)_{n\in\mathbb{N}}$ lies in $C = \big\{x \in \mathcal{H} \bigm| f(x) = 0\big\}$ and $y_n \to -e_0$. On the other hand, $-e_0 \notin C$, since $f(-e_0) = -1$. As a result, the hyperplane C is not closed. In fact, as will be proved in Example 8.42, C is dense in $\mathcal{H}$.

2.4 Strong and Weak Topologies

The metric topology of $(\mathcal{H}, d)$ is called the *strong topology* (or *norm topology*) of $\mathcal{H}$. Thus, a net $(x_a)_{a\in A}$ in $\mathcal{H}$ converges strongly to a point x if $\|x_a - x\| \to 0$; in symbols, $x_a \to x$. When used without modifiers, topological notions in $\mathcal{H}$ (closedness, openness, neighborhood, continuity, compactness, convergence, etc.) will always be understood with respect to the strong topology.

Fact 2.28 *Let U and V be closed linear subspaces of $\mathcal{H}$ such that V has finite dimension or finite codimension. Then $U + V$ is a closed linear subspace.*

In addition to the strong topology, a very important alternative topology can be introduced.

Definition 2.29 The family of all finite intersections of open half-spaces of $\mathcal{H}$ forms the base of the *weak topology* of $\mathcal{H}$; $\mathcal{H}^{\mathrm{weak}}$ denotes the resulting topological space.

Lemma 2.30 *$\mathcal{H}^{\mathrm{weak}}$ is a Hausdorff space.*

Proof. Suppose that x and y are distinct points in $\mathcal{H}$. Set $u = x - y$ and $w = (x+y)/2$. Then $\big\{z \in \mathcal{H} \bigm| \langle z - w \mid u\rangle > 0\big\}$ and $\big\{z \in \mathcal{H} \bigm| \langle z - w \mid u\rangle < 0\big\}$ are disjoint weak neighborhoods of x and y, respectively. $\square$

A subset of $\mathcal{H}$ is weakly open if it is a union of finite intersections of open half-spaces. If $\mathcal{H}$ is infinite-dimensional, nonempty intersections of finitely many open half-spaces are unbounded and, therefore, nonempty weakly open sets are unbounded. A net $(x_a)_{a\in A}$ in $\mathcal{H}$ *converges weakly* to a point $x \in \mathcal{H}$ if, for every $u \in \mathcal{H}$, $\langle x_a \mid u\rangle \to \langle x \mid u\rangle$; in symbols, $x_a \rightharpoonup x$. Moreover (see Section 1.7), a subset C of $\mathcal{H}$ is *weakly closed* if the weak limit of every weakly convergent net in C is also in C, and *weakly compact* if every net in C has a weak cluster point in C. Likewise (see Section 1.11), a subset C of $\mathcal{H}$ is *weakly sequentially closed* if the weak limit of every weakly convergent sequence in C is also in C, and *weakly sequentially compact* if every sequence in C has a weak sequential cluster point in C. Finally, let D be a nonempty subset of $\mathcal{H}$, let $\mathcal{K}$ be a real Hilbert space, let $T\colon D \to \mathcal{K}$, and let $f\colon \mathcal{H} \to [-\infty, +\infty]$. Then T is *weakly continuous* if it is continuous with respect to the weak topologies on $\mathcal{H}$ and $\mathcal{K}$, i.e., if, for every net $(x_a)_{a\in A}$ in D such that $x_a \rightharpoonup x \in D$, we have $Tx_a \rightharpoonup Tx$. Likewise, f is *weakly lower semicontinuous* at $x \in \mathcal{H}$ if, for every net $(x_a)_{a\in A}$ in $\mathcal{H}$ such that $x_a \rightharpoonup x$, we have $f(x) \leqslant \underline{\lim}\, f(x_a)$.

Remark 2.31 Strong and weak convergence of a net $(x_a)_{a\in A}$ in $\mathcal{H}$ to a point x in $\mathcal{H}$ can be interpreted in geometrical terms: $x_a \to x$ means that $d_{\{x\}}(x_a) \to 0$ whereas, by (2.27), $x_a \rightharpoonup x$ means that $d_C(x_a) \to 0$ for every closed hyperplane C containing x.

Example 2.32 Suppose that $\mathcal{H}$ is infinite-dimensional, let $(x_n)_{n\in\mathbb{N}}$ be an orthonormal sequence in $\mathcal{H}$, and let u be a point in $\mathcal{H}$. Bessel's inequality yields $\sum_{n\in\mathbb{N}} |\langle x_n \mid u\rangle|^2 \leqslant \|u\|^2$, hence $\langle x_n \mid u\rangle \to 0$. Thus $x_n \rightharpoonup 0$. However, $\|x_n\| \equiv 1$ and therefore $x_n \not\to 0$. Actually, $(x_n)_{n\in\mathbb{N}}$ has no Cauchy subsequence since, for any two distinct positive integers n and m, we have $\|x_n - x_m\|^2 = \|x_n\|^2 + \|x_m\|^2 = 2$. This also shows that the unit sphere $\big\{x \in \mathcal{H} \;\big|\; \|x\| = 1\big\}$ is closed but not weakly sequentially closed.

Suppose that $\mathcal{H}$ is infinite-dimensional. As seen in Example 2.32, an orthonormal sequence in $\mathcal{H}$ has no strongly convergent subsequence. Hence, it follows from Fact 1.39 that the closed unit ball of $\mathcal{H}$ is not compact. This property characterizes infinite-dimensional Hilbert spaces.

Fact 2.33 *The following are equivalent:*

(i) *$\mathcal{H}$ is finite-dimensional.*
(ii) *The closed unit ball $B(0;1)$ of $\mathcal{H}$ is compact.*
(iii) *The weak topology of $\mathcal{H}$ coincides with its strong topology.*
(iv) *The weak topology of $\mathcal{H}$ is metrizable.*

In striking contrast, compactness of closed balls always holds in the weak topology. This fundamental and deep result is known as the *Banach–Alaoglu–Bourbaki theorem.*

Fact 2.34 (Banach–Alaoglu–Bourbaki) *The closed unit ball $B(0;1)$ of $\mathcal{H}$ is weakly compact.*

Fact 2.35 (See [192, p. 181] and [2, Theorems 6.30&6.34]) The weak topology of the closed unit ball $B(0;1)$ of $\mathcal{H}$ is metrizable if and only if $\mathcal{H}$ is separable.

Lemma 2.36 *Let C be a subset of $\mathcal{H}$. Then C is weakly compact if and only if it is weakly closed and bounded.*

Proof. First, suppose that C is weakly compact. Then Lemma 1.12 and Lemma 2.30 assert that C is weakly closed. Now set $\mathcal{C} = \{\langle x \mid \cdot\rangle\}_{x\in C} \subset \mathcal{B}(\mathcal{H},\mathbb{R})$ and take $u \in \mathcal{H}$. Then $\langle \cdot \mid u\rangle$ is weakly continuous. By Lemma 1.20, $\{\langle x \mid u\rangle\}_{x\in C}$ is a compact subset of $\mathbb{R}$, and it is therefore bounded by Lemma 1.41. Hence, $\mathcal{C}$ is pointwise bounded, and Lemma 2.22 implies that $\sup_{x\in C} \|x\| < +\infty$, i.e., that C is bounded. Conversely, suppose that C is weakly closed and bounded, say $C \subset B(0;\rho)$ for some $\rho \in \mathbb{R}_{++}$. By Fact 2.34, $B(0;\rho)$ is weakly compact. Using Lemma 1.12 in $\mathcal{H}^{\mathrm{weak}}$, we deduce that C is weakly compact. ☐

The following important fact states that weak compactness and weak sequential compactness coincide.

Fact 2.37 (Eberlein–Šmulian) *Let C be a subset of $\mathcal{H}$. Then C is weakly compact if and only if it is weakly sequentially compact.*

Corollary 2.38 *Let C be a subset of $\mathcal{H}$. Then the following are equivalent:*

(i) *C is weakly compact.*
(ii) *C is weakly sequentially compact.*
(iii) *C is weakly closed and bounded.*

Proof. Combine Lemma 2.36 and Fact 2.37. ☐

Lemma 2.39 *Let C be a bounded subset of $\mathcal{H}$. Then C is weakly closed if and only if it is weakly sequentially closed.*

Proof. If C is weakly closed, it is weakly sequentially closed. Conversely, suppose that C is weakly sequentially closed. By assumption, there exists $\rho \in \mathbb{R}_{++}$ such that $C \subset B(0;\rho)$. Since $B(0;\rho)$ is weakly sequentially compact by Fact 2.34 and Fact 2.37, it follows from Lemma 2.30 and Lemma 1.34 that C is weakly sequentially compact. In turn, appealing once more to Fact 2.37, we obtain the weak compactness of C and therefore its weak closedness by applying Lemma 1.12 in $\mathcal{H}^{\mathrm{weak}}$. ☐

Remark 2.40 As will be seen in Example 3.33, weakly sequentially closed sets need not be weakly closed.

Lemma 2.41 *Let $\mathcal{K}$ be a real Hilbert space, and let $T\colon \mathcal{H} \to \mathcal{K}$ be a continuous affine operator. Then T is weakly continuous.*

Proof. Set $L\colon x \mapsto Tx - T0$, let $x \in \mathcal{H}$, let $y \in \mathcal{K}$, and let $(x_a)_{a\in A}$ be a net in $\mathcal{H}$ such that $x_a \rightharpoonup x$. Then $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$ and $\langle x_a \mid L^*y\rangle \to \langle x \mid L^*y\rangle$. Hence, $\langle Lx_a \mid y\rangle \to \langle Lx \mid y\rangle$, i.e., $Lx_a \rightharpoonup Lx$. We conclude that $Tx_a = T0 + Lx_a \rightharpoonup T0 + Lx = Tx$. ☐

Lemma 2.42 *The norm of $\mathcal{H}$ is weakly lower semicontinuous, i.e., for every net $(x_a)_{a\in A}$ in $\mathcal{H}$ and every x in $\mathcal{H}$, we have*

$$x_a \rightharpoonup x \quad \Rightarrow \quad \|x\| \leqslant \varliminf \|x_a\|. \tag{2.30}$$

Proof. Take a net $(x_a)_{a\in A}$ in $\mathcal{H}$ and a point x in $\mathcal{H}$ such that $x_a \rightharpoonup x$. Then, by Cauchy–Schwarz, $\|x\|^2 = \lim |\langle x_a \mid x\rangle| \leqslant \varliminf \|x_a\|\,\|x\|$. ☐

Lemma 2.43 *Let $\mathcal{G}$ and $\mathcal{K}$ be real Hilbert spaces and let $T\colon \mathcal{H}\oplus\mathcal{G} \to \mathcal{K}$ be a bilinear operator such that*

$$(\exists\, \beta \in \mathbb{R}_{++})(\forall x \in \mathcal{H})(\forall u \in \mathcal{G}) \quad \|T(x,u)\| \leqslant \beta\|x\|\,\|u\|. \tag{2.31}$$

Let $(x_a)_{a\in A}$ be a net in $\mathcal{H}$, let $(u_a)_{a\in A}$ be a net in $\mathcal{G}$, let $x \in \mathcal{H}$, and let $u \in \mathcal{G}$. Suppose that $(x_a)_{a\in A}$ is bounded, that $x_a \rightharpoonup x$, and that $u_a \to u$. Then $T(x_a,u_a) \rightharpoonup T(x,u)$.

Proof. Since $\sup_{a\in A}\|x_a\| < +\infty$ and $\|u_a - u\| \to 0$, we have $\|T(x_a, u_a - u)\| \leqslant \beta(\sup_{b\in A}\|x_b\|)\|u_a - u\| \to 0$. On the other hand, $T(\cdot,u) \in \mathcal{B}(\mathcal{H},\mathcal{K})$ by Fact 2.21. Thus, $T(\cdot,u)$ is weakly continuous by Lemma 2.41. In turn, $T(x_a - x, u) \rightharpoonup T(0,u) = 0$. Altogether $T(x_a,u_a) - T(x,u) = T(x_a, u_a - u) + T(x_a - x, u) \rightharpoonup 0$. ☐

Lemma 2.44 *Let $(x_a)_{a\in A}$ and $(u_a)_{a\in A}$ be nets in $\mathcal{H}$, and let x and u be points in $\mathcal{H}$. Suppose that $(x_a)_{a\in A}$ is bounded, that $x_a \rightharpoonup x$, and that $u_a \to u$. Then $\langle x_a \mid u_a\rangle \to \langle x \mid u\rangle$.*

Proof. Apply Lemma 2.43 to $\mathcal{G} = \mathcal{H}$, $\mathcal{K} = \mathbb{R}$, and $F = \langle \cdot \mid \cdot\rangle$. ☐

2.5 Weak Convergence of Sequences

Lemma 2.45 *Let $(x_n)_{n\in\mathbb{N}}$ be a bounded sequence in $\mathcal{H}$. Then $(x_n)_{n\in\mathbb{N}}$ possesses a weakly convergent subsequence.*

Proof. First, recall from Lemma 2.30 that $\mathcal{H}^{\text{weak}}$ is a Hausdorff space. Now set $\rho = \sup_{n\in\mathbb{N}}\|x_n\|$ and $C = B(0;\rho)$. Fact 2.34 and Fact 2.37 imply that C is weakly sequentially compact. Since $(x_n)_{n\in\mathbb{N}}$ lies in C, the claim follows from Definition 1.33. ☐

Lemma 2.46 *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly if and only if it is bounded and possesses at most one weak sequential cluster point.*

Proof. Suppose that $x_n \rightharpoonup x \in \mathcal{H}$. Then it follows from Lemma 2.30 and Fact 1.9 that x is the unique weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$. Moreover, for every $u \in \mathcal{H}$, $\langle x_n \mid u\rangle \to \langle x \mid u\rangle$ and therefore $\sup_{n\in\mathbb{N}} |\langle x_n \mid u\rangle| < +\infty$. Upon applying Lemma 2.22 to the sequence of continuous linear functionals $(\langle x_n \mid \cdot\rangle)_{n\in\mathbb{N}}$, we obtain the boundedness of $(\|x_n\|)_{n\in\mathbb{N}}$. Conversely, suppose that $(x_n)_{n\in\mathbb{N}}$ is bounded and possesses at most one weak sequential cluster point. Then Lemma 2.45 asserts that it possesses exactly one weak sequential cluster point. Moreover, it follows from Fact 2.34 and Fact 2.37 that $(x_n)_{n\in\mathbb{N}}$ lies in a weakly sequentially compact set. Therefore, appealing to Lemma 2.30, we apply Lemma 1.35 in $\mathcal{H}^{\text{weak}}$ to obtain the conclusion. ☐

Lemma 2.47 *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ and let C be a nonempty subset of $\mathcal{H}$. Suppose that, for every $x \in C$, $(\|x_n - x\|)_{n\in\mathbb{N}}$ converges and that every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ belongs to C. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in C.*

Proof. By assumption, $(x_n)_{n\in\mathbb{N}}$ is bounded. Therefore, in view of Lemma 2.46, it is enough to show that $(x_n)_{n\in\mathbb{N}}$ cannot have two distinct weak sequential cluster points in C. To this end, let x and y be weak sequential cluster points of $(x_n)_{n\in\mathbb{N}}$ in C, say $x_{k_n} \rightharpoonup x$ and $x_{l_n} \rightharpoonup y$. Since x and y belong to C, the sequences $(\|x_n - x\|)_{n\in\mathbb{N}}$ and $(\|x_n - y\|)_{n\in\mathbb{N}}$ converge. In turn, since

$$(\forall n \in \mathbb{N}) \quad 2\,\langle x_n \mid x - y\rangle = \|x_n - y\|^2 - \|x_n - x\|^2 + \|x\|^2 - \|y\|^2, \tag{2.32}$$

$(\langle x_n \mid x - y\rangle)_{n\in\mathbb{N}}$ converges as well, say $\langle x_n \mid x - y\rangle \to \ell$. Passing to the limit along $(x_{k_n})_{n\in\mathbb{N}}$ and along $(x_{l_n})_{n\in\mathbb{N}}$ yields, respectively, $\ell = \langle x \mid x - y\rangle = \langle y \mid x - y\rangle$. Therefore, $\|x - y\|^2 = 0$ and hence $x = y$. ☐

Proposition 2.48 *Suppose that $(y_n)_{n\in\mathbb{N}}$ is an orthonormal sequence in $\mathcal{H}$ and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that $x_n - y_n \to 0$. Then $x_n \rightharpoonup 0$.*

Proof. This follows from Example 2.32. ☐

The next result provides a partial converse to Proposition 2.48.

Proposition 2.49 *Suppose that $(x_n)_{n\in\mathbb{N}}$ is a sequence in $\mathcal{H}$ such that $x_n \rightharpoonup 0$ and $(\forall n \in \mathbb{N})$ $\|x_n\| = 1$. Then there exist an orthonormal sequence $(y_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ and a subsequence $(x_{k_n})_{n\in\mathbb{N}}$ of $(x_n)_{n\in\mathbb{N}}$ such that $x_{k_n} - y_n \to 0$.*

Proof. Let $(\varepsilon_n)_{n\in\mathbb{N}}$ be a sequence in $]0, 1/2[$ such that $\varepsilon_n \to 0$. Set $V = \overline{\operatorname{span}}\,\{x_n\}_{n\in\mathbb{N}}$ and let $(e_n)_{n\in\mathbb{N}}$ be an orthonormal basis of V. Let $l_0 \in \mathbb{N}$. Since $x_n \rightharpoonup 0$, there exists $k_0 \in \mathbb{N}$ such that $u_0 = \sum_{i=0}^{l_0} \langle x_{k_0} \mid e_i\rangle\, e_i \in B(0; \varepsilon_0)$. Because $(e_n)_{n\in\mathbb{N}}$ is an orthonormal basis of V, there exists $l_1 \in \mathbb{N}$ such that $l_1 > l_0$ and $w_0 = \sum_{i\geqslant l_1+1} \langle x_{k_0} \mid e_i\rangle\, e_i \in B(0; \varepsilon_0)$. We continue in this fashion and thus obtain inductively two strictly increasing sequences $(l_n)_{n\in\mathbb{N}}$ and $(k_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$ such that

$$(\forall n \in \mathbb{N}) \quad u_n = \sum_{i=0}^{l_n} \langle x_{k_n} \mid e_i\rangle\, e_i \in B(0; \varepsilon_n) \tag{2.33}$$

and

$$(\forall n \in \mathbb{N}) \quad w_n = \sum_{i \geqslant l_{n+1}+1} \langle x_{k_n} \mid e_i \rangle \, e_i \in B(0; \varepsilon_n). \tag{2.34}$$

Now set

$$(\forall n \in \mathbb{N}) \quad v_n = \sum_{i=l_n+1}^{l_{n+1}} \langle x_{k_n} \mid e_i \rangle \, e_i = x_{k_n} - u_n - w_n. \tag{2.35}$$

Then $(\forall n \in \mathbb{N})$ $1 = \|x_{k_n}\| \geqslant \|v_n\| \geqslant \|x_{k_n}\| - \|u_n\| - \|w_n\| \geqslant 1 - 2\varepsilon_n > 0$ and $\|x_{k_n} - v_n\| \leqslant \|u_n\| + \|w_n\| \leqslant 2\varepsilon_n$. Moreover, $(v_n)_{n\in\mathbb{N}}$ is an orthogonal sequence since $(e_n)_{n\in\mathbb{N}}$ is an orthonormal basis of V. Finally, set $(\forall n \in \mathbb{N})$ $y_n = v_n/\|v_n\|$. Then $(y_n)_{n\in\mathbb{N}}$ is an orthonormal sequence in $\mathcal{H}$ and $(\forall n \in \mathbb{N})$ $\|x_{k_n} - y_n\| \leqslant \|x_{k_n} - v_n\| + \|v_n - y_n\| = \|x_{k_n} - v_n\| + (1 - \|v_n\|) \leqslant 4\varepsilon_n \to 0$. ☐

Proposition 2.50 *Let $(e_i)_{i\in I}$ be a family in $\mathcal{H}$ such that $\overline{\operatorname{span}}\,\{e_i\}_{i\in I} = \mathcal{H}$, let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$, and let x be a point in $\mathcal{H}$. Then the following are equivalent:*

(i) $x_n \rightharpoonup x$.
(ii) $(x_n)_{n\in\mathbb{N}}$ *is bounded and* $(\forall i \in I)$ $\langle x_n \mid e_i \rangle \to \langle x \mid e_i \rangle$ *as* $n \to +\infty$.

Proof. (i)⇒(ii): Lemma 2.46.

(ii)⇒(i): Set $(y_n)_{n\in\mathbb{N}} = (x_n - x)_{n\in\mathbb{N}}$. Lemma 2.45 asserts that $(y_n)_{n\in\mathbb{N}}$ possesses a weak sequential cluster point y, say $y_{k_n} \rightharpoonup y$. In view of Lemma 2.46, it suffices to show that $y = 0$. For this purpose, fix $\varepsilon \in \mathbb{R}_{++}$. Then there exists a finite subset J of I such that $\|y - z\| \sup_{n\in\mathbb{N}} \|y_{k_n}\| \leqslant \varepsilon$, where $z = \sum_{j\in J} \langle y \mid e_j \rangle \, e_j$. Thus, by Cauchy–Schwarz,

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad |\langle y_{k_n} \mid y \rangle| &\leqslant |\langle y_{k_n} \mid y - z \rangle| + |\langle y_{k_n} \mid z \rangle| \\ &\leqslant \varepsilon + \sum_{j\in J} |\langle y \mid e_j \rangle| \, |\langle y_{k_n} \mid e_j \rangle| . \end{aligned}\tag{2.36}$$

Hence $\overline{\lim}\, |\langle y_{k_n} \mid y \rangle| \leqslant \varepsilon$. Letting $\varepsilon \downarrow 0$ yields $\|y\|^2 = \lim \langle y_{k_n} \mid y \rangle = 0$. ☐

Lemma 2.51 *Let $(x_n)_{n\in\mathbb{N}}$ and $(u_n)_{n\in\mathbb{N}}$ be sequences in $\mathcal{H}$, and let x and u be points in $\mathcal{H}$. Then the following hold:*

(i) $\big[x_n \rightharpoonup x$ *and* $\overline{\lim}\, \|x_n\| \leqslant \|x\|\big] \;\Leftrightarrow\; x_n \to x$.
(ii) *Suppose that $\mathcal{H}$ is finite-dimensional. Then* $x_n \rightharpoonup x \;\Leftrightarrow\; x_n \to x$.
(iii) *Suppose that $x_n \rightharpoonup x$ and $u_n \to u$. Then* $\langle x_n \mid u_n \rangle \to \langle x \mid u \rangle$.

Proof. (i): Suppose that $x_n \rightharpoonup x$ and that $\overline{\lim}\, \|x_n\| \leqslant \|x\|$. Then it follows from Lemma 2.42 that $\|x\| \leqslant \underline{\lim}\, \|x_n\| \leqslant \overline{\lim}\, \|x_n\| \leqslant \|x\|$, hence $\|x_n\| \to \|x\|$. In turn, $\|x_n - x\|^2 = \|x_n\|^2 - 2\langle x_n \mid x \rangle + \|x\|^2 \to 0$. Conversely, suppose that $x_n \to x$. Then $\|x_n\| \to \|x\|$ by continuity of the norm. On the other hand, $x_n \rightharpoonup x$ since for every $n \in \mathbb{N}$ and every $u \in \mathcal{H}$, the Cauchy–Schwarz inequality yields $0 \leqslant |\langle x_n - x \mid u \rangle| \leqslant \|x_n - x\| \, \|u\|$.

(ii): Set $\dim \mathcal{H} = m$ and let $(e_k)_{1 \leqslant k \leqslant m}$ be an orthonormal basis of $\mathcal{H}$. Now assume that $x_n \rightharpoonup x$. Then $\|x_n - x\|^2 = \sum_{k=1}^m |\langle x_n - x \mid e_k \rangle|^2 \to 0$.

(iii): Combine Lemma 2.44 and Lemma 2.46. ☐

The combination of Lemma 2.42 and Lemma 2.51(i) yields the following characterization of strong convergence.

Corollary 2.52 *Let $(x_n)_{n \in \mathbb{N}}$ be a sequence in $\mathcal{H}$ and let x be in $\mathcal{H}$. Then $x_n \to x \Leftrightarrow \big[x_n \rightharpoonup x \text{ and } \|x_n\| \to \|x\|\big]$.*

We conclude this section with a consequence of Ostrowski's theorem (Theorem 1.49).

Lemma 2.53 *Suppose that $\mathcal{H}$ is finite-dimensional and let $(x_n)_{n \in \mathbb{N}}$ be a bounded sequence in $\mathcal{H}$ such that $x_n - x_{n+1} \to 0$. Then the set of cluster points of $(x_n)_{n \in \mathbb{N}}$ is compact and connected.*

2.6 Differentiability

In this section, $\mathcal{K}$ is a real Banach space.

Definition 2.54 Let C be a nonempty subset of $\mathcal{H}$, let $T : C \to \mathcal{K}$, and suppose that $x \in C$ is such that $(\forall y \in \mathcal{H})(\exists\, \alpha \in \mathbb{R}_{++})\ [x, x + \alpha y] \subset C$. Then T is *Gâteaux differentiable* at x if there exists an operator $\mathsf{D}T(x) \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, called the *Gâteaux derivative* of T at x, such that

$$(\forall y \in \mathcal{H}) \quad \mathsf{D}T(x)y = \lim_{\alpha \downarrow 0} \frac{T(x + \alpha y) - T(x)}{\alpha}. \tag{2.37}$$

Moreover, T is Gâteaux differentiable on C if it is Gâteaux differentiable at every point in C. Higher-order Gâteaux derivatives are defined inductively. Thus, the *second Gâteaux derivative* of T at x is the operator $\mathsf{D}^2T(x) \in \mathcal{B}(\mathcal{H}, \mathcal{B}(\mathcal{H}, \mathcal{K}))$ that satisfies

$$(\forall y \in \mathcal{H}) \quad \mathsf{D}^2T(x)y = \lim_{\alpha \downarrow 0} \frac{\mathsf{D}T(x + \alpha y) - \mathsf{D}T(x)}{\alpha}. \tag{2.38}$$

The Gâteaux derivative $\mathsf{D}T(x)$ in Definition 2.54 is unique whenever it exists (Exercise 2.22). Moreover, since $\mathsf{D}T(x)$ is linear, for every $y \in \mathcal{H}$, we have $\mathsf{D}T(x)y = -\mathsf{D}T(x)(-y)$, and we can therefore replace (2.37) by

$$(\forall y \in \mathcal{H}) \quad \mathsf{D}T(x)y = \lim_{0 \neq \alpha \to 0} \frac{T(x + \alpha y) - T(x)}{\alpha}. \tag{2.39}$$

Remark 2.55 Let C be a subset of $\mathcal{H}$, let $f : C \to \mathbb{R}$, and suppose that f is Gâteaux differentiable at $x \in C$. Then, by Fact 2.24, there exists a unique vector $\nabla f(x) \in \mathcal{H}$ such that

$$(\forall y \in \mathcal{H}) \quad \mathsf{D}f(x)y = \langle y \mid \nabla f(x)\rangle . \tag{2.40}$$

We call $\nabla f(x)$ the *Gâteaux gradient* of f at x. If f is Gâteaux differentiable on C, the *gradient operator* is $\nabla f\colon C \to \mathcal{H}\colon x \mapsto \nabla f(x)$. Likewise, if f is twice Gâteaux differentiable at x, we can identify $\mathsf{D}^2 f(x)$ with an operator $\nabla^2 f(x) \in \mathcal{B}(\mathcal{H})$ in the sense that

$$(\forall y \in \mathcal{H})(\forall z \in \mathcal{H}) \quad (\mathsf{D}^2 f(x)y)z = \big\langle z \mid \nabla^2 f(x)y\big\rangle . \tag{2.41}$$

If the convergence in (2.39) is uniform with respect to y on bounded sets, then $x \in \operatorname{int} C$ and we obtain the following notion.

Definition 2.56 Let $x \in \mathcal{H}$, let $C \in \mathcal{V}(x)$, and let $T\colon C \to \mathcal{K}$. Then T is *Fréchet differentiable* at x if there exists an operator $\mathsf{D}T(x) \in \mathcal{B}(\mathcal{H},\mathcal{K})$, called the *Fréchet derivative* of T at x, such that

$$\lim_{0\neq\|y\|\to 0} \frac{\|T(x+y) - Tx - \mathsf{D}T(x)y\|}{\|y\|} = 0. \tag{2.42}$$

Moreover, T is Fréchet differentiable on C if it is Fréchet differentiable at every point in C. Higher-order Fréchet derivatives are defined inductively. Thus, the *second Fréchet derivative* of T at x is the operator $\mathsf{D}^2 T(x) \in \mathcal{B}\big(\mathcal{H}, \mathcal{B}(\mathcal{H},\mathcal{K})\big)$ that satisfies

$$\lim_{0\neq\|y\|\to 0} \frac{\|\mathsf{D}T(x+y) - \mathsf{D}Tx - \mathsf{D}^2 T(x)y\|}{\|y\|} = 0. \tag{2.43}$$

The *Fréchet gradient* of a function $f\colon C \to \mathbb{R}$ at $x \in C$ is defined as in Remark 2.55. Here are some examples.

Example 2.57 Let $L \in \mathcal{B}(\mathcal{H})$, let $u \in \mathcal{H}$, let $x \in \mathcal{H}$, and set $f\colon \mathcal{H} \to \mathbb{R}\colon y \mapsto \langle Ly \mid y\rangle - \langle y \mid u\rangle$. Then f is twice Fréchet differentiable on $\mathcal{H}$ with $\nabla f(x) = (L+L^*)x - u$ and $\nabla^2 f(x) = L + L^*$.

Proof. Take $y \in \mathcal{H}$. Since

$$\begin{aligned} f(x+y) - f(x) &= \langle Lx \mid y\rangle + \langle Ly \mid x\rangle + \langle Ly \mid y\rangle - \langle y \mid u\rangle \\ &= \langle y \mid (L+L^*)x\rangle - \langle y \mid u\rangle + \langle Ly \mid y\rangle , \end{aligned} \tag{2.44}$$

we have

$$\big| f(x+y) - f(x) - \langle y \mid (L+L^*)x - u\rangle \big| = |\langle Ly \mid y\rangle| \leqslant \|L\|\,\|y\|^2. \tag{2.45}$$

In view of (2.42), f is Fréchet differentiable at x with $\nabla f(x) = (L+L^*)x - u$. In turn, (2.43) yields $\nabla^2 f(x) = L + L^*$. ☐

Proposition 2.58 *Let $f\colon \mathcal{H} \to \mathbb{R}$ be Gâteaux differentiable, let $L \in \mathcal{B}(\mathcal{H})$, and suppose that $\nabla f = L$. Then $L = L^*$, $f\colon x \mapsto f(0) + (1/2)\langle Lx \mid x\rangle$, and f is twice Fréchet differentiable.*

Proof. Fix $x \in \mathcal{H}$ and set $\phi\colon \mathbb{R} \to \mathbb{R}\colon t \mapsto f(tx)$. Then $(\forall t \in \mathbb{R})\ \phi'(t) = \langle x \mid \nabla f(tx)\rangle = \langle x \mid L(tx)\rangle = t\,\langle x \mid Lx\rangle$. It follows that $f(x) - f(0) = \phi(1) - \phi(0) = \int_0^1 \phi'(t)\,dt = \int_0^1 t\,\langle Lx \mid x\rangle\,dt = (1/2)\,\langle Lx \mid x\rangle$. We deduce from Example 2.57 that f is twice Fréchet differentiable and that $L = \nabla f = (L + L^*)/2$. Hence, $L^* = L$. ☐

Example 2.59 Let $F\colon \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ be a symmetric bilinear form such that, for some $\beta \in \mathbb{R}_+$,

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad |F(x,y)| \leqslant \beta\|x\|\,\|y\|, \tag{2.46}$$

let $\ell \in \mathcal{B}(\mathcal{H}, \mathbb{R})$, let $x \in \mathcal{H}$, and set $f\colon \mathcal{H} \to \mathbb{R}\colon y \mapsto (1/2)F(y,y) - \ell(y)$. Then f is Fréchet differentiable on $\mathcal{H}$ with $\mathsf{D}f(x) = F(x,\cdot) - \ell$.

Proof. Take $y \in \mathcal{H}$. Then,

$$\begin{aligned} f(x+y) - f(x) &= \tfrac{1}{2}F(x+y, x+y) - \ell(x+y) - \tfrac{1}{2}F(x,x) + \ell(x) \\ &= \tfrac{1}{2}F(y,y) + F(x,y) - \ell(y). \end{aligned} \tag{2.47}$$

Consequently, (2.46) yields

$$2\big|f(x+y) - f(x) - \big(F(x,y) - \ell(y)\big)\big| = |F(y,y)| \leqslant \beta\|y\|^2, \tag{2.48}$$

and we infer from (2.42) and (2.46) that $\mathsf{D}f(x)y = F(x,y) - \ell(y)$. ☐

Example 2.60 Let $\mathcal{K}$ be a real Hilbert space, let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, let $r \in \mathcal{K}$, let $x \in \mathcal{H}$, and set $f\colon \mathcal{H} \to \mathbb{R}\colon y \mapsto \|Ly - r\|^2$. Then f is twice Fréchet differentiable on $\mathcal{H}$ with $\nabla f(x) = 2L^*(Lx - r)$ and $\nabla^2 f(x) = 2L^*L$.

Proof. Set $F\colon \mathcal{H} \times \mathcal{H} \to \mathbb{R}\colon (y,z) \mapsto (1/2)\,\langle L^*Ly \mid z\rangle$, $\ell\colon \mathcal{H} \to \mathbb{R}\colon y \mapsto \langle y \mid L^*r\rangle$, and $\alpha = (1/2)\|r\|^2$. Then $(\forall y \in \mathcal{H})\ f(y) = 2(F(y,y) - \ell(y) + \alpha)$. Hence we derive from Example 2.59 that $\nabla f(x) = 2L^*(Lx - r)$, and from (2.43) that $\nabla^2 f(x) = 2L^*L$. ☐

Lemma 2.61 *Let $x \in \mathcal{H}$, let $C \in \mathcal{V}(x)$, and let $T\colon C \to \mathcal{K}$. Suppose that T is Fréchet differentiable at x. Then the following hold:*

(i) *T is Gâteaux differentiable at x and the two derivatives coincide.*
(ii) *T is continuous at x.*

Proof. Denote the Fréchet derivative of T at x by L_x.

(i): Let $\alpha \in \mathbb{R}_{++}$ and $y \in \mathcal{H} \smallsetminus \{0\}$. Then

$$\left\| \frac{T(x+\alpha y) - Tx}{\alpha} - L_x y \right\| = \|y\|\,\frac{\|T(x+\alpha y) - Tx - L_x(\alpha y)\|}{\|\alpha y\|} \tag{2.49}$$

converges to 0 as $\alpha \downarrow 0$.

(ii): Fix $\varepsilon \in \mathbb{R}_{++}$. By (2.42), we can find $\delta \in]0, \varepsilon/(\varepsilon + \|L_x\|)]$ such that $(\forall y \in B(0;\delta))\ \|T(x+y) - Tx - L_x y\| \leqslant \varepsilon\|y\|$. Thus $(\forall y \in B(0;\delta))\ \|T(x+y) - Tx\| \leqslant \|T(x+y) - Tx - L_x y\| + \|L_x y\| \leqslant \varepsilon\|y\| + \|L_x\|\,\|y\| \leqslant \delta(\varepsilon + \|L_x\|) \leqslant \varepsilon$. It follows that T is continuous at x. ☐

Fact 2.62 (See [146, Proposition 5.1.8]) Let $T\colon \mathcal{H} \to \mathcal{K}$ and let $x \in \mathcal{H}$. Suppose that the Gâteaux derivative of T exists on a neighborhood of x and that $\mathsf{D}T$ is continuous at x. Then T is Fréchet differentiable at x.

Fact 2.63 (See [146, Theorem 5.1.11]) Let $x \in \mathcal{H}$, let U be a neighborhood of x, let $\mathcal{G}$ be a real Banach space, let $T\colon U \to \mathcal{G}$, let V be a neighborhood of Tx such that $T(U) \subset V$, and let $R\colon V \to K$. Suppose that T is Gâteaux differentiable at x and that R is Fréchet differentiable at Tx. Then $R \circ T$ is Gâteaux differentiable at x and $\mathsf{D}(R \circ T)(x) = (\mathsf{D}R(Tx)) \circ \mathsf{D}T(x)$. If T is Fréchet differentiable at x, then so is $R \circ T$.

Item (i) in the next result is known as the *descent lemma*.

Lemma 2.64 *Let U be a nonempty open convex subset of $\mathcal{H}$, let $\beta \in \mathbb{R}_{++}$, let $f\colon U \to \mathbb{R}$ be a Fréchet differentiable function such that ∇f is β-Lipschitz continuous on U, and let x and y be in U. Then the following hold:*

(i) $|f(y) - f(x) - \langle y - x \mid \nabla f(x)\rangle\,| \leqslant (\beta/2)\|y - x\|^2$.
(ii) $|\,\langle x - y \mid \nabla f(x) - \nabla f(y)\rangle\,| \leqslant \beta\|y - x\|^2$.

Proof. (i): Set $\phi\colon [0,1] \to \mathbb{R}\colon t \mapsto f(x + t(y - x))$. Then, by Cauchy–Schwarz,

$$\begin{aligned}
\big|f(y) - f(x) - \langle y - x \mid \nabla f(x)\rangle\,\big| \\
&= \left|\int_0^1 \phi'(t)\,dt - \langle y - x \mid \nabla f(x)\rangle\right| \\
&= \left|\int_0^1 \langle y - x \mid \nabla f(x + t(y-x)) - \nabla f(x)\rangle\,dt\right| \\
&\leqslant \int_0^1 \|y - x\|\beta\|t(y - x)\|\,dt \\
&= \frac{\beta}{2}\|y - x\|^2, \qquad (2.50)
\end{aligned}$$

as claimed.

(ii): This follows from the Cauchy–Schwarz inequality. ☐

Example 2.65 Suppose that $\mathcal{H} \neq \{0\}$ and let $f\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \|x\|$. Then $f = \sqrt{\|\cdot\|^2}$ and, since Example 2.60 asserts that $\|\cdot\|^2$ is Fréchet differentiable with gradient operator $\nabla\|\cdot\|^2 = 2\mathrm{Id}$, it follows from Fact 2.63 that f is Fréchet differentiable on $\mathcal{H} \smallsetminus \{0\}$ with $(\forall x \in \mathcal{H} \smallsetminus \{0\})\ \nabla f(x) = x/\|x\|$. On the other hand, f is not Gâteaux differentiable at $x = 0$ since, although the limit in (2.37) exists, it is not linear with respect to y: $(\forall y \in \mathcal{H})\ \lim_{\alpha \downarrow 0} (\|0 + \alpha y\| - \|0\|)/\alpha = \|y\|$.

Fact 2.66 (See [146, Proposition 5.1.22]) Let $x \in \mathcal{H}$, let U be a neighborhood of x, let $\mathcal{K}$ be a real Banach space, and let $T\colon U \to \mathcal{K}$. Suppose that T is twice Fréchet differentiable at x. Then $(\forall (y,z) \in \mathcal{H}\times\mathcal{H})\ (\mathsf{D}^2T(x)y)z = (\mathsf{D}^2T(x)z)y$.

Example 2.67 Let $x \in \mathcal{H}$, let U be a neighborhood of x, and let $f\colon U \to \mathbb{R}$. Suppose that f is twice Fréchet differentiable at x. Then, in view of Fact 2.66 and (2.41), $\nabla^2 f(x)$ is self-adjoint.

Exercises

Exercise 2.1 Let x and y be points in $\mathcal{H}$. Show that the following are equivalent:

(i) $\|y\|^2 + \|x-y\|^2 = \|x\|^2$.
(ii) $\|y\|^2 = \langle x \mid y\rangle$.
(iii) $\langle y \mid x-y\rangle = 0$.
(iv) $(\forall \alpha \in [-1,1])\ \|y\| \leqslant \|\alpha x + (1-\alpha)y\|$.
(v) $(\forall \alpha \in \mathbb{R})\ \|y\| \leqslant \|\alpha x + (1-\alpha)y\|$.
(vi) $\|2y - x\| = \|x\|$.

Exercise 2.2 Consider $\mathcal{X} = \mathbb{R}^2$ with the norms $\|\cdot\|_1\colon \mathcal{X} \to \mathbb{R}_+\colon (\xi_1,\xi_2) \mapsto |\xi_1| + |\xi_2|$ and $\|\cdot\|_\infty\colon \mathcal{X} \to \mathbb{R}_+\colon (\xi_1,\xi_2) \mapsto \max\{|\xi_1|,|\xi_2|\}$. Show that neither norm satisfies the parallelogram identity.

Exercise 2.3 Let x and y be points in $\mathcal{H}$, and let α and β be real numbers. Show that

$$\|\alpha x + \beta y\|^2 + \alpha\beta\|x-y\|^2 = \alpha(\alpha+\beta)\|x\|^2 + \beta(\alpha+\beta)\|y\|^2. \tag{2.51}$$

Exercise 2.4 Set

$$\varDelta\colon \mathcal{H}\times\mathcal{H} \to \mathbb{R}_+\colon (x,y) \mapsto \left\| \frac{x}{1+\|x\|} - \frac{y}{1+\|y\|} \right\|. \tag{2.52}$$

Show that $(\mathcal{H}, \varDelta)$ is a metric space.

Exercise 2.5 Define $\varDelta$ as in Exercise 2.4, let $(x_n)_{n\in\mathbb{N}}$ be a bounded sequence in $\mathcal{H}$, and let $x \in \mathcal{H}$. Show that $x_n \to x$ if and only if $\varDelta(x_n, x) \to 0$.

Exercise 2.6 Construct a monotone operator $T \in \mathcal{B}(\mathcal{H})$ such that T is not self-adjoint.

Exercise 2.7 Suppose that $\mathcal{H} \neq \{0\}$ and define on $\mathcal{H} \smallsetminus \{0\}$ a relation by $x \equiv y \Leftrightarrow x \in \mathbb{R}_{++}y$. Show that $\equiv$ is an equivalence relation. For every $x \in \mathcal{H}\smallsetminus\{0\}$, let $[x] = \big\{y \in \mathcal{H} \smallsetminus \{0\} \;\big|\; x \equiv y\big\}$ be the corresponding equivalence class. The quotient set $\operatorname{hzn}\mathcal{H} = \big\{[x] \;\big|\; x \in \mathcal{H} \smallsetminus \{0\}\big\}$ is the *horizon* of $\mathcal{H}$ and

$\operatorname{csm} \mathcal{H} = \mathcal{H} \cup \operatorname{hzn} \mathcal{H}$ is the *cosmic closure* of $\mathcal{H}$. Show that the function Δ of Exercise 2.4 extends to a distance on $\operatorname{csm} \mathcal{H}$ by defining

$$(\forall [x] \in \operatorname{hzn} \mathcal{H})(\forall y \in \mathcal{H}) \quad \Delta([x], y) = \Delta(y, [x]) = \left\| \frac{x}{\|x\|} - \frac{y}{1 + \|y\|} \right\| \tag{2.53}$$

and

$$\begin{aligned}(\forall [x] \in \operatorname{hzn} \mathcal{H})(\forall [y] \in \operatorname{hzn} \mathcal{H}) \quad \Delta([x], [y]) &= \Delta([y], [x]) \\ &= \left\| \frac{x}{\|x\|} - \frac{y}{\|y\|} \right\|. \end{aligned} \tag{2.54}$$

Exercise 2.8 Consider Exercise 2.7 and its notation. Let $(x_n)_{n \in \mathbb{N}}$ be a sequence in $\mathcal{H}$ and let $[x] \in \operatorname{hzn} \mathcal{H}$. Show that $\Delta(x_n, [x]) \to 0$ if and only if there exists a sequence $(\lambda_n)_{n \in \mathbb{N}}$ in $\mathbb{R}_{++}$ such that $\lambda_n \to 0$ and $\lambda_n x_n \to x$.

Exercise 2.9 Consider Exercise 2.7 and its notation. Let $([x_n])_{n \in \mathbb{N}}$ be a sequence in $\operatorname{hzn} \mathcal{H}$ and let $[x] \in \operatorname{hzn} \mathcal{H}$. Show that $\Delta([x_n], [x]) \to 0$ if and only if there exists a sequence $(\lambda_n)_{n \in \mathbb{N}}$ in $\mathbb{R}_{++}$ such that $\lambda_n x_n \to x$.

Exercise 2.10 Suppose that $\mathcal{H}$ is finite-dimensional and consider Exercise 2.7 and its notation. Show that $\operatorname{csm} \mathcal{H}$ is compact and sequentially compact with respect to the distance Δ.

Exercise 2.11 Let N be a strictly positive integer, set $I = \{1, \ldots, N\}$, and suppose that $(x_i)_{i \in I}$ are points in $\mathcal{H}$ such that $(\forall i \in I)\ \|x_i\| = 1$. Show the following:

(i) $\| \sum_{i \in I} x_i \|^2 = N + 2 \sum_{1 \leqslant i < j \leqslant N} \langle x_i \mid x_j \rangle$.
(ii) Suppose that, for every $(i, j) \in I \times I$ such that $i \neq j$ we have $\langle x_i \mid x_j \rangle = -1/(N-1)$. Then $\sum_{i \in I} x_i = 0$.
(iii) Suppose that $\sum_{i \in I} x_i = 0$. Then $2 \sum_{1 \leqslant i < j \leqslant N-1} \langle x_i \mid x_j \rangle = 2 - N$.
(iv) Suppose $N = 3$. Then $x_1 + x_2 + x_3 = 0$ if and only if $\langle x_1 \mid x_2 \rangle = \langle x_1 \mid x_3 \rangle = \langle x_2 \mid x_3 \rangle = -1/2$.

Exercise 2.12 Let x and y be points in $\mathcal{H}$, and let α and β be real numbers in $\mathbb{R}_+$. Show that $4 \langle \alpha x - \beta y \mid y - x \rangle \leqslant \alpha \|y\|^2 + \beta \|x\|^2$.

Exercise 2.13 Let x and y be in $\mathcal{H}$, and let α and β be in $\mathbb{R}$. Show that

$$\begin{aligned}\alpha(1-\alpha)\|\beta x + (1-\beta) y\|^2 + \beta(1-\beta)\|\alpha x - (1-\alpha) y\|^2 \\ = (\alpha + \beta - 2\alpha\beta)\big(\alpha\beta\|x\|^2 + (1-\alpha)(1-\beta)\|y\|^2\big). \end{aligned} \tag{2.55}$$

Exercise 2.14 Let x, y, and z be points in $\mathcal{H}$ such that $\|2x - y - z\| = \|2y - x - z\| = \|2z - x - y\|$. Show that $\|x - y\| = \|y - z\| = \|z - x\|$.

Exercise 2.15 Suppose that $\mathcal{H}$ is infinite-dimensional. Show that every weakly compact set has an empty weak interior.

Exercise 2.16 Provide an unbounded convergent net in $\mathbb{R}$ and compare with Lemma 2.46.

Exercise 2.17 Construct a sequence in $\mathcal{H}$ that converges weakly and possesses a strong sequential cluster point, but that does not converge strongly.

Exercise 2.18 Let C be a subset of $\mathcal{H}$ such that $(\forall n \in \mathbb{N})$ $C \cap B(0;n)$ is weakly sequentially closed. Show that C is weakly sequentially closed and compare with Lemma 1.40.

Exercise 2.19 Show that the conclusion of Lemma 2.51(iii) fails if the strong convergence of $(u_n)_{n\in\mathbb{N}}$ is replaced by weak convergence.

Exercise 2.20 (Opial's condition) Let $(x_n)_{n\in\mathbb{N}}$ be a weakly convergent sequence in $\mathcal{H}$ and let $x \in \mathcal{H}$. Show that $x_n \rightharpoonup x$ if and only if

$$(\forall y \in \mathcal{H}) \quad \underline{\lim}\, \|x_n - y\|^2 = \|x - y\|^2 + \underline{\lim}\, \|x_n - x\|^2. \tag{2.56}$$

In particular, if $x_n \rightharpoonup x$ and $y \in \mathcal{H} \smallsetminus \{x\}$, then $\underline{\lim}\, \|x_n - y\| > \underline{\lim}\, \|x_n - x\|$. This implication is known as *Opial's condition.*

Exercise 2.21 Suppose that $\mathcal{H}$ is infinite-dimensional and let $(e_n)_{n\in\mathbb{N}}$ be an orthonormal sequence in $\mathcal{H}$. Construct a bounded sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ such that $x_n - x_{n+1} \to 0$ and the set of strong cluster points of $(x_n)_{n\in\mathbb{N}}$ is $\{e_0, -e_0\}$. Compare to the Ostrowski results (Theorem 1.49 and Lemma 2.53).

Exercise 2.22 Show that if the derivative $\mathsf{D}T(x)$ exists in Definition 2.54, then it is unique.

Exercise 2.23 Let D be a nonempty open interval in $\mathbb{R}$, let $f\colon D \to \mathbb{R}$, and let $x \in D$. Show that the notions of Gâteaux and Fréchet differentiability of f at x coincide with classical differentiability, and that the Gâteaux and Fréchet derivatives coincide with the classical derivative

$$f'(x) = \lim_{0 \neq h \to 0} \frac{f(x+h) - f(x)}{h}. \tag{2.57}$$

Exercise 2.24 Consider the function

$$f\colon \mathbb{R}^2 \to \mathbb{R}\colon (\xi_1, \xi_2) \mapsto \begin{cases} \dfrac{\xi_1^2 \xi_2^4}{\xi_1^4 + \xi_2^8}, & \text{if } (\xi_1, \xi_2) \neq (0,0); \\ 0, & \text{if } (\xi_1, \xi_2) = (0,0). \end{cases} \tag{2.58}$$

Show that f is Gâteaux differentiable, but not continuous, at $(0,0)$.

Exercise 2.25 Consider the function

$$f\colon \mathbb{R}^2 \to \mathbb{R}\colon x = (\xi_1, \xi_2) \mapsto \begin{cases} \dfrac{\xi_1 \xi_2^4}{\xi_1^2 + \xi_2^4}, & \text{if } (\xi_1, \xi_2) \neq (0,0); \\ 0, & \text{if } (\xi_1, \xi_2) = (0,0). \end{cases} \tag{2.59}$$

Show that f is Fréchet differentiable at $(0,0)$ and that ∇f is not continuous at $(0,0)$. Conclude that the converse of Fact 2.62 does not hold.

Exercise 2.26 Consider the function

$$f\colon \mathbb{R}^2 \to \mathbb{R}\colon (\xi_1,\xi_2) \mapsto \begin{cases} \dfrac{\xi_1\xi_2^3}{\xi_1^2+\xi_2^4}, & \text{if } (\xi_1,\xi_2) \neq (0,0);\\ 0, & \text{if } (\xi_1,\xi_2) = (0,0). \end{cases} \tag{2.60}$$

Show that, at $(0,0)$, f is continuous and Gâteaux differentiable, but not Fréchet differentiable.

Exercise 2.27 Consider the function

$$f\colon \mathbb{R}^2 \to \mathbb{R}\colon (\xi_1,\xi_2) \mapsto \begin{cases} \dfrac{\xi_1\xi_2^2}{\xi_1^2+\xi_2^2}, & \text{if } (\xi_1,\xi_2) \neq (0,0);\\ 0, & \text{if } (\xi_1,\xi_2) = (0,0). \end{cases} \tag{2.61}$$

Show that f is continuous and that, at $(0,0)$, the limit on the right-hand side of (2.37) exists but it is not linear as a function of (η_1,η_2). Conclude that f is not Gâteaux differentiable at $(0,0)$.

Chapter 3
Convex Sets

In this chapter we introduce the fundamental notion of the convexity of a set and establish various properties. The key result is Theorem 3.16, which asserts that every nonempty closed convex subset C of $\mathcal{H}$ is a Chebyshev set, i.e., that every point in $\mathcal{H}$ possesses a unique best approximation from C, and which provides a characterization of this best approximation.

3.1 Basic Properties and Examples

Definition 3.1 A subset C of $\mathcal{H}$ is *convex* if $(\forall \alpha \in\,]0,1[)\ \alpha C + (1-\alpha)C = C$ or, equivalently, if

$$(\forall x \in C)(\forall y \in C) \quad]x,y[\, \subset C. \tag{3.1}$$

In particular, $\mathcal{H}$ and $\varnothing$ are convex.

Example 3.2 In each of the following cases, C is a convex subset of $\mathcal{H}$.

(i) C is a ball.
(ii) C is an affine subspace.
(iii) C is a half-space.
(iv) $C = \bigcap_{i \in I} C_i$, where $(C_i)_{i \in I}$ is a family of convex subsets of $\mathcal{H}$.

The intersection property (iv) of Example 3.2 justifies the following definition.

Definition 3.3 Let C be a subset of $\mathcal{H}$. The *convex hull* of C is the intersection of all the convex subsets of $\mathcal{H}$ containing C, i.e., the smallest convex subset of $\mathcal{H}$ containing C. It is denoted by $\operatorname{conv} C$. The *closed convex hull* of C is the smallest closed convex subset of $\mathcal{H}$ containing C. It is denoted by $\overline{\operatorname{conv}}\, C$.

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-48311-5_31

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_3

The proof of the following simple fact is left as Exercise 3.1.

Proposition 3.4 *Let C be a subset of $\mathcal{H}$ and let D be the set of all* convex combinations *of points in C, i.e.,*

$$D = \left\{ \sum_{i\in I} \alpha_i x_i \;\middle|\; I \text{ finite}, (x_i) \in C^I,\, (\alpha_i)_{i\in I} \in \left]0,1\right]^I,\, \sum_{i\in I} \alpha_i = 1 \right\}. \tag{3.2}$$

Then $D = \operatorname{conv} C$.

Proposition 3.5 *Let $\mathcal{K}$ be a real Hilbert space, let $T\colon \mathcal{H} \to \mathcal{K}$ be an affine operator, and let C and D be convex subsets of $\mathcal{H}$ and $\mathcal{K}$, respectively. Then $T(C)$ and $T^{-1}(D)$ are convex subsets of $\mathcal{K}$ and $\mathcal{H}$, respectively.*

Proof. It follows from (1.12) that $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ T(]x,y[) =]Tx,Ty[$. Now take two points in $T(C)$, say Tx and Ty, where x and y are in C. By convexity, $]x,y[\subset C$ and, therefore, $]Tx,Ty[= T(]x,y[) \subset T(C)$. Thus, $T(C)$ is convex. Finally, let x and y be two points in $T^{-1}(D)$. Then Tx and Ty are in D and, by convexity, $T(]x,y[) =]Tx,Ty[\subset D$. Therefore $]x,y[\subset T^{-1}(T(]x,y[)) \subset T^{-1}(D)$, which proves the convexity of $T^{-1}(D)$. ☐

Proposition 3.6 *Let $(C_i)_{i\in I}$ be a totally ordered finite family of m convex subsets of $\mathcal{H}$. Then the following hold:*

(i) $\bigtimes_{i\in I} C_i$ *is convex.*
(ii) $(\forall (\alpha_i)_{i\in I} \in \mathbb{R}^m)$ $\sum_{i\in I} \alpha_i C_i$ *is convex.*

Proof. (i): Straightforward.

(ii): This is a consequence of (i) and Proposition 3.5 since $\sum_{i\in I} \alpha_i C_i = L\big(\bigtimes_{i\in I} C_i\big)$, where $L\colon \mathcal{H}^m \to \mathcal{H}\colon (x_i)_{i\in I} \mapsto \sum_{i\in I} \alpha_i x_i$ is linear. ☐

Proposition 3.7 *Let C be a nonempty convex subset of $\mathcal{H}$, and let $\beta \in \mathbb{R}_+$. Suppose that $(\forall x \in C)$ $\|x\| = \beta$. Then C is a singleton.*

Proof. Clearly, $C = \{0\}$ if $\beta = 0$. Now assume that $\beta > 0$, and suppose to the contrary that C is not a singleton. Let x and y be distinct points in C, and set $z = (x+y)/2$. On the one hand, since C is convex, $\|z\| = \beta$. On the other hand, Corollary 2.16 yields $\|z\| < \|x\| = \beta$. Altogether, we reach a contradiction. ☐

3.2 Best Approximation Properties

Definition 3.8 Let C be a nonempty subset of $\mathcal{H}$, let $x \in \mathcal{H}$, and let $p \in C$. Then p is a *best approximation* to x from C (or a *projection* of x onto C) if $\|x - p\| = d_C(x)$. If every point in $\mathcal{H}$ has at least one projection onto C, then C is *proximinal*. If every point in $\mathcal{H}$ has exactly one projection onto C, then

C is a *Chebyshev set*. In this case, the *projector* (or *projection operator*) onto C is the operator, denoted by P_C, that maps every point in $\mathcal{H}$ to its unique projection onto C.

Our first example of a proximinal set concerns low rank approximations.

Example 3.9 (Eckart–Young theorem) Suppose that $\mathcal{H}$ is the matrix space of Example 2.4 and that $m = \min\{M, N\} \geqslant 2$. Let $q \in \{1, \ldots, m-1\}$, and set $C = \big\{A \in \mathcal{H} \;\big|\; \operatorname{rank} A \leqslant q\big\}$. Then C is a proximinal set but not a Chebyshev set. More precisely, let $A \in \mathcal{H}$ be such that $r = \operatorname{rank} A > q$, let

$$\sigma_1(A) \geqslant \cdots \geqslant \sigma_r(A) > \sigma_{r+1}(A) = \cdots = \sigma_m(A) = 0 \tag{3.3}$$

be the singular values of A, and let $A = U\Sigma V^{\mathsf{T}}$ be the singular value decomposition of A. Thus, $U \in \mathbb{R}^{M\times M}$ and $V \in \mathbb{R}^{N\times N}$ are orthogonal matrices, and $\Sigma = \operatorname{Diag}(\sigma_1(A), \ldots, \sigma_m(A)) \in \mathbb{R}^{M\times N}$ is the matrix with diagonal entries $(\sigma_1(A), \ldots, \sigma_m(A))$ and zero off-diagonal entries. Then

$$P = U\Sigma_q V^{\mathsf{T}}, \quad \text{where} \quad \Sigma_q = \operatorname{Diag}\big(\sigma_1(A), \ldots, \sigma_q(A), 0, \ldots, 0\big), \tag{3.4}$$

is a projection of A onto C, and it is unique if and only if $\sigma_{q+1}(A) \neq \sigma_q(A)$.

Proof. Let B be a matrix in C and set $X = U^{\mathsf{T}}BV = [\xi_{i,j}]_{\substack{1\leqslant i\leqslant M,\\ 1\leqslant j\leqslant N}}$. Then

$$\begin{aligned}\|A-B\|_{\mathsf{F}}^2 &= \|U(\Sigma - X)V^{\mathsf{T}}\|_{\mathsf{F}}^2\\ &= \operatorname{tra}\big((\Sigma - X)^{\mathsf{T}}(\Sigma - X)\big)\\ &= \sum_{i=1}^{r} |\sigma_i(A) - \xi_{i,i}|^2 + \sum_{i=r+1}^{m} |\xi_{i,i}|^2 + \sum_{\substack{1\leqslant i\leqslant M,\\ 1\leqslant j\leqslant N,\\ i\neq j}} |\xi_{i,j}|^2\end{aligned} \tag{3.5}$$

is minimal when the second and third sums are zero and the first sum is minimal. Hence $X = \operatorname{Diag}(\xi_{1,1}, \ldots, \xi_{r,r}, 0, \ldots, 0)$ and, since $\operatorname{rank} X = \operatorname{rank} B \leqslant q$, $(\xi_{i,i})_{1\leqslant i\leqslant r}$ has at most q nonzero components. Thus, in view of (3.3), the minimum of $\sum_{i=1}^r |\sigma_i(A) - \xi_{i,i}|^2$ over the set of vectors $(\xi_{i,i})_{1\leqslant i\leqslant r}$ with at most q nonzero components is attained when

$$(\forall i \in \{1, \ldots, r\}) \quad \xi_{i,i} = \begin{cases} \sigma_i(A), & \text{if } 1 \leqslant i \leqslant q;\\ 0, & \text{if } q+1 \leqslant i \leqslant r,\end{cases} \tag{3.6}$$

and this solution is unique unless $\sigma_{q+1}(A) = \sigma_q(A)$. ☐

Example 3.10 Let $\{e_i\}_{i\in I}$ be a finite orthonormal set in $\mathcal{H}$, let $V = \operatorname{span}\{e_i\}_{i\in I}$, and let $x \in \mathcal{H}$. Then V is a Chebyshev set,

$$P_V x = \sum_{i\in I} \langle x \mid e_i\rangle\, e_i, \quad \text{and} \quad d_V(x) = \sqrt{\|x\|^2 - \sum_{i\in I} |\langle x \mid e_i\rangle|^2}. \tag{3.7}$$

Proof. For every family $(\alpha_i)_{i\in I}$ in $\mathbb{R}$, we have

$$\begin{aligned}\Big\|x - \sum_{i\in I}\alpha_i e_i\Big\|^2 &= \|x\|^2 - 2\Big\langle x \;\Big|\; \sum_{i\in I}\alpha_i e_i\Big\rangle + \Big\|\sum_{i\in I}\alpha_i e_i\Big\|^2\\ &= \|x\|^2 - 2\sum_{i\in I}\alpha_i\,\langle x \mid e_i\rangle + \sum_{i\in I}|\alpha_i|^2\\ &= \|x\|^2 - \sum_{i\in I}|\langle x \mid e_i\rangle|^2 + \sum_{i\in I}|\alpha_i - \langle x \mid e_i\rangle|^2. \end{aligned} \tag{3.8}$$

Therefore, the function $(\alpha_i)_{i\in I} \mapsto \|x - \sum_{i\in I}\alpha_i e_i\|^2$ admits $(\langle x \mid e_i\rangle)_{i\in I}$ as its unique minimizer, and its minimum value is $\|x\|^2 - \sum_{i\in I}|\langle x \mid e_i\rangle|^2$. ☐

Remark 3.11 Let C be a nonempty subset of $\mathcal{H}$.

(i) Since $\overline{C} = \big\{x\in\mathcal{H} \;\big|\; d_C(x) = 0\big\}$, no point in $\overline{C}\smallsetminus C$ has a projection onto C. A proximinal set (in particular a Chebyshev set) must therefore be closed.

(ii) If C is a finite-dimensional linear subspace of $\mathcal{H}$, then it is a Chebyshev set and it is hence closed by (i). This follows from Example 3.10, since C possesses a finite orthonormal basis by Gram–Schmidt.

Proposition 3.12 *Suppose that $\mathcal{H}$ is finite-dimensional and let C be a Chebyshev subset of $\mathcal{H}$. Then P_C is continuous.*

Proof. Let $x\in\mathcal{H}$ and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that $x_n \to x$. By Example 1.48, d_C is continuous and thus

$$\|x_n - P_C x_n\| = d_C(x_n) \to d_C(x) = \|x - P_C x\|. \tag{3.9}$$

Hence, $(P_C x_n)_{n\in\mathbb{N}}$ is bounded. Let y be a cluster point of $(P_C x_n)_{n\in\mathbb{N}}$, say $P_C x_{k_n} \to y$. Then Remark 3.11(i) asserts that $y\in C$, and (3.9) implies that $\|x_{k_n} - P_C x_{k_n}\| \to \|x-y\| = d_C(x)$. It follows that $y = P_C x$ is the only cluster point of the bounded sequence $(P_C x_n)_{n\in\mathbb{N}}$. Therefore, $P_C x_n \to P_C x$. ☐

In connection with Remark 3.11(i), the next example shows that closedness is not sufficient to guarantee proximinality.

Example 3.13 Suppose that $\mathcal{H}$ is infinite-dimensional, let $(e_n)_{n\in\mathbb{N}}$ be an orthonormal sequence in $\mathcal{H}$, and let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence in $]1,+\infty[$ such that $\alpha_n \downarrow 1$. Set $(\forall n\in\mathbb{N})\ x_n = \alpha_n e_n$ and $C = \{x_n\}_{n\in\mathbb{N}}$. Then, for any two distinct points x_n and x_m in C, we have $\|x_n - x_m\|^2 = \|x_n\|^2 + \|x_m\|^2 > \|e_n\|^2 + \|e_m\|^2 = 2$. Therefore, every convergent sequence in C is eventually constant and C is thus closed. However, 0 has no projection onto C, since $(\forall n\in\mathbb{N})\ d_C(0) = 1 < \alpha_n = \|0 - x_n\|$.

Proposition 3.14 *Let C be a nonempty weakly closed subset of $\mathcal{H}$. Then C is proximinal.*

Proof. Suppose that $x \in \mathcal{H} \smallsetminus C$ and let $z \in C$. Set $D = C \cap B(x; \|x - z\|)$ and $f : \mathcal{H} \to \mathbb{R} : y \mapsto \|x - y\|$. It is enough to show that f has a minimizer over D, for it will be a projection of x onto C. Since C is weakly closed and $B(x; \|x - z\|)$ is weakly compact by Fact 2.34, it follows from Lemma 1.12 and Lemma 2.30 that D is weakly compact and, by construction, nonempty. Hence, since f is weakly lower semicontinuous by Lemma 2.42, applying Theorem 1.29 in $\mathcal{H}^{\text{weak}}$ yields the existence of a minimizer of f over D. ☐

Corollary 3.15 *Suppose that $\mathcal{H}$ is finite-dimensional. Let C be a nonempty subset of $\mathcal{H}$. Then C is proximinal if and only if it is closed.*

Proof. An immediate consequence of Remark 3.11(i), Proposition 3.14, and Fact 2.33. ☐

Every Chebyshev set is a proximinal set. However, a proximinal set is not necessarily a Chebyshev set: for instance, in $\mathcal{H} = \mathbb{R}$, the projections of 0 onto $C = \{-1, 1\}$ are 1 and -1. A fundamental result is that nonempty closed convex sets are Chebyshev sets.

Theorem 3.16 (Projection theorem) *Let C be a nonempty closed convex subset of $\mathcal{H}$. Then C is a Chebyshev set and, for every x and every p in $\mathcal{H}$,*

$$p = P_C x \quad \Leftrightarrow \quad \big[p \in C \quad \text{and} \quad (\forall y \in C) \quad \langle y - p \mid x - p \rangle \leqslant 0 \big]. \tag{3.10}$$

Proof. Let $x \in \mathcal{H}$. Then it is enough to show that x possesses a unique projection onto C, and that this projection is characterized by (3.10). By definition of d_C (see (1.47)), there exists a sequence $(y_n)_{n\in\mathbb{N}}$ in C such that $d_C(x) = \lim \|y_n - x\|$. Now take m and n in $\mathbb{N}$. Since C is convex, $(y_n + y_m)/2 \in C$ and therefore $\|x - (y_n + y_m)/2\| \geqslant d_C(x)$. It follows from Apollonius's identity (Lemma 2.12(iv)) that

$$\begin{aligned} \|y_n - y_m\|^2 &= 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4\|x - (y_n + y_m)/2\|^2 \\ &\leqslant 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4d_C^2(x). \end{aligned} \tag{3.11}$$

Letting m and n go to $+\infty$, we obtain that $(y_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. It therefore converges to some point $p \in C$, since C is complete as a closed subset of $\mathcal{H}$. The continuity of $\|\cdot - x\|$ then yields $\lim \|y_n - x\| = \|p - x\|$, hence $d_C(x) = \|p - x\|$. This shows the existence of p. Now suppose that $q \in C$ satisfies $d_C(x) = \|q - x\|$. Then $(p + q)/2 \in C$ and, by Apollonius's identity, $\|p - q\|^2 = 2\|p - x\|^2 + 2\|q - x\|^2 - 4\|x - (p + q)/2\|^2 = 4d_C^2(x) - 4\|x - (p+q)/2\|^2 \leqslant 0$. This implies that $p = q$ and shows uniqueness. Finally, for every $y \in C$ and $\alpha \in [0, 1]$, set $y_\alpha = \alpha y + (1 - \alpha)p$, which belongs to C by convexity. Lemma 2.13(i) yields

$$\begin{aligned} \|x - p\| = d_C(x) &\Leftrightarrow (\forall y \in C)(\forall \alpha \in [0, 1]) \ \ \|x - p\| \leqslant \|x - y_\alpha\| \\ &\Leftrightarrow (\forall y \in C)(\forall \alpha \in [0, 1]) \ \ \|x - p\| \leqslant \|x - p - \alpha(y - p)\| \\ &\Leftrightarrow (\forall y \in C) \ \ \langle y - p \mid x - p \rangle \leqslant 0, \end{aligned} \tag{3.12}$$

which establishes the characterization. ☐

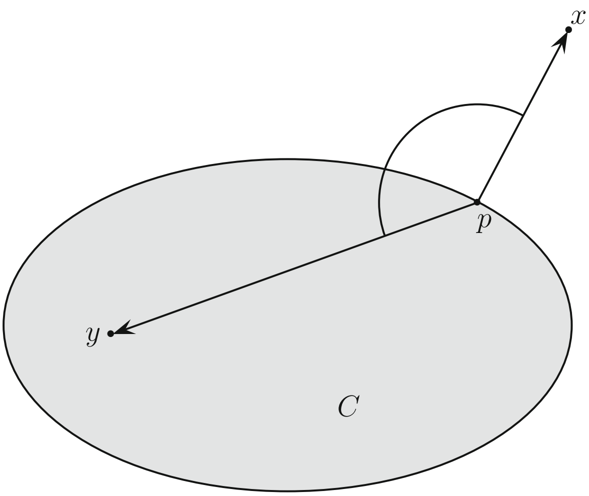

Fig. 3.1 Projection onto a nonempty closed convex set C in the Euclidean plane. The characterization (3.10) states that $p \in C$ is the projection of x onto C if and only if the vectors $x - p$ and $y - p$ form a right or obtuse angle for every $y \in C$.

Remark 3.17 Theorem 3.16 states that every nonempty closed convex set is a Chebyshev set. Conversely, as seen above, a Chebyshev set must be nonempty and closed. The famous *Chebyshev problem* asks whether every Chebyshev set must indeed be convex. The answer is affirmative if $\mathcal{H}$ is finite-dimensional (see Corollary 21.16), but is unknown otherwise. For a discussion, see [147].

The following example is obtained by checking (3.10) (further examples will be provided in Chapter 29).

Example 3.18 Let $\rho \in \mathbb{R}_{++}$ and set $C = B(0;\rho)$. Then

$$(\forall x \in \mathcal{H}) \quad P_C x = \frac{\rho}{\max\big\{\|x\|, \rho\big\}}\, x \quad \text{and} \quad d_C(x) = \max\big\{\|x\| - \rho, 0\big\}. \tag{3.13}$$

Proposition 3.19 *Let C be a nonempty closed convex subset of $\mathcal{H}$, and let x and y be in $\mathcal{H}$. Then $P_{y+C}x = y + P_C(x-y)$.*

Proof. It is clear that $y + P_C(x-y) \in y + C$. Using Theorem 3.16, we obtain

$$\begin{aligned}(\forall z \in C) \quad & \langle (y+z) - (y + P_C(x-y)) \mid x - (y + P_C(x-y))\rangle \\ &= \langle z - P_C(x-y) \mid (x-y) - P_C(x-y)\rangle \\ &\leqslant 0,\end{aligned} \tag{3.14}$$

and we conclude that $P_{y+C}x = y + P_C(x-y)$. $\square$

A decreasing sequence of nonempty closed convex sets can have an empty intersection (consider the sequence $([n,+\infty[)_{n\in\mathbb{N}}$ in $\mathcal{H}=\mathbb{R}$). Next, we show that the boundedness of the sets prevents such a situation.

Proposition 3.20 *Let $(C_n)_{n\in\mathbb{N}}$ be a sequence of nonempty bounded closed convex subsets of $\mathcal{H}$ such that $(\forall n\in\mathbb{N})\ C_{n+1}\subset C_n$. Then $\bigcap_{n\in\mathbb{N}} C_n \neq \varnothing$.*

Proof. Using Theorem 3.16, we define $(\forall n\in\mathbb{N})\ p_n = P_{C_n}0$. The assumptions imply that $(\|p_n\|)_{n\in\mathbb{N}}$ is increasing and bounded, hence convergent. For every m and n in $\mathbb{N}$ such that $m\leqslant n$, since $(p_n+p_m)/2\in C_m$, Lemma 2.12 yields $\|p_n-p_m\|^2 = 2(\|p_n\|^2+\|p_m\|^2)-4\|(p_n+p_m)/2\|^2 \leqslant 2(\|p_n\|^2-\|p_m\|^2)\to 0$ as $\min\{m,n\}\to+\infty$. Hence, $(p_n)_{n\in\mathbb{N}}$ is a Cauchy sequence and, in turn, $p_n\to p$ for some $p\in\mathcal{H}$. For every $n\in\mathbb{N}$, $(p_k)_{k\geqslant n}$ lies in C_n and hence $p\in C_n$ since C_n is closed. Therefore, $p\in\bigcap_{n\in\mathbb{N}} C_n$. ☐

Proposition 3.21 *Let C be a nonempty closed convex subset of $\mathcal{H}$ and let $x\in\mathcal{H}$. Then $(\forall\lambda\in\mathbb{R}_+)\ P_C(P_Cx+\lambda(x-P_Cx)) = P_Cx$.*

Proof. Let $\lambda\in\mathbb{R}_+$ and $y\in C$. We derive from Theorem 3.16 that $\langle y-P_Cx \mid (P_Cx+\lambda(x-P_Cx))-P_Cx\rangle = \lambda\,\langle y-P_Cx\mid x-P_Cx\rangle \leqslant 0$ and in turn that $P_C(P_Cx+\lambda(x-P_Cx)) = P_Cx$. ☐

Projectors onto affine subspaces have additional properties.

Corollary 3.22 *Let C be a closed affine subspace of $\mathcal{H}$. Then the following hold:*

(i) *Let x and p be in $\mathcal{H}$. Then $p=P_Cx$ if and only if*

$$p\in C \quad \textit{and} \quad (\forall y\in C)(\forall z\in C) \quad \langle y-z\mid x-p\rangle = 0. \tag{3.15}$$

(ii) *P_C is an affine operator.*

Proof. Let $x\in\mathcal{H}$.

(i): Let $y\in C$ and $z\in C$. By (1.2), $2P_Cx-y = 2P_Cx+(1-2)y\in C$ and (3.10) therefore yields $\langle y-P_Cx\mid x-P_Cx\rangle\leqslant 0$ and $\langle y-P_Cx\mid x-P_Cx\rangle = \langle(2P_Cx-y)-P_Cx\mid x-P_Cx\rangle\leqslant 0$. Altogether, $\langle y-P_Cx\mid x-P_Cx\rangle = 0$. Likewise, $\langle z-P_Cx\mid x-P_Cx\rangle = 0$. By subtraction, $\langle y-z\mid x-P_Cx\rangle = 0$. Conversely, it is clear that (3.15) implies the right-hand side of (3.10).

(ii): Let $y\in\mathcal{H}$ and $\alpha\in\mathbb{R}$, and set $z=\alpha x+(1-\alpha)y$ and $p=\alpha P_Cx+(1-\alpha)P_Cy$. We derive from (i) and (1.2) that $p\in C$. Now fix u and v in C. Then we also derive from (i) that $\langle u-v\mid z-p\rangle = \alpha\,\langle u-v\mid x-P_Cx\rangle + (1-\alpha)\,\langle u-v\mid y-P_Cy\rangle = 0$. Altogether, it follows from (i) that $p=P_Cz$. ☐

Example 3.23 Suppose that $u\in\mathcal{H}\smallsetminus\{0\}$, let $\eta\in\mathbb{R}$, and set $C=\big\{x\in\mathcal{H}\ \big|\ \langle x\mid u\rangle=\eta\big\}$. Then

$$(\forall x\in\mathcal{H}) \quad P_Cx = x+\frac{\eta-\langle x\mid u\rangle}{\|u\|^2}\,u, \tag{3.16}$$

and therefore $d_C(x) = |\langle x \mid u\rangle - \eta|/\|u\|$.

Proof. Check that (3.15) is satisfied to get (3.16). $\square$

Next, we collect basic facts about projections onto linear subspaces.

Corollary 3.24 *Let V be a closed linear subspace of $\mathcal{H}$ and let $x \in \mathcal{H}$. Then the following hold:*

(i) *Let $p \in \mathcal{H}$. Then $p = P_V x$ if and only if $(p, x-p) \in V \times V^{\perp}$.*
(ii) $\|P_V x\|^2 = \langle P_V x \mid x\rangle$.
(iii) $P_V \in \mathcal{B}(\mathcal{H})$, *$\|P_V\| = 1$ if $V \neq \{0\}$, and $\|P_V\| = 0$ if $V = \{0\}$.*
(iv) $V^{\perp\perp} = V$.
(v) $P_{V^\perp} = \mathrm{Id} - P_V$.
(vi) $P_V^* = P_V$.
(vii) $\|x\|^2 = \|P_V x\|^2 + \|P_{V^\perp} x\|^2 = d_V^2(x) + d_{V^\perp}^2(x)$.

Proof. (i): A special case of Corollary 3.22(i).

(ii): We deduce from (i) that $\langle x - P_V x \mid P_V x\rangle = 0$.

(iii): Let $y \in \mathcal{H}$ and $\alpha \in \mathbb{R}$, and set $z = \alpha x + y$ and $p = \alpha P_V x + P_V y$. Then (i) yields $p \in V$, $x - P_V x \in V^\perp$, and $y - P_V y \in V^\perp$; hence $z - p = \alpha(x - P_V x) + (y - P_V y) \in V^\perp$. Altogether (i) yields $p = P_V z$. This shows the linearity of P_V. The other assertions follow from (2.19), (ii), and Cauchy–Schwarz.

(iv): The inclusion $V \subset V^{\perp\perp}$ is clear. Conversely, suppose that $x \in V^{\perp\perp}$. It follows from (i) that $\langle P_V x \mid x - P_V x\rangle = 0$. Likewise, since $x \in V^{\perp\perp}$ and $x - P_V x \in V^\perp$ by (i), we have $\langle x \mid x - P_V x\rangle = 0$. Thus, $\|x - P_V x\|^2 = \langle x - P_V x \mid x - P_V x\rangle = 0$, i.e., $x = P_V x \in V$. We conclude that $V^{\perp\perp} \subset V$.

(v): By (i), $(\mathrm{Id} - P_V)x$ lies in the closed linear subspace $V^\perp$. On the other hand, $P_V x \in V \Rightarrow (\forall v \in V^\perp)\ \langle x - (\mathrm{Id} - P_V)x \mid v\rangle = \langle P_V x \mid v\rangle = 0 \Rightarrow x - (\mathrm{Id} - P_V)x \perp V^\perp$. Altogether, in view of (i), we conclude that $P_{V^\perp} x = (\mathrm{Id} - P_V)x$.

(vi): Take $y \in \mathcal{H}$. Then (v) yields $\langle P_V x \mid y\rangle = \langle P_V x \mid P_V y + P_{V^\perp} y\rangle = \langle P_V x \mid P_V y\rangle = \langle P_V x + P_{V^\perp} x \mid P_V y\rangle = \langle x \mid P_V y\rangle$.

(vii): By (i) and (v), $\|x\|^2 = \|(x - P_V x) + P_V x\|^2 = \|x - P_V x\|^2 + \|P_V x\|^2 = \|P_{V^\perp} x\|^2 + \|P_V x\|^2 = \|x - P_V x\|^2 + \|x - P_{V^\perp} x\|^2 = d_V^2(x) + d_{V^\perp}^2(x)$. $\square$

Proposition 3.25 *Let C be a nonempty subset of $\mathcal{H}$, let $V = \overline{\mathrm{span}}\, C$, and let*

$$\Pi_C \colon \mathcal{H} \to 2^C \colon x \mapsto \big\{ p \in C \;\big|\; \|x - p\| = d_C(x) \big\} \tag{3.17}$$

be the set-valued projector onto C. Then $\Pi_C = \Pi_C \circ P_V$. Consequently, C is a proximinal subset of $\mathcal{H}$ if and only if C is a proximinal subset of V.

Proof. Let $x \in \mathcal{H}$ and $p \in C$. In view of Corollary 3.24(vii)&(iii),

$$(\forall z \in C) \quad \|x - z\|^2 = \|P_V x - z\|^2 + \|P_{V^\perp} x\|^2. \tag{3.18}$$

Therefore $p \in \Pi_C x \Leftrightarrow (\forall z \in C)\ \|x-p\|^2 \leqslant \|x-z\|^2 \Leftrightarrow (\forall z \in C)\ \|P_V x-p\|^2 + \|P_{V^\perp} x\|^2 \leqslant \|P_V x - z\|^2 + \|P_{V^\perp} x\|^2 \Leftrightarrow (\forall z \in C)\ \|P_V x - p\|^2 \leqslant \|P_V x - z\|^2 \Leftrightarrow p \in \Pi_C(P_V x)$. ☐

An important application of Corollary 3.24 is the notion of least-squares solutions to linear equations.

Definition 3.26 Let $\mathcal{K}$ be a real Hilbert space, let $T \in \mathcal{B}(\mathcal{H},\mathcal{K})$, let $y \in \mathcal{K}$, and let $x \in \mathcal{H}$. Then x is a *least-squares solution* to the equation $Tz = y$ if

$$\|Tx - y\| = \min_{z \in \mathcal{H}} \|Tz - y\|. \tag{3.19}$$

Proposition 3.27 *Let $\mathcal{K}$ be a real Hilbert space, let $T \in \mathcal{B}(\mathcal{H},\mathcal{K})$ be such that* $\operatorname{ran} T$ *is closed, and let $y \in \mathcal{K}$. Then the equation $Tz = y$ has at least one least-squares solution. Moreover, for every $x \in \mathcal{H}$, the following are equivalent:*

(i) *x is a least-squares solution.*
(ii) $Tx = P_{\operatorname{ran} T}\, y$.
(iii) $T^*Tx = T^*y$ *(normal equation).*

Proof. (i)⇔(ii): Since $\operatorname{ran} T$ is a closed linear subspace, Theorem 3.16 asserts that $P_{\operatorname{ran} T}\, y$ is a well-defined point in $\operatorname{ran} T$. Now fix $x \in \mathcal{H}$. Then

$$\begin{aligned}(\forall z \in \mathcal{H})\ \ \|Tx - y\| \leqslant \|Tz - y\| &\Leftrightarrow (\forall r \in \operatorname{ran} T)\ \ \|Tx - y\| \leqslant \|r - y\| \\ &\Leftrightarrow Tx = P_{\operatorname{ran} T}\, y. \end{aligned} \tag{3.20}$$

Hence, the set of solutions to (3.19) is the nonempty set $T^{-1}(\{P_{\operatorname{ran} T}\, y\})$.

(ii)⇔(iii): We derive from Corollary 3.24(i) that

$$\begin{aligned} Tx = P_{\operatorname{ran} T}\, y &\Leftrightarrow (\forall r \in \operatorname{ran} T)\ \ \langle r \mid Tx - y\rangle = 0 \\ &\Leftrightarrow (\forall z \in \mathcal{H})\ \ \langle Tz \mid Tx - y\rangle = 0 \\ &\Leftrightarrow (\forall z \in \mathcal{H})\ \ \langle z \mid T^*(Tx - y)\rangle = 0 \\ &\Leftrightarrow T^*Tx = T^*y, \end{aligned} \tag{3.21}$$

which completes the proof. ☐

The set of least-squares solutions $\big\{x \in \mathcal{H} \;\big|\; T^*Tx = T^*y\big\}$ in Proposition 3.27 is a closed affine subspace. Hence, by Theorem 3.16, it possesses a unique minimal norm element, which will be denoted by $T^\dagger y$.

Definition 3.28 Let $\mathcal{K}$ be a real Hilbert space, let $T \in \mathcal{B}(\mathcal{H},\mathcal{K})$ be such that $\operatorname{ran} T$ is closed, and, for every $y \in \mathcal{K}$, set $C_y = \big\{x \in \mathcal{H} \;\big|\; T^*Tx = T^*y\big\}$. The *generalized* (or *Moore–Penrose*) *inverse* of T is $T^\dagger \colon \mathcal{K} \to \mathcal{H} \colon y \mapsto P_{C_y} 0$.

Example 3.29 Let $\mathcal{K}$ be a real Hilbert space and let $T \in \mathcal{B}(\mathcal{H},\mathcal{K})$ be such that T^*T is invertible. Then $T^\dagger = (T^*T)^{-1}T^*$.

Proposition 3.30 *Let $\mathcal{K}$ be a real Hilbert space and let $T \in \mathcal{B}(\mathcal{H},\mathcal{K})$ be such that* $\operatorname{ran} T$ *is closed. Then the following hold:*

(i) $(\forall y \in \mathcal{K})$ $\big\{x \in \mathcal{H} \bigm| T^*Tx = T^*y\big\} \cap (\ker T)^\perp = \{T^\dagger y\}$.
(ii) $P_{\operatorname{ran} T} = TT^\dagger$.
(iii) $P_{\ker T} = \mathrm{Id} - T^*T^{*\dagger}$.
(iv) $T^\dagger \in \mathcal{B}(\mathcal{K}, \mathcal{H})$.
(v) $\operatorname{ran} T^\dagger = \operatorname{ran} T^*$.
(vi) $P_{\operatorname{ran} T^\dagger} = T^\dagger T$.

Proof. (i): Fix $y \in \mathcal{K}$ and $z \in C_y$. Then it follows from Fact 2.25(vi) that $C_y = \big\{x \in \mathcal{H} \bigm| T^*Tx = T^*Tz\big\} = \big\{x \in \mathcal{H} \bigm| x - z \in \ker T^*T = \ker T\big\} = z + \ker T$. Hence, since C_y is a closed affine subspace, it follows from (3.10) and (3.15) that

$$\begin{aligned} z = T^\dagger y &\Leftrightarrow z = P_{C_y} 0 \\ &\Leftrightarrow (\forall x \in C_y) \quad \langle x - z \mid 0 - z\rangle = 0 \\ &\Leftrightarrow (\forall x \in z + \ker T) \quad \langle x - z \mid z\rangle = 0 \\ &\Leftrightarrow z \perp \ker T. \end{aligned} \tag{3.22}$$

(ii): Fix $y \in \mathcal{K}$. Since $T^\dagger y$ is a least-squares solution, Proposition 3.27(ii) yields $T(T^\dagger y) = P_{\operatorname{ran} T}\, y$.

(iii): It follows from Corollary 3.24(v), Fact 2.25(iv), Fact 2.26, and (ii) that $P_{\ker T} = \mathrm{Id} - P_{(\ker T)^\perp} = \mathrm{Id} - P_{\overline{\operatorname{ran}}\, T^*} = \mathrm{Id} - P_{\operatorname{ran} T^*} = \mathrm{Id} - T^*T^{*\dagger}$.

(iv): Fix y_1 and y_2 in $\mathcal{K}$, fix $\alpha \in \mathbb{R}$, and set $x = \alpha T^\dagger y_1 + T^\dagger y_2$. To establish the linearity of $T^\dagger$, we must show that $x = T^\dagger(\alpha y_1 + y_2)$, i.e., by (i), that $T^*Tx = T^*(\alpha y_1 + y_2)$ and that $x \perp \ker T$. Since, by (i), $T^*TT^\dagger y_1 = T^*y_1$ and $T^*TT^\dagger y_2 = T^*y_2$, it follows immediately from the linearity of T^*T that $T^*Tx = \alpha T^*TT^\dagger y_1 + T^*TT^\dagger y_2 = \alpha T^*y_1 + T^*y_2$. On the other hand, since (i) also implies that $T^\dagger y_1$ and $T^\dagger y_2$ lie in the linear subspace $(\ker T)^\perp$, so does their linear combination x. This shows the linearity of $T^\dagger$. It remains to show that $T^\dagger$ is bounded. It follows from (i) that $(\forall y \in \mathcal{K})$ $T^\dagger y \in (\ker T)^\perp$. Hence, Fact 2.26 asserts that there exists $\alpha \in \mathbb{R}_{++}$ such that $(\forall y \in \mathcal{K})$ $\|TT^\dagger y\| \geqslant \alpha\|T^\dagger y\|$. Consequently, we derive from (ii) and Corollary 3.24(iii) that

$$(\forall y \in \mathcal{K}) \quad \alpha\|T^\dagger y\| \leqslant \|TT^\dagger y\| = \|P_{\operatorname{ran} T}\, y\| \leqslant \|y\|, \tag{3.23}$$

which establishes the boundedness of $T^\dagger$.

(v): Fact 2.25(iv) and Fact 2.26 yield $(\ker T)^\perp = \overline{\operatorname{ran}}\, T^* = \operatorname{ran} T^*$. Hence (i) implies that $\operatorname{ran} T^\dagger \subset \operatorname{ran} T^*$. Conversely, take $x \in \operatorname{ran} T^*$ and set $y = Tx$. Then $T^*y = T^*Tx \in \operatorname{ran} T^* = (\ker T)^\perp$ and (i) yields $x = T^\dagger y \in \operatorname{ran} T^\dagger$.

(vi): By (ii), Fact 2.25(iv), Fact 2.26, and (v), $TT^\dagger T = P_{\operatorname{ran} T} T = T = T(P_{\ker T} + P_{(\ker T)^\perp}) = TP_{\operatorname{ran} T^*} = TP_{\operatorname{ran} T^\dagger}$. Hence $T(T^\dagger T - P_{\operatorname{ran} T^\dagger}) = 0$, i.e., $\operatorname{ran}(T^\dagger T - P_{\operatorname{ran} T^\dagger}) \subset \ker T$. On the other hand, $\operatorname{ran}(T^\dagger T - P_{\operatorname{ran} T^\dagger}) \subset \operatorname{ran} T^\dagger = \operatorname{ran} T^* = (\ker T)^\perp$. Altogether, $\operatorname{ran}(T^\dagger T - P_{\operatorname{ran} T^\dagger}) \subset \ker T \cap (\ker T)^\perp = \{0\}$, which implies that $T^\dagger T = P_{\operatorname{ran} T^\dagger}$. ☐

Proposition 3.31 (Moore–Desoer–Whalen) *Let $\mathcal{K}$ be a real Hilbert space, let $T \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ be such that* $\operatorname{ran} T$ *is closed, and let $\widetilde{T} \in \mathcal{B}(\mathcal{K}, \mathcal{H})$ be such that* $\operatorname{ran} \widetilde{T}$ *is closed. Then the following are equivalent:*

(i) $\widetilde{T} = T^\dagger$.
(ii) $T\widetilde{T} = P_{\operatorname{ran} T}$ *and* $\widetilde{T}T = P_{\operatorname{ran} \widetilde{T}}$.
(iii) $\widetilde{T}T|_{(\ker T)^\perp} = \mathrm{Id}$ *and* $\widetilde{T}|_{(\operatorname{ran} T)^\perp} = 0$.

Proof. (i)⇒(ii): See Proposition 3.30(ii)&(vi).

(ii)⇒(iii): Since $T = P_{\operatorname{ran} T}T = (T\widetilde{T})T$, we have $T(\mathrm{Id} - \widetilde{T}T) = 0$ and thus $\operatorname{ran}(\mathrm{Id} - \widetilde{T}T) \subset \ker T$. Hence, for every $x \in (\ker T)^\perp$, $\|x\|^2 \geqslant \|P_{\operatorname{ran} \widetilde{T}}x\|^2 = \|\widetilde{T}Tx\|^2 = \|\widetilde{T}Tx - x\|^2 + \|x\|^2$ and therefore $\widetilde{T}Tx = x$. Furthermore, $\widetilde{T} = P_{\operatorname{ran} \widetilde{T}}\widetilde{T} = (\widetilde{T}T)\widetilde{T} = \widetilde{T}(T\widetilde{T}) = \widetilde{T}P_{\operatorname{ran} T}$, which implies that $\widetilde{T}|_{(\operatorname{ran} T)^\perp} = 0$.

(iii)⇒(i): Take $y \in \mathcal{K}$, and set $y_1 = P_{\operatorname{ran} T}y$ and $y_2 = P_{(\operatorname{ran} T)^\perp}y$. Then there exists $x_1 \in (\ker T)^\perp$ such that $y_1 = Tx_1$. Hence

$$\widetilde{T}y = \widetilde{T}y_1 + \widetilde{T}y_2 = \widetilde{T}y_1 = \widetilde{T}Tx_1 = x_1. \tag{3.24}$$

It follows that $T\widetilde{T}y = Tx_1 = y_1 = P_{\operatorname{ran} T}y$ and Proposition 3.27 yields $T^*T\widetilde{T}y = T^*y$. Since (3.24) asserts that $\widetilde{T}y = x_1 \in (\ker T)^\perp$, we deduce from Proposition 3.30(i) that $\widetilde{T}y = T^\dagger y$. ☐

Corollary 3.32 *Let $\mathcal{K}$ be a real Hilbert space and let $T \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ be such that* $\operatorname{ran} T$ *is closed. Then the following hold:*

(i) $T^{\dagger\dagger} = T$.
(ii) $T^{\dagger *} = T^{*\dagger}$.
(iii) $T^*T^{*\dagger} = T^\dagger T$.
(iv) $P_{\ker T} = \mathrm{Id} - T^\dagger T$.

Proof. (i): Proposition 3.30(v) and Fact 2.26 assert that $\operatorname{ran} T^* = \operatorname{ran} T^\dagger$ is closed. Hence, it follows from Proposition 3.31 that

$$TT^\dagger = P_{\operatorname{ran} T} \quad \text{and} \quad T^\dagger T = P_{\operatorname{ran} T^\dagger}. \tag{3.25}$$

Combining Proposition 3.31 applied to $T^\dagger$ with (3.25) yields $T = T^{\dagger\dagger}$.

(ii): It follows from (3.25), Corollary 3.24(vi), and Proposition 3.30(v) that $T^{\dagger *}T^* = P^*_{\operatorname{ran} T} = P_{\operatorname{ran} T} = P_{\operatorname{ran}(T^{**})} = P_{\operatorname{ran}(T^{*\dagger})}$ and that $T^*T^{\dagger *} = P_{\operatorname{ran} T^\dagger} = P_{\operatorname{ran} T^*}$. Therefore, by Proposition 3.31, $T^{*\dagger} = T^{\dagger *}$.

(iii): As seen in the proof of (i), $\operatorname{ran} T^\dagger$ is closed, and it follows from (ii), Proposition 3.30(vi), and Corollary 3.24(vi) that $T^*T^{*\dagger} = T^*T^{\dagger *} = (T^\dagger T)^* = P^*_{\operatorname{ran} T^\dagger} = P_{\operatorname{ran} T^\dagger} = T^\dagger T$.

(iv): This follows from Proposition 3.30(iii) and (iii). ☐

3.3 Topological Properties

Since a Hilbert space is a metric space, the notions of closedness and sequential closedness coincide for the strong topology (see Section 1.12). The following example illustrates the fact that more care is required for the weak topology.

Example 3.33 Suppose that $\mathcal{H}$ is infinite-dimensional, let $(e_n)_{n\in\mathbb{N}}$ be an orthonormal sequence in $\mathcal{H}$, and set $C = \{\alpha_n e_n\}_{n\in\mathbb{N}}$, where $(\alpha_n)_{n\in\mathbb{N}}$ is a sequence in $[1, +\infty[$ such that $\sum_{n\in\mathbb{N}} 1/\alpha_n^2 = +\infty$ and $\alpha_n \uparrow +\infty$ (e.g., $(\forall n \in \mathbb{N})$ $\alpha_n = \sqrt{n+1}$). Then C is closed and weakly sequentially closed, but not weakly closed. In fact, 0 belongs to the weak closure of C but not to C.

Proof. Since the distance between two distinct points in C is at least $\sqrt{2}$, every Cauchy sequence in C is eventually constant and C is therefore closed. It follows from Lemma 2.46 that every weakly convergent sequence in C is bounded and hence, since $\alpha_n \uparrow +\infty$, eventually constant. Thus, C is weakly sequentially closed. Now let V be a weak neighborhood of 0. Then there exist $\varepsilon \in \mathbb{R}_{++}$ and a finite family $(u_i)_{i\in I}$ in $\mathcal{H}$ such that

$$U = \big\{x \in \mathcal{H} \;\big|\; (\forall i \in I)\ |\langle x \mid u_i\rangle| < \varepsilon\big\} \subset V. \tag{3.26}$$

Set $(\forall n \in \mathbb{N})$ $\zeta_n = \sum_{i\in I} |\langle u_i \mid e_n\rangle|$. We have $(\forall i \in I)$ $(|\langle u_i \mid e_n\rangle|)_{n\in\mathbb{N}} \in \ell^2(\mathbb{N})$. Consequently, the sequence $(\zeta_n)_{n\in\mathbb{N}}$ belongs to the vector space $\ell^2(\mathbb{N})$. Hence $\sum_{n\in\mathbb{N}} \zeta_n^2 < +\infty = \sum_{n\in\mathbb{N}} 1/\alpha_n^2$, and the set $\mathbb{M} = \big\{n \in \mathbb{N} \;\big|\; \zeta_n < \varepsilon/\alpha_n\big\}$ therefore contains infinitely many elements. Since $(\forall i \in I)(\forall n \in \mathbb{M})$ $|\langle u_i \mid \alpha_n e_n\rangle| = \alpha_n |\langle u_i \mid e_n\rangle| \leqslant \alpha_n\zeta_n < \varepsilon$, we deduce that

$$\{\alpha_n e_n\}_{n\in\mathbb{M}} \subset U \subset V. \tag{3.27}$$

Thus, every weak neighborhood of 0 contains elements from C. Therefore, 0 belongs to the weak closure of C. Finally, it is clear that $0 \notin C$. ☐

Next, we show that the distinctions illustrated in Example 3.33 among the various types of closure disappear for convex sets.

Theorem 3.34 *Let C be a convex subset of $\mathcal{H}$. Then the following are equivalent:*

(i) *C is weakly sequentially closed.*
(ii) *C is sequentially closed.*
(iii) *C is closed.*
(iv) *C is weakly closed.*

Proof. Suppose that C is nonempty (otherwise the conclusion is clear).
(i)⇒(ii): This follows from Corollary 2.52.
(ii)⇔(iii): Since $\mathcal{H}$ is a metric space, see Section 1.12.

(iii)⇒(iv): Take a net $(x_a)_{a\in A}$ in C that converges weakly to some point $x \in \mathcal{H}$. Then (3.10) yields

$$(\forall a \in A) \quad \langle x_a - P_C x \mid x - P_C x\rangle \leqslant 0. \tag{3.28}$$

Since $x_a \rightharpoonup x$, passing to the limit in (3.28) yields $\|x - P_C x\|^2 \leqslant 0$, hence $x = P_C x \in C$.

(iv)⇒(i): Clear. □

Corollary 3.35 *Let C be a nonempty closed convex subset of $\mathcal{H}$, and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in C that converges weakly to $x \in \mathcal{H}$. Then $x \in C$.*

Corollary 3.36 (Mazur's lemma) *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ that converges weakly to $x \in \mathcal{H}$. Then there exists a sequence $(y_n)_{n\in\mathbb{N}}$ of convex combinations of the elements of $(x_n)_{n\in\mathbb{N}}$ which converges strongly to x. More precisely, we have $y_n = \sum_{k=0}^{n} \alpha_{k,n} x_k \to x$, where $(\forall n \in \mathbb{N})$ $(\alpha_{k,n})_{0\leqslant k\leqslant n} \in [0,1]^{n+1}$ and $\sum_{k=0}^{n} \alpha_{k,n} = 1$.*

Proof. Let $C = \operatorname{conv}\{x_n\}_{n\in\mathbb{N}}$. Then $\overline{\operatorname{conv}}\, C$ is closed and convex and therefore weakly sequentially closed and sequentially closed by Theorem 3.34. Hence $x \in \overline{\operatorname{conv}}\, C$ and there exists a sequence $(z_n)_{n\in\mathbb{N}}$ in $\operatorname{conv} C$ such that $z_n \to x$, i.e., there exists a sequence $(\mathbb{K}_n)_{n\in\mathbb{N}}$ of finite ordered subsets of $\mathbb{N}$ such that $z_n = \sum_{k\in\mathbb{K}_n} \gamma_{k,n} x_k \to x$, where $(\forall n \in \mathbb{N})$ $(\gamma_{k,n})_{k\in\mathbb{K}_n} \in [0,1]^{\mathbb{K}_n}$ and $\sum_{k\in\mathbb{K}_n} \gamma_{k,n} = 1$. To show that $(z_n)_{n\in\mathbb{N}}$ can be written as claimed, let us set $(\forall n \in \mathbb{N})$ $C_n = \operatorname{conv}\{x_0, \ldots, x_n\}$ and $k_n = \max \mathbb{K}_n$. Then $(C_n)_{n\in\mathbb{N}}$ is an increasing sequence of sets and therefore $(d_{C_n}(x))_{n\in\mathbb{N}}$ is decreasing. On the other hand, $(\forall n \in \mathbb{N})$ $z_n \in C_{k_n}$. Hence $d_{C_{k_n}}(x) \leqslant \|z_n - x\| \to 0$ and therefore $d_{C_n}(x) \to 0$. Now let $\varepsilon \in \mathbb{R}_{++}$. Then there exists $N \in \mathbb{N}$ such that $(\forall n \in \{N, N+1, \ldots\})$ $d_{C_n}(x) \leqslant \varepsilon/2$. Hence $(\forall n \in \{N, N+1, \ldots\})$ $(\exists\, y_n \in C_n)$ $\|y_n - x\| \leqslant \varepsilon$. We conclude that $y_n \to x$. □

Theorem 3.37 *Let C be a bounded closed convex subset of $\mathcal{H}$. Then C is weakly compact and weakly sequentially compact.*

Proof. Theorem 3.34 asserts that C is weakly closed. In turn, it follows from Lemma 2.36 that C is weakly compact. Therefore, by Fact 2.37, C is weakly sequentially compact. □

Corollary 3.38 *Let C be a nonempty bounded closed convex subset of $\mathcal{H}$, and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in C. Then $(x_n)_{n\in\mathbb{N}}$ possesses a weakly convergent subsequence, and every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ lies in C.*

Proposition 3.39 *Let m be a strictly positive integer, set $I = \{1, \ldots, m\}$, let $(C_i)_{i\in I}$ be a family of convex subsets of $\mathcal{H}$, and set $C = \operatorname{conv} \bigcup_{i\in I} C_i$. Then the following hold:*

(i) *Suppose that, for every $i \in I$, C_i is compact. Then C is compact.*

(ii) *Suppose that, for every $i \in I$, C_i is weakly compact. Then C is weakly compact.*

Proof. The simplex $S = \big\{(\eta_1,\ldots,\eta_m) \in \mathbb{R}^m_+ \;\big|\; \sum_{i\in I}\eta_i = 1\big\}$ is compact. Set $\boldsymbol{\mathcal{H}} = \mathcal{H}^m$, $\boldsymbol{\mathcal{K}} = \mathbb{R}^m$, $D = C_1 \times \cdots \times C_m$, and define $T\colon \boldsymbol{\mathcal{H}} \oplus \boldsymbol{\mathcal{K}} \to \mathcal{H}\colon (x_1,\ldots,x_m,\eta_1,\ldots,\eta_m) \mapsto \sum_{i\in I}\eta_i x_i$. Then T is bilinear,

$$(\forall \boldsymbol{x} \in \boldsymbol{\mathcal{H}})(\forall \boldsymbol{y} \in \boldsymbol{\mathcal{K}}) \quad \|T(\boldsymbol{x},\boldsymbol{y})\| \leqslant \|\boldsymbol{x}\|\,\|\boldsymbol{y}\|, \tag{3.29}$$

hence continuous by Fact 2.21.

(i): Since D and S are compact, so is $D \times S$ and, by continuity of T, it follows from Lemma 1.20 that $C = T(D \times S)$ is compact.

(ii): Since D is weakly compact and S is compact, $D \times S$ is compact in $\boldsymbol{\mathcal{H}}^{\mathrm{weak}} \times \boldsymbol{\mathcal{K}}^{\mathrm{strong}}$. Now let $T(\boldsymbol{x}_a,\boldsymbol{y}_a)_{a\in A}$ be a net in $C = T(D \times S)$. By Fact 1.11, there exist a subnet $(\boldsymbol{x}_{k(b)},\boldsymbol{y}_{k(b)})_{b\in B}$ of $(\boldsymbol{x}_a,\boldsymbol{y}_a)_{a\in A}$ and a point $(\boldsymbol{x},\boldsymbol{y}) \in D \times S$ such that $\boldsymbol{x}_a \rightharpoonup \boldsymbol{x}$ and $\boldsymbol{y}_a \to \boldsymbol{y}$. In turn, we derive from the boundedness of D, (3.29), and Lemma 2.43 that $T(\boldsymbol{x}_{k(b)},\boldsymbol{y}_{k(b)}) \rightharpoonup T(\boldsymbol{x},\boldsymbol{y}) \in C$. Thus, C is weakly compact by Fact 1.11. ☐

As the next two examples show, the sum of two closed convex sets is not necessarily closed, even in Euclidean spaces and even in the case of closed linear subspaces in infinite-dimensional spaces.

Example 3.40 Suppose that $\mathcal{H} = \mathbb{R}^2$, let $C = \mathbb{R} \times \{0\}$, and let $D = \big\{(\xi_1,\xi_2) \in \mathbb{R}^2_{++} \;\big|\; \xi_1\xi_2 \geqslant 1\big\}$. Then C and D are closed and convex but $C + D = \mathbb{R} \times \mathbb{R}_{++}$ is open.

Example 3.41 Suppose that $\mathcal{H}$ is infinite-dimensional and let $(e_n)_{n\in\mathbb{N}}$ be an orthonormal sequence in $\mathcal{H}$. Set

$$C = \overline{\operatorname{span}}\,\{e_{2n}\}_{n\in\mathbb{N}} \;\text{ and }\; D = \overline{\operatorname{span}}\,\{\cos(\theta_n)e_{2n} + \sin(\theta_n)e_{2n+1}\}_{n\in\mathbb{N}}, \tag{3.30}$$

where $(\theta_n)_{n\in\mathbb{N}}$ is a sequence in $]0,\pi/2]$ such that $\sum_{n\in\mathbb{N}}\sin^2(\theta_n) < +\infty$. Then $C \cap D = \{0\}$ and $C + D$ is a linear subspace of $\mathcal{H}$ that is not closed.

Proof. It follows from (3.30) that the elements of C and D are of the form $\sum_{n\in\mathbb{N}}\gamma_n e_{2n}$ and $\sum_{n\in\mathbb{N}}\delta_n(\cos(\theta_n)e_{2n} + \sin(\theta_n)e_{2n+1})$, respectively, where $(\gamma_n)_{n\in\mathbb{N}} \in \ell^2(\mathbb{N})$ and $(\delta_n)_{n\in\mathbb{N}} \in \ell^2(\mathbb{N})$. It is straightforward that $C \cap D = \{0\}$. Now let

$$x = \sum_{n\in\mathbb{N}}\sin(\theta_n)e_{2n+1} \tag{3.31}$$

and observe that $x \in \overline{C + D}$. Assume that $x \in C + D$. Then there exist sequences $(\gamma_n)_{n\in\mathbb{N}}$ and $(\delta_n)_{n\in\mathbb{N}}$ in $\ell^2(\mathbb{N})$ such that

$$(\forall n \in \mathbb{N}) \quad 0 = \gamma_n + \delta_n\cos(\theta_n) \;\text{ and }\; \sin(\theta_n) = \delta_n\sin(\theta_n). \tag{3.32}$$

Thus $\delta_n \equiv 1$ and $\gamma_n = -\cos(\theta_n) \to -1$, which is impossible since $(\gamma_n)_{n\in\mathbb{N}} \in \ell^2(\mathbb{N})$. ☐

Proposition 3.42 *Let C and D be nonempty closed convex subsets of $\mathcal{H}$ such that D is bounded. Then $C + D$ is nonempty, closed, and convex.*

Proof. The convexity of $C+D$ follows from Proposition 3.6(ii). To show that $C + D$ is closed, take a convergent sequence in $C + D$, say $x_n + y_n \to z$, where $(x_n)_{n\in\mathbb{N}}$ lies in C, $(y_n)_{n\in\mathbb{N}}$ lies in D, and $z \in \mathcal{H}$. It then follows from Corollary 3.38 that there exists a subsequence $(y_{k_n})_{n\in\mathbb{N}}$ converging weakly to a point $y \in D$. Therefore $x_{k_n} \rightharpoonup z - y$ and, in view of Corollary 3.35, we have $z - y \in C$. We conclude that $z \in C + D$. ☐

Remark 3.43 Suppose that $\mathcal{H}$ is infinite-dimensional, let C be a convex subset of $\mathcal{H}$, and let D be a bounded closed convex subset of $\mathcal{H}$ such that $C + D$ is closed. Then C is not necessarily closed. For instance, let C be a hyperplane which is dense in $\mathcal{H}$ (see Example 8.42(v)), and let D be a closed ball. Then $C + D = \mathcal{H}$ is closed but C is not.

Proposition 3.44 *Let C be a convex subset of $\mathcal{H}$. Then*

$$(\forall x \in \operatorname{int} C)(\forall y \in \overline{C}) \quad [x, y[\; \subset \operatorname{int} C. \tag{3.33}$$

Proof. Suppose that $x \in \operatorname{int} C$ and that $y \in \overline{C}$. If $x = y$, the conclusion is trivial. Now assume that $x \neq y$ and fix $z \in [x, y[$, say $z = \alpha x+(1-\alpha)y$, where $\alpha \in \,]0, 1]$. Since $x \in \operatorname{int} C$, there exists $\varepsilon \in \mathbb{R}_{++}$ such that $B(x; \varepsilon(2 - \alpha)/\alpha) \subset C$. On the other hand, since $y \in \overline{C}$, we have $y \in C + B(0; \varepsilon)$. Therefore, by convexity,

$$\begin{aligned} B(z; \varepsilon) &= \alpha x + (1 - \alpha)y + B(0; \varepsilon)\\ &\subset \alpha x + (1 - \alpha)\big(C + B(0; \varepsilon)\big) + B(0; \varepsilon)\\ &= \alpha B(x; \varepsilon(2 - \alpha)/\alpha) + (1 - \alpha)C\\ &\subset \alpha C + (1 - \alpha)C\\ &= C. \end{aligned} \tag{3.34}$$

Hence $z \in \operatorname{int} C$. ☐

Proposition 3.45 *Let C be a convex subset of $\mathcal{H}$. Then the following hold:*

(i) $\overline{C}$ *is convex.*
(ii) $\operatorname{int} C$ *is convex.*
(iii) *Suppose that* $\operatorname{int} C \neq \varnothing$. *Then* $\operatorname{int} C = \operatorname{int} \overline{C}$ *and* $\overline{C} = \overline{\operatorname{int} C}$.

Proof. (i): Take x and y in $\overline{C}$, and $\alpha \in \,]0, 1[$. Then there exist sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ in C such that $x_n \to x$ and $y_n \to y$. By convexity, $C \ni \alpha x_n + (1 - \alpha)y_n \to \alpha x + (1 - \alpha)y$ and, therefore, $\alpha x + (1 - \alpha)y \in \overline{C}$.

(ii): Take x and y in $\operatorname{int} C$. Then, since $y \in \overline{C}$, Proposition 3.44 implies that $]x, y[\; \subset [x, y[\; \subset \operatorname{int} C$.

(iii): It is clear that $\operatorname{int} C \subset \operatorname{int} \overline{C}$. Conversely, let $y \in \operatorname{int} \overline{C}$. Then we can find $\varepsilon \in \mathbb{R}_{++}$ such that $B(y; \varepsilon) \subset \overline{C}$. Now take $x \in \operatorname{int} C$ and $\alpha \in \mathbb{R}_{++}$

such that $x \neq y$ and $y + \alpha(y - x) \in B(y; \varepsilon)$. Then Proposition 3.44 implies that $y \in [x, y + \alpha(y - x)[\subset \operatorname{int} C$ and we deduce that $\operatorname{int} \overline{C} \subset \operatorname{int} C$. Thus, $\operatorname{int} C = \operatorname{int} \overline{C}$. It is also clear that $\overline{\operatorname{int} C} \subset \overline{C}$. Now take $x \in \operatorname{int} C$, $y \in \overline{C}$, and define $(\forall \alpha \in]0, 1])$ $y_\alpha = \alpha x + (1 - \alpha)y$. Proposition 3.44 implies that $(y_\alpha)_{\alpha \in]0,1]}$ lies in $[x, y[\subset \operatorname{int} C$. Hence $y = \lim_{\alpha \downarrow 0} y_\alpha \in \overline{\operatorname{int} C}$. Therefore $\overline{C} \subset \overline{\operatorname{int} C}$, and we conclude that $\overline{C} = \overline{\operatorname{int} C}$. ☐

Proposition 3.46 *Let C be a subset of $\mathcal{H}$. Then* $\operatorname{conv} \overline{C} \subset \overline{\operatorname{conv} C} = \overline{\operatorname{conv}}\, C$.

Proof. Since $C \subset \operatorname{conv} C$, we have $\overline{C} \subset \overline{\operatorname{conv} C}$ and hence Proposition 3.45(i) yields $\operatorname{conv} \overline{C} \subset \operatorname{conv} \overline{\operatorname{conv} C} = \overline{\operatorname{conv} C}$. Furthermore, since $C \subset \overline{\operatorname{conv}}\, C$, we have $\operatorname{conv} C \subset \operatorname{conv} \overline{\operatorname{conv}}\, C = \overline{\operatorname{conv}}\, C$ and therefore $\overline{\operatorname{conv} C} \subset \overline{\overline{\operatorname{conv}}\, C} = \overline{\operatorname{conv}}\, C$. However, since $\overline{\operatorname{conv} C}$ is a closed convex set containing C, we have $\overline{\operatorname{conv}}\, C \subset \overline{\operatorname{conv} C}$. ☐

The following example shows that in general $\operatorname{conv} \overline{C} \neq \overline{\operatorname{conv} C}$.

Example 3.47 Suppose that $\mathcal{H} = \mathbb{R}^2$ and set $C = \operatorname{epi}(1 + |\cdot|)^{-1}$. Then $C = \overline{C}$, $\operatorname{conv} \overline{C} = \mathbb{R} \times \mathbb{R}_{++}$, and $\overline{\operatorname{conv} C} = \mathbb{R} \times \mathbb{R}_+$.

Proposition 3.48 *Let C and D be convex subsets of $\mathcal{H}$ such that C is closed and $C \cap \operatorname{int} D \neq \varnothing$. Then $\overline{C \cap \operatorname{int} D} = C \cap \overline{D}$.*

Proof. Proposition 3.45(iii) yields $\overline{C \cap \operatorname{int} D} \subset \overline{C} \cap \overline{\operatorname{int} D} = C \cap \overline{D}$. To show the reverse inclusion, fix $x \in C \cap \operatorname{int} D$ and $y \in C \cap \overline{D}$. By convexity and Proposition 3.44, $[x, y] \subset C$ and $[x, y[\subset \operatorname{int} D$. Therefore, $(\forall \alpha \in]0, 1])$ $z_\alpha = \alpha x + (1 - \alpha)y \in C \cap \operatorname{int} D$. Consequently $y = \lim_{\alpha \downarrow 0} z_\alpha \in \overline{C \cap \operatorname{int} D}$, and we conclude that $C \cap \overline{D} \subset \overline{C \cap \operatorname{int} D}$. ☐

3.4 Separation

Definition 3.49 Let C and D be subsets of $\mathcal{H}$. Then C and D are *separated* if (see Figure 3.2)

$$\big(\exists\, u \in \mathcal{H} \smallsetminus \{0\}\big) \quad \sup \langle C \mid u \rangle \leqslant \inf \langle D \mid u \rangle\,, \tag{3.35}$$

and *strongly separated* if the above inequality is strict. Moreover, a point $x \in \mathcal{H}$ is separated from D if the sets $\{x\}$ and D are separated; likewise, x is strongly separated from D if $\{x\}$ and D are strongly separated.

Theorem 3.50 *Let C be a nonempty closed convex subset of $\mathcal{H}$ and suppose that $x \in \mathcal{H} \smallsetminus C$. Then x is strongly separated from C.*

Proof. Set $u = x - P_C x$ and fix $y \in C$. Then $u \neq 0$ and (3.10) yields $\langle y - x + u \mid u \rangle \leqslant 0$, i.e., $\langle y - x \mid u \rangle \leqslant -\|u\|^2$. Hence $\sup \langle C - x \mid u \rangle \leqslant -\|u\|^2 < 0$. ☐

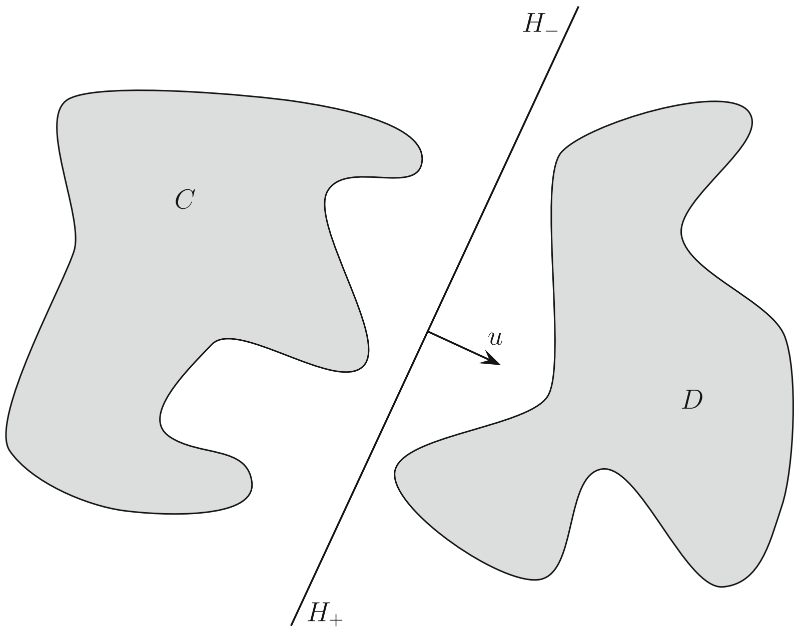

Fig. 3.2 The sets C and D are separated: by (3.35), there exist $u \in \mathcal{H} \smallsetminus \{0\}$ and $\eta \in \mathbb{R}$ such that C is contained in the half-space $H_- = \{x \in \mathcal{H} \mid \langle x \mid u \rangle \leqslant \eta\}$ and D is contained in the half-space $H_+ = \{x \in \mathcal{H} \mid \langle x \mid u \rangle \geqslant \eta\}$.

Corollary 3.51 *Let C and D be nonempty subsets of $\mathcal{H}$ such that $C \cap D = \varnothing$ and $C - D$ is closed and convex. Then C and D are strongly separated.*

Proof. Since $0 \notin C - D$, Theorem 3.50 asserts that the vector 0 is strongly separated from $C - D$. However, it follows from Definition 3.49 that C and D are strongly separated if and only if 0 is strongly separated from $C - D$. $\square$

Corollary 3.52 *Let C and D be nonempty closed convex subsets of $\mathcal{H}$ such that $C \cap D = \varnothing$ and D is bounded. Then C and D are strongly separated.*

Proof. This follows from Corollary 3.51 and Proposition 3.42. $\square$

Theorem 3.53 *Suppose that $\mathcal{H}$ is finite-dimensional, and let C and D be nonempty closed convex subsets of $\mathcal{H}$ such that $C \cap D = \varnothing$. Then C and D are separated.*

Proof. Set $(\forall n \in \mathbb{N})\ D_n = D \cap B(0; n)$. By Corollary 3.52,

$$(\forall n \in \mathbb{N})(\exists\, u_n \in \mathcal{H}) \quad \|u_n\| = 1 \text{ and } \sup \langle C \mid u_n \rangle < \inf \langle D_n \mid u_n \rangle . \tag{3.36}$$

There exist a subsequence $(u_{k_n})_{n\in\mathbb{N}}$ of $(u_n)_{n\in\mathbb{N}}$ and $u \in \mathcal{H}$ such that $u_{k_n} \to u$ and $\|u\| = 1$. Now let $(x, y) \in C \times D$. Then eventually $y \in D_{k_n}$ and $\langle x \mid u_{k_n} \rangle < \langle y \mid u_{k_n} \rangle$. Passing to the limit, we conclude that $\langle x \mid u \rangle \leqslant \langle y \mid u \rangle$. $\square$

We conclude this section by pointing out that separation of nonintersecting closed convex sets is not necessarily achievable in infinite-dimensional spaces.

Example 3.54 Suppose that $\mathcal{H}$ is infinite-dimensional. Then there exist two closed affine subspaces that do not intersect and that are not separated.

Proof. Let C and D be as in Example 3.41 and fix $z \in \overline{C + D} \smallsetminus (C + D)$. Define two closed affine subspaces by $U = C+(C+D)^{\perp}$ and $V = z+D$. Then $U \cap V = \varnothing$ and, since Corollary 3.24(v) implies that $\overline{C + D}+(C+D)^{\perp} = \mathcal{H}$, $U-V = (C+D)-z+(C+D)^{\perp}$ is dense in $\mathcal{H}$. Now suppose that $u \in \mathcal{H}$ satisfies $\inf \langle U \mid u\rangle \geqslant \sup \langle V \mid u\rangle$. Then $\inf \langle U - V \mid u\rangle \geqslant 0$, and hence $\inf \langle \mathcal{H} \mid u\rangle \geqslant 0$. This implies that $u = 0$ and therefore that the separation of U and V is impossible. $\square$

Exercises

Exercise 3.1 Prove Proposition 3.4.

Exercise 3.2 Let I be a nonempty finite set and, for every $i \in I$, let C_i be a subset of a real Hilbert space $\mathcal{H}_i$. Show that $\operatorname{conv}(\bigtimes_{i\in I} C_i) = \bigtimes_{i\in I} \operatorname{conv}(C_i)$.

Exercise 3.3 Let C be a nonempty convex subset of $\mathcal{H}$ and let α and β be in $\mathbb{R}_+$. Show that $\alpha C + \beta C = (\alpha + \beta)C$ and that this property fails if C is not convex.

Exercise 3.4 Let $(C_n)_{n\in\mathbb{N}}$ be a sequence of convex subsets of $\mathcal{H}$ such that $(\forall n \in \mathbb{N})\ C_n \subset C_{n+1}$, and set $C = \bigcup_{n\in\mathbb{N}} C_n$.

(i) Show that C is convex.
(ii) Find an example in which the sets $(C_n)_{n\in\mathbb{N}}$ are closed and C is not closed.

Exercise 3.5 Let C be a subset of $\mathcal{H}$. Show that the *convex kernel* of C, which is defined by

$$\big\{x \in C \;\big|\; (\forall y \in C)\ \left]x, y\right[\subset C\big\}, \tag{3.37}$$

is convex.

Exercise 3.6 A subset C of $\mathcal{H}$ is *midpoint convex* if

$$(\forall x \in C)(\forall y \in C) \quad \frac{x + y}{2} \in C. \tag{3.38}$$

(i) Suppose that $\mathcal{H} = \mathbb{R}$ and let C be the set of rational numbers. Show that C is midpoint convex but not convex.
(ii) Suppose that C is closed and midpoint convex. Show that C is convex.

Exercise 3.7 Consider the setting of Proposition 3.5 and suppose that, in addition, T is continuous and C and D are closed. Show that $T^{-1}(D)$ is closed and find an example in which $T(C)$ is not closed.

Exercise 3.8 Let C and D be nonempty closed convex subsets of $\mathcal{H}$ such that $C \subset D$, and let $x \in \mathcal{H}$. Suppose that $P_D x \in C$. Show that $P_C x = P_D x$.

Exercise 3.9 Consider the setting of Proposition 3.6 and suppose that, in addition, each C_i is closed. Show that the set $\bigtimes_{i=1}^m C_i$ in item (i) is closed and find an example in which the set $\sum_{i=1}^m \alpha_i C_i$ in item (ii) is not closed.

Exercise 3.10 Let $(e_n)_{n\in\mathbb{N}}$ be an orthonormal sequence in $\mathcal{H}$, let $V = \overline{\operatorname{span}}\,\{e_n\}_{n\in\mathbb{N}}$, and let $x \in \mathcal{H}$. Show that $P_V x = \sum_{n\in\mathbb{N}} \langle x \mid e_n\rangle\, e_n$ and that $d_V(x) = \sqrt{\|x\|^2 - \sum_{n\in\mathbb{N}} |\langle x \mid e_n\rangle|^2}$.

Exercise 3.11 Let V be a closed linear subspace of $\mathcal{H}$. Show that $P_V^\dagger = P_V$.

Exercise 3.12 (Penrose) Let $\mathcal{K}$ be a real Hilbert space and let $T \in \mathcal{B}(\mathcal{H},\mathcal{K})$ be such that $\operatorname{ran} T$ is closed. Show the following: $TT^\dagger = (TT^\dagger)^*$, $T^\dagger T = (T^\dagger T)^*$, $TT^\dagger T = T$, and $T^\dagger T T^\dagger = T^\dagger$.

Exercise 3.13 Let $\mathcal{K}$ be a real Hilbert space and let $T \in \mathcal{B}(\mathcal{H},\mathcal{K})$ be such that $\operatorname{ran} T$ is closed. Show that $\ker T^\dagger = \ker T^*$.

Exercise 3.14 Let $\mathcal{K}$ be a real Hilbert space and let $T \in \mathcal{B}(\mathcal{H},\mathcal{K})$ be such that $\operatorname{ran} T$ is closed. Show that $T^\dagger = P_{\operatorname{ran} T^*} \circ T^{-1} \circ P_{\operatorname{ran} T}$, where all operators are understood in the set-valued sense.

Exercise 3.15 Let C be a nonempty subset of $\mathcal{H}$ consisting of vectors of equal norm, let $x \in \mathcal{H}$ and $p \in C$. Prove that p is a projection of x onto C if and only if $\langle x \mid p\rangle = \sup \langle x \mid C\rangle$.

Exercise 3.16 Suppose that $\mathcal{H} = \mathbb{R}^N$, set $I = \{1,\ldots,N\}$, and denote the group of permutations on I by G. For every $x = (\xi_i)_{i\in I} \in \mathcal{H}$ and $\pi \in G$, set $x_\pi = (\xi_{\pi(i)})_{i\in I}$ and write $x_\downarrow = x_\pi$ if $\xi_{\pi(1)} \geqslant \cdots \geqslant \xi_{\pi(N)}$. Now let $y \in \mathcal{H}$, set $C = \big\{y_\pi \;\big|\; \pi \in G\big\}$, and let $x \in \mathcal{H}$. Use (2.17) to show that $\big\{(y_\downarrow)_{\pi^{-1}} \;\big|\; \pi \in G, x_\pi = x_\downarrow\big\} \subset P_C x$.

Exercise 3.17 Provide a subset C of $\mathcal{H}$ that is not weakly closed and such that, for every $n \in \mathbb{N}$, $C \cap B(0;n)$ is weakly closed. Compare with Lemma 1.40 and with Exercise 2.18.

Exercise 3.18 Let C be a convex subset of $\mathcal{H}$ such that $(\forall n \in \mathbb{N})$ $C \cap B(0;n)$ is weakly closed. Show that C is weakly closed. Compare with Exercise 3.17.

Exercise 3.19 Show that the closed convex hull of a weakly compact subset of $\mathcal{H}$ is weakly compact. In contrast, provide an example of a compact subset of $\mathcal{H}$ the convex hull of which is not closed.

Exercise 3.20 Suppose that $\mathcal{H}$ is finite-dimensional, let $x_0 \in \mathcal{H}$, and let $\beta \in \mathbb{R}_{++}$. Show that there exist $\alpha \in \mathbb{R}_{++}$ and a finite family $(y_i)_{i\in I}$ in $\mathcal{H}$ such that $B(x_0;\alpha) \subset \operatorname{conv}\{y_i\}_{i\in I} \subset B(x_0;\beta)$.

Exercise 3.21 Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ and let C be a nonempty subset of $\mathcal{H}$. Suppose that, for every $y \in C$, $(\|x_n - y\|)_{n\in\mathbb{N}}$ converges. Show that this property remains true for every $y \in \overline{\operatorname{conv}}\, C$. Use this result to obtain an extension of Lemma 2.47.

Exercise 3.22 (Rådström's cancellation) Let C, D, and E be subsets of $\mathcal{H}$. Suppose that D is nonempty and bounded, that E is closed and convex, and that $C + D \subset E + D$. Show that $C \subset E$ and that this inclusion fails if E is not convex.

Exercise 3.23 Find a subset C of $\mathbb{R}$ such that $\operatorname{int} C$ is nonempty and convex, $\overline{C}$ is convex, $\operatorname{int} C \neq \operatorname{int} \overline{C}$, and $\overline{C} \neq \overline{\operatorname{int} C}$. Compare with Proposition 3.45(iii).

Exercise 3.24 Find a nonempty compact subset C of $\mathcal{H}$ and a point $x \in \mathcal{H} \smallsetminus C$ that cannot be separated from C. Compare with Corollary 3.51.

Exercise 3.25 Find two nonempty closed convex subsets C and D of $\mathbb{R}^2$ such that $C \cap D = \varnothing$, and C and D are not strongly separated. Compare with Corollary 3.52 and Theorem 3.53.

Exercise 3.26 Let C and D be nonempty subset of $\mathcal{H}$ that are separated. Show that C and D form an *extremal system*, i.e., $(\forall \varepsilon \in \mathbb{R}_{++})(\exists\, z \in B(0;\varepsilon))$ $(C - z) \cap D = \varnothing$.

Chapter 4
Convexity and Notions of Nonexpansiveness

Nonexpansive operators are Lipschitz continuous operators with Lipschitz constant 1. They play a central role in applied mathematics because many problems in nonlinear analysis reduce to finding fixed points of nonexpansive operators. In this chapter, we discuss nonexpansiveness and several variants. The properties of the fixed point sets of nonexpansive operators are investigated, in particular in terms of convexity.

4.1 Nonexpansive and Firmly Nonexpansive Operators

Definition 4.1 Let D be a nonempty subset of $\mathcal{H}$ and let $T\colon D \to \mathcal{H}$. Then T is

(i) *firmly nonexpansive* if

$$(\forall x \in D)(\forall y \in D) \quad \|Tx - Ty\|^2 + \|(\mathrm{Id} - T)x - (\mathrm{Id} - T)y\|^2 \leqslant \|x - y\|^2; \tag{4.1}$$

(ii) *nonexpansive* if it is Lipschitz continuous with constant 1, i.e.,

$$(\forall x \in D)(\forall y \in D) \quad \|Tx - Ty\| \leqslant \|x - y\|; \tag{4.2}$$

(iii) *strictly nonexpansive* if

$$(\forall x \in D)(\forall y \in D) \quad x \neq y \quad \Rightarrow \quad \|Tx - Ty\| < \|x - y\|; \tag{4.3}$$

(iv) *firmly quasinonexpansive* if

$$(\forall x \in D)(\forall y \in \operatorname{Fix} T) \quad \|Tx - y\|^2 + \|Tx - x\|^2 \leqslant \|x - y\|^2; \tag{4.4}$$

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_4

(v) *quasinonexpansive* if

$$(\forall x \in D)(\forall y \in \operatorname{Fix} T) \quad \|Tx - y\| \leqslant \|x - y\|; \tag{4.5}$$

(vi) and *strictly quasinonexpansive* if

$$(\forall x \in D \smallsetminus \operatorname{Fix} T)(\forall y \in \operatorname{Fix} T) \quad \|Tx - y\| < \|x - y\|. \tag{4.6}$$

Concerning Definition 4.1, let us point out the following implications: (4.1)⇒(4.2)⇒(4.5); (4.1)⇒(4.4)⇒(4.6)⇒(4.5); (4.3)⇒(4.6).

Proposition 4.2 *Let D be a nonempty subset of $\mathcal{H}$ and let $T\colon D \to \mathcal{H}$. Then the following are equivalent:*

(i) *T is firmly quasinonexpansive.*
(ii) *$2T - \operatorname{Id}$ is quasinonexpansive.*
(iii) *$(\forall x \in D)(\forall y \in \operatorname{Fix} T)$ $\|Tx - y\|^2 \leqslant \langle x - y \mid Tx - y\rangle$.*
(iv) *$(\forall x \in D)(\forall y \in \operatorname{Fix} T)$ $\langle y - Tx \mid x - Tx\rangle \leqslant 0$.*
(v) *$(\forall x \in D)(\forall y \in \operatorname{Fix} T)$ $\|Tx - x\|^2 \leqslant \langle y - x \mid Tx - x\rangle$.*

Proof. Let $x \in \mathcal{H}$ and suppose that $y \in \operatorname{Fix} T$. Then the equivalences of items (i)–(iv) follow from (4.4) and Lemma 2.17(ii) applied to $(x-y, Tx-y)$.
(iv)⇔(v): Clear. □

Proposition 4.3 *Let D be a nonempty subset of $\mathcal{H}$, let $T\colon D \to \mathcal{H}$ be firmly quasinonexpansive, let $\lambda \in \mathbb{R}$, and set $R = \operatorname{Id} + \lambda(T - \operatorname{Id})$. Let $x \in \mathcal{H}$ and suppose that $y \in \operatorname{Fix} T$. Then $\|Rx - y\|^2 \leqslant \|x - y\|^2 - \lambda(2 - \lambda)\|Tx - x\|^2$.*

Proof. By the equivalence of (i) and (v) in Proposition 4.2,

$$\begin{aligned}\|Rx - y\|^2 &= \|x - y\|^2 - 2\lambda \langle y - x \mid Tx - x\rangle + \lambda^2 \|Tx - x\|^2 \\ &\leqslant \|x - y\|^2 - \lambda(2 - \lambda)\|Tx - x\|^2, \end{aligned} \tag{4.7}$$

as claimed. □

In the case of firmly nonexpansive operators, Proposition 4.2 can be refined as follows.

Proposition 4.4 *Let D be a nonempty subset of $\mathcal{H}$ and let $T\colon D \to \mathcal{H}$. Then the following are equivalent:*

(i) *T is firmly nonexpansive.*
(ii) *$\operatorname{Id} - T$ is firmly nonexpansive.*
(iii) *$2T - \operatorname{Id}$ is nonexpansive.*
(iv) *$(\forall x \in D)(\forall y \in D)$ $\|Tx - Ty\|^2 \leqslant \langle x - y \mid Tx - Ty\rangle$.*
(v) *$(\forall x \in D)(\forall y \in D)$ $0 \leqslant \langle Tx - Ty \mid (\operatorname{Id} - T)x - (\operatorname{Id} - T)y\rangle$.*
(vi) *$(\forall x \in D)(\forall y \in D)(\forall \alpha \in [0,1])$ $\|Tx - Ty\| \leqslant \|\alpha(x - y) + (1 - \alpha)(Tx - Ty)\|$.*

Proof. Let x and y be in D. Then the equivalences of items (i)–(v) follow from (4.1) and Lemma 2.17(ii) applied to $(x-y, Tx-Ty)$.

(v)⇔(vi): Use Lemma 2.13(i). ☐

Corollary 4.5 *Let $T \in \mathcal{B}(\mathcal{H})$. Then the following are equivalent:*

(i) *T is firmly nonexpansive.*
(ii) $\|2T - \mathrm{Id}\| \leqslant 1$.
(iii) $(\forall x \in \mathcal{H})\ \|Tx\|^2 \leqslant \langle x \mid Tx \rangle$.
(iv) *T^* is firmly nonexpansive.*
(v) *$T + T^* - 2T^*T$ is monotone.*

Proof. (i)⇔(ii)⇔(iii): This follows from the equivalences (i)⇔(iii)⇔(iv) in Proposition 4.4.

(i)⇔(iv): Since $\|2T^* - \mathrm{Id}\| = \|(2T - \mathrm{Id})^*\| = \|2T - \mathrm{Id}\|$, this follows from the equivalence (i)⇔(ii).

(iii)⇔(v): Indeed, (iii) ⇔ $(\forall x \in \mathcal{H})\ \langle x \mid (T - T^*T)x \rangle \geqslant 0$ ⇔ $(\forall x \in \mathcal{H})$ $\langle x \mid (T - T^*T)x + (T - T^*T)^*x \rangle \geqslant 0$ ⇔ $(\forall x \in \mathcal{H})\ \langle x \mid (T + T^* - 2T^*T)x \rangle \geqslant 0$ ⇔ (v). ☐

Proposition 4.6 *Let D be a nonempty subset of $\mathcal{H}$, let $(T_i)_{i\in I}$ be a finite family of firmly nonexpansive operators from D to $\mathcal{H}$, let $(\omega_i)_{i\in I}$ be real numbers in $]0,1]$ such that $\sum_{i\in I} \omega_i = 1$, set $T = \sum_{i\in I} \omega_i T_i$, and let x and y be in D. Then*

$$
\begin{aligned}
&\|Tx - Ty\|^2 \\
&\quad = \sum_{i\in I} \omega_i \|T_i x - T_i y\|^2 - \frac{1}{2}\sum_{i\in I}\sum_{j\in I} \omega_i\omega_j \|T_i x - T_i y - T_j x + T_j y\|^2 &\quad (4.8)\\
&\quad \leqslant \sum_{i\in I} \omega_i \big(\|x-y\|^2 - \|(\mathrm{Id} - T_i)x - (\mathrm{Id} - T_i)y\|^2\big) \\
&\qquad - \frac{1}{2}\sum_{i\in I}\sum_{j\in I} \omega_i\omega_j \|T_i x - T_i y - T_j x + T_j y\|^2 &\quad (4.9)\\
&\quad = \|x-y\|^2 - \|(\mathrm{Id} - T)x - (\mathrm{Id} - T)y\|^2 \\
&\qquad - \sum_{i\in I}\sum_{j\in I} \omega_i\omega_j \|T_i x - T_i y - T_j x + T_j y\|^2 &\quad (4.10)\\
&\quad \leqslant \|x-y\|^2 - \|(\mathrm{Id} - T)x - (\mathrm{Id} - T)y\|^2. &\quad (4.11)
\end{aligned}
$$

Proof. A consequence of (4.1) and Lemma 2.14(ii). ☐

Example 4.7 It follows from (4.11) in Proposition 4.6 and (4.1) that every convex combination of firmly nonexpansive operators is likewise.

Proposition 4.8 (Zarantonello) *Let D be a nonempty convex subset of $\mathcal{H}$, let $T\colon D \to \mathcal{H}$, let $(x_i)_{i\in I}$ be a finite family in D, let $(\alpha_i)_{i\in I}$ be a finite family in $\mathbb{R}$ such that $\sum_{i\in I} \alpha_i = 1$, and set $y = \sum_{i\in I} \alpha_i x_i$. Then the following hold:*

(i) $\|Ty - \sum_{i\in I} \alpha_i T x_i\|^2 + \sum_{i\in I} \alpha_i \langle Ty - Tx_i \mid (\mathrm{Id} - T)y - (\mathrm{Id} - T)x_i \rangle$
$= \sum_{i\in I} \sum_{j\in I} \alpha_i\alpha_j \langle Tx_i - Tx_j \mid (\mathrm{Id} - T)x_i - (\mathrm{Id} - T)x_j \rangle /2.$

(ii) *Suppose that T is firmly nonexpansive and that $(\alpha_i)_{i\in I}$ lies in $]0,1]$. Then*

$$\Big\|Ty - \sum_{i\in I} \alpha_i T x_i\Big\|^2 \leqslant \frac{1}{2} \sum_{i\in I} \sum_{j\in I} \alpha_i\alpha_j \langle Tx_i - Tx_j \mid (\mathrm{Id} - T)x_i - (\mathrm{Id} - T)x_j \rangle .$$

(iii) $\|Ty - \sum_{i\in I} \alpha_i T x_i\|^2 + \sum_{i\in I} \alpha_i \big(\|y - x_i\|^2 - \|Ty - Tx_i\|^2\big)$
$= \sum_{i\in I} \sum_{j\in I} \alpha_i\alpha_j \big(\|x_i - x_j\|^2 - \|Tx_i - Tx_j\|^2\big)/2.$

(iv) *Suppose that T is nonexpansive and that $(\alpha_i)_{i\in I}$ lies in $]0,1]$. Then*

$$\Big\|Ty - \sum_{i\in I} \alpha_i T x_i\Big\|^2 \leqslant \frac{1}{2} \sum_{i\in I} \sum_{j\in I} \alpha_i\alpha_j \big(\|x_i - x_j\|^2 - \|Tx_i - Tx_j\|^2\big).$$

Proof. (i): This follows from Lemma 2.14(i) when applied to $(Tx_i - Ty)_{i\in I}$ and $((\mathrm{Id} - T)x_i - (\mathrm{Id} - T)y)_{i\in I}$.

(ii): Combine (i) and Proposition 4.4.

(iii): This follows from (i), applied to $(T + \mathrm{Id})/2$.

(iv): This follows from (iii). ☐

Proposition 4.9 *Let D be a nonempty subset of $\mathcal{H}$, let m be a strictly positive integer, set $I = \{1,\dots,m\}$, let $(T_i)_{i\in I}$ be a family of nonexpansive operators from D to $\mathcal{H}$, and let $(\omega_i)_{i\in I}$ be real numbers in $]0,1]$ such that $\sum_{i\in I} \omega_i = 1$. Then the following hold:*

(i) $\sum_{i\in I} \omega_i T_i \colon D \to \mathcal{H}$ *is nonexpansive.*

(ii) *Suppose that $(\forall i \in I)$ $\operatorname{ran} T_i \subset D$. Then $T_1 \cdots T_m \colon D \to D$ is nonexpansive.*

Proof. A straightforward consequence of (4.2). ☐

4.2 Cocoercive Operators

Definition 4.10 Let D be a nonempty subset of $\mathcal{H}$, let $T\colon D \to \mathcal{H}$, and let $\beta \in \mathbb{R}_{++}$. Then T is *β-cocoercive* (or *β-inverse strongly monotone*) if βT is firmly nonexpansive, i.e.,

$$(\forall x \in D)(\forall y \in D) \quad \langle x - y \mid Tx - Ty \rangle \geqslant \beta \|Tx - Ty\|^2. \tag{4.12}$$

Proposition 4.11 *Let D be a nonempty subset of $\mathcal{H}$ and let $T\colon D \to \mathcal{H}$. Then T is nonexpansive if and only if* $\mathrm{Id} - T$ *is* $(1/2)$*-cocoercive.*

Proof. Proposition 4.4 yields the equivalences T is nonexpansive $\Leftrightarrow$ $-T = 2((\mathrm{Id} - T)/2) - \mathrm{Id}$ is nonexpansive $\Leftrightarrow$ $(\mathrm{Id} - T)/2$ is firmly nonexpansive $\Leftrightarrow$ $\mathrm{Id} - T$ is $(1/2)$-cocoercive. $\square$

Proposition 4.12 *Let $(\mathcal{K}_i)_{i\in I}$ be a finite family of real Hilbert spaces. For every $i \in I$, suppose that $L_i \in \mathcal{B}(\mathcal{H}, \mathcal{K}_i) \smallsetminus \{0\}$, let $\beta_i \in \mathbb{R}_{++}$, and let $T_i\colon \mathcal{K}_i \to \mathcal{K}_i$ be β_i-cocoercive. Set*

$$T = \sum_{i\in I} L_i^* T_i L_i \quad \text{and} \quad \beta = \frac{1}{\displaystyle\sum_{i\in I} \frac{\|L_i\|^2}{\beta_i}}. \tag{4.13}$$

Then T is β-cocoercive.

Proof. Set $(\forall i \in I)$ $\alpha_i = \beta\|L_i\|^2/\beta_i$. Then $\sum_{i\in I} \alpha_i = 1$ and, using (4.12), Fact 2.25(ii), and Lemma 2.14(ii), we obtain

$$\begin{aligned}
(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \langle x - y \mid Tx - Ty\rangle &= \sum_{i\in I} \langle x - y \mid L_i^* T_i L_i x - L_i^* T_i L_i y\rangle \\
&= \sum_{i\in I} \langle L_i x - L_i y \mid T_i L_i x - T_i L_i y\rangle \\
&\geqslant \sum_{i\in I} \beta_i \|T_i L_i x - T_i L_i y\|^2 \\
&\geqslant \sum_{i\in I} \frac{\beta_i}{\|L_i\|^2} \|L_i^* T_i L_i x - L_i^* T_i L_i y\|^2 \\
&= \beta \sum_{i\in I} \alpha_i \left\| \frac{1}{\alpha_i}(L_i^* T_i L_i x - L_i^* T_i L_i y)\right\|^2 \\
&\geqslant \beta \left\| \sum_{i\in I} (L_i^* T_i L_i x - L_i^* T_i L_i y)\right\|^2 \\
&= \beta\|Tx - Ty\|^2,
\end{aligned} \tag{4.14}$$

which concludes the proof. $\square$

Corollary 4.13 *Let $\mathcal{K}$ be a real Hilbert space, let $T\colon \mathcal{K} \to \mathcal{K}$ be firmly nonexpansive, and let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ be such that $\|L\| \leqslant 1$. Then L^*TL is firmly nonexpansive.*

Proof. We assume that $L \neq 0$. Applying Proposition 4.12 with $I = \{1\}$ and $\beta_1 = 1$, we obtain that L^*TL is $\|L\|^{-2}$-cocoercive and hence firmly nonexpansive since $\|L\|^{-2} \geqslant 1$. $\square$

Example 4.14 Let $T\colon \mathcal{H} \to \mathcal{H}$ be firmly nonexpansive and let V be a closed linear subspace of $\mathcal{H}$. Then it follows from Corollary 3.24(iii)&(vi) and Corollary 4.13 that $P_V T P_V$ is firmly nonexpansive.

Remark 4.15 Let D be a nonempty subset of $\mathcal{H}$, let $\beta \in \mathbb{R}_{++}$, and let $T\colon D \to \mathcal{H}$ be β-cocoercive. Then it follows from (4.12) and the Cauchy–Schwarz inequality that T is Lipschitz continuous with constant $1/\beta$. However, this constant is not necessarily tight. For instance, suppose that $\mathcal{H} = D = \mathbb{R}^3$ and set

$$B = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix}. \tag{4.15}$$

Then, for every $x = (\xi_1, \xi_2, \xi_3) \in \mathbb{R}^3$, we have $\|Bx\|^2 = 2\,\langle Bx \mid x\rangle = 3\|x\|^2 - (\xi_1 + \xi_2 + \xi_3)^2$. Therefore, B is $1/2$-cocoercive and $\sqrt{3}$-Lipschitz continuous.

4.3 Projectors onto Convex Sets

Proposition 4.16 *Let C be a nonempty closed convex subset of $\mathcal{H}$. Then the projector P_C is firmly nonexpansive.*

Proof. Fix x and y in $\mathcal{H}$. Theorem 3.16 asserts that $\langle P_C y - P_C x \mid x - P_C x\rangle \leqslant 0$ and that $\langle P_C x - P_C y \mid y - P_C y\rangle \leqslant 0$. Adding these two inequalities yields $\|P_C x - P_C y\|^2 \leqslant \langle x - y \mid P_C x - P_C y\rangle$. The claim therefore follows from Proposition 4.4. ☐

Example 4.17 Suppose that $\mathcal{H} \neq \{0\}$, let $\rho \in \mathbb{R}_{++}$, and let $\alpha \in\,]0,1]$. Then the following hold:

(i) Define the *soft thresholder* at level ρ by

$$(\forall x \in \mathcal{H}) \quad T_1 x = \begin{cases} \big(1 - \rho/\|x\|\big)x, & \text{if } \|x\| > \rho; \\ 0, & \text{if } \|x\| \leqslant \rho. \end{cases} \tag{4.16}$$

Then T_1 is firmly nonexpansive.

(ii) Define the *hard thresholder* at level ρ by

$$(\forall x \in \mathcal{H}) \quad T_2 x = \begin{cases} \alpha x, & \text{if } \|x\| > \rho; \\ 0, & \text{if } \|x\| \leqslant \rho. \end{cases} \tag{4.17}$$

Then T_2 is quasinonexpansive but not nonexpansive for $\alpha < 1$, and it is not quasinonexpansive for $\alpha = 1$.

(iii) Define

$$(\forall x \in \mathcal{H}) \quad T_3 x = \begin{cases} \big(1 - 2\rho/\|x\|\big)x, & \text{if } \|x\| > \rho; \\ -x, & \text{if } \|x\| \leqslant \rho. \end{cases} \tag{4.18}$$

Then T_3 is nonexpansive but not firmly nonexpansive.

Proof. (i): In view of Example 3.18, $T_1 = \mathrm{Id} - P_{B(0;\rho)}$. Hence, it follows from Proposition 4.16 and Proposition 4.4 that T_1 is firmly nonexpansive.

(ii): Suppose that $\alpha < 1$. Then 0 is the unique fixed point of T_2 and $(\forall x \in \mathcal{H})\ \|T_2 x\| \leqslant \|x\|$. Thus, T_2 is quasinonexpansive but not nonexpansive, since it is not continuous. Now suppose that $\alpha = 1$, take $x \in B(0;\rho) \smallsetminus \{0\}$, and set $y = 2\rho x/\|x\|$. Then $y \in \operatorname{Fix} T_2$ but $\|T_2 x - y\| = 2\rho > 2\rho - \|x\| = \|x - y\|$. Thus, T_2 is not quasinonexpansive.

(iii): We derive from Proposition 4.4 that $T_3 = 2T_1 - \mathrm{Id}$ is nonexpansive. Now, take $x \in \mathcal{H}$ such that $\|x\| = \rho$ and set $y = -x$. Then the inequality in (4.1) fails for T_3 and, therefore, T_3 is not firmly nonexpansive. □

Corollary 4.18 *Let C be a nonempty closed convex subset of $\mathcal{H}$. Then* $\mathrm{Id} - P_C$ *is firmly nonexpansive and* $2P_C - \mathrm{Id}$ *is nonexpansive.*

Proof. A consequence of Proposition 4.16 and Proposition 4.4. □

Proposition 4.16 implies that projectors are continuous. In the affine case, weak continuity also holds.

Proposition 4.19 *Let C be a closed affine subspace of $\mathcal{H}$. Then the following hold:*

(i) *P_C is weakly continuous.*
(ii) $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ \|P_C x - P_C y\|^2 = \langle x - y \mid P_C x - P_C y\rangle$.

Proof. (i): Combine Lemma 2.41 and Corollary 3.22(ii).

(ii): Use Corollary 3.22(i) instead of Theorem 3.16 in the proof of Proposition 4.16. □

Let us illustrate the fact that projectors may fail to be weakly continuous.

Example 4.20 Suppose that $\mathcal{H}$ is infinite-dimensional and set $C = B(0;1)$. Then P_C is not weakly continuous.

Proof. Let $(e_n)_{n\in\mathbb{N}}$ be an orthonormal sequence and set $(\forall n \in \mathbb{N})$ $x_n = e_1 + e_{2n}$. Then, as seen in Example 2.32, $x_n \rightharpoonup e_1$. However, it follows from Example 3.18 that $P_C x_n = x_n/\sqrt{2} \rightharpoonup e_1/\sqrt{2} \neq e_1 = P_C e_1$. □

We conclude this section with a sufficient condition for strong convergence to a projection.

Proposition 4.21 *Let C be a nonempty closed convex subset of $\mathcal{H}$, let $z \in \mathcal{H}$, and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that $\varlimsup \|x_n - z\| \leqslant \|P_C z - z\|$. Suppose that every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ belongs to C. Then $x_n \to P_C z$.*

Proof. Let x be a weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$, say $x_{k_n} \rightharpoonup x \in C$. Lemma 2.42 yields $\|x - z\| \leqslant \varliminf \|x_{k_n} - z\| \leqslant \varlimsup \|x_{k_n} - z\| \leqslant \varlimsup \|x_n - z\| \leqslant \|P_C z - z\| \leqslant \|x - z\|$. Hence $x = P_C z$ is the unique weak sequential cluster point of the bounded sequence $(x_n)_{n\in\mathbb{N}}$ and it follows from Lemma 2.46 that $x_n \rightharpoonup P_C z$, i.e., $x_n - z \rightharpoonup P_C z - z$. In view of Corollary 2.52, the proof is complete. □

4.4 Fixed Points of Nonexpansive Operators

The projection operator P_C onto a nonempty closed convex subset C of $\mathcal{H}$ is (firmly) nonexpansive by Proposition 4.16 with

$$\operatorname{Fix} P_C = C, \tag{4.19}$$

which is closed and convex. The following results extend this observation.

Proposition 4.22 *Let D be a nonempty convex subset of $\mathcal{H}$ and let $T\colon D \to \mathcal{H}$ be quasinonexpansive. Then $\operatorname{Fix} T$ is convex.*

Proof. Assume that x and y are in $\operatorname{Fix} T$, let $\alpha \in \left]0,1\right[$, and set $z = \alpha x + (1-\alpha)y$. Then $z \in D$ and, by Corollary 2.15,

$$\begin{aligned}
\|Tz - z\|^2 &= \|\alpha(Tz - x) + (1-\alpha)(Tz - y)\|^2 \\
&= \alpha\|Tz - x\|^2 + (1-\alpha)\|Tz - y\|^2 - \alpha(1-\alpha)\|x - y\|^2 \\
&\leqslant \alpha\|z - x\|^2 + (1-\alpha)\|z - y\|^2 - \alpha(1-\alpha)\|x - y\|^2 \\
&= \|\alpha(z - x) + (1-\alpha)(z - y)\|^2 \\
&= 0.
\end{aligned} \tag{4.20}$$

Therefore $z \in \operatorname{Fix} T$. ☐

Proposition 4.23 *Let D be a nonempty subset of $\mathcal{H}$ and let $T\colon D \to \mathcal{H}$ be quasinonexpansive. Then the following hold:*

(i) $\operatorname{Fix} T = \bigcap_{x \in D} \big\{ y \in D \;\big|\; \langle y - Tx \mid x - Tx \rangle \leqslant (1/2)\|Tx - x\|^2 \big\}$.

(ii) *Suppose that D is closed and convex. Then* $\operatorname{Fix} T$ *is closed and convex.*

Proof. (i): Set $C = \bigcap_{x \in D} \big\{ y \in D \;\big|\; \langle y - Tx \mid x - Tx \rangle \leqslant (1/2)\|Tx - x\|^2 \big\}$. Let $x \in D$ and assume that $y \in \operatorname{Fix} T$. Then $\|Tx - y\|^2 \leqslant \|x - y\|^2$, hence $\|Tx - x\|^2 + 2\langle Tx - x \mid x - y\rangle + \|x - y\|^2 \leqslant \|x - y\|^2$ and therefore $\|Tx - x\|^2 + 2\langle Tx - x \mid x - Tx\rangle + 2\langle Tx - x \mid Tx - y\rangle \leqslant 0$, i.e., $2\langle Tx - x \mid Tx - y\rangle \leqslant \|Tx - x\|^2$. This shows that $\operatorname{Fix} T \subset C$. Conversely, assume that $x \in C$. Then $x \in \big\{ y \in D \;\big|\; \langle y - Tx \mid x - Tx \rangle \leqslant (1/2)\|Tx - x\|^2 \big\}$ and therefore $\|Tx - x\|^2 = \langle x - Tx \mid x - Tx\rangle \leqslant (1/2)\|Tx - x\|^2$, which implies that $Tx = x$. Thus, $C \subset \operatorname{Fix} T$.

(ii): By (i), $\operatorname{Fix} T$ is the intersection of the closed convex set D with an intersection of closed half-spaces. It is therefore closed and convex. ☐

Corollary 4.24 *Let D be a nonempty closed convex subset of $\mathcal{H}$ and let $T\colon D \to \mathcal{H}$ be nonexpansive. Then* $\operatorname{Fix} T$ *is closed and convex.*

Corollary 4.25 *Let D be a nonempty subset of $\mathcal{H}$ and let $T\colon D \to \mathcal{H}$ be firmly quasinonexpansive. Then*

$$\operatorname{Fix} T = \bigcap_{x \in D} \big\{ y \in D \;\big|\; \langle y - Tx \mid x - Tx \rangle \leqslant 0 \big\}. \tag{4.21}$$

Proof. Set $R = 2T - \mathrm{Id}$. Then $\operatorname{Fix} T = \operatorname{Fix} R$ and Proposition 4.2 asserts that R is quasinonexpansive. However, Proposition 4.23(i) yields

$$\begin{aligned}
\operatorname{Fix} R &= \bigcap_{x\in D} \big\{ y \in D \;\big|\; \langle y - Rx \mid x - Rx\rangle \leqslant (1/2)\|Rx - x\|^2 \big\} \\
&= \bigcap_{x\in D} \big\{ y \in D \;\big|\; \langle y - Tx + x - Tx \mid 2(x - Tx)\rangle \leqslant 2\|Tx - x\|^2 \big\} \\
&= \bigcap_{x\in D} \big\{ y \in D \;\big|\; \langle y - Tx \mid x - Tx\rangle \leqslant 0 \big\},
\end{aligned} \tag{4.22}$$

which establishes (4.21). ☐

The following notion describes operators T for which $\operatorname{gra}(\mathrm{Id} - T)$ is (sequentially) closed in $\mathcal{H}^{\mathrm{weak}} \times \mathcal{H}^{\mathrm{strong}}$.

Definition 4.26 Let D be a nonempty weakly sequentially closed subset of $\mathcal{H}$, let $T\colon D \to \mathcal{H}$, and let $u \in \mathcal{H}$. Then T is *demiclosed at* u if, for every sequence $(x_n)_{n\in\mathbb{N}}$ in D and every $x \in D$ such that $x_n \rightharpoonup x$ and $Tx_n \to u$, we have $Tx = u$. In addition, T is *demiclosed* if it is demiclosed at every point in D.

Theorem 4.27 (Browder's demiclosedness principle) *Let D be a nonempty weakly sequentially closed subset of $\mathcal{H}$ and let $T\colon D \to \mathcal{H}$ be a nonexpansive operator. Then* $\mathrm{Id} - T$ *is demiclosed.*

Proof. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in D, let $x \in D$, and let $u \in \mathcal{H}$. Suppose that $x_n \rightharpoonup x$ and that $x_n - Tx_n \to u$. For every $n \in \mathbb{N}$, it follows from the nonexpansiveness of T that

$$\begin{aligned}
\|x - Tx - u\|^2 &= \|x_n - Tx - u\|^2 - \|x_n - x\|^2 - 2\,\langle x_n - x \mid x - Tx - u\rangle \\
&= \|x_n - Tx_n - u\|^2 + 2\,\langle x_n - Tx_n - u \mid Tx_n - Tx\rangle \\
&\quad + \|Tx_n - Tx\|^2 - \|x_n - x\|^2 - 2\,\langle x_n - x \mid x - Tx - u\rangle \\
&\leqslant \|x_n - Tx_n - u\|^2 + 2\,\langle x_n - Tx_n - u \mid Tx_n - Tx\rangle \\
&\quad - 2\,\langle x_n - x \mid x - Tx - u\rangle .
\end{aligned} \tag{4.23}$$

However, by assumption, $x_n - Tx_n - u \to 0$, $x_n - x \rightharpoonup 0$, and hence $Tx_n - Tx \rightharpoonup x - Tx - u$. Taking the limit as $n \to +\infty$ in (4.23) and appealing to Lemma 2.51(iii), we obtain $x - Tx = u$. ☐

Corollary 4.28 *Let D be a nonempty closed and convex subset of $\mathcal{H}$, let $T\colon D \to \mathcal{H}$ be nonexpansive, let $(x_n)_{n\in\mathbb{N}}$ be a sequence in D, and let x be a point in D. Suppose that $x_n \rightharpoonup x$ and that $x_n - Tx_n \to 0$. Then $x \in \operatorname{Fix} T$.*

Proof. Since Theorem 3.34 asserts that D is weakly sequentially closed, the result follows from Theorem 4.27. ☐

The set of fixed points of a nonexpansive operator may be empty (consider a translation by a nonzero vector). The following theorem gives a condition that guarantees the existence of fixed points.

Theorem 4.29 (Browder–Göhde–Kirk) *Let D be a nonempty bounded closed convex subset of $\mathcal{H}$ and let $T\colon D \to D$ be a nonexpansive operator. Then $\operatorname{Fix} T \neq \varnothing$.*

Proof. It follows from Theorem 3.34 that D is weakly sequentially closed, and from Theorem 3.37 that it is weakly sequentially compact. Now fix $x_0 \in D$ and a sequence $(\alpha_n)_{n\in\mathbb{N}}$ in $]0,1]$ such that $\alpha_0 = 1$ and $\alpha_n \downarrow 0$. For every $n \in \mathbb{N}$, the operator $T_n\colon D \to D\colon x \mapsto \alpha_n x_0 + (1-\alpha_n)Tx$ is Lipschitz continuous with constant $1-\alpha_n$, and it therefore possesses a fixed point $x_n \in D$ by Theorem 1.50. Moreover, for every $n \in \mathbb{N}$, $\|x_n - Tx_n\| = \|T_n x_n - Tx_n\| = \alpha_n\|x_0 - Tx_n\| \leqslant \alpha_n \operatorname{diam} D$. Hence $x_n - Tx_n \to 0$. On the other hand, since $(x_n)_{n\in\mathbb{N}}$ lies in D, by weak sequential compactness we can extract a weakly convergent subsequence, say $x_{k_n} \rightharpoonup x \in D$. Since $x_{k_n} - Tx_{k_n} \to 0$, Corollary 4.28 asserts that $x \in \operatorname{Fix} T$. ☐

The proof of Theorem 4.29 rests on Lipschitz continuous operators and their unique fixed points. These fixed points determine a curve, which we investigate in more detail in the next result.

Proposition 4.30 (Approximating curve) *Let D be a nonempty closed convex subset of $\mathcal{H}$ and let $T\colon D \to D$ be a nonexpansive operator. Then, for every $\varepsilon \in\,]0,1[$ and every $x \in D$, there exists a unique point $x_\varepsilon \in D$ such that*

$$x_\varepsilon = \varepsilon x + (1-\varepsilon)Tx_\varepsilon. \tag{4.24}$$

Set, for every $\varepsilon \in\,]0,1[$, $T_\varepsilon\colon D \to D\colon x \mapsto x_\varepsilon$, and let $x \in D$. Then the following hold:

(i) $(\forall \varepsilon \in\,]0,1[)$ $T_\varepsilon = \varepsilon \mathrm{Id} + (1-\varepsilon)TT_\varepsilon = (\mathrm{Id} - (1-\varepsilon)T)^{-1} \circ \varepsilon \mathrm{Id}$.
(ii) $(\forall \varepsilon \in\,]0,1[)$ *T_ε is firmly nonexpansive.*
(iii) $(\forall \varepsilon \in\,]0,1[)$ $\operatorname{Fix} T_\varepsilon = \operatorname{Fix} T$.
(iv) $(\forall \varepsilon \in\,]0,1[)$ $\varepsilon(x - Tx_\varepsilon) = x_\varepsilon - Tx_\varepsilon = (1-\varepsilon)^{-1}\varepsilon(x - x_\varepsilon)$.
(v) *Suppose that* $\operatorname{Fix} T = \varnothing$. *Then* $\lim_{\varepsilon\downarrow 0}\|x_\varepsilon\| = +\infty$.
(vi) $(\forall \varepsilon \in\,]0,1[)(\forall y \in \operatorname{Fix} T)$ $\|x - x_\varepsilon\|^2 + \|x_\varepsilon - y\|^2 \leqslant \|x - y\|^2$.
(vii) *Suppose that* $\operatorname{Fix} T \neq \varnothing$. *Then* $\lim_{\varepsilon\downarrow 0} x_\varepsilon = P_{\operatorname{Fix} T}\, x$.
(viii) $(\forall \varepsilon \in\,]0,1[)(\forall \delta \in\,]0,1[)$

$$\left(\frac{\varepsilon-\delta}{1-\varepsilon}\right)^2 \|x_\varepsilon - x\|^2 + \delta(2-\delta)\|x_\delta - x_\varepsilon\|^2 \leqslant 2\frac{\varepsilon-\delta}{1-\varepsilon}\left\langle x_\varepsilon - x \mid x_\delta - x_\varepsilon\right\rangle.$$

(ix) $(\forall \varepsilon \in\,]0,1[)(\forall \delta \in\,]0,\varepsilon[)$ $\|x - x_\varepsilon\|^2 + \|x_\varepsilon - x_\delta\|^2 \leqslant \|x - x_\delta\|^2$.
(x) *The function* $]0,1[\, \to \mathbb{R}_+\colon \varepsilon \mapsto \|x - x_\varepsilon\|$ *is decreasing.*
(xi) *The curve* $]0,1[\, \to \mathcal{H}\colon \varepsilon \mapsto x_\varepsilon$ *is continuous.*

(xii) *If* $x \in \operatorname{Fix} T$, *then* $x_\varepsilon \equiv x$ *is constant; otherwise,* $(x_\varepsilon)_{\varepsilon \in]0,1[}$ *is an injective curve.*

Proof. Let $\varepsilon \in]0,1[$. By Theorem 1.50, the operator $D \to D \colon z \mapsto \varepsilon x + (1-\varepsilon)Tz$ has a unique fixed point. Hence, x_ε is unique, and T_ε is therefore well defined.

(i): The first identity is clear from (4.24). Furthermore, $\varepsilon \mathrm{Id} = T_\varepsilon - (1-\varepsilon)TT_\varepsilon = (\mathrm{Id} - (1-\varepsilon)T)T_\varepsilon$, and therefore, since $\mathrm{Id} - (1-\varepsilon)T$ is injective, we obtain $T_\varepsilon = (\mathrm{Id} - (1-\varepsilon)T)^{-1} \circ \varepsilon \mathrm{Id}$.

(ii): Let $y \in D$. Then

$$\begin{aligned} x - y &= \varepsilon^{-1}\big((x_\varepsilon - (1-\varepsilon)Tx_\varepsilon) - (y_\varepsilon - (1-\varepsilon)Ty_\varepsilon)\big) \\ &= \varepsilon^{-1}\big((x_\varepsilon - y_\varepsilon) - (1-\varepsilon)(Tx_\varepsilon - Ty_\varepsilon)\big). \end{aligned} \tag{4.25}$$

Using (4.24), (4.25), and Cauchy–Schwarz, we deduce that

$$\begin{aligned} &\langle T_\varepsilon x - T_\varepsilon y \mid (\mathrm{Id} - T_\varepsilon)x - (\mathrm{Id} - T_\varepsilon)y \rangle \\ &\quad = \langle x_\varepsilon - y_\varepsilon \mid (1-\varepsilon)(x - Tx_\varepsilon) - (1-\varepsilon)(y - Ty_\varepsilon) \rangle \\ &\quad = (1-\varepsilon)\,\langle x_\varepsilon - y_\varepsilon \mid (x-y) - (Tx_\varepsilon - Ty_\varepsilon) \rangle \\ &\quad = (1-\varepsilon)\varepsilon^{-1}\,\langle x_\varepsilon - y_\varepsilon \mid (x_\varepsilon - y_\varepsilon) - (Tx_\varepsilon - Ty_\varepsilon) \rangle \\ &\quad \geqslant (\varepsilon^{-1} - 1)\big(\|x_\varepsilon - y_\varepsilon\|^2 - \|x_\varepsilon - y_\varepsilon\|\,\|Tx_\varepsilon - Ty_\varepsilon\|\big) \\ &\quad = (\varepsilon^{-1} - 1)\|x_\varepsilon - y_\varepsilon\|\big(\|x_\varepsilon - y_\varepsilon\| - \|Tx_\varepsilon - Ty_\varepsilon\|\big) \\ &\quad \geqslant 0. \end{aligned} \tag{4.26}$$

Hence, by Proposition 4.4, T_ε is firmly nonexpansive.

(iii): Let $x \in D$. Suppose first that $x \in \operatorname{Fix} T$. Then $x = \varepsilon x + (1-\varepsilon)Tx$ and hence $x = x_\varepsilon$ by uniqueness of x_ε. It follows that $T_\varepsilon x = x$ and therefore that $x \in \operatorname{Fix} T_\varepsilon$. Conversely, assume that $x \in \operatorname{Fix} T_\varepsilon$. Then $x = x_\varepsilon = \varepsilon x + (1-\varepsilon)Tx_\varepsilon = x + (1-\varepsilon)(Tx - x)$ and thus $x = Tx$, i.e., $x \in \operatorname{Fix} T$.

(iv): This follows from (4.24).

(v): Suppose that there exists a sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ in $]0,1[$ such that $\varepsilon_n \downarrow 0$ and $(x_{\varepsilon_n})_{n\in\mathbb{N}}$ is bounded. We derive from Lemma 2.45 that there exists a weak sequential cluster point y of $(x_{\varepsilon_n})_{n\in\mathbb{N}}$. In turn, by (iv) and Corollary 4.28, $y \in \operatorname{Fix} T$.

(vi): Assume that $y \in \operatorname{Fix} T$. By (iii), $y = T_\varepsilon y = y_\varepsilon$ and, by (ii),

$$\|x-y\|^2 \geqslant \|x_\varepsilon - y_\varepsilon\|^2 + \|(x - x_\varepsilon) - (y - y_\varepsilon)\|^2 = \|x_\varepsilon - y\|^2 + \|x - x_\varepsilon\|^2. \tag{4.27}$$

(vii): Let $(\varepsilon_n)_{n\in\mathbb{N}}$ be a sequence in $]0,1[$ such that $\varepsilon_n \downarrow 0$, and set $(\forall n \in \mathbb{N})$ $z_n = x_{\varepsilon_n}$. By (vi), $(z_n)_{n\in\mathbb{N}}$ is bounded. Thus, using (iv), we see that $z_n - Tz_n \to 0$. Let z be a weak sequential cluster point of $(z_n)_{n\in\mathbb{N}}$, say $z_{k_n} \rightharpoonup z$. Theorem 4.27 implies that $z \in \operatorname{Fix} T$. In view of (vi), we obtain $\overline{\lim}\,\|x - z_{k_n}\|^2 \leqslant \|x - z\|^2$. Since $x - z_{k_n} \rightharpoonup x - z$, Lemma 2.51(i) yields $x - z_{k_n} \to x - z$. Thus $z_{k_n} \to z$. Again, using (vi), we see that

$$(\forall n \in \mathbb{N}) \quad \|x - z_{k_n}\|^2 + \|z_{k_n} - y\|^2 \leqslant \|x - y\|^2. \tag{4.28}$$

Taking the limit as $n \to +\infty$, we deduce that $\|x-z\|^2 + \|z-y\|^2 \leqslant \|x-y\|^2$. Hence $(\forall y \in \operatorname{Fix} T)$ $\langle y-z \mid x-z\rangle \leqslant 0$. It now follows from Theorem 3.16 that $z = P_{\operatorname{Fix} T}x$. Therefore, $z_n \to P_{\operatorname{Fix} T}x$ and hence $x_\varepsilon \to P_{\operatorname{Fix} T}\, x$ as $\varepsilon \downarrow 0$.

(viii): Let $\delta \in \left]0,1\right[$ and set $y_\varepsilon = x_\varepsilon - x$ and $y_\delta = x_\delta - x$. Since $y_\delta = y_\varepsilon + x_\delta - x_\varepsilon$, using (4.24), we obtain

$$\begin{aligned}
\|x_\delta - x_\varepsilon\|^2 &\geqslant \|Tx_\delta - Tx_\varepsilon\|^2 \\
&= \left\| \frac{x_\delta - \delta x}{1-\delta} - \frac{x_\varepsilon - \varepsilon x}{1-\varepsilon} \right\|^2 \\
&= \left\| \frac{y_\delta}{1-\delta} - \frac{y_\varepsilon}{1-\varepsilon} \right\|^2 \\
&= \frac{1}{(1-\delta)^2} \left\| \frac{\delta-\varepsilon}{1-\varepsilon} y_\varepsilon + x_\delta - x_\varepsilon \right\|^2 \\
&= \frac{1}{(1-\delta)^2} \Bigg(\left(\frac{\delta-\varepsilon}{1-\varepsilon}\right)^2 \|y_\varepsilon\|^2 + 2\frac{\delta-\varepsilon}{1-\varepsilon} \langle y_\varepsilon \mid x_\delta - x_\varepsilon\rangle \\
&\qquad + \|x_\delta - x_\varepsilon\|^2 \Bigg).
\end{aligned} \tag{4.29}$$

Therefore,

$$\left(\frac{\varepsilon-\delta}{1-\varepsilon}\right)^2 \|y_\varepsilon\|^2 + \delta(2-\delta)\|x_\delta - x_\varepsilon\|^2 \leqslant 2\frac{\varepsilon-\delta}{1-\varepsilon} \langle y_\varepsilon \mid x_\delta - x_\varepsilon\rangle, \tag{4.30}$$

which is the desired inequality.

(viii)⇒(ix): Let $\delta \in \left]0,\varepsilon\right[$. Then $\langle x_\varepsilon - x \mid x_\delta - x_\varepsilon\rangle \geqslant 0$. In turn, $\|x_\delta - x\|^2 = \|x_\delta - x_\varepsilon\|^2 + 2\langle x_\delta - x_\varepsilon \mid x_\varepsilon - x\rangle + \|x_\varepsilon - x\|^2 \geqslant \|x_\delta - x_\varepsilon\|^2 + \|x_\varepsilon - x\|^2$.

(ix)⇒(x): Clear.

(xi): We derive from (viii) and Cauchy–Schwarz that

$$(\forall \delta \in \left]0,\varepsilon\right[) \quad \|x_\delta - x_\varepsilon\| \leqslant \frac{2(\varepsilon-\delta)}{\delta(2-\delta)(1-\varepsilon)}\|x_\varepsilon - x\|. \tag{4.31}$$

Hence, $\|x_\delta - x_\varepsilon\| \downarrow 0$ as $\delta \uparrow \varepsilon$, and therefore the curve $\left]0,1\right[\to \mathcal{H}\colon \varepsilon \mapsto x_\varepsilon$ is left-continuous. Likewise, we have

$$(\forall \delta \in \left]\varepsilon,1\right[) \quad \|x_\delta - x_\varepsilon\| \leqslant \frac{2(\delta-\varepsilon)}{\delta(2-\delta)(1-\varepsilon)}\|x_\varepsilon - x\|, \tag{4.32}$$

so that $\|x_\delta - x_\varepsilon\| \downarrow 0$ as $\delta \downarrow \varepsilon$. Thus, the curve $\left]0,1\right[\to \mathcal{H}\colon \varepsilon \mapsto x_\varepsilon$ is right-continuous.

(xii): If $x \in \operatorname{Fix} T$, then $x \in \operatorname{Fix} T_\varepsilon$ by (iii) and hence $x = T_\varepsilon x = x_\varepsilon$. Now assume that $x \notin \operatorname{Fix} T$. If $\delta \in \left]0,\varepsilon\right[$ and $x_\varepsilon = x_\delta$, then (viii) yields $T_\varepsilon x = x_\varepsilon = x$ and hence $x \in \operatorname{Fix} T_\varepsilon$, which is impossible in view of (iii). Hence the curve $(x_\varepsilon)_{\varepsilon\in\left]0,1\right[}$ is injective in that case. ☐

Proposition 4.31 *Let $T_1 \colon \mathcal{H} \to \mathcal{H}$ and $T_2 \colon \mathcal{H} \to \mathcal{H}$ be firmly nonexpansive and set $T = T_1(2T_2 - \mathrm{Id}) + \mathrm{Id} - T_2$. Then the following hold:*

(i) $2T - \mathrm{Id} = (2T_1 - \mathrm{Id})(2T_2 - \mathrm{Id})$.
(ii) *T is firmly nonexpansive.*
(iii) $\mathrm{Fix}\, T = \mathrm{Fix}(2T_1 - \mathrm{Id})(2T_2 - \mathrm{Id})$.
(iv) *Suppose that T_1 is the projector onto a closed affine subspace. Then* $\mathrm{Fix}\, T = \{x \in \mathcal{H} \mid T_1 x = T_2 x\}$.

Proof. (i): Expand.

(ii): Proposition 4.4 asserts that $2T_1 - \mathrm{Id}$ and $2T_1 - \mathrm{Id}$ are nonexpansive. Therefore, $(2T_1 - \mathrm{Id})(2T_2 - \mathrm{Id})$ is nonexpansive and so is $2T - \mathrm{Id}$ by (i). In turn, T is firmly nonexpansive.

(iii): By (i), $\mathrm{Fix}\, T = \mathrm{Fix}(2T - \mathrm{Id}) = \mathrm{Fix}(2T_1 - \mathrm{Id})(2T_2 - \mathrm{Id})$.

(iv): Suppose that $T_1 = P_C$, where C is a closed affine subspace of $\mathcal{H}$, and let $x \in \mathcal{H}$. It follows from Proposition 4.16 that T_1 is firmly nonexpansive and from Corollary 3.22(ii) that $x \in \mathrm{Fix}\, T \Leftrightarrow x = P_C(2T_2x + (1-2)x) + x - T_2x \Leftrightarrow T_2x = 2P_C(T_2x) + (1-2)P_Cx \in C \Leftrightarrow P_C(T_2x) = T_2x = 2P_C(T_2x) + (1-2)P_Cx \Leftrightarrow T_2x = P_Cx$. ☐

Corollary 4.32 *Let T_1 be the projector onto a closed linear subspace of $\mathcal{H}$, let $T_2 \colon \mathcal{H} \to \mathcal{H}$ be firmly nonexpansive, and set $T = T_1T_2 + (\mathrm{Id} - T_1)(\mathrm{Id} - T_2)$. Then T is firmly nonexpansive and* $\mathrm{Fix}\, T = \{x \in \mathcal{H} \mid T_1x = T_2x\}$.

Proof. Since $T = T_1(2T_2 - \mathrm{Id}) + \mathrm{Id} - T_2$, the result follows from Proposition 4.31. ☐

4.5 Averaged Nonexpansive Operators

Definition 4.33 Let D be a nonempty subset of $\mathcal{H}$, let $T \colon D \to \mathcal{H}$ be nonexpansive, and let $\alpha \in \left]0,1\right[$. Then T is *averaged* with constant α, or *α-averaged*, if there exists a nonexpansive operator $R \colon D \to \mathcal{H}$ such that $T = (1-\alpha)\mathrm{Id} + \alpha R$.

Remark 4.34 Let D be a nonempty subset of $\mathcal{H}$ and let $T \colon D \to \mathcal{H}$.

(i) If T is averaged, then it is nonexpansive.
(ii) If T is nonexpansive, it is not necessarily averaged: consider $T = -\mathrm{Id} \colon \mathcal{H} \to \mathcal{H}$ when $\mathcal{H} \neq \{0\}$.
(iii) It follows from Proposition 4.4 that T is firmly nonexpansive if and only if it is 1/2-averaged.
(iv) Let $\beta \in \mathbb{R}_{++}$. Then it follows from (iii) that T is β-cocoercive if and only if βT is 1/2-averaged.

Proposition 4.35 *Let D be a nonempty subset of $\mathcal{H}$, let $T \colon D \to \mathcal{H}$ be nonexpansive, and let $\alpha \in \left]0,1\right[$. Then the following are equivalent:*

(i) *T is α-averaged.*
(ii) $(1-1/\alpha)\mathrm{Id}+(1/\alpha)T$ *is nonexpansive.*
(iii) $(\forall x \in D)(\forall y \in D)\ \|Tx-Ty\|^2 \leqslant \|x-y\|^2-\dfrac{1-\alpha}{\alpha}\|(\mathrm{Id}-T)x-(\mathrm{Id}-T)y\|^2.$
(iv) $(\forall x \in D)(\forall y \in D)\ \|Tx-Ty\|^2+(1-2\alpha)\|x-y\|^2 \leqslant 2(1-\alpha)\langle x-y \mid Tx-Ty\rangle.$

Proof. Let x and y be in D, and apply Lemma 2.17(i) to $(x-y, Tx-Ty)$. ☐

Remark 4.36 It follows from the implication (i)⇒(iii) in Proposition 4.35 that averaged operators are strictly quasinonexpansive.

Remark 4.37 It follows from Proposition 4.35(iii) that if $T\colon D \to \mathcal{H}$ is α-averaged with $\alpha \in \left]0,1/2\right]$, then T is firmly nonexpansive.

Proposition 4.38 *Let D be a nonempty subset of $\mathcal{H}$, let $\rho \in \left]0,1\right[$, let $T\colon D \to \mathcal{H}$ be ρ-Lipschitz continuous, and set $\alpha=(\rho+1)/2$. Then T is α-averaged.*

Proof. The operator

$$\left(1-\frac{1}{\alpha}\right)\mathrm{Id}+\frac{1}{\alpha}T=\frac{\rho-1}{\rho+1}\mathrm{Id}+\frac{2}{\rho+1}T \tag{4.33}$$

is Lipschitz continuous with constant $(1-\rho)/(1+\rho)+2\rho/(1+\rho)=1$. In view of Proposition 4.35, T is α-averaged. ☐

Proposition 4.39 *Let D be a nonempty subset of $\mathcal{H}$, let $T\colon D \to \mathcal{H}$, let $\beta \in \mathbb{R}_{++}$, and let $\gamma \in \left]0,2\beta\right[$. Then T is β-cocoercive if and only if* $\mathrm{Id}-\gamma T$ *is $\gamma/(2\beta)$-averaged.*

Proof. It follows from Definition 4.10, Proposition 4.4, and Proposition 4.35 that T is β-cocoercive ⇔ βT is firmly nonexpansive ⇔ $\mathrm{Id}-\beta T$ is firmly nonexpansive ⇔ $(1-2\beta/\gamma)\mathrm{Id}+(2\beta/\gamma)(\mathrm{Id}-\gamma T)=2(\mathrm{Id}-\beta T)-\mathrm{Id}$ is nonexpansive ⇔ $\mathrm{Id}-\gamma T$ is $\gamma/(2\beta)$-averaged. ☐

We now describe operations that preserve averagedness.

Proposition 4.40 *Let D be a nonempty subset of $\mathcal{H}$, let $T\colon D \to \mathcal{H}$, let $\alpha \in \left]0,1\right[$, and let $\lambda \in \left]0,1/\alpha\right[$. Then T is α-averaged if and only if* $(1-\lambda)\mathrm{Id}+\lambda T$ *is $\lambda\alpha$-averaged.*

Proof. Set $R=(1-\alpha^{-1})\mathrm{Id}+\alpha^{-1}T$. Then the conclusion follows from the identities $T=(1-\alpha)\mathrm{Id}+\alpha R$ and $(1-\lambda)\mathrm{Id}+\lambda T=(1-\lambda\alpha)\mathrm{Id}+\lambda\alpha R$. ☐

Corollary 4.41 *Let D be a nonempty subset of $\mathcal{H}$, let $T\colon D \to \mathcal{H}$, and let $\lambda \in \left]0,2\right[$. Then T is firmly nonexpansive if and only if $(1-\lambda)\mathrm{Id}+\lambda T$ is $\lambda/2$-averaged.*

Proof. Set $\alpha=1/2$ in Proposition 4.40 and use Remark 4.34(iii). ☐

Proposition 4.42 *Let D be a nonempty subset of $\mathcal{H}$, let $(T_i)_{i\in I}$ be a finite family of nonexpansive operators from D to $\mathcal{H}$, let $(\omega_i)_{i\in I}$ be real numbers in $]0,1]$ such that $\sum_{i\in I}\omega_i = 1$, and let $(\alpha_i)_{i\in I}$ be real numbers in $]0,1[$ such that, for every $i \in I$, T_i is α_i-averaged, and set $\alpha = \sum_{i\in I}\omega_i\alpha_i$. Then $\sum_{i\in I}\omega_i T_i$ is α-averaged.*

Proof. For every $i \in I$, there exists a nonexpansive operator $R_i\colon D \to \mathcal{H}$ such that $T_i = (1-\alpha_i)\mathrm{Id} + \alpha_i R_i$. Now set $R = \sum_{i\in I}(\omega_i\alpha_i/\alpha)R_i$. Then Proposition 4.9(i) asserts that R is nonexpansive and

$$\sum_{i\in I}\omega_i T_i = \sum_{i\in I}\omega_i(1-\alpha_i)\mathrm{Id} + \sum_{i\in I}\omega_i\alpha_i R_i = (1-\alpha)\mathrm{Id} + \alpha R. \tag{4.34}$$

We conclude that T is α-averaged. ☐

Remark 4.43 In view of Remark 4.34(iii), setting $\alpha_i \equiv 1/2$ in Proposition 4.42 we recover Example 4.7, which states that a convex combination of firmly nonexpansive operators is firmly nonexpansive.

The next result shows that the compositions of averaged operators are averaged.

Proposition 4.44 *Let D be a nonempty subset of $\mathcal{H}$, let $\alpha_1 \in]0,1[$, let $\alpha_2 \in]0,1[$, let $T_1\colon D \to D$ be α_1-averaged, and let $T_2\colon D \to D$ be α_2-averaged. Set*

$$T = T_1T_2 \quad \text{and} \quad \alpha = \frac{\alpha_1+\alpha_2-2\alpha_1\alpha_2}{1-\alpha_1\alpha_2}. \tag{4.35}$$

Then $\alpha \in]0,1[$ and T is α-averaged.

Proof. Since $\alpha_1(1-\alpha_2) < (1-\alpha_2)$, we have $\alpha_1+\alpha_2 < 1+\alpha_1\alpha_2$ and, therefore, $\alpha \in]0,1[$. Now let $x \in D$ and $y \in D$. It follows from Proposition 4.35 that

$$\begin{aligned}\|T_1T_2x - T_1T_2y\|^2 &\leqslant \|T_2x - T_2y\|^2 - \frac{1-\alpha_1}{\alpha_1}\|(\mathrm{Id}-T_1)T_2x - (\mathrm{Id}-T_1)T_2y\|^2\\ &\leqslant \|x-y\|^2 - \frac{1-\alpha_2}{\alpha_2}\|(\mathrm{Id}-T_2)x - (\mathrm{Id}-T_2)y\|^2\\ &\quad - \frac{1-\alpha_1}{\alpha_1}\|(\mathrm{Id}-T_1)T_2x - (\mathrm{Id}-T_1)T_2y\|^2.\end{aligned} \tag{4.36}$$

Now set

$$\tau = \frac{1-\alpha_1}{\alpha_1} + \frac{1-\alpha_2}{\alpha_2}. \tag{4.37}$$

By Corollary 2.15, we have

$$\begin{aligned}&\frac{1-\alpha_1}{\tau\alpha_1}\|(\mathrm{Id}-T_1)T_2x - (\mathrm{Id}-T_1)T_2y\|^2 + \frac{1-\alpha_2}{\tau\alpha_2}\|(\mathrm{Id}-T_2)y - (\mathrm{Id}-T_2)x\|^2\\ &= \left\|\frac{1-\alpha_1}{\tau\alpha_1}\big((\mathrm{Id}-T_1)T_2x - (\mathrm{Id}-T_1)T_2y\big) + \frac{1-\alpha_2}{\tau\alpha_2}\big((\mathrm{Id}-T_2)y - (\mathrm{Id}-T_2)x\big)\right\|^2\end{aligned}$$

$$
\begin{aligned}
&+\frac{(1-\alpha_1)(1-\alpha_2)}{\tau^2\alpha_1\alpha_2}\|T_1T_2y-T_1T_2x-y+x\|^2\\
\geqslant\ &\frac{(1-\alpha_1)(1-\alpha_2)}{\tau^2\alpha_1\alpha_2}\|(\mathrm{Id}-T_1T_2)x-(\mathrm{Id}-T_1T_2)y\|^2. \qquad (4.38)
\end{aligned}
$$

Combining (4.36), (4.38), and (4.35) yields

$$
\begin{aligned}
&\|T_1T_2x-T_1T_2y\|^2\\
&\quad\leqslant\|x-y\|^2-\frac{(1-\alpha_1)(1-\alpha_2)}{\tau\alpha_1\alpha_2}\|(\mathrm{Id}-T_1T_2)\,x-(\mathrm{Id}-T_1T_2)\,y\|^2\\
&\quad=\|x-y\|^2-\frac{1-\alpha_1-\alpha_2+\alpha_1\alpha_2}{\alpha_1+\alpha_2-2\alpha_1\alpha_2}\|(\mathrm{Id}-T_1T_2)\,x-(\mathrm{Id}-T_1T_2)\,y\|^2\\
&\quad=\|x-y\|^2-\frac{1-\alpha}{\alpha}\|(\mathrm{Id}-T)\,x-(\mathrm{Id}-T)\,y\|^2. \qquad (4.39)
\end{aligned}
$$

In view of Proposition 4.35, we conclude that T is α-averaged. ☐

Example 4.45 Taking $\alpha_1=\alpha_2=1/2$ in Proposition 4.44 and invoking Remark 4.34(iii), we obtain that the composition of two firmly nonexpansive operators is $2/3$-averaged. However, it need not be firmly nonexpansive: suppose that $\mathcal{H}=\mathbb{R}^2$, set $U=\mathbb{R}(1,1)$, and set $V=\mathbb{R}(1,0)$. Then P_U and P_V are firmly nonexpansive by Proposition 4.16, but P_UP_V is not.

Next, we consider the finite composition of averaged operators.

Proposition 4.46 *Let D be a nonempty subset of $\mathcal{H}$, let $m\geqslant 2$ be an integer, set $I=\{1,\ldots,m\}$, let $(T_i)_{i\in I}$ be a family of operators from D to D, and let $(\alpha_i)_{i\in I}$ be real numbers in $]0,1[$ such that, for every $i\in I$, T_i is α_i-averaged. Set*

$$
T=T_1\cdots T_m \quad\textit{and}\quad \alpha=\frac{1}{1+\dfrac{1}{\displaystyle\sum_{i=1}^m\frac{\alpha_i}{1-\alpha_i}}}. \qquad (4.40)
$$

Then T is α-averaged.

Proof. We proceed by induction on $k\in\{2,\ldots,m\}$. To this end, let us set $(\forall k\in\{2,\ldots,m\})$ $\beta_k=(1+(\sum_{i=1}^k\alpha_i/(1-\alpha_i))^{-1})^{-1}$. By Proposition 4.44, the claim is true for $k=2$ since

$$
\frac{\alpha_1+\alpha_2-2\alpha_1\alpha_2}{1-\alpha_1\alpha_2}=\frac{1}{1+\dfrac{1}{\dfrac{\alpha_1}{1-\alpha_1}+\dfrac{\alpha_2}{1-\alpha_2}}}=\beta_2. \qquad (4.41)
$$

Now assume that, for some $k\in\{2,\ldots,m-1\}$, $T_1\cdots T_k$ is β_k-averaged. Then we deduce from Proposition 4.44 and (4.41) that the averagedness constant of $(T_1\cdots T_k)T_{k+1}$ is

$$\frac{1}{1+\cfrac{1}{\cfrac{1}{\beta_k^{-1}-1}+\cfrac{\alpha_{k+1}}{1-\alpha_{k+1}}}} = \frac{1}{1+\cfrac{1}{\left(\displaystyle\sum_{i=1}^{k}\frac{\alpha_i}{1-\alpha_i}\right)+\cfrac{\alpha_{k+1}}{1-\alpha_{k+1}}}} = \beta_{k+1}, \tag{4.42}$$

which concludes the induction argument. ☐

4.6 Common Fixed Points

The first proposition concerns the fixed point set of convex combinations of quasinonexpansive operators and the second one that of compositions of strictly quasinonexpansive operators.

Proposition 4.47 *Let D be a nonempty subset of $\mathcal{H}$, let $(T_i)_{i\in I}$ be a finite family of quasinonexpansive operators from D to $\mathcal{H}$ such that $\bigcap_{i\in I}\operatorname{Fix}T_i \neq \varnothing$, and let $(\omega_i)_{i\in I}$ be strictly positive real numbers such that $\sum_{i\in I}\omega_i = 1$. Then $\operatorname{Fix}\sum_{i\in I}\omega_i T_i = \bigcap_{i\in I}\operatorname{Fix}T_i$.*

Proof. Set $T = \sum_{i\in I}\omega_i T_i$. It is clear that $\bigcap_{i\in I}\operatorname{Fix}T_i \subset \operatorname{Fix}T$. To prove the reverse inclusion, let $y \in \bigcap_{i\in I}\operatorname{Fix}T_i$. Then (4.5) yields

$$\begin{aligned}(\forall i\in I)(\forall x\in D)\quad 2\langle T_i x - x \mid x-y\rangle &= \|T_i x - y\|^2 - \|T_i x - x\|^2 - \|x-y\|^2\\ &\leqslant -\|T_i x - x\|^2.\end{aligned} \tag{4.43}$$

Now let $x \in \operatorname{Fix}T$. Then we derive from (4.43) that

$$0 = 2\langle Tx - x \mid x-y\rangle = 2\sum_{i\in I}\omega_i\langle T_i x - x \mid x-y\rangle \leqslant -\sum_{i\in I}\omega_i\|T_i x - x\|^2. \tag{4.44}$$

Therefore $\sum_{i\in I}\omega_i\|T_i x - x\|^2 = 0$, and we conclude that $x \in \bigcap_{i\in I}\operatorname{Fix}T_i$. ☐

Corollary 4.48 *Let D be a nonempty subset of $\mathcal{H}$, let $(T_i)_{i\in I}$ be a finite family of firmly quasinonexpansive operators from D to $\mathcal{H}$, let $(\omega_i)_{i\in I}$ be real numbers in $]0,1]$ such that $\sum_{i\in I}\omega_i = 1$, and set $T = \sum_{i\in I}\omega_i T_i$. Then T is firmly quasinonexpansive.*

Proof. Let $x \in D$ and suppose that $y \in \operatorname{Fix}T$. Then, by Proposition 4.47, $(\forall i \in I)$ $y \in \operatorname{Fix}T_i$. Therefore, appealing to the equivalence (i)⇔(v) in Proposition 4.2, we obtain $(\forall i\in I)$ $\|T_i x - x\|^2 \leqslant \langle y - x \mid T_i x - x\rangle$. In view of Lemma 2.14(ii), we conclude that $\|Tx - x\|^2 \leqslant \sum_{i\in I}\omega_i\|T_i x - x\|^2 \leqslant \sum_{i\in I}\omega_i\langle y-x \mid T_i x - x\rangle = \langle y - x \mid Tx - x\rangle$. ☐

Proposition 4.49 *Let D be a nonempty subset of $\mathcal{H}$, and let T_1 and T_2 be quasinonexpansive operators from D to D. Suppose that T_1 or T_2 is strictly quasinonexpansive, and that $\operatorname{Fix}T_1 \cap \operatorname{Fix}T_2 \neq \varnothing$. Then the following hold:*

(i) $\operatorname{Fix} T_1T_2 = \operatorname{Fix} T_1 \cap \operatorname{Fix} T_2$.
(ii) T_1T_2 *is quasinonexpansive.*
(iii) *Suppose that* T_1 *and* T_2 *are strictly quasinonexpansive. Then* T_1T_2 *is strictly quasinonexpansive.*

Proof. (i): It is clear that $\operatorname{Fix} T_1 \cap \operatorname{Fix} T_2 \subset \operatorname{Fix} T_1T_2$. Now let $x \in \operatorname{Fix} T_1T_2$ and let $y \in \operatorname{Fix} T_1 \cap \operatorname{Fix} T_2$. We consider three cases.
(a) $T_2x \in \operatorname{Fix} T_1$. Then $T_2x = T_1T_2x = x \in \operatorname{Fix} T_1 \cap \operatorname{Fix} T_2$.
(b) $x \in \operatorname{Fix} T_2$. Then $T_1x = T_1T_2x = x \in \operatorname{Fix} T_1 \cap \operatorname{Fix} T_2$.
(c) $T_2x \notin \operatorname{Fix} T_1$ and $x \notin \operatorname{Fix} T_2$. Since T_1 or T_2 is strictly quasinonexpansive, at least one of the inequalities in $\|x-y\| = \|T_1T_2x-y\| \leqslant \|T_2x-y\| \leqslant \|x-y\|$ is strict, which is impossible.

(ii): Let $x \in D$ and let $y \in \operatorname{Fix} T_1T_2 = \operatorname{Fix} T_1 \cap \operatorname{Fix} T_2$. Then $\|T_1T_2x-y\| \leqslant \|T_2x-y\| \leqslant \|x-y\|$, and therefore T_1T_2 is quasinonexpansive.

(iii): Let $x \in D \smallsetminus \operatorname{Fix} T_1T_2$ and let $y \in \operatorname{Fix} T_1T_2 = \operatorname{Fix} T_1 \cap \operatorname{Fix} T_2$. If $x \notin \operatorname{Fix} T_2$, then $\|T_1T_2x-y\| \leqslant \|T_2x-y\| < \|x-y\|$. Finally, if $x \in \operatorname{Fix} T_2$, then $x \notin \operatorname{Fix} T_1$ (for otherwise $T_1x = x \in \operatorname{Fix} T_1 \cap \operatorname{Fix} T_2$, which is impossible) and hence $\|T_1T_2x-y\| < \|x-y\|$. ☐

Corollary 4.50 *Let* D *be a nonempty subset of* $\mathcal{H}$, *let* m *be a strictly positive integer, set* $I = \{1,\dots,m\}$, *let* $(T_i)_{i\in I}$ *be a family of strictly quasinonexpansive operators from* D *to* D *such that* $\bigcap_{i\in I} \operatorname{Fix} T_i \neq \varnothing$, *and set* $T = T_1 \cdots T_m$. *Then* T *is strictly quasinonexpansive and* $\operatorname{Fix} T = \bigcap_{i\in I} \operatorname{Fix} T_i$.

Proof. We proceed by strong induction on m. The result is clear for $m = 1$ and, for $m = 2$, by Proposition 4.49. Now suppose that $m \geqslant 2$ and that the result holds for up to m operators. Let $(T_i)_{1\leqslant i\leqslant m+1}$ be a family of strictly quasinonexpansive operators from D to D such that $\bigcap_{i=1}^{m+1} \operatorname{Fix} T_i \neq \varnothing$. Set $R_1 = T_1 \cdots T_m$ and $R_2 = T_{m+1}$. Then R_2 is strictly quasinonexpansive with $\operatorname{Fix} R_2 = \operatorname{Fix} T_{m+1}$, and, by the induction hypothesis, R_1 is strictly quasinonexpansive with $\operatorname{Fix} R_1 = \bigcap_{i=1}^{m} \operatorname{Fix} T_i$. Therefore, by Proposition 4.49(iii)&(i), $R_1R_2 = T_1 \cdots T_{m+1}$ is strictly quasinonexpansive and $\operatorname{Fix} T_1 \cdots T_{m+1} = \operatorname{Fix} R_1R_2 = \operatorname{Fix} R_1 \cap \operatorname{Fix} R_2 = \bigcap_{i=1}^{m+1} \operatorname{Fix} T_i$. ☐

Corollary 4.51 *Let* D *be a nonempty subset of* $\mathcal{H}$, *let* m *be a strictly positive integer, set* $I = \{1,\dots,m\}$, *let* $(T_i)_{i\in I}$ *be a family of averaged nonexpansive operators from* D *to* D *such that* $\bigcap_{i\in I} \operatorname{Fix} T_i \neq \varnothing$, *and set* $T = T_1 \cdots T_m$. *Then* $\operatorname{Fix} T = \bigcap_{i\in I} \operatorname{Fix} T_i$.

Proof. In view of Remark 4.36, this follows from Corollary 4.50. ☐

Exercises

Exercise 4.1 Let D be a nonempty subset of $\mathcal{H}$, let $T\colon D \to \mathcal{H}$ be firmly nonexpansive, and let x and y be in D. Show that the following are equivalent:

(i) $\|Tx - Ty\|^2 + \|(\mathrm{Id} - T)x - (\mathrm{Id} - T)y\|^2 = \|x - y\|^2$.
(ii) $\|Tx - Ty\|^2 = \langle x - y \mid Tx - Ty\rangle$.
(iii) $\langle Tx - Ty \mid (\mathrm{Id} - T)x - (\mathrm{Id} - T)y\rangle = 0$.
(iv) $(\forall \alpha \in \mathbb{R})$ $\|Tx - Ty\| \leqslant \|\alpha(x - y) + (1 - \alpha)(Tx - Ty)\|$.
(v) $\|(2T - \mathrm{Id})x - (2T - \mathrm{Id})y\| = \|x - y\|$.

Exercise 4.2 Let U be a nonempty open interval in $\mathbb{R}$, let D be a closed interval contained in U, and suppose that $\widetilde{T}\colon U \to \mathbb{R}$ is differentiable on U. Set $T = \widetilde{T}|_D$. Show the following:

(i) T is firmly nonexpansive if and only if $\operatorname{ran} T' \subset [0, 1]$.
(ii) T is nonexpansive if and only if $\operatorname{ran} T' \subset [-1, 1]$.

Exercise 4.3 Let D be a nonempty subset of $\mathbb{R}$, and let $T\colon D \to \mathbb{R}$. Show that T is firmly nonexpansive if and only if T is nonexpansive and increasing. Provided that $\operatorname{ran} T \subset D$, deduce that if T is firmly nonexpansive, then so is $T \circ T$.

Exercise 4.4 Let D be a nonempty subset of $\mathcal{H}$ and let $(T_i)_{i\in I}$ be a finite family of firmly nonexpansive mappings from D to $\mathcal{H}$, let $(\omega_i)_{i\in I}$ be real numbers in $]0, 1]$ such that $\sum_{i\in I} \omega_i = 1$, and set $T = \sum_{i\in I} \omega_i T_i$. Let x and y be points in D such that $Tx = Ty$. Show that $(\forall i \in I)$ $T_i x = T_i y$. Deduce that, if some T_i is injective, then so is T.

Exercise 4.5 Suppose that $\mathcal{H} \neq \{0\}$. Without using Example 4.17, show that every firmly nonexpansive operator is nonexpansive, but not vice versa. Furthermore, show that every nonexpansive operator is quasinonexpansive, but not vice versa.

Exercise 4.6 Let $T \in \mathcal{B}(\mathcal{H})$. Show that the following are equivalent:

(i) T is nonexpansive.
(ii) $\|T\| \leqslant 1$.
(iii) T is quasinonexpansive.

Exercise 4.7 Let $T \in \mathcal{B}(\mathcal{H})$.

(i) Suppose that

$$(\exists\, \alpha \in \mathbb{R}_{++})(\forall x \in \mathcal{H}) \quad \langle Tx \mid x\rangle \geqslant \alpha\|x\|^2. \tag{4.45}$$

Show that $\alpha\|T\|^{-2}T$ is firmly nonexpansive.
(ii) Suppose that $\mathcal{H} = \mathbb{R}^3$ and set $T\colon (\xi_1, \xi_2, \xi_3) \mapsto (1/2)(\xi_1 - \xi_2, \xi_1 + \xi_2, 0)$. Show that T does not satisfy (4.45), is not self-adjoint, but is firmly nonexpansive.

Exercise 4.8 Let $T \in \mathcal{B}(\mathcal{H})$ be nonexpansive. Show that $\operatorname{Fix} T = \operatorname{Fix} T^*$.

Exercise 4.9 Let $T \in \mathcal{B}(\mathcal{H})$ be nonexpansive and such that $\operatorname{ran}(\mathrm{Id} - T)$ is closed. Show that there exists $\beta \in \mathbb{R}_{++}$ such that $(\forall x \in \mathcal{H})$ $\beta d_{\operatorname{Fix} T}(x) \leqslant \|x - Tx\|$.

Exercise 4.10 Let $\mathcal{K}$ be a real Hilbert space, let $T\colon \mathcal{K} \to \mathcal{K}$ be firmly nonexpansive, let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ be such that $\|L\| \leqslant 1$, let $\bar{x} \in \mathcal{H}$, and let $\bar{z} \in \mathcal{K}$. Show that $x \mapsto \bar{x} + L^*T(\bar{z} + Lx)$ is firmly nonexpansive.

Exercise 4.11 As seen in Proposition 4.4, if $T\colon \mathcal{H} \to \mathcal{H}$ is firmly nonexpansive, then so is $\mathrm{Id} - T$. By way of examples, show that if C is a nonempty closed convex subset of $\mathcal{H}$, then $\mathrm{Id} - P_C$ may or may not be a projector.

Exercise 4.12 Let C and D be closed linear subspaces of $\mathcal{H}$ such that $C \subset D$. Show that $P_C = P_D P_C = P_C P_D$.

Exercise 4.13 Let $T\colon \mathcal{H} \to \mathcal{H}$ be firmly nonexpansive, let $\lambda \in \mathbb{R}$, and set $T_\lambda = \lambda T + (1 - \lambda)\mathrm{Id}$. Show that T_λ is nonexpansive for $\lambda \in [0, 2]$ and that this interval is the largest possible with this property.

Exercise 4.14 Let D be a nonempty subset of $\mathcal{H}$, let $(T_i)_{i\in I}$ be a finite family of quasinonexpansive operators from D to $\mathcal{H}$ such that $\bigcap_{i\in I} \operatorname{Fix} T_i \neq \varnothing$, and let $(\omega_i)_{i\in I}$ be strictly positive real numbers such that $\sum_{i\in I} \omega_i = 1$. Show that $\sum_{i\in I} \omega_i T_i$ is quasinonexpansive.

Exercise 4.15 Let D be a nonempty subset of $\mathcal{H}$, and let T_1 and T_2 be firmly nonexpansive operators from D to $\mathcal{H}$. Show that $T_1 - T_2$ and $\mathrm{Id} - T_1 - T_2$ are nonexpansive.

Exercise 4.16 Let T, T_1, and T_2 be operators from $\mathcal{H}$ to $\mathcal{H}$.

(i) Show that, if T is firmly nonexpansive, then $T \circ T$ may fail to be firmly nonexpansive even when $\operatorname{Fix} T \neq \varnothing$. Compare with Exercise 4.3.
(ii) Show that, if T_1 and T_2 are both nonexpansive, then so is $T_2 \circ T_1$.
(iii) Show that, if T is quasinonexpansive, then $T \circ T$ may fail to be quasinonexpansive even when $\mathcal{H} = \mathbb{R}$ and $\operatorname{Fix} T \neq \varnothing$.

Exercise 4.17 Provide the details for the counterexample in Example 4.45.

Exercise 4.18 Let D be a nonempty compact subset of $\mathcal{H}$ and suppose that $T\colon D \to \mathcal{H}$ is firmly nonexpansive and that $\mathrm{Id} - T$ is injective. Show that for every $\delta \in \mathbb{R}_{++}$, there exists $\beta \in [0, 1[$ such that, if x and y belong to D and $\|x - y\| \geqslant \delta$, then $\|Tx - Ty\| \leqslant \beta\|x - y\|$. In addition, provide, for $\mathcal{H} = \mathbb{R}$, a set D and an operator T such that the hypothesis holds and such that, for every $\beta \in [0, 1[$, T is not Lipschitz continuous with constant β.

Exercise 4.19 Prove Corollary 4.25 using Proposition 4.4(v).

Exercise 4.20 Let D be a nonempty convex subset of $\mathcal{H}$ and let $T\colon D \to \mathcal{H}$ be firmly nonexpansive. Show that $\operatorname{Fix} T$ is convex using Proposition 4.8(ii).

Exercise 4.21 Let D be a nonempty convex subset of $\mathcal{H}$ and let $T\colon D \to \mathcal{H}$ be nonexpansive. Show that $\operatorname{Fix} T$ is convex using Proposition 4.8(iv).

Exercise 4.22 Let $T\colon \mathcal{H} \to \mathcal{H}$ be a firmly nonexpansive operator such that $C = \operatorname{ran} T = \operatorname{Fix} T \neq \varnothing$. Show that $T = P_C$.

Exercise 4.23 Let D be a nonempty closed convex subset of $\mathcal{H}$, and let $T\colon D \to \mathcal{H}$ be firmly nonexpansive. Show that, for every $y \in \mathcal{H}$, $T^{-1}y$ is closed and convex.

Exercise 4.24 Provide a simple proof of Theorem 4.27 in the case when $\mathcal{H}$ is finite-dimensional.

Exercise 4.25 Using Example 4.20, find a firmly nonexpansive operator $T\colon \mathcal{H} \to \mathcal{H}$, a sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$, and two points x and u in $\mathcal{H}$ such that $x_n \rightharpoonup x$, $x_n - Tx_n \rightharpoonup u$, but $x - Tx \neq u$. Compare to Theorem 4.27 and conclude that $\operatorname{gra}(\mathrm{Id} - T)$ is not sequentially closed in $\mathcal{H}^{\mathrm{weak}} \times \mathcal{H}^{\mathrm{weak}}$.

Exercise 4.26 Show that each of the following assumptions on D in Theorem 4.29 is necessary: boundedness, closedness, convexity.

Exercise 4.27 Use items (ix) and (x) of Proposition 4.30 to prove Theorem 4.29 without using Corollary 4.28.

Exercise 4.28 Let D be a nonempty subset of $\mathcal{H}$, let m be an integer such that $m \geqslant 2$, and let $(T_i)_{1\leqslant i\leqslant m}$ be a finite family of m quasinonexpansive operators from D to D such that $\bigcap_{i=1}^m \operatorname{Fix} T_i \neq \varnothing$ and $m-1$ of these operators are strictly quasinonexpansive. Show that $T_1 \cdots T_m$ is quasinonexpansive and $\operatorname{Fix} T_1 \cdots T_m = \bigcap_{i=1}^m \operatorname{Fix} T_i$.

Chapter 5
Fejér Monotonicity and Fixed Point Iterations

A sequence is Fejér monotone with respect to a set C if no point of the sequence is strictly farther from any point in C than its predecessor. Such sequences possess attractive properties that simplify the analysis of their asymptotic behavior. In this chapter, we provide the basic theory for Fejér monotone sequences and apply it to obtain in a systematic fashion convergence results for various classical iterations involving (quasi)nonexpansive operators.

5.1 Fejér Monotone Sequences

The following notion is central in the study of various iterative methods, in particular in connection with the construction of fixed points of nonexpansive operators.

Definition 5.1 Let C be a nonempty subset of $\mathcal{H}$ and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$. Then $(x_n)_{n\in\mathbb{N}}$ is *Fejér monotone* with respect to C if

$$(\forall x \in C)(\forall n \in \mathbb{N}) \quad \|x_{n+1} - x\| \leqslant \|x_n - x\|. \tag{5.1}$$

Example 5.2 Let $(x_n)_{n\in\mathbb{N}}$ be a bounded sequence in $\mathbb{R}$ that is increasing (respectively decreasing). Then $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to $[\sup\{x_n\}_{n\in\mathbb{N}}, +\infty[$ (respectively $]-\infty, \inf\{x_n\}_{n\in\mathbb{N}}]$).

Example 5.3 Let D be a nonempty subset of $\mathcal{H}$, let $T\colon D \to D$ be a quasinonexpansive operator such that $\operatorname{Fix} T \neq \varnothing$, and let $x_0 \in D$. Set $(\forall n \in \mathbb{N})$ $x_{n+1} = Tx_n$. Then $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to $\operatorname{Fix} T$.

We start with some basic properties.

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-48311-5_31

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_5

Proposition 5.4 *Let C be a nonempty subset of $\mathcal{H}$ and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$. Suppose that $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to C. Then the following hold:*

(i) *$(x_n)_{n\in\mathbb{N}}$ is bounded.*
(ii) *For every $x \in C$, $(\|x_n - x\|)_{n\in\mathbb{N}}$ converges.*
(iii) *$(d_C(x_n))_{n\in\mathbb{N}}$ is decreasing and converges.*
(iv) *Let $m \in \mathbb{N}$ and let $n \in \mathbb{N}$. Then $\|x_{n+m} - x_n\| \leqslant 2d_C(x_n)$.*

Proof. (i): Let $x \in C$. Then (5.1) implies that $(x_n)_{n\in\mathbb{N}}$ lies in $B(x; \|x_0 - x\|)$.
(ii): Clear from (5.1).
(iii): Taking the infimum in (5.1) over $x \in C$ yields $(\forall n \in \mathbb{N})\ d_C(x_{n+1}) \leqslant d_C(x_n)$.
(iv): Let $x \in C$. We derive from (5.1) that $\|x_{n+m} - x_n\| \leqslant \|x_{n+m} - x\| + \|x_n - x\| \leqslant 2\|x_n - x\|$. Taking the infimum over $x \in C$ yields the claim. ☐

The next result concerns weak convergence.

Theorem 5.5 *Let C be a nonempty subset of $\mathcal{H}$ and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$. Suppose that $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to C and that every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ belongs to C. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in C.*

Proof. The result follows from Proposition 5.4(ii) and Lemma 2.47. ☐

Example 5.6 Suppose that $\mathcal{H}$ is infinite-dimensional and let $(x_n)_{n\in\mathbb{N}}$ be an orthonormal sequence in $\mathcal{H}$. Then $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to $\{0\}$. As seen in Example 2.32, $x_n \rightharpoonup 0$ but $x_n \not\to 0$.

While a Fejér monotone sequence with respect to a closed convex set C need not converge strongly, its "shadow" on C always does.

Proposition 5.7 *Let C be a nonempty closed convex subset of $\mathcal{H}$ and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$. Suppose that $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to C. Then the* shadow sequence $(P_C x_n)_{n\in\mathbb{N}}$ *converges strongly to a point z in C and $(\forall x \in C)\ \lim \|x_n - z\|^2 \leqslant \|x - z\|^2 + \lim \|x_n - x\|^2$.*

Proof. It follows from (5.1) and (3.10) that, for every m and n in $\mathbb{N}$,

$$\begin{aligned}
\|P_C x_n - P_C x_{n+m}\|^2 &= \|P_C x_n - x_{n+m}\|^2 + \|x_{n+m} - P_C x_{n+m}\|^2 \\
&\quad + 2\langle P_C x_n - x_{n+m} \mid x_{n+m} - P_C x_{n+m}\rangle \\
&\leqslant \|P_C x_n - x_n\|^2 + d_C^2(x_{n+m}) \\
&\quad + 2\langle P_C x_n - P_C x_{n+m} \mid x_{n+m} - P_C x_{n+m}\rangle \\
&\quad + 2\langle P_C x_{n+m} - x_{n+m} \mid x_{n+m} - P_C x_{n+m}\rangle \\
&\leqslant d_C^2(x_n) - d_C^2(x_{n+m}). \qquad (5.2)
\end{aligned}$$

Consequently, since $(d_C(x_n))_{n\in\mathbb{N}}$ was seen in Proposition 5.4(iii) to converge, $(P_C x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in the complete set C. Now, set $z = \lim P_C x_n$ and $\beta = \sup_{n\in\mathbb{N}} \|x_n - z\|$, let $x \in C$, and let $n \in \mathbb{N}$. By Proposition 5.4(i), $\beta < +\infty$. Hence, by Theorem 3.16 and Cauchy-Schwarz,

$$\begin{aligned}\|x_n - x\|^2 &= \|x_n - P_C x_n\|^2 + \|P_C x_n - x\|^2 + 2\,\langle x_n - P_C x_n \mid P_C x_n - x\rangle \\ &\geqslant \|x_n - P_C x_n\|^2 + \|P_C x_n - x\|^2 \\ &= \|x_n - z\|^2 + \|z - P_C x_n\|^2 + 2\,\langle x_n - z \mid z - P_C x_n\rangle \\ &\quad + \|P_C x_n - x\|^2 \\ &\geqslant \|x_n - z\|^2 + (1-2\beta)\|z - P_C x_n\|^2 + \|P_C x_n - x\|^2. \end{aligned} \tag{5.3}$$

The inequality follows by taking the limit and using Proposition 5.4(ii). □

Corollary 5.8 *Let C be a nonempty closed convex subset of $\mathcal{H}$, let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$, and let $x \in C$. Suppose that $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to C and that $x_n \rightharpoonup x$. Then $P_C x_n \to x$.*

Proof. By Proposition 5.7, $(P_C x_n)_{n\in\mathbb{N}}$ converges strongly to some point $y \in C$. Hence, since $x - P_C x_n \to x - y$ and $x_n - P_C x_n \rightharpoonup x - y$, it follows from Theorem 3.16 and Lemma 2.51(iii) that $0 \geqslant \langle x - P_C x_n \mid x_n - P_C x_n\rangle \to \|x - y\|^2$. Thus, $x = y$. □

For sequences that are Fejér monotone with respect to closed affine subspaces, Proposition 5.7 can be strengthened.

Proposition 5.9 *Let C be a closed affine subspace of $\mathcal{H}$ and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$. Suppose that $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to C. Then the following hold:*

(i) $(\forall n \in \mathbb{N})$ $P_C x_n = P_C x_0$.
(ii) *Suppose that every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ belongs to C. Then $x_n \rightharpoonup P_C x_0$.*

Proof. (i): Fix $n \in \mathbb{N}$, $\alpha \in \mathbb{R}$, and set $y_\alpha = \alpha P_C x_0 + (1-\alpha) P_C x_n$. Since C is an affine subspace, $y_\alpha \in C$, and it therefore follows from Corollary 3.22(i) and (5.1) that

$$\begin{aligned}\alpha^2 \|P_C x_n - P_C x_0\|^2 &= \|P_C x_n - y_\alpha\|^2 \\ &\leqslant \|x_n - P_C x_n\|^2 + \|P_C x_n - y_\alpha\|^2 \\ &= \|x_n - y_\alpha\|^2 \\ &\leqslant \|x_0 - y_\alpha\|^2 \\ &= \|x_0 - P_C x_0\|^2 + \|P_C x_0 - y_\alpha\|^2 \\ &= d_C^2(x_0) + (1-\alpha)^2 \|P_C x_n - P_C x_0\|^2. \end{aligned} \tag{5.4}$$

Consequently, $(2\alpha - 1)\|P_C x_n - P_C x_0\|^2 \leqslant d_C^2(x_0)$ and, letting $\alpha \to +\infty$, we conclude that $P_C x_n = P_C x_0$.

(ii): Combine Theorem 5.5, Corollary 5.8, and (i). □

We now turn our attention to strong convergence properties.

Proposition 5.10 (Raik) *Let C be a subset of $\mathcal{H}$ such that $\operatorname{int} C \neq \varnothing$ and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$. Suppose that $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to C. Then $(x_n)_{n\in\mathbb{N}}$ converges strongly to a point in $\mathcal{H}$ and $\sum_{n\in\mathbb{N}} \|x_{n+1} - x_n\| < +\infty$.*

Proof. Take $x \in \operatorname{int} C$ and $\rho \in \mathbb{R}_{++}$ such that $B(x;\rho) \subset C$. Define a sequence $(z_n)_{n\in\mathbb{N}}$ in $B(x;\rho)$ by

$$(\forall n \in \mathbb{N}) \quad z_n = \begin{cases} x, & \text{if } x_{n+1} = x_n; \\ x - \rho \dfrac{x_{n+1} - x_n}{\|x_{n+1} - x_n\|}, & \text{otherwise.} \end{cases} \tag{5.5}$$

Then (5.1) yields $(\forall n \in \mathbb{N})$ $\|x_{n+1} - z_n\|^2 \leqslant \|x_n - z_n\|^2$ and, after expanding, we obtain

$$(\forall n \in \mathbb{N}) \quad \|x_{n+1} - x\|^2 \leqslant \|x_n - x\|^2 - 2\rho\|x_{n+1} - x_n\|. \tag{5.6}$$

Thus, $\sum_{n\in\mathbb{N}} \|x_{n+1} - x_n\| \leqslant \|x_0 - x\|^2/(2\rho)$ and $(x_n)_{n\in\mathbb{N}}$ is therefore a Cauchy sequence. ☐

Theorem 5.11 *Let C be a nonempty closed convex subset of $\mathcal{H}$ and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$. Suppose that $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to C. Then the following are equivalent:*

(i) $(x_n)_{n\in\mathbb{N}}$ *converges strongly to a point in C.*
(ii) $(x_n)_{n\in\mathbb{N}}$ *possesses a strong sequential cluster point in C.*
(iii) $\underline{\lim}\, d_C(x_n) = 0$.

Proof. (i)⇒(ii): Clear.
(ii)⇒(iii): Suppose that $x_{k_n} \to x \in C$. Then $d_C(x_{k_n}) \leqslant \|x_{k_n} - x\| \to 0$.
(iii)⇒(i): Proposition 5.4(iii) implies that $d_C(x_n) \to 0$. Hence, $x_n - P_C x_n \to 0$ and (i) follows from Proposition 5.7. ☐

The next result concerns linear convergence.

Theorem 5.12 *Let C be a nonempty closed convex subset of $\mathcal{H}$ and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$. Suppose that $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to C and that, for some $\kappa \in [0,1[$,*

$$(\forall n \in \mathbb{N}) \quad d_C(x_{n+1}) \leqslant \kappa d_C(x_n). \tag{5.7}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges linearly to a point $x \in C$; more precisely,

$$(\forall n \in \mathbb{N}) \quad \|x_n - x\| \leqslant 2\kappa^n d_C(x_0). \tag{5.8}$$

Proof. Theorem 5.11 and (5.7) imply that $(x_n)_{n\in\mathbb{N}}$ converges strongly to some point $x \in C$. On the other hand, Proposition 5.4(iv) yields

$$(\forall n \in \mathbb{N})(\forall m \in \mathbb{N}) \quad \|x_n - x_{n+m}\| \leqslant 2d_C(x_n). \tag{5.9}$$

Letting $m \to +\infty$ in (5.9), we conclude that $\|x_n - x\| \leqslant 2d_C(x_n)$. ☐

We conclude this section by investigating the asymptotic behavior of iterations of firmly quasinonexpansive operators.

Proposition 5.13 *Let $(T_n)_{n\in\mathbb{N}}$ be a family of firmly quasinonexpansive operators from $\mathcal{H}$ to $\mathcal{H}$ such that $C = \bigcap_{n\in\mathbb{N}} \operatorname{Fix} T_n \neq \varnothing$, and let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,2[$. Let $x_0 \in \mathcal{H}$ and set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n(T_n x_n - x_n). \tag{5.10}$$

Then the following hold:

(i) $(\forall x \in C)(\forall n \in \mathbb{N})$ $\|x_{n+1} - x\|^2 \leqslant \|x_n - x\|^2 - \lambda_n(2-\lambda_n)\|T_n x_n - x_n\|^2$.
(ii) $(x_n)_{n\in\mathbb{N}}$ *is Fejér monotone with respect to C.*
(iii) $(\forall n \in \mathbb{N})$ $d_C^2(x_{n+1}) \leqslant d_C^2(x_0) - \sum_{k=0}^{n} \lambda_k(2-\lambda_k)\|T_k x_k - x_k\|^2$.
(iv) $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n)\|T_n x_n - x_n\|^2 \leqslant d_C^2(x_0)$.
(v) $\sum_{n\in\mathbb{N}} (2/\lambda_n - 1)\|x_{n+1} - x_n\|^2 \leqslant d_C^2(x_0)$.
(vi) *Suppose that every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ belongs to C. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in C.*
(vii) *Suppose that $\operatorname{int} C \neq \varnothing$. Then $(x_n)_{n\in\mathbb{N}}$ converges strongly; the strong limit lies in C provided that some strong sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ lies in C, or that every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ belongs to C.*
(viii) *Suppose that $\underline{\lim}\, d_C(x_n) = 0$. Then $(x_n)_{n\in\mathbb{N}}$ converges strongly to a point in C.*

Proof. (i): Apply Proposition 4.3.

(ii): This follows from (i).

(iii): Let $x \in C$ and let $n \in \mathbb{N}$. It follows from (i) that

$$\|x_{n+1} - x\|^2 \leqslant \|x_0 - x\|^2 - \sum_{k=0}^{n} \lambda_k(2-\lambda_k)\|T_k x_k - x_k\|^2. \tag{5.11}$$

Taking the infimum over $x \in C$ yields the result.

(iv): A consequence of (iii).

(v): This follows from (iv) and (5.10).

(vi): This follows from (ii) and Theorem 5.5.

(vii): Combine (ii), (vi), and Proposition 5.10.

(viii): Combine (ii) and Theorem 5.11. ☐

5.2 Krasnosel'skiĭ–Mann Iteration

Given a nonexpansive operator T, the sequence generated by the Banach–Picard iteration $x_{n+1} = Tx_n$ of (1.69) may fail to produce a fixed point of T. A simple illustration of this situation is $T = -\mathrm{Id}$ and $x_0 \neq 0$. In this case, however, it is clear that the *asymptotic regularity* property $x_n - Tx_n \to 0$ does not hold. As we shall now see, this property is critical.

Theorem 5.14 *Let D be a nonempty closed convex subset of $\mathcal{H}$, let $T\colon D \to D$ be a nonexpansive operator such that $\operatorname{Fix} T \neq \varnothing$, and let $x_0 \in D$. Set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = Tx_n \tag{5.12}$$

and suppose that $x_n - Tx_n \to 0$. Then the following hold:

(i) *$(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{Fix} T$.*
(ii) *Suppose that $D = -D$ and that T is odd: $(\forall x \in D)\ T(-x) = -Tx$. Then $(x_n)_{n\in\mathbb{N}}$ converges strongly to a point in $\operatorname{Fix} T$.*

Proof. By Example 5.3, $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to $\operatorname{Fix} T$.

(i): Let x be a weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$, say $x_{k_n} \rightharpoonup x$. Since $Tx_{k_n} - x_{k_n} \to 0$, Corollary 4.28 asserts that $x \in \operatorname{Fix} T$. Appealing to Theorem 5.5, the assertion is proved.

(ii): Since $D = -D$ is convex, $0 \in D$ and, since T is odd, $0 \in \operatorname{Fix} T$. Therefore, by Fejér monotonicity, $(\forall n \in \mathbb{N})\ \|x_{n+1}\| \leqslant \|x_n\|$. Thus, there exists $\ell \in \mathbb{R}_+$ such that $\|x_n\| \downarrow \ell$. Now let $m \in \mathbb{N}$. Then, for every $n \in \mathbb{N}$,

$$\|x_{n+1+m} + x_{n+1}\| = \|Tx_{n+m} - T(-x_n)\| \leqslant \|x_{n+m} + x_n\|, \tag{5.13}$$

and, by the parallelogram identity,

$$\|x_{n+m} + x_n\|^2 = 2\big(\|x_{n+m}\|^2 + \|x_n\|^2\big) - \|x_{n+m} - x_n\|^2. \tag{5.14}$$

However, since $Tx_n - x_n \to 0$, we have $\lim_n \|x_{n+m} - x_n\| = 0$. Therefore, since $\|x_n\| \downarrow \ell$, (5.13) and (5.14) yield $\|x_{n+m} + x_n\| \downarrow 2\ell$ as $n \to +\infty$. In turn, we derive from (5.14) that $\|x_{n+m} - x_n\|^2 \leqslant 4\|x_n\|^2 - 4\ell^2 \to 0$ as $n \to +\infty$. Thus, $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence and $x_n \to x$ for some $x \in D$. Since $x_{n+1} \to x$ and $x_{n+1} = Tx_n \to Tx$, we have $x \in \operatorname{Fix} T$. ☐

We now turn our attention to a relaxation variant, known as the *Krasnosel'skiĭ–Mann algorithm*.

Theorem 5.15 (Groetsch) *Let D be a nonempty closed convex subset of $\mathcal{H}$, let $T\colon D \to D$ be a nonexpansive operator such that $\operatorname{Fix} T \neq \varnothing$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0,1]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(1-\lambda_n) = +\infty$, and let $x_0 \in D$. Set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n\big(Tx_n - x_n\big). \tag{5.15}$$

Then the following hold:

(i) $(x_n)_{n\in\mathbb{N}}$ *is Fejér monotone with respect to* $\operatorname{Fix} T$.
(ii) $(Tx_n - x_n)_{n\in\mathbb{N}}$ *converges strongly to* 0.
(iii) $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* $\operatorname{Fix} T$.

Proof. Since $x_0 \in D$ and D is convex, (5.15) produces a well-defined sequence in D.

(i): It follows from Corollary 2.15 and the nonexpansiveness of T that, for every $y \in \operatorname{Fix} T$ and every $n \in \mathbb{N}$,

$$\begin{aligned}\|x_{n+1} - y\|^2 &= \|(1-\lambda_n)(x_n - y) + \lambda_n(Tx_n - y)\|^2\\ &= (1-\lambda_n)\|x_n - y\|^2 + \lambda_n\|Tx_n - Ty\|^2\\ &\quad - \lambda_n(1-\lambda_n)\|Tx_n - x_n\|^2\\ &\leqslant \|x_n - y\|^2 - \lambda_n(1-\lambda_n)\|Tx_n - x_n\|^2. \end{aligned}\tag{5.16}$$

Hence, $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to $\operatorname{Fix} T$.

(ii): We derive from (5.16) that $\sum_{n\in\mathbb{N}} \lambda_n(1-\lambda_n)\|Tx_n - x_n\|^2 \leqslant \|x_0 - y\|^2$. Since $\sum_{n\in\mathbb{N}} \lambda_n(1-\lambda_n) = +\infty$, we have $\underline{\lim}\, \|Tx_n - x_n\| = 0$. However, for every $n \in \mathbb{N}$,

$$\begin{aligned}\|Tx_{n+1} - x_{n+1}\| &= \|Tx_{n+1} - Tx_n + (1-\lambda_n)(Tx_n - x_n)\|\\ &\leqslant \|x_{n+1} - x_n\| + (1-\lambda_n)\|Tx_n - x_n\|\\ &= \|Tx_n - x_n\|. \end{aligned}\tag{5.17}$$

Consequently, $(\|Tx_n - x_n\|)_{n\in\mathbb{N}}$ converges and we must have $Tx_n - x_n \to 0$.

(iii): Let x be a weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$, say $x_{k_n} \rightharpoonup x$. Then it follows from Corollary 4.28 that $x \in \operatorname{Fix} T$. In view of Theorem 5.5, the proof is complete. ☐

Proposition 5.16 *Let* $\alpha \in\,]0,1[$, *let* $T\colon \mathcal{H} \to \mathcal{H}$ *be an* α*-averaged operator such that* $\operatorname{Fix} T \neq \varnothing$, *let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $[0, 1/\alpha]$ *such that* $\sum_{n\in\mathbb{N}} \lambda_n(1-\alpha\lambda_n) = +\infty$, *and let* $x_0 \in \mathcal{H}$. *Set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n\bigl(Tx_n - x_n\bigr). \tag{5.18}$$

Then the following hold:

(i) $(x_n)_{n\in\mathbb{N}}$ *is Fejér monotone with respect to* $\operatorname{Fix} T$.
(ii) $(Tx_n - x_n)_{n\in\mathbb{N}}$ *converges strongly to* 0.
(iii) $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* $\operatorname{Fix} T$.

Proof. Set $R = (1 - 1/\alpha)\mathrm{Id} + (1/\alpha)T$ and $(\forall n \in \mathbb{N})$ $\mu_n = \alpha\lambda_n$. Then $\operatorname{Fix} R = \operatorname{Fix} T$ and R is nonexpansive by Proposition 4.35. In addition, we rewrite (5.18) as $(\forall n \in \mathbb{N})$ $x_{n+1} = x_n + \mu_n\bigl(Rx_n - x_n\bigr)$. Since $(\mu_n)_{n\in\mathbb{N}}$ lies in $[0,1]$ and $\sum_{n\in\mathbb{N}} \mu_n(1-\mu_n) = +\infty$, the results follow from Theorem 5.15. ☐

Corollary 5.17 *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a firmly nonexpansive operator such that* $\operatorname{Fix} T \neq \varnothing$*, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0,2]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$, and let $x_0 \in \mathcal{H}$. Set $(\forall n \in \mathbb{N})$ $x_{n+1} = x_n + \lambda_n\big(Tx_n - x_n\big)$. Then the following hold:*

(i) *$(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to* $\operatorname{Fix} T$.
(ii) *$(Tx_n - x_n)_{n\in\mathbb{N}}$ converges strongly to* 0.
(iii) *$(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in* $\operatorname{Fix} T$.

Proof. In view of Remark 4.34(iii), apply Proposition 5.16 with $\alpha = 1/2$. □

Example 5.18 Let $T\colon \mathcal{H} \to \mathcal{H}$ be a firmly nonexpansive operator such that $\operatorname{Fix} T \neq \varnothing$, let $x_0 \in \mathcal{H}$, and set $(\forall n \in \mathbb{N})$ $x_{n+1} = Tx_n$. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{Fix} T$.

The following type of iterative method involves a mix of compositions and convex combinations of nonexpansive operators.

Corollary 5.19 *Let $(T_i)_{i\in I}$ be a finite family of nonexpansive operators from $\mathcal{H}$ to $\mathcal{H}$ such that $\bigcap_{i\in I} \operatorname{Fix} T_i \neq \varnothing$, and let $(\alpha_i)_{i\in I}$ be real numbers in $]0,1[$ such that, for every $i \in I$, T_i is α_i-averaged. Let p be a strictly positive integer, for every $k \in \{1,\ldots,p\}$, let m_k be a strictly positive integer and $\omega_k \in \,]0,1]$, and suppose that $\mathrm{i}\colon \{(k,l) \mid k \in \{1,\ldots,p\},\ l \in \{1,\ldots,m_k\}\} \to I$ is surjective and that $\sum_{k=1}^{p} \omega_k = 1$. Set*

$$\alpha = \sum_{k=1}^{p} \omega_k \rho_k, \quad \textit{where} \ \big(\forall k \in \{1,\ldots,p\}\big) \quad \rho_k = \frac{1}{1 + \dfrac{1}{\displaystyle\sum_{i=1}^{m_k} \frac{\alpha_{\mathrm{i}(k,i)}}{1-\alpha_{\mathrm{i}(k,i)}}}}, \tag{5.19}$$

and let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0,1/\alpha]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(1-\alpha\lambda_n) = +\infty$. Furthermore, let $x_0 \in \mathcal{H}$ and set

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n \left(\sum_{k=1}^{p} \omega_k T_{\mathrm{i}(k,1)} \cdots T_{\mathrm{i}(k,m_k)} x_n - x_n \right). \tag{5.20}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\bigcap_{i\in I} \operatorname{Fix} T_i$.

Proof. For every $k \in \{1,\ldots,p\}$, set $I_k = \{\mathrm{i}(k,1),\ldots,\mathrm{i}(k,m_k)\}$ and $R_k = T_{\mathrm{i}(k,1)} \cdots T_{\mathrm{i}(k,m_k)}$, and let $T = \sum_{k=1}^{p} \omega_k R_k$. Then (5.20) reduces to (5.18) and, in view of Proposition 5.16, it suffices to show that T is α-averaged and that $\operatorname{Fix} T = \bigcap_{i\in I} \operatorname{Fix} T_i$. For every $k \in \{1,\ldots,p\}$, it follows from Proposition 4.46 and (5.19) that R_k is ρ_k-averaged and, from Corollary 4.51 that $\operatorname{Fix} R_k = \bigcap_{i\in I_k} \operatorname{Fix} T_i$. In turn, we derive from Proposition 4.42 and (5.19) that T is α-averaged and, from Proposition 4.47 and the surjectivity of i, that $\operatorname{Fix} T = \bigcap_{k=1}^{p} \operatorname{Fix} R_k = \bigcap_{k=1}^{p} \bigcap_{i\in I_k} \operatorname{Fix} T_i = \bigcap_{i\in I} \operatorname{Fix} T_i$. □

Remark 5.20 It follows from Remark 4.34(iii) that Corollary 5.19 is applicable to firmly nonexpansive operators and, a fortiori, to projection operators by Proposition 4.16.

Corollary 5.19 provides an algorithm to solve a *convex feasibility problem*, i.e., to find a point in the intersection of a family of closed convex sets. Here are two more examples.

Example 5.21 (String-averaged relaxed projections) Let $(C_i)_{i\in I}$ be a finite family of closed convex sets such that $C = \bigcap_{i\in I} C_i \neq \varnothing$. For every $i \in I$, let $\beta_i \in \left]0, 2\right[$ and set $T_i = (1-\beta_i)\mathrm{Id} + \beta_i P_{C_i}$. Let p be a strictly positive integer. For every $k \in \{1,\ldots,p\}$, let m_k be a strictly positive integer and $\omega_k \in \left]0, 1\right]$, and suppose that $\mathrm{i}\colon \{(k,l) \mid k \in \{1,\ldots,p\}, l \in \{1,\ldots,m_k\}\} \to I$ is surjective and that $\sum_{k=1}^{p} \omega_k = 1$. Furthermore, let $x_0 \in \mathcal{H}$ and set

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \sum_{k=1}^{p} \omega_k T_{\mathrm{i}(k,1)} \cdots T_{\mathrm{i}(k,m_k)} x_n. \tag{5.21}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in C.

Proof. For every $i \in I$, set $\alpha_i = \beta_i/2 \in \left]0, 1\right[$. Since, for every $i \in I$, Proposition 4.16 asserts that P_{C_i} is firmly nonexpansive, Corollary 4.41 implies that T_i is α_i-averaged. Using the notation (5.19), we note that, for every $k \in \{1,\ldots,p\}$, $\rho_k \in \left]0, 1\right[$ and thus that $\alpha \in \left]0, 1\right[$. Altogether, the result follows from Corollary 5.19 with $\lambda_n \equiv 1$. ☐

Example 5.22 (Parallel projection algorithm) Let $(C_i)_{i\in I}$ be a finite family of closed convex subsets of $\mathcal{H}$ such that $C = \bigcap_{i\in I} C_i \neq \varnothing$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0, 2]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$, let $(\omega_i)_{i\in I}$ be strictly positive real numbers such that $\sum_{i\in I} \omega_i = 1$, and let $x_0 \in \mathcal{H}$. Set

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n\Bigg(\sum_{i\in I} \omega_i P_i x_n - x_n\Bigg) \tag{5.22}$$

where, for every $i \in I$, P_i denotes the projector onto C_i. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in C.

Proof. This is an application of Corollary 5.17(iii) with $T = \sum_{i\in I} \omega_i P_i$. Indeed, since the operators $(P_i)_{i\in I}$ are firmly nonexpansive by Proposition 4.16, their convex combination T is also firmly nonexpansive by Example 4.7. Moreover, Proposition 4.47 asserts that $\operatorname{Fix} T = \bigcap_{i\in I} \operatorname{Fix} P_i = \bigcap_{i\in I} C_i = C$. Alternatively, apply Corollary 5.19 with $(\forall k \in \{1,\ldots,p\} = I)$ $m_k = 1$, $\mathrm{i}(k,1) = k$, and $\alpha_{\mathrm{i}(k,1)} = 1/2$. ☐

5.3 Iterating Compositions of Averaged Operators

Our first result concerns the asymptotic behavior of iterates of a composition of averaged nonexpansive operators with possibly no common fixed point.

Theorem 5.23 *Let D be a nonempty closed convex subset of $\mathcal{H}$, let m be a strictly positive integer, set $I = \{1, \ldots, m\}$, let $(T_i)_{i\in I}$ be a family of nonexpansive operators from D to D such that $\operatorname{Fix}(T_1 \cdots T_m) \neq \varnothing$, and let $(\alpha_i)_{i\in I}$ be real numbers in $]0,1[$ such that, for every $i \in I$, T_i is α_i-averaged. Let $x_0 \in D$ and set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = T_1 \cdots T_m x_n. \tag{5.23}$$

Then $x_n - T_1 \cdots T_m x_n \to 0$, and there exist points $y_1 \in \operatorname{Fix} T_1 \cdots T_m$, $y_2 \in \operatorname{Fix} T_2 \cdots T_m T_1$, …, $y_m \in \operatorname{Fix} T_m T_1 \cdots T_{m-1}$ such that

$$x_n \rightharpoonup y_1 = T_1 y_2, \tag{5.24}$$

$$T_m x_n \rightharpoonup y_m = T_m y_1, \tag{5.25}$$

$$T_{m-1} T_m x_n \rightharpoonup y_{m-1} = T_{m-1} y_m, \tag{5.26}$$

$$\vdots$$

$$T_3 \cdots T_m x_n \rightharpoonup y_3 = T_3 y_4, \tag{5.27}$$

$$T_2 \cdots T_m x_n \rightharpoonup y_2 = T_2 y_3. \tag{5.28}$$

Proof. Set $T = T_1 \cdots T_m$ and $(\forall i \in I)$ $\beta_i = (1-\alpha_i)/\alpha_i$. Now take $y \in \operatorname{Fix} T$. The equivalence (i)⇔(iii) in Proposition 4.35 yields

$$\begin{aligned}\|x_{n+1} - y\|^2 &= \|Tx_n - Ty\|^2 \\ &\leqslant \|T_2 \cdots T_m x_n - T_2 \cdots T_m y\|^2 \\ &\quad - \beta_1 \|(\mathrm{Id} - T_1) T_2 \cdots T_m x_n - (\mathrm{Id} - T_1) T_2 \cdots T_m y\|^2 \\ &\leqslant \|x_n - y\|^2 - \beta_m \|(\mathrm{Id} - T_m) x_n - (\mathrm{Id} - T_m) y\|^2 \\ &\quad - \beta_{m-1} \|(\mathrm{Id} - T_{m-1}) T_m x_n - (\mathrm{Id} - T_{m-1}) T_m y\|^2 - \cdots \\ &\quad - \beta_2 \|(\mathrm{Id} - T_2) T_3 \cdots T_m x_n - (\mathrm{Id} - T_2) T_3 \cdots T_m y\|^2 \\ &\quad - \beta_1 \|(\mathrm{Id} - T_1) T_2 \cdots T_m x_n - (T_2 \cdots T_m y - y)\|^2. \end{aligned} \tag{5.29}$$

Therefore, $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to $\operatorname{Fix} T$ and

$$(\mathrm{Id} - T_m) x_n - (\mathrm{Id} - T_m) y \to 0, \tag{5.30}$$

$$(\mathrm{Id} - T_{m-1}) T_m x_n - (\mathrm{Id} - T_{m-1}) T_m y \to 0, \tag{5.31}$$

$$\vdots$$

$$(\mathrm{Id} - T_2) T_3 \cdots T_m x_n - (\mathrm{Id} - T_2) T_3 \cdots T_m y \to 0, \tag{5.32}$$

$$(\mathrm{Id} - T_1) T_2 \cdots T_m x_n - (T_2 \cdots T_m y - y) \to 0. \tag{5.33}$$

Upon adding (5.30)–(5.33), we obtain $x_n - Tx_n \to 0$. Hence, since T is nonexpansive as a composition of nonexpansive operators, it follows from Theorem 5.14(i) that $(x_n)_{n\in\mathbb{N}}$ converges weakly to some point $y_1 \in \operatorname{Fix} T$, which provides (5.24). On the other hand, (5.30) yields $T_m x_n - x_n \to T_m y_1 - y_1$. So altogether $T_m x_n \rightharpoonup T_m y_1 = y_m$, and we obtain (5.25). In turn, since (5.31) asserts that $T_{m-1}T_m x_n - T_m x_n \to T_{m-1}y_m - y_m$, we obtain $T_{m-1}T_m x_n \rightharpoonup T_{m-1}y_m = y_{m-1}$, hence (5.26). Continuing this process, we arrive at (5.28). ☐

As noted in Remark 5.20, results on averaged nonexpansive operators apply in particular to firmly nonexpansive operators and projectors onto convex sets. Thus, by specializing Theorem 5.23 to convex projectors, we obtain the iterative method described in the next corollary, which is known as the POCS (Projections Onto Convex Sets) algorithm in the signal recovery literature.

Corollary 5.24 (POCS algorithm) *Let m be a strictly positive integer, set $I = \{1,\ldots,m\}$, let $(C_i)_{i\in I}$ be a family of nonempty closed convex subsets of $\mathcal{H}$, let $(P_i)_{i\in I}$ denote their respective projectors, and let $x_0 \in \mathcal{H}$. Suppose that* $\operatorname{Fix}(P_1\cdots P_m) \neq \varnothing$ *and set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = P_1\cdots P_m x_n. \tag{5.34}$$

Then there exists $(y_1,\ldots,y_m) \in C_1 \times\cdots\times C_m$ such that $x_n \rightharpoonup y_1 = P_1 y_2$, $P_m x_n \rightharpoonup y_m = P_m y_1$, $P_{m-1}P_m x_n \rightharpoonup y_{m-1} = P_{m-1}y_m$, …, $P_3\cdots P_m x_n \rightharpoonup y_3 = P_3 y_4$, and $P_2\cdots P_m x_n \rightharpoonup y_2 = P_2 y_3$.

Proof. This follows from Proposition 4.16 and Theorem 5.23. ☐

Remark 5.25 In Corollary 5.24, suppose that, for some $j \in I$, C_j is bounded. Then $\operatorname{Fix}(P_1\cdots P_m) \neq \varnothing$. Indeed, consider the circular composition of the m projectors given by $T = P_j\cdots P_m P_1 \cdots P_{j-1}$. Then Proposition 4.16 asserts that T is a nonexpansive operator that maps the nonempty bounded closed convex set C_j to itself. Hence, it follows from Theorem 4.29 that there exists a point $x \in C_j$ such that $Tx = x$.

The next corollary describes a periodic projection method to solve a convex feasibility problem.

Corollary 5.26 (Bregman) *Let m be a strictly positive integer, set $I = \{1,\ldots,m\}$, let $(C_i)_{i\in I}$ be a family of closed convex subsets of $\mathcal{H}$ such that $C = \bigcap_{i\in I} C_i \neq \varnothing$, let $(P_i)_{i\in I}$ denote their respective projectors, and let $x_0 \in \mathcal{H}$. Set $(\forall n \in \mathbb{N})$ $x_{n+1} = P_1\cdots P_m x_n$. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in C.*

Proof. Using Corollary 5.24, Proposition 4.16, and Corollary 4.51, we obtain $x_n \rightharpoonup y_1 \in \operatorname{Fix}(P_1\cdots P_m) = \bigcap_{i\in I}\operatorname{Fix} P_i = C$. Alternatively, this is a special case of Example 5.21 with $p = 1$, $\omega_1 = 1$, $m_1 = m$, and $(\forall l \in I)$ $\beta_l = 1$ and $\mathrm{i}(1,l) = l$. ☐

Remark 5.27 If, in Corollary 5.26, all the sets are closed affine subspaces, so is C and we derive from Proposition 5.9(i) that $x_n \rightharpoonup P_C x_0$. As will be seen in Corollary 5.30, the convergence is actually strong in this case. In striking contrast, the example constructed in [206] provides a closed hyperplane and a closed convex cone in $\ell^2(\mathbb{N})$ for which alternating projections converge weakly but not strongly.

The next result will help us obtain a sharper form of Corollary 5.26 for closed affine subspaces.

Proposition 5.28 *Let $T \in \mathcal{B}(\mathcal{H})$ be nonexpansive and let $x_0 \in \mathcal{H}$. Set $V = \operatorname{Fix} T$ and $(\forall n \in \mathbb{N})$ $x_{n+1} = Tx_n$. Then $x_n \to P_V x_0 \Leftrightarrow x_n - x_{n+1} \to 0$.*

Proof. If $x_n \to P_V x_0$, then $x_n - x_{n+1} \to P_V x_0 - P_V x_0 = 0$. Conversely, suppose that $x_n - x_{n+1} \to 0$. We derive from Theorem 5.14(ii) that there exists $v \in V$ such that $x_n \to v$. In turn, Proposition 5.9(i) yields $P_V x_0 = P_V x_n \to P_V v = v$ and therefore $v = P_V x_0$. □

Example 5.29 Let $T \in \mathcal{B}(\mathcal{H})$ be averaged and let $x_0 \in \mathcal{H}$. Set $V = \operatorname{Fix} T$ and $(\forall n \in \mathbb{N})$ $x_{n+1} = Tx_n$. Then $x_n \to P_V x_0$.

Proof. Combine Proposition 5.16(ii) and Proposition 5.28. □

Corollary 5.30 (von Neumann–Halperin) *Let m be a strictly positive integer, set $I = \{1, \ldots, m\}$, let $(C_i)_{i \in I}$ be a family of closed affine subspaces of $\mathcal{H}$ such that $C = \bigcap_{i \in I} C_i \neq \varnothing$, let $(P_i)_{i \in I}$ denote their respective projectors, let $x_0 \in \mathcal{H}$, and set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = P_1 \cdots P_m x_n. \tag{5.35}$$

Then $x_n \to P_C x_0$.

Proof. Set $T = P_1 \cdots P_m$. Then T is nonexpansive and $\operatorname{Fix} T = C$ by Corollary 4.51. We first assume that, for every $i \in I$, C_i is a linear subspace. Then T is odd, and Theorem 5.23 implies that $x_n - Tx_n \to 0$. Thus, by Proposition 5.28, $x_n \to P_C x_0$. We now turn our attention to the general affine case. Since $C \neq \varnothing$, there exists $y \in C$ such that, for every $i \in I$, $C_i = y + V_i$, i.e., V_i is the closed linear subspace parallel to C_i, and $C = y + V$, where $V = \bigcap_{i \in I} V_i$. Proposition 3.19 asserts that, for every $x \in \mathcal{H}$, $P_C x = P_{y+V} x = y + P_V(x - y)$ and $(\forall i \in I)$ $P_i x = P_{y+V_i} x = y + P_{V_i}(x - y)$. Using these identities repeatedly, we obtain

$$(\forall n \in \mathbb{N}) \quad x_{n+1} - y = (P_{V_1} \cdots P_{V_m})(x_n - y). \tag{5.36}$$

Invoking the already verified linear case, we get $x_n - y \to P_V(x_0 - y)$ and conclude that $x_n \to y + P_V(x_0 - y) = P_C x_0$. □

5.4 Quasi-Fejér Monotone Sequences

We study a relaxation of Definition 5.1. The following technical fact will be required. We denote by $\ell_+^1(\mathbb{N})$ the set of sequences $(\xi_n)_{n\in\mathbb{N}}$ in $\mathbb{R}_+$ such that $\sum_{n\in\mathbb{N}} \xi_n < +\infty$.

Lemma 5.31 *Let $(\alpha_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$ be sequences in $\mathbb{R}_+$, and let $(\gamma_n)_{n\in\mathbb{N}}$ and $(\varepsilon_n)_{n\in\mathbb{N}}$ be sequences in $\ell_+^1(\mathbb{N})$ such that*

$$(\forall n \in \mathbb{N}) \quad \alpha_{n+1} \leqslant (1+\gamma_n)\alpha_n - \beta_n + \varepsilon_n. \tag{5.37}$$

Then $(\alpha_n)_{n\in\mathbb{N}}$ converges and $\sum_{n\in\mathbb{N}} \beta_n < +\infty$.

Proof. Set $\varepsilon = \sum_{n\in\mathbb{N}} \varepsilon_n$, $\sigma = \sum_{n\in\mathbb{N}} \gamma_n$, $\gamma = \prod_{n\in\mathbb{N}}(1+\gamma_n)$, and

$$(\forall n \in \mathbb{N}) \quad \delta_n = \alpha_n - \sum_{k=0}^{n-1}(\gamma_k\alpha_k + \varepsilon_k). \tag{5.38}$$

Then, since $\sum_{n\in\mathbb{N}} \gamma_n < +\infty \Leftrightarrow \prod_{n\in\mathbb{N}}(1+\gamma_n) < +\infty$ (see also Exercise 17.10(iii)), we have $\varepsilon < +\infty$, $\sigma < +\infty$, and $\gamma < +\infty$. Furthermore, (5.37) yields

$$(\forall n \in \mathbb{N}) \quad \alpha_{n+1} \leqslant \alpha_0 \prod_{k=0}^{n}(1+\gamma_k) + \sum_{k=0}^{n} \varepsilon_k \prod_{l=k+1}^{n}(1+\gamma_l) \leqslant \gamma(\alpha_0+\varepsilon). \tag{5.39}$$

In turn, we derive from (5.38) and (5.37) that

$$(\forall n \in \mathbb{N}) \quad -\sigma\gamma(\alpha_0+\varepsilon)-\varepsilon \leqslant -\sum_{k=0}^{n}(\gamma_k\alpha_k+\varepsilon_k) \leqslant \delta_{n+1} \leqslant \delta_n-\beta_n \leqslant \delta_n. \tag{5.40}$$

Therefore $\delta = \lim \delta_n$ exists and so does $\lim \alpha_n = \lim \delta_n + \sum_{n\in\mathbb{N}}(\gamma_n\alpha_n + \varepsilon_n)$. Furthermore, $(\forall n \in \mathbb{N})$ $\sum_{k=0}^{n} \beta_k \leqslant \delta_0 - \delta_{n+1}$. Thus, $\sum_{k\in\mathbb{N}} \beta_k \leqslant \delta_0 - \delta$. ☐

Definition 5.32 Let C be a nonempty subset of $\mathcal{H}$ and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$. Then $(x_n)_{n\in\mathbb{N}}$ is *quasi-Fejér monotone* with respect to C if

$$(\forall x \in C)(\exists\, (\varepsilon_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N}))(\forall n \in \mathbb{N}) \quad \|x_{n+1}-x\|^2 \leqslant \|x_n-x\|^2+\varepsilon_n. \tag{5.41}$$

Many of the properties investigated in Section 5.1 remain true for quasi-Fejér sequences. Thus, the following result extends Proposition 5.4 and Theorem 5.5.

Theorem 5.33 *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ and let C be a nonempty subset of $\mathcal{H}$ such that $(x_n)_{n\in\mathbb{N}}$ is quasi-Fejér monotone with respect to C. Then the following hold:*

(i) *For every* $x \in C$, $(\|x_n - x\|)_{n\in\mathbb{N}}$ *converges.*
(ii) $(x_n)_{n\in\mathbb{N}}$ *is bounded.*
(iii) $(d_C(x_n))_{n\in\mathbb{N}}$ *converges.*
(iv) *Suppose that every weak sequential cluster point of* $(x_n)_{n\in\mathbb{N}}$ *belongs to* C. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* C.

Proof. (i): This follows from Lemma 5.31.

(ii): Clear from (i).

(iii): Taking the infimum in (5.41) over $x \in C$ yields $(\forall n \in \mathbb{N})\ d_C^2(x_{n+1}) \leqslant d_C^2(x_n) + \varepsilon_n$. Hence, the claim follows from Lemma 5.31.

(iv): Combine (i) and Lemma 2.47. ☐

The above theorem allows us to derive variants of the weak convergence results for fixed point methods under perturbations. As an illustration, here is a version of Theorem 5.15 in which an error sequence $(e_n)_{n\in\mathbb{N}}$ modeling approximate computations of the sequence $(Tx_n)_{n\in\mathbb{N}}$ is allowed in the Krasnosel'skiĭ–Mann iteration (5.15).

Proposition 5.34 *Let* $T\colon \mathcal{H} \to \mathcal{H}$ *be a nonexpansive operator such that* $\operatorname{Fix} T \neq \varnothing$, *let* $x_0 \in \mathcal{H}$, *let* $(e_n)_{n\in\mathbb{N}}$ *be a sequence in* $\mathcal{H}$, *and let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $[0,1]$ *such that* $\sum_{n\in\mathbb{N}} \lambda_n(1-\lambda_n) = +\infty$ *and* $\sum_{n\in\mathbb{N}} \lambda_n\|e_n\| < +\infty$. *Set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n\big(Tx_n + e_n - x_n\big). \tag{5.42}$$

Then the following hold:

(i) $(x_n)_{n\in\mathbb{N}}$ *is quasi-Fejér monotone with respect to* $\operatorname{Fix} T$.
(ii) $(Tx_n - x_n)_{n\in\mathbb{N}}$ *converges strongly to* 0.
(iii) $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* $\operatorname{Fix} T$.

Proof. (i): Let $x \in \operatorname{Fix} T$ and set $\beta = \sup_{n\in\mathbb{N}} \|x_n - x\|$. By nonexpansiveness of T, we have

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad \|x_{n+1} - x\| &\leqslant (1-\lambda_n)\|x_n - x\| + \lambda_n\|Tx_n - Tx\| + \lambda_n\|e_n\| \\ &\leqslant \|x_n - x\| + \lambda_n\|e_n\|. \end{aligned} \tag{5.43}$$

Hence, Lemma 5.31 implies that $\beta < +\infty$. Now set $(\forall n \in \mathbb{N})\ y_n = x_n + \lambda_n(Tx_n - x_n)$. Then we derive from Corollary 2.15 that

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad \|y_n - x\|^2 &= \|(1-\lambda_n)(x_n - x) + \lambda_n(Tx_n - x)\|^2 \\ &= (1-\lambda_n)\|x_n - x\|^2 + \lambda_n\|Tx_n - Tx\|^2 \\ &\quad - \lambda_n(1-\lambda_n)\|Tx_n - x_n\|^2 \\ &\leqslant \|x_n - x\|^2 - \lambda_n(1-\lambda_n)\|Tx_n - x_n\|^2 \qquad (5.44) \\ &\leqslant \beta^2. \qquad (5.45)\end{aligned}$$

Upon setting $\delta = 2\beta + \sum_{n\in\mathbb{N}} \lambda_n\|e_n\|$ and using (5.44), the Cauchy–Schwarz inequality, and (5.45), we obtain

$$
\begin{aligned}
(\forall n \in \mathbb{N}) \quad \|x_{n+1} - x\|^2 &= \|y_n - x\|^2 + 2\lambda_n \langle y_n - x \mid e_n \rangle + \lambda_n^2 \|e_n\|^2 \\
&\leqslant \|x_n - x\|^2 - \lambda_n(1-\lambda_n)\|Tx_n - x_n\|^2 \\
&\quad + (2\|y_n - x\| + \lambda_n\|e_n\|)\lambda_n\|e_n\| \\
&\leqslant \|x_n - x\|^2 - \lambda_n(1-\lambda_n)\|Tx_n - x_n\|^2 \\
&\quad + \delta\lambda_n\|e_n\|.
\end{aligned} \tag{5.46}
$$

This shows that $(x_n)_{n\in\mathbb{N}}$ is quasi-Fejér monotone with respect to $\operatorname{Fix} T$.

(ii): By (5.46) and Lemma 5.31, $\sum_{n\in\mathbb{N}} \lambda_n(1-\lambda_n)\|Tx_n - x_n\|^2 < +\infty$. Thus, we conclude using the same argument as in the proof of Theorem 5.15(ii).

(iii): It follows from (ii) and Corollary 4.28 that every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ is in $\operatorname{Fix} T$. Hence, the conclusion follows from (i) and Theorem 5.33(iv). ☐

Remark 5.35 In Proposition 5.34, suppose that there exist $\alpha \in \mathbb{R}_{++}$ and $\kappa \in \left]0,1\right]$ such that, for n large enough, $\|e_n\| \leqslant \alpha(1 + \sqrt{1 - 1/n})/n^\kappa$. Such a scenario certainly permits $\sum_{n\in\mathbb{N}} \|e_n\| = +\infty$. Yet, upon adopting the relaxation pattern $\lambda_n = (1 - \sqrt{1 - 1/n})/2$ for n large, the weak convergence conditions $\sum_{n\in\mathbb{N}} \lambda_n(1-\lambda_n) = +\infty$ and $\sum_{n\in\mathbb{N}} \lambda_n\|e_n\| < +\infty$ are satisfied.

5.5 Nonlinear Ergodic Theorems

Theorem 5.33(iv) states that a sequence that is quasi-Fejér monotone with respect to a set C converges weakly to a point in C if all its weak sequential cluster points lie in C. This weak convergence principle is generalized in the following theorem, which concerns a sequence that is asymptotically close to a quasi-Fejér sequence in some suitable sense, and the weak sequential cluster points of which lie in C.

Theorem 5.36 *Let $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ be sequences in $\mathcal{H}$, let C be a nonempty subset of $\mathcal{H}$, and set $(\forall m \in \mathbb{N})$ $C_m = \overline{\operatorname{conv}}\,\{y_k\}_{k\geqslant m}$. Suppose that the following hold:*

(i) *$(y_n)_{n\in\mathbb{N}}$ is quasi-Fejér monotone with respect to C.*
(ii) $(\forall m \in \mathbb{N})$ $d_{C_m}(x_n) \to 0$.
(iii) *Every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ is in C.*

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in C.

Proof. We derive from (i) and Theorem 5.33(ii) that $(y_n)_{n\in\mathbb{N}}$ is bounded. Hence, for every $m \in \mathbb{N}$, C_m is bounded and (ii) implies that $(x_n)_{n\in\mathbb{N}}$ is bounded as well. In view of (iii) and Lemma 2.46, it suffices to show that $(x_n)_{n\in\mathbb{N}}$ has at most one weak sequential cluster point in C. To this end, suppose that x and z are points in C, and that $(k_n)_{n\in\mathbb{N}}$ and $(l_n)_{n\in\mathbb{N}}$ are strictly increasing sequences in $\mathbb{N}$ such that $x_{k_n} \rightharpoonup x$ and $x_{l_n} \rightharpoonup z$. Using (i), Theorem 5.33(i), and the expansion

$$(\forall n \in \mathbb{N}) \quad 2\langle y_n \mid x - z\rangle = \|y_n - z\|^2 - \|y_n - x\|^2 + \|x\|^2 - \|z\|^2, \tag{5.47}$$

we derive that $(\langle y_n \mid x - z\rangle)_{n\in\mathbb{N}}$ converges to some $\ell \in \mathbb{R}$. Now fix $\varepsilon \in \mathbb{R}_{++}$, define a closed and convex set

$$Q = \big\{y \in \mathcal{H} \ \big|\ |\langle y \mid x - z\rangle - \ell| \leqslant \varepsilon\big\}, \tag{5.48}$$

and take $m \in \mathbb{N}$ sufficiently large so that $(y_n)_{n\geqslant m}$ lies in Q. Then $C_m \subset Q$ and we deduce from (ii) that, for $n \in \mathbb{N}$ sufficiently large,

$$\begin{cases} \|x_n - P_Q x_n\| = d_Q(x_n) \leqslant d_{C_m}(x_n) \leqslant \varepsilon, \\ |\langle P_Q x_n \mid x - z\rangle - \ell| \leqslant \varepsilon. \end{cases} \tag{5.49}$$

It follows that, for $n \in \mathbb{N}$ sufficiently large,

$$\begin{aligned} |\langle x_n \mid x - z\rangle - \ell| &\leqslant \|x_n - P_Q x_n\|\, \|x - z\| + |\langle P_Q x_n \mid x - z\rangle - \ell| \\ &\leqslant \varepsilon(\|x - z\| + 1). \end{aligned} \tag{5.50}$$

Taking the limit along $(k_n)_{n\in\mathbb{N}}$ yields $|\langle x \mid x - z\rangle - \ell| \leqslant \varepsilon(\|x - z\| + 1)$, i.e., since ε can be taken arbitrarily small, $\langle x \mid x - z\rangle = \ell$. Likewise, taking the limit in (5.50) along $(l_n)_{n\in\mathbb{N}}$ yields $\langle z \mid x - z\rangle = \ell$. Consequently, $\|x - z\|^2 = \langle x \mid x - z\rangle - \langle z \mid x - z\rangle = \ell - \ell = 0$, and thus $x = z$. ☐

The idea underlying ergodic theorems in nonlinear analysis is to transform a nonconvergent sequence $(y_n)_{n\in\mathbb{N}}$ into a convergent sequence $(x_n)_{n\in\mathbb{N}}$ constructed by a moving average of the elements of $(y_n)_{n\in\mathbb{N}}$. Our next result is an application of Theorem 5.36 which provides an abstract principle describing such a mechanism.

Corollary 5.37 *Let $(y_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ which is quasi-Fejér monotone with respect to a nonempty subset C of $\mathcal{H}$ and, for every $n \in \mathbb{N}$, let $(\alpha_{n,k})_{0\leqslant k\leqslant n}$ be real numbers in $[0,1]$. Suppose that*

$$\begin{cases} (\forall n \in \mathbb{N}) \quad \displaystyle\sum_{k=0}^{n} \alpha_{n,k} = 1, \\ (\forall k \in \mathbb{N}) \quad \displaystyle\lim_{n\to+\infty} \alpha_{n,k} = 0, \end{cases} \tag{5.51}$$

set

$$(\forall n \in \mathbb{N}) \quad x_n = \sum_{k=0}^{n} \alpha_{n,k}\, y_k, \tag{5.52}$$

and suppose that every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ is in C. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in C.

Proof. It suffices to verify condition (ii) in Theorem 5.36. Fix $m \in \mathbb{N}$ and set $C_m = \overline{\operatorname{conv}}\,\{y_k\}_{k \geqslant m}$. Note that, in view of Theorem 5.33(ii), $\mu = \sup_{k\in\mathbb{N}} \|y_k\| < +\infty$. Let $m \leqslant n \in \mathbb{N}$ and set

$$z_n = \sum_{k=m}^{n} \beta_{n,k} y_k, \quad \text{where} \quad (\forall k \in \{m,\ldots,n\}) \quad \beta_{n,k} = \frac{\alpha_{n,k}}{\displaystyle\sum_{l=m}^{n} \alpha_{n,l}}. \tag{5.53}$$

Then, $\sum_{k=m}^{n} \beta_{n,k} = 1$ and therefore $z_n \in C_m$. Hence, (5.52), (5.53), and (5.51) yield

$$\begin{aligned} d_{C_m}(x_n) &\leqslant \|x_n - z_n\| \\ &= \left\| \sum_{k=0}^{m-1} \alpha_{n,k} y_k - \sum_{k=m}^{n} (\beta_{n,k} - \alpha_{n,k}) y_k \right\| \\ &\leqslant \mu \left(\sum_{k=0}^{m-1} \alpha_{n,k} + 1 - \sum_{k=m}^{n} \alpha_{n,k} \right) \\ &= 2\mu \sum_{k=0}^{m-1} \alpha_{n,k}. \end{aligned} \tag{5.54}$$

Since, for every $k \in \mathbb{N}$, $\lim_{n\to+\infty} \alpha_{n,k} = 0$, it follows that $\lim_{n\to+\infty} d_{C_m}(x_n) = 0$, which completes the proof. ☐

We conclude this chapter with Baillon's nonlinear ergodic theorem, which employs the Cesàro average of successive approximations generated by a nonexpansive operator.

Example 5.38 (Baillon) Let D be a nonempty closed convex subset of $\mathcal{H}$, let $T\colon D \to D$ be a nonexpansive operator such that $\operatorname{Fix} T \neq \varnothing$, and let $x_0 \in D$. Set

$$(\forall n \in \mathbb{N}) \quad x_n = \frac{1}{n+1} \sum_{k=0}^{n} T^k x_0. \tag{5.55}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{Fix} T$.

Proof. Set $C = \operatorname{Fix} T$ and $(\forall k \in \mathbb{N})$ $y_k = T^k x_0$. By Example 5.3, $(y_k)_{k\in\mathbb{N}}$ is Fejér monotone with respect to C. Now set $(\forall n \in \mathbb{N})(\forall k \in \{0,\ldots,n\})$ $\alpha_{n,k} = 1/(n+1)$. Then (5.51) is satisfied and, in view of Corollary 5.37, it remains to show that every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ is in C. According to Corollary 4.28, it is enough to show that $x_n - Tx_n \to 0$. Set $(\forall n \in \mathbb{N})$ $z_n = (n+1)^{-1} \sum_{k=0}^{n} y_{k+1}$. Then we derive from Proposition 4.8(iv) and Proposition 5.4(iv) that, for every $n \in \mathbb{N}$,

$$\begin{aligned}
\|Tx_n - z_n\|^2 &\leqslant \frac{1}{2(n+1)^2} \sum_{k=0}^{n} \sum_{l=0}^{n} \big(\|y_k - y_l\|^2 - \|y_{k+1} - y_{l+1}\|^2\big) \\
&= \frac{1}{(n+1)^2} \sum_{k=1}^{n} \big(\|y_k - y_0\|^2 - \|y_k - y_{n+1}\|^2\big) \\
&\leqslant \frac{1}{(n+1)^2} \sum_{k=1}^{n} \|y_k - y_0\|^2 \\
&\leqslant \frac{2nd_C^2(x_0)}{(n+1)^2}. \qquad (5.56)
\end{aligned}$$

Thus, $Tx_n - z_n \to 0$. On the other hand, it follows from Proposition 5.4(iv) that, for every $n \in \mathbb{N}$,

$$\|z_n - x_n\| = \frac{1}{n+1}\left\| \sum_{k=0}^{n} (y_{k+1} - y_k) \right\| = \frac{\|y_{n+1} - y_0\|}{n+1} \leqslant \frac{2d_C(x_0)}{n+1}. \qquad (5.57)$$

Therefore, $z_n - x_n \to 0$ and we conclude that $x_n - Tx_n \to 0$. ☐

Exercises

Exercise 5.1 Find a nonexpansive operator $T\colon \mathcal{H} \to \mathcal{H}$ that is not firmly nonexpansive and such that, for every $x_0 \in \mathcal{H} \smallsetminus \operatorname{Fix} T$, the sequence $(T^n x_0)_{n\in\mathbb{N}}$ converges weakly but not strongly to a fixed point of T.

Exercise 5.2 Construct a non-Cauchy sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathbb{R}$ that is asymptotically regular, i.e., $x_n - x_{n+1} \to 0$.

Exercise 5.3 Let C be a nonempty subset of $\mathcal{H}$, let D be a closed affine subspace of $\mathcal{H}$ such that $C \subset D$, and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that, for every $x \in C$, $(\|x_n - x\|)_{n\in\mathbb{N}}$ converges and every weak sequential cluster point of $(P_D x_n)_{n\in\mathbb{N}}$ belongs to C. Show that $(P_D x_n)_{n\in\mathbb{N}}$ converges weakly to a point in C.

Exercise 5.4 Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ that is Fejér monotone with respect to a nonempty closed convex subset C of $\mathcal{H}$, and let x be a weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$. Show that $P_C x_n \to P_C x$.

Exercise 5.5 Suppose that $\mathcal{H}$ is finite-dimensional, let C and D be nonempty closed convex subsets of $\mathcal{H}$ such that $C \subset D$, let $(x_n)_{n\in\mathbb{N}}$ be Fejér monotone with respect to C, and suppose that all the cluster points of $(P_D x_n)_{n\in\mathbb{N}}$ lie in C. Show that $(P_D x_n)_{n\in\mathbb{N}}$ converges to $\lim P_C x_n$.

Exercise 5.6 Find an alternative proof of Theorem 5.5 based on Corollary 5.8 in the case when C is closed and convex.

Exercise 5.7 Let C be a nonempty subset of $\mathcal{H}$ and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ that is Fejér monotone with respect to C. Show that $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to $\overline{\mathrm{conv}}\, C$.

Exercise 5.8 Let $T\colon \mathcal{H} \to \mathcal{H}$ be a nonexpansive operator such that $\operatorname{Fix} T \neq \varnothing$, and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that the following hold:

(i) For every $x \in \operatorname{Fix} T$, $(\|x_n - x\|)_{n\in\mathbb{N}}$ converges.
(ii) $x_n - Tx_n \to 0$.

Show that $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{Fix} T$.

Exercise 5.9 Let $m \geqslant 2$ be an integer, set $I = \{1, \ldots, m\}$, let $(C_i)_{i\in I}$ be a family of closed convex subsets of $\mathcal{H}$ such that $C = \bigcap_{i\in I} C_i \neq \varnothing$, and let $(P_i)_{i\in I}$ be their respective projectors. Derive parts (ii) and (iii) from (i) and Theorem 5.5, and then also from Corollary 5.19.

(i) Let $i \in I$, let $x \in C_i$, and let $y \in \mathcal{H}$. Show that $\|P_i y - x\|^2 \leqslant \|y - x\|^2 - \|P_i y - y\|^2$.
(ii) Set $x_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \frac{1}{m}\big(P_1 x_n + P_1 P_2 x_n + \cdots + P_1 \cdots P_m x_n\big). \tag{5.58}$$

(a) Let $x \in C$ and $n \in \mathbb{N}$. Show that $\|x_{n+1} - x\|^2 \leqslant \|x_n - x\|^2 - (1/m)\sum_{i\in I} \|P_i x_n - x_n\|^2$.
(b) Let x be a weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$. Show that $x \in C$.
(c) Show that $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in C.

(iii) Set $x_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \frac{1}{m-1}\big(P_1 P_2 x_n + P_2 P_3 x_n + \cdots + P_{m-1} P_m x_n\big). \tag{5.59}$$

(a) Let $x \in C$ and $n \in \mathbb{N}$. Show that $\|x_{n+1} - x\|^2 \leqslant \|x_n - x\|^2 - \sum_{i=1}^{m-1}(\|P_{i+1} x_n - x_n\|^2 + \|P_i P_{i+1} x_n - P_{i+1} x_n\|^2)/(m-1)$.
(b) Let x be a weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$. Show that $x \in C$.
(c) Show that $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in C.

Exercise 5.10 Let $T\colon \mathcal{H} \to \mathcal{H}$ be firmly nonexpansive and such that $\operatorname{Fix} T \neq \varnothing$. Let $x_0 \in \mathcal{H}$ and set $(\forall n \in \mathbb{N})$ $x_{n+1} = Tx_n$. Show that there exists a constant $\gamma \in \mathbb{R}_{++}$ such that $(\forall n \in \mathbb{N})$ $\|x_{n+1} - x_n\|^2 \leqslant \gamma/(n+1)$.

Exercise 5.11 Suppose that $\mathcal{H} \neq \{0\}$. In connection with Example 5.38, provide a nonexpansive operator $T\colon \mathcal{H} \to \mathcal{H}$ and a point $x \in \mathcal{H}$ such that $\operatorname{Fix} T \neq \varnothing$ and $(T^k x)_{k\in\mathbb{N}}$ does not converge.

Chapter 6
Convex Cones and Generalized Interiors

The notion of a convex cone, which lies between that of a linear subspace and that of a convex set, is the main topic of this chapter. It has been very fruitful in many branches of nonlinear analysis. For instance, closed convex cones provide decompositions analogous to the well-known orthogonal decomposition based on closed linear subspaces. They also arise naturally in convex analysis in the local study of a convex set via the tangent cone and the normal cone operators, and they are central in the analysis of various extensions of the notion of an interior that will be required in later chapters.

6.1 Convex Cones

Recall from (1.1) that a subset C of $\mathcal{H}$ is a cone if $C = \mathbb{R}_{++}C$. Hence, $\mathcal{H}$ is a cone and the intersection of a family of cones is a cone. The following notions are therefore well defined.

Definition 6.1 Let C be a subset of $\mathcal{H}$. The *conical hull* of C is the intersection of all the cones in $\mathcal{H}$ containing C, i.e., the smallest cone in $\mathcal{H}$ containing C. It is denoted by $\operatorname{cone} C$. The *closed conical hull* of C is the smallest closed cone in $\mathcal{H}$ containing C. It is denoted by $\overline{\operatorname{cone}}\, C$.

Proposition 6.2 *Let C be a subset of $\mathcal{H}$. Then the following hold:*

(i) $\operatorname{cone} C = \mathbb{R}_{++}C$.
(ii) $\overline{\operatorname{cone} C} = \overline{\operatorname{cone}}\, C$.
(iii) $\operatorname{cone}(\operatorname{conv} C) = \operatorname{conv}(\operatorname{cone} C)$ *is the smallest convex cone containing C.*
(iv) $\overline{\operatorname{cone}}\,(\operatorname{conv} C) = \overline{\operatorname{conv}}\,(\operatorname{cone} C)$ *is the smallest closed convex cone containing C.*

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-48311-5_31

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_6

Proof. We assume that $C \neq \varnothing$.

(i): Set $D = \mathbb{R}_{++}C$. Then D is a cone and $C \subset D$. Therefore $\operatorname{cone} C \subset \operatorname{cone} D = D$. Conversely, take $y \in D$, say $y = \lambda x$, where $\lambda \in \mathbb{R}_{++}$ and $x \in C$. Then $x \in \operatorname{cone} C$ and therefore $y = \lambda x \in \operatorname{cone} C$. Thus, $D \subset \operatorname{cone} C$.

(ii): Since $\overline{\operatorname{cone}}\, C$ is a closed cone and $C \subset \overline{\operatorname{cone}}\, C$, we have $\overline{\operatorname{cone} C} \subset \overline{\operatorname{cone}(\overline{\operatorname{cone}}\, C)} = \overline{\operatorname{cone}}\, C$. Conversely, since the closure of a cone is a cone, we have $\overline{\operatorname{cone}}\, C \subset \overline{\operatorname{cone} C}$.

(iii): Take $x \in \operatorname{cone}(\operatorname{conv} C)$. Proposition 3.4 and (i) imply the existence of $\lambda \in \mathbb{R}_{++}$, of a finite family $(\alpha_i)_{i\in I}$ in $\mathbb{R}_{++}$, and of a family $(x_i)_{i\in I}$ in C such that $\sum_{i\in I} \alpha_i = 1$ and $x = \lambda \sum_{i\in I} \alpha_i x_i$. Thus, $x = \sum_{i\in I} \alpha_i(\lambda x_i) \in \operatorname{conv}(\operatorname{cone} C)$. Therefore, $\operatorname{cone}(\operatorname{conv} C) \subset \operatorname{conv}(\operatorname{cone} C)$. Conversely, take $x \in \operatorname{conv}(\operatorname{cone} C)$. Proposition 3.4 and (i) imply the existence of finite families $(\alpha_i)_{i\in I}$ in $\mathbb{R}_{++}$, $(\lambda_i)_{i\in I}$ in $\mathbb{R}_{++}$, and $(x_i)_{i\in I}$ in C such that $\sum_{i\in I} \alpha_i = 1$ and $x = \sum_{i\in I} \alpha_i \lambda_i x_i$. Set $(\forall i \in I)\ \beta_i = \alpha_i\lambda_i$, and $\lambda = \sum_{i\in I} \beta_i$. Then $\sum_{i\in I} \beta_i \lambda^{-1} x_i \in \operatorname{conv} C$ and hence $x = \lambda \sum_{i\in I} \beta_i \lambda^{-1} x_i \in \operatorname{cone}(\operatorname{conv} C)$. Now let K be the smallest convex cone containing C. Since K is a convex cone and $C \subset K$, we have $\operatorname{conv}(\operatorname{cone} C) \subset \operatorname{conv}(\operatorname{cone} K) = K$. On the other hand, $\operatorname{conv}(\operatorname{cone} C) = \operatorname{cone}(\operatorname{conv} C)$ is also a convex cone containing C and, therefore, $K \subset \operatorname{conv}(\operatorname{cone} C)$.

(iv): It follows from (iii) that $\overline{\operatorname{cone}}\,(\operatorname{conv} C) = \overline{\operatorname{conv}}\,(\operatorname{cone} C)$. Denote this set by D and denote the smallest closed convex cone containing C by K. Then D is closed and contains C. Proposition 3.46 and (ii) imply that D is a convex cone. Thus $K \subset D$. On the other hand, (iii) yields $\operatorname{cone}(\operatorname{conv} C) = \operatorname{conv}(\operatorname{cone} C) \subset K$. Taking closures, we deduce that $D \subset K$. Altogether, $K = D$. ☐

Convex cones are of particular importance due to their ubiquity in convex analysis. We record two simple propositions the proofs of which we leave as Exercise 6.2 and Exercise 6.3.

Proposition 6.3 *Let C be a subset of $\mathcal{H}$. Then the following hold:*

(i) *Suppose that C is a cone. Then C is convex if and only if $C + C \subset C$.*
(ii) *Suppose that C is convex and that $0 \in C$. Then C is a cone if and only if $C + C \subset C$.*

Proposition 6.4 *Let C be a nonempty convex subset of $\mathcal{H}$. Then the following hold:*

(i) $\operatorname{span} C = \operatorname{cone} C - \operatorname{cone} C = \operatorname{cone} C + \operatorname{cone}(-C)$.
(ii) *Suppose that $C = -C$. Then* $\operatorname{span} C = \operatorname{cone} C$.

Next, we introduce two important properties of convex cones.

Definition 6.5 Let K be a convex cone in $\mathcal{H}$. Then K is *pointed* if $K \cap (-K) \subset \{0\}$, and K is *solid* if $\operatorname{int} K \neq \varnothing$.

Note that $\{0\}$ is the only pointed linear subspace and $\mathcal{H}$ is the only solid linear subspace. The next examples illustrate the fact that various important cones are pointed or solid.

Example 6.6 Suppose that $u \in \mathcal{H} \smallsetminus \{0\}$ and set $K = \big\{x \in \mathcal{H} \;\big|\; \langle x \mid u\rangle \leqslant 0\big\}$. Then K is a solid convex cone, and it is not pointed if $\dim \mathcal{H} > 1$.

Proof. It is straightforward to check that K is a convex cone and that $\{u\}^{\perp} \subset K$. Moreover, since Cauchy–Schwarz implies that $B(-u; \|u\|) \subset K$, K is solid. Finally, take $x \in \{u\}^{\perp}$ such that $x \neq 0$. Then $\{0\} \neq \operatorname{span}\{x\} \subset K \cap (-K)$. Thus, K is not pointed. ☐

Example 6.7 Let I be a nonempty set, suppose that $\mathcal{H} = \ell^2(I)$, and let $(e_i)_{i\in I}$ be the standard unit vectors (see (2.8)). Then

$$\ell^2_+(I) = \big\{(\xi_i)_{i\in I} \in \ell^2(I) \;\big|\; (\forall i \in I)\ \xi_i \geqslant 0\big\} = \overline{\operatorname{cone}}\operatorname{conv}\{e_i\}_{i\in I} \tag{6.1}$$

is a nonempty closed convex pointed cone, and so is $\ell^2_-(I) = -\ell^2_+(I)$. Furthermore, $\ell^2_+(I)$ is solid if and only if I is finite. In particular, the positive orthant $\mathbb{R}^N_+$ is a closed convex cone in $\mathbb{R}^N$ that is pointed and solid.

Proof. It is clear that $\ell^2_+(I) = \bigcap_{i\in I} \big\{x \in \ell^2(I) \;\big|\; \langle x \mid e_i\rangle \geqslant 0\big\}$ is a nonempty pointed closed convex cone. Hence, since $\{e_i\}_{i\in I} \subset \ell^2_+(I) \subset \overline{\operatorname{cone}}\operatorname{conv}\{e_i\}_{i\in I}$, we obtain (6.1). If I is finite, then $B((1)_{i\in I}; 1) \subset \ell^2_+(I)$, and hence $\ell^2_+(I)$ is solid. Now assume that I is infinite and that $\ell^2_+(I)$ is solid. Then there exist $x = (\xi_i)_{i\in I} \in \ell^2_+(I)$ and $\varepsilon \in \mathbb{R}_{++}$ such that $B(x; 2\varepsilon) \subset \ell^2_+(I)$. Since I is infinite, there exists $j \in I$ such that $\xi_j \leqslant \varepsilon$. On the one hand, $(\eta_i)_{i\in I} = x - 2\varepsilon e_j \in B(x; 2\varepsilon) \subset \ell^2_+(I)$. On the other hand, $\eta_j = \xi_j - 2\varepsilon \leqslant -\varepsilon$, which implies that $(\eta_i)_{i\in I} \notin \ell^2_+(I)$. We therefore arrive at a contradiction. ☐

Proposition 6.8 *Let $\{x_i\}_{i\in I}$ be a nonempty finite subset of $\mathcal{H}$ and set*

$$K = \sum_{i\in I} \mathbb{R}_+ x_i. \tag{6.2}$$

Then K is the smallest closed convex cone containing $\{x_i\}_{i\in I} \cup \{0\}$.

Proof. We claim that

$$\operatorname{cone}\big(\operatorname{conv}\big(\{x_i\}_{i\in I} \cup \{0\}\big)\big) = K. \tag{6.3}$$

Set $C = \operatorname{cone}\big(\operatorname{conv}(\{x_i\}_{i\in I} \cup \{0\})\big)$ and let $x \in C$. Then there exist $\lambda \in \mathbb{R}_{++}$ and a family $(\alpha_i)_{i\in I}$ in $\mathbb{R}_+$ such that $\sum_{i\in I} \alpha_i \leqslant 1$ and $x = \lambda \sum_{i\in I} \alpha_i x_i$. Hence $x = \sum_{i\in I} (\lambda\alpha_i) x_i \in \sum_{i\in I} \mathbb{R}_+ x_i$ and thus $C \subset K$. Conversely, let $x \in K$. Then there exists a family $(\alpha_i)_{i\in I}$ in $\mathbb{R}_+$ such that $x = \sum_{i\in I} \alpha_i x_i$. If $\alpha_i \equiv 0$, then $x = 0 \in C$; otherwise, set $\lambda = \sum_{i\in I} \alpha_i$ and observe that $x = \lambda \sum_{i\in I} (\alpha_i/\lambda) x_i \in C$. Therefore, $K \subset C$ and (6.3) follows. Using Proposition 6.2(iii), we deduce that K is the smallest convex cone containing $\{x_i\}_{i\in I} \cup \{0\}$.

In view of (6.3) and Proposition 6.2(iv), it remains to verify that K is closed. To do so, we consider two alternatives.

(a) $\{x_i\}_{i\in I}$ is linearly independent: Set $V = \operatorname{span}\{x_i\}_{i\in I}$. Then $\{x_i\}_{i\in I}$ is a basis of V and $K \subset V$. Now let $z \in \overline{K}$ and let $(z_n)_{n\in\mathbb{N}}$ be a sequence in K such that $z_n \to z$. Then $z \in V$ and hence there exists $\{\alpha_i\}_{i\in I} \subset \mathbb{R}$ such that $z = \sum_{i\in I} \alpha_i x_i$. However, for every $n \in \mathbb{N}$, there exists $\{\alpha_{n,i}\}_{i\in I} \subset \mathbb{R}_+$ such that $z_n = \sum_{i\in I} \alpha_{n,i} x_i$. Since $(z_n)_{n\in\mathbb{N}}$ converges to z in V and $\{x_i\}_{i\in I}$ is a basis of V, we have $(\forall i \in I)\ 0 \leqslant \alpha_{n,i} \to \alpha_i$. Thus $\min_{i\in I} \alpha_i \geqslant 0$ and therefore $z \in K$.

(b) $\{x_i\}_{i\in I}$ is linearly dependent: Then there exists $\{\beta_i\}_{i\in I} \subset \mathbb{R}$ such that

$$\sum_{i\in I} \beta_i x_i = 0 \quad \text{and} \quad J = \big\{ i \in I \;\big|\; \beta_i < 0 \big\} \neq \varnothing. \tag{6.4}$$

Fix $z \in K$, say $z = \sum_{i\in I} \alpha_i x_i$, where $\{\alpha_i\}_{i\in I} \subset \mathbb{R}_+$, and set $(\forall i \in I)$ $\delta_i = \alpha_i - \gamma\beta_i$, where $\gamma = \max_{i\in J}\{\alpha_i/\beta_i\}$. Then $\gamma \leqslant 0$, $\{\delta_i\}_{i\in I} \subset \mathbb{R}_+$, and $z = \sum_{i\in I} \delta_i x_i$. Moreover, if $j \in J$ satisfies $\alpha_j/\beta_j = \gamma$, then $\delta_j = 0$ and therefore $z = \sum_{i\in I\smallsetminus\{j\}} \delta_i x_i$. Thus, we obtain the decomposition

$$K = \bigcup_{j\in I} K_j, \quad \text{where} \quad (\forall j \in I) \quad K_j = \sum_{i\in I\smallsetminus\{j\}} \mathbb{R}_+ x_i. \tag{6.5}$$

If the families $(\{x_i\}_{i\in I\smallsetminus\{j\}})_{j\in I}$ are linearly independent, it follows from (a) that the sets $(K_j)_{j\in I}$ are closed and that K is therefore closed. Otherwise, for every $j \in I$ for which $\{x_i\}_{i\in I\smallsetminus\{j\}}$ is linearly dependent, we reapply the decomposition procedure to K_j recursively until it can be expressed as a union of cones of the form $\sum_{i\in I\smallsetminus I'} \mathbb{R}_+ x_i$, where $\{x_i\}_{i\in I\smallsetminus I'}$ is linearly independent. We thus obtain a decomposition of K as a finite union of closed sets. ☐

6.2 Generalized Interiors

The interior of a subset C of $\mathcal{H}$ can be expressed as

$$\operatorname{int} C = \big\{ x \in C \;\big|\; (\exists\, \rho \in \mathbb{R}_{++}) \quad B(0;\rho) \subset C - x \big\}. \tag{6.6}$$

This formulation suggests several weaker notions of interiority.

Definition 6.9 Let C be a convex subset of $\mathcal{H}$. The *core* of C is

$$\operatorname{core} C = \big\{ x \in C \;\big|\; \operatorname{cone}(C - x) = \mathcal{H} \big\}; \tag{6.7}$$

the *strong relative interior* of C is

$$\operatorname{sri} C = \big\{ x \in C \;\big|\; \operatorname{cone}(C - x) = \overline{\operatorname{span}}\,(C - x) \big\}; \tag{6.8}$$

the *relative interior* of C is

$$\operatorname{ri} C = \big\{x \in C \bigm| \operatorname{cone}(C - x) = \operatorname{span}(C - x)\big\}; \tag{6.9}$$

and the *quasirelative interior* of C is

$$\operatorname{qri} C = \big\{x \in C \bigm| \overline{\operatorname{cone}}\,(C - x) = \overline{\operatorname{span}}\,(C - x)\big\}. \tag{6.10}$$

In addition, we use the notation $\overline{\operatorname{ri}}\, C = \overline{\operatorname{ri} C}$ and $\overline{\operatorname{qri}}\, C = \overline{\operatorname{qri} C}$.

Example 6.10 Let C be a nonempty convex subset of $\mathcal{H}$ such that $C = -C$. Then the following hold:

(i) $0 \in \operatorname{ri} C$.
(ii) Suppose that $\operatorname{span} C$ is closed. Then $0 \in \operatorname{sri} C$.

Proof. By Proposition 6.4(ii), $\operatorname{cone} C = \operatorname{span} C$. ☐

For every convex subset C of $\mathcal{H}$, since $\operatorname{cone} C \subset \operatorname{span} C \subset \overline{\operatorname{span}}\, C$, we have

$$\operatorname{int} C \subset \operatorname{core} C \subset \operatorname{sri} C \subset \operatorname{ri} C \subset \operatorname{qri} C \subset C. \tag{6.11}$$

As we now illustrate, each of the inclusions in (6.11) can be strict.

Example 6.11

(i) Example 8.42(iii) will provide a convex set C such that $\operatorname{int} C = \varnothing$ and $0 \in \operatorname{core} C$. In contrast, Proposition 6.12 and Fact 6.13 provide common instances when $\operatorname{int} C = \operatorname{core} C$.
(ii) Let C be a proper closed linear subspace of $\mathcal{H}$. Then $\operatorname{core} C = \varnothing$ and $\operatorname{sri} C = C$.
(iii) Suppose that $\mathcal{H}$ is infinite-dimensional and separable, let $(e_n)_{n\in\mathbb{N}}$ be an orthonormal basis of $\mathcal{H}$, and set

$$C = \bigg\{\sum_{n\in\mathbb{N}} \xi_n e_n \biggm| (\forall n \in \mathbb{N})\ \ |\xi_n| \leqslant \frac{1}{4^n}\bigg\}. \tag{6.12}$$

Then C is closed, convex, and $C = -C$. Hence, by Proposition 6.4(ii), $\operatorname{span} C = \operatorname{cone} C$. Since $\{4^{-n}e_n\}_{n\in\mathbb{N}} \subset C$, we see that $\overline{\operatorname{span}}\, C = \mathcal{H}$. Now set $x = \sum_{n\in\mathbb{N}} 2^{-n}e_n$. Then $x \in \overline{\operatorname{span}}\, C$ and, if we had $x \in \operatorname{cone} C$, then there would exist $\beta \in \mathbb{R}_{++}$ such that $(\forall n \in \mathbb{N})\ 2^{-n} \leqslant \beta 4^{-n}$, which is impossible. Hence $x \notin \operatorname{cone} C$ and thus

$$\operatorname{cone} C = \operatorname{span} C \neq \overline{\operatorname{span}}\, C = \mathcal{H}. \tag{6.13}$$

Therefore, $0 \in (\operatorname{ri} C) \smallsetminus (\operatorname{sri} C)$.
(iv) Suppose that $\mathcal{H}$ is infinite-dimensional and separable, let $(e_n)_{n\in\mathbb{N}}$ be an orthonormal basis of $\mathcal{H}$, and set

$$C = \bigg\{\sum_{n\in\mathbb{N}} \xi_n e_n \biggm| (\forall n \in \mathbb{N})\ \ -\frac{1}{2^n} \leqslant \xi_n \leqslant \frac{1}{4^n}\bigg\}. \tag{6.14}$$

Then C is closed and convex and, since $\{e_n, -e_n\}_{n\in\mathbb{N}} \subset \operatorname{cone} C$, we have $\overline{\operatorname{cone}}\, C = \overline{\operatorname{span}}\, C = \mathcal{H}$. Moreover, arguing as in (iii), we note that $x = -\sum_{n\in\mathbb{N}} 2^{-n} e_n \in C \subset \operatorname{cone} C$, while $-x \notin \operatorname{cone} C$. Finally, $\sum_{n\in\mathbb{N}} 2^{-n/2} e_n \in (\overline{\operatorname{span}}\, C) \smallsetminus (\operatorname{span} C)$. Altogether,

$$\operatorname{cone} C \neq \operatorname{span} C \neq \overline{\operatorname{span}}\, C = \overline{\operatorname{cone}}\, C = \mathcal{H} \tag{6.15}$$

and, therefore, $0 \in (\operatorname{qri} C) \smallsetminus (\operatorname{ri} C)$.

(v) Suppose that $\mathcal{H} = \ell^2(\mathbb{R})$, let $(e_\rho)_{\rho\in\mathbb{R}}$ denote the standard unit vectors, and set $C = \ell^2_+(\mathbb{R})$, i.e.,

$$C = \left\{ \sum_{\rho\in\mathbb{R}} \xi_\rho e_\rho \in \ell^2(\mathbb{R}) \;\middle|\; (\forall \rho \in \mathbb{R})\;\; \xi_\rho \geqslant 0 \right\}. \tag{6.16}$$

Then C is closed and convex. Fix $x = \sum_{\rho\in\mathbb{R}} \xi_\rho e_\rho \in C$. Since $\{\rho \in \mathbb{R} \mid \xi_\rho \neq 0\}$ is countable, there exists $\gamma \in \mathbb{R}$ such that $\xi_\gamma = 0$. Note that the γ-coordinate of every vector in $\operatorname{cone}(C - x)$ is positive and the same is true for $\overline{\operatorname{cone}}\,(C - x)$. On the other hand, since $x + e_\gamma \in C$, we have $e_\gamma \in (C - x)$ and therefore $-e_\gamma \in \operatorname{span}(C - x) \subset \overline{\operatorname{span}}\,(C - x)$. Altogether, $-e_\gamma \in (\overline{\operatorname{span}}\,(C - x)) \smallsetminus (\overline{\operatorname{cone}}\,(C - x))$ and consequently $C \smallsetminus (\operatorname{qri} C) = C$.

Proposition 6.12 *Let C be a convex subset of $\mathcal{H}$, and suppose that one of the following holds:*

(i) $\operatorname{int} C \neq \varnothing$.
(ii) *C is closed.*
(iii) *$\mathcal{H}$ is finite-dimensional.*

Then $\operatorname{int} C = \operatorname{core} C$.

Proof. Let $x \in \operatorname{core} C$. It suffices to show that $x \in \operatorname{int} C$. After subtracting x from C and replacing C by $(-C) \cap C$, we assume that $x = 0$ and that $C = -C$. Thus, it is enough to assume that $0 \in \operatorname{core} C$ and to show that $0 \in \operatorname{int} C$.

(i): Take $y \in \operatorname{int} C$. Since $C = -C$, $-y \in \operatorname{int} C$, and Proposition 3.45(ii) yields $0 \in [-y, y] \subset \operatorname{int} C$.

(ii): Since $\bigcup_{n\in\mathbb{N}} nC = \mathcal{H}$, Lemma 1.44(i) yields $\operatorname{int} C \neq \varnothing$. Now apply (i).

(iii): Let $(e_i)_{i\in I}$ be an orthonormal basis of $\mathcal{H}$. There exists $\varepsilon \in \mathbb{R}_{++}$ such that $D = \operatorname{conv}\{-\varepsilon e_i, +\varepsilon e_i\}_{i\in I} \subset C$. Since $B(0; \varepsilon/\sqrt{\dim \mathcal{H}}) \subset D$, the proof is complete. ☐

The following results refine Proposition 6.12(ii) and provide further information on generalized interiors.

Fact 6.13 (See [329, Corollary 13.2]) Let C and D be closed convex subsets of $\mathcal{H}$. Then $\operatorname{int}(C - D) = \operatorname{core}(C - D)$.

Fact 6.14 *Let C be a nonempty convex subset of $\mathcal{H}$.*

(i) *Suppose that $\mathcal{H}$ is finite-dimensional. Then $\operatorname{ri} C$ is the interior of C relative to $\operatorname{aff} C$ and $\operatorname{ri} C \neq \varnothing$. Moreover, $\overline{\operatorname{ri} C} = \overline{C}$, $\operatorname{ri} \overline{C} = \operatorname{ri} C$, and $\operatorname{sri} C = \operatorname{ri} C = \operatorname{qri} C$.*
(ii) *Suppose that $\mathcal{H}$ is separable and that C is closed. Then $C = \overline{\operatorname{qri}} C$ and, in particular, $\operatorname{qri} C \neq \varnothing$.*
(iii) *Suppose that $\operatorname{int} C \neq \varnothing$. Then $\operatorname{int} C = \operatorname{core} C = \operatorname{sri} C = \operatorname{ri} C = \operatorname{qri} C$.*
(iv) *Let $\mathcal{K}$ be a finite-dimensional real Hilbert space, let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, and suppose that $\operatorname{qri} C \neq \varnothing$. Then $\operatorname{ri} L(C) = L(\operatorname{qri} C)$.*
(v) *Suppose that $\mathcal{H}$ is finite-dimensional and let D be a convex subset of $\mathcal{H}$ such that $(\operatorname{ri} C) \cap (\operatorname{ri} D) \neq \varnothing$. Then $\operatorname{ri}(C \cap D) = (\operatorname{ri} C) \cap (\operatorname{ri} D)$.*

Proof. (i): It follows from (6.7) and (6.9) that $\operatorname{ri} C$ is the core relative to $\operatorname{aff} C$. However, as seen in Proposition 6.12(iii), since $\mathcal{H}$ is finite-dimensional, the notions of core and interior coincide. Furthermore, $\operatorname{ri} C \neq \varnothing$ by [313, Theorem 6.2]. Next, since $\operatorname{ri} C$ is nonempty and coincides with the interior of C relative to $\operatorname{aff} C$, the identities $\overline{\operatorname{ri} C} = \overline{C}$ and $\operatorname{ri} \overline{C} = \operatorname{ri} C$ follow from Proposition 3.45(iii). We also observe that $\operatorname{sri} C = \operatorname{ri} C$ since finite-dimensional linear subspaces are closed. Now assume that $x \in \operatorname{qri} C$. Then $\overline{\operatorname{cone}}\,(C - x) = \overline{\operatorname{span}}\,(C - x) = \operatorname{span}(C - x)$, and hence $\operatorname{ri}(\operatorname{cone}(C - x)) = \operatorname{ri}(\overline{\operatorname{cone}}\,(C - x)) = \operatorname{ri}(\operatorname{span}(C - x)) = \operatorname{span}(C - x)$. It follows that $\operatorname{cone}(C - x) = \operatorname{span}(C - x)$, i.e., $x \in \operatorname{ri} C$. Altogether, $\operatorname{sri} C = \operatorname{ri} C = \operatorname{qri} C$.

(ii): See [69, Proposition 2.12 and Theorem 2.19] or [365, Lemma 2.7].

(iii): [69, Corollary 2.14] implies that $\operatorname{qri} C = \operatorname{int} C$ (see also [365, Lemma 2.8] when C is closed). The identities thus follow from (6.11).

(iv): See [69, Proposition 2.10].

(v): [313, Theorem 6.5]. ☐

Corollary 6.15 *Suppose that $\mathcal{H}$ is finite-dimensional, let $\mathcal{K}$ be a finite-dimensional real Hilbert space, let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, and let C and D be nonempty convex subsets of $\mathcal{H}$. Then the following hold:*

(i) $\operatorname{ri} L(C) = L(\operatorname{ri} C)$.
(ii) $\operatorname{ri}(C - D) = (\operatorname{ri} C) - (\operatorname{ri} D)$.

Proof. (i): It follows from Fact 6.14(i) that $\operatorname{ri} C = \operatorname{qri} C \neq \varnothing$. Hence, Fact 6.14(iv) yields $\operatorname{ri} L(C) = L(\operatorname{ri} C)$.

(ii): Set $L \colon \mathcal{H} \times \mathcal{H} \to \mathcal{H} \colon (x, y) \mapsto x - y$. It follows from (i) that $\operatorname{ri}(C - D) = \operatorname{ri} L(C \times D) = L(\operatorname{ri}(C \times D)) = L((\operatorname{ri} C) \times (\operatorname{ri} D)) = (\operatorname{ri} C) - (\operatorname{ri} D)$. ☐

Proposition 6.16 *Let C be a convex subset of $\mathcal{H}$ such that $\operatorname{int} C \neq \varnothing$ and $0 \in C$. Then $\operatorname{int}(\operatorname{cone} C) = \operatorname{cone}(\operatorname{int} C)$.*

Proof. It is clear that $\operatorname{cone}(\operatorname{int} C) \subset \operatorname{cone} C$ and that $\operatorname{cone}(\operatorname{int} C)$ is open, since it is a union of open sets. Hence

$$\operatorname{cone}(\operatorname{int} C) \subset \operatorname{int}(\operatorname{cone} C). \tag{6.17}$$

To prove the reverse inclusion, take $x \in \operatorname{int}(\operatorname{cone} C)$. We must show that

$$x \in \operatorname{cone}(\operatorname{int} C). \tag{6.18}$$

Since $x \in \operatorname{int}(\operatorname{cone} C)$, there exist $\varepsilon_0 \in \mathbb{R}_{++}$, $\gamma \in \mathbb{R}_{++}$, and $x_1 \in C$ such that $B(x;\varepsilon_0) \subset \operatorname{cone} C$ and $x = \gamma x_1 \in \gamma C$. If $x_1 \in \operatorname{int} C$, then (6.18) holds. We therefore assume that $x_1 \in C \smallsetminus (\operatorname{int} C)$. Fix $y_1 \in \operatorname{int} C$ and set $y = \gamma y_1 \in \gamma C$. Since $x_1 \neq y_1$, we have $x \neq y$. Now set $\varepsilon = \varepsilon_0/\|x-y\|$. Then $x + \varepsilon(x-y) \in B(x;\varepsilon_0) \subset \operatorname{cone} C$, and hence there exists $\rho \in \mathbb{R}_{++}$ such that $x+\varepsilon(x-y) \in \rho C$. Set $\mu = \max\{\gamma,\rho\} > 0$. Because C is convex and $0 \in C$, we have $(\gamma C) \cup (\rho C) = \mu C$. On the other hand, the inclusions $x \in \gamma C \subset \mu C$, $y \in \gamma C \subset \mu C$, and $x + \varepsilon(x-y) \in \rho C \subset \mu C$ yield

$$x/\mu \in C, \quad y/\mu \in C, \quad \text{and} \quad x/\mu + \varepsilon(x/\mu - y/\mu) \in C. \tag{6.19}$$

We claim that

$$y/\mu \in \operatorname{int} C. \tag{6.20}$$

If $y_1 = 0$, then $y = \gamma y_1 = 0$ and thus $y/\mu = 0 = y_1 \in \operatorname{int} C$; otherwise, $y_1 \neq 0$ and Proposition 3.44 yields $y/\mu = (\gamma/\mu)y_1 \in \left]0, y_1\right] \subset \operatorname{int} C$. Hence (6.20) holds. Now set $\lambda = 1/(1+\varepsilon) \in \left]0,1\right[$. Then $\lambda\varepsilon = 1-\lambda$ and, since $x/\mu + \varepsilon(x/\mu - y/\mu) \in C$ by (6.19) and $y/\mu \in \operatorname{int} C$ by (6.20), it follows from Proposition 3.44 that $x/\mu = \lambda(x/\mu + \varepsilon(x/\mu - y/\mu)) + (1-\lambda)(y/\mu) \in \operatorname{int} C$. Thus $x \in \mu \operatorname{int} C$ and (6.18) follows. ☐

Proposition 6.17 *Let C be a convex subset of $\mathcal{H}$ such that $\operatorname{int} C \neq \varnothing$ and $0 \in C$. Then the following are equivalent:*

(i) $0 \in \operatorname{int} C$.
(ii) $\operatorname{cone}(\operatorname{int} C) = \mathcal{H}$.
(iii) $\operatorname{cone} C = \mathcal{H}$.
(iv) $\overline{\operatorname{cone}}\, C = \mathcal{H}$.

Proof. (i)⇒(ii)⇒(iii)⇒(iv): Clear.

(iv)⇒(ii): Since Proposition 6.2(iii) asserts that $\operatorname{cone} C$ is convex, Proposition 3.45(iii) and Proposition 6.16 imply that

$$\mathcal{H} = \operatorname{int} \mathcal{H} = \operatorname{int}(\overline{\operatorname{cone}}\, C) = \operatorname{int}(\operatorname{cone} C) = \operatorname{cone}(\operatorname{int} C). \tag{6.21}$$

(ii)⇒(i): We have $0 \in \operatorname{cone}(\operatorname{int} C)$ and thus $0 \in \lambda \operatorname{int} C$, for some $\lambda \in \mathbb{R}_{++}$. We conclude that $0 \in \operatorname{int} C$. ☐

The next example illustrates the fact that items (i) and (iv) in Proposition 6.17 are no longer equivalent when the assumption on the interior is dropped.

Example 6.18 Suppose that $\mathcal{H}$ is infinite-dimensional, let $(e_n)_{n\in\mathbb{N}}$ be an orthonormal sequence in $\mathcal{H}$, and set $S = \overline{\operatorname{conv}}\,\{\pm 2^{-n} e_n\}_{n\in\mathbb{N}}$ and $C = S + S^{\perp}$. Then C is closed and convex, $0 \in C$, $\operatorname{int} C = \varnothing$, and $\overline{\operatorname{cone}}\, C = \mathcal{H}$.

Proof. Since $\operatorname{span}\{\pm e_n\}_{n\in\mathbb{N}} + S^\perp \subset \operatorname{cone} C$, we have $\overline{\operatorname{cone}}\, C = \mathcal{H}$. Furthermore, $0 \in [-e_0, e_0] \subset S \subset C$. Now assume that $0 \in \operatorname{int} C$ and let $m \in \mathbb{N}$. Then there exists $\varepsilon \in \mathbb{R}_{++}$ such that $B(0;\varepsilon) \subset C$ and $\varepsilon e_m \in \overline{\operatorname{conv}}\,\{\pm 2^{-n} e_n\}_{n\in\mathbb{N}}$. Hence $\varepsilon = \langle \varepsilon e_m \mid e_m\rangle \in \langle \overline{\operatorname{conv}}\,\{\pm 2^{-n} e_n\}_{n\in\mathbb{N}} \mid e_m\rangle$. Since $\langle \cdot \mid e_m\rangle$ is continuous and linear, it follows that

$$\varepsilon \in \overline{\operatorname{conv}}\,\big\langle \{\pm 2^{-n} e_n\}_{n\in\mathbb{N}} \mid e_m \big\rangle = [-2^{-m}, 2^{-m}]. \tag{6.22}$$

Thus, $\varepsilon \leqslant 2^{-m}$, which is impossible since m is arbitrary. Therefore, $0 \notin \operatorname{int} C$ and hence $\operatorname{int} C = \varnothing$ by Proposition 6.17. ☐

The property that the origin lies in the strong relative interior of composite sets will be central in several places in this book. The next proposition provides sufficient conditions under which it is satisfied.

Proposition 6.19 *Let C be a convex subset of $\mathcal{H}$, let $\mathcal{K}$ be a real Hilbert space, let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$, and let D be a convex subset of $\mathcal{K}$. Suppose that one of the following holds:*

(i) *$D - L(C)$ is a closed linear subspace.*
(ii) *C and D are linear subspaces and one of the following holds:*
   (a) *$D + L(C)$ is closed.*
   (b) *D is closed, and $L(C)$ is finite-dimensional or finite-codimensional.*
   (c) *D is finite-dimensional or finite-codimensional, and $L(C)$ is closed.*
(iii) *D is a cone and $D - \operatorname{cone} L(C)$ is a closed linear subspace.*
(iv) *$D = L(C)$ and $\operatorname{span} D$ is closed.*
(v) *$0 \in \operatorname{core}(D - L(C))$.*
(vi) *$0 \in \operatorname{int}(D - L(C))$.*
(vii) *$D \cap \operatorname{int} L(C) \neq \varnothing$ or $L(C) \cap \operatorname{int} D \neq \varnothing$.*
(viii) *$\mathcal{K}$ is finite-dimensional and $(\operatorname{ri} D) \cap (\operatorname{ri} L(C)) \neq \varnothing$.*
(ix) *$\mathcal{K}$ is finite-dimensional and $(\operatorname{ri} D) \cap L(\operatorname{qri} C) \neq \varnothing$.*
(x) *$\mathcal{H}$ and $\mathcal{K}$ are finite-dimensional and $(\operatorname{ri} D) \cap L(\operatorname{ri} C) \neq \varnothing$.*

Then $0 \in \operatorname{sri}(D - L(C))$.

Proof. (i): We have $D - L(C) \subset \operatorname{cone}(D - L(C)) \subset \operatorname{span}(D - L(C)) \subset \overline{\operatorname{span}}\,(D - L(C))$. Hence, since the assumption implies that $D - L(C) = \overline{\operatorname{span}}\,(D - L(C))$, we obtain $\operatorname{cone}(D - L(C)) = \overline{\operatorname{span}}\,(D - L(C))$, and (6.8) yields $0 \in \operatorname{sri}(D - L(C))$.

(ii)(a): Since D and C are linear subspaces, so is $D - L(C) = D + L(C)$. Now apply (i).

(ii)(b)&(ii)(c): In view of Fact 2.28, this follows from (ii)(a).

(iii): Since $\overline{\operatorname{span}}\,(D - L(C)) \subset \overline{\operatorname{span}}\,(D - \operatorname{cone} L(C)) = D - \operatorname{cone} L(C) = \operatorname{cone}(D - L(C)) \subset \overline{\operatorname{span}}\,(D - L(C))$, we have $\operatorname{cone}(D - L(C)) = \overline{\operatorname{span}}\,(D - L(C))$, and (6.8) yields $0 \in \operatorname{sri}(D - L(C))$.

(iv): Since $D - L(C) = D - D = -(D - D) = -(D - L(C))$, Proposition 6.4(ii) yields $\operatorname{cone}(D-L(C)) = \operatorname{cone}(D-D) = \operatorname{span}(D-D) = \operatorname{span} D = \overline{\operatorname{span}}\, D = \overline{\operatorname{span}}\,(D - D) = \overline{\operatorname{span}}\,(D - L(C))$.

(v)&(vi): See (6.11).

(vii): Suppose that $y \in D \cap \operatorname{int} L(C)$, say $B(y;\rho) \subset L(C)$ for some $\rho \in \mathbb{R}_{++}$. Then $B(0;\rho) = y - B(y;\rho) \subset D - L(C)$ and therefore $0 \in \operatorname{int}(D - L(C))$. Now apply (vi). The second condition is handled analogously.

(viii): By Fact 6.14(i), $\operatorname{ri}(D - L(C)) = \operatorname{sri}(D - L(C))$. On the other hand, we derive from Corollary 6.15(ii) that $(\operatorname{ri} D) \cap (\operatorname{ri} L(C)) \neq \varnothing \Leftrightarrow 0 \in (\operatorname{ri} D) - (\operatorname{ri} L(C)) = \operatorname{ri}(D - L(C))$.

(ix): In view of Fact 6.14(iv), this follows from (viii).

(x): In view of Fact 6.14(i), this follows from (ix). ☐

Proposition 6.20 *Let m be an integer such that $m \geqslant 2$, set $I = \{1, \ldots, m\}$, and let $(C_i)_{i\in I}$ be convex subsets of $\mathcal{H}$ such that one of the following holds:*

(i) *For every $i \in \{2, \ldots, m\}$, $C_i - \bigcap_{j=1}^{i-1} C_j$ is a closed linear subspace.*
(ii) *The sets $(C_i)_{i\in I}$ are linear subspaces and, for every $i \in \{2, \ldots, m\}$, $C_i + \bigcap_{j=1}^{i-1} C_j$ is closed.*
(iii) $C_m \cap \bigcap_{i=1}^{m-1} \operatorname{int} C_i \neq \varnothing$.
(iv) *$\mathcal{H}$ is finite-dimensional and $\bigcap_{i\in I} \operatorname{ri} C_i \neq \varnothing$.*

Then

$$0 \in \bigcap_{i=2}^{m} \operatorname{sri}\Big(C_i - \bigcap_{j=1}^{i-1} C_j\Big). \tag{6.23}$$

Proof. We apply several items from Proposition 6.19 with $\mathcal{K} = \mathcal{H}$ and $L = \operatorname{Id}$. The fact that each condition implies (6.23) is justified as follows.

(i): Use Proposition 6.19(i).

(ii): Use Proposition 6.19(ii)(a).

(iii): Use Proposition 6.19(vii).

(iv): Use Proposition 6.19(viii) and Fact 6.14(v). ☐

Proposition 6.21 *Let $\mathcal{K}$ be a real Hilbert space, let C be a nonempty convex subset of $\mathcal{H}$, let D be a nonempty convex subset of $\mathcal{K}$, let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$, let $A \subset \mathcal{H}$, and let $B \subset \mathcal{K}$. Then the following hold:*

(i) $0 \in \operatorname{sri}(D - L(C))$ *if and only if* $0 \in \operatorname{sri}(C \times D - \operatorname{gra} L)$.
(ii) *Suppose that $A \subset C \subset \overline{\operatorname{conv}}\, A$, $B \subset D \subset \overline{\operatorname{conv}}\, B$, and* $\operatorname{cone}(B-L(A)) = \overline{\operatorname{span}}\,(B - L(A))$. *Then* $0 \in \operatorname{sri}(D - L(C))$ *and* $0 \in \operatorname{sri}(C \times D - \operatorname{gra} L)$.

Proof. (i): Set $K = \operatorname{cone}(D - L(C))$ and $\boldsymbol{K} = \operatorname{cone}(C \times D - \operatorname{gra} L)$. Since C and D are convex sets and L is a linear operator, the cones K and $\boldsymbol{K}$ are convex.

First, assume that $0 \in \operatorname{sri} K$, and let $\boldsymbol{x} \in \boldsymbol{K}$. Then there exist $(x,y) \in C \times D$, $z \in \mathcal{H}$, and $\gamma \in \mathbb{R}_{++}$ such that $\boldsymbol{x} = \gamma(x - z, y - Lz) = \gamma(0, y - Lx) + \gamma(x - z, L(x - z))$. Since $\gamma(y - Lx) \in K = \operatorname{cone} K = \overline{\operatorname{span}}\, K$, there exists

$(\gamma_1, x_1, y_1) \in \mathbb{R}_{++} \times C \times D$ such that $-\gamma(y - Lx) = \gamma_1(y_1 - Lx_1)$. It follows that

$$\begin{aligned}
-\boldsymbol{x} &= \gamma_1(0, y_1 - Lx_1) - \gamma\big(x - z, L(x - z)\big) \\
&= \gamma_1\Big((x_1, y_1) - \Big(x_1 + \gamma\gamma_1^{-1}(x - z), L\big(x_1 + \gamma\gamma_1^{-1}(x - z)\big)\Big)\Big) \\
&\in \boldsymbol{K}.
\end{aligned} \tag{6.24}$$

This implies that $-\boldsymbol{K} \subset \boldsymbol{K}$ and thus that $\boldsymbol{K}$ is a linear subspace. Now let $\overline{\boldsymbol{x}} = (\overline{x}, \overline{y}) \in \overline{\boldsymbol{K}}$. Then there exist sequences $(\gamma_n)_{n\in\mathbb{N}}$ in $\mathbb{R}_{++}$, $(x_n)_{n\in\mathbb{N}}$ in C, $(y_n)_{n\in\mathbb{N}}$ in D, and $(z_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ such that $\boldsymbol{x}_n = \gamma_n(x_n - z_n, y_n - Lz_n) \to (\overline{x}, \overline{y}) = \overline{\boldsymbol{x}}$. Hence $\gamma_n(x_n - z_n) \to \overline{x}$, which yields $\gamma_n(Lx_n - Lz_n) \to L\overline{x}$ since $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$. It follows that $\gamma_n(y_n - Lx_n) = \gamma_n(y_n - Lz_n) - \gamma_n(Lx_n - Lz_n) \to \overline{y} - L\overline{x}$. On the other hand, $(\gamma_n(y_n - Lx_n))_{n\in\mathbb{N}}$ lies in K, which is a closed linear subspace by assumption. Altogether, there exists $(\tilde{\gamma}, \tilde{x}, \tilde{y}) \in \mathbb{R}_{++} \times C \times D$ such that $\overline{y} - L\overline{x} = \tilde{\gamma}(\tilde{y} - L\tilde{x})$. Hence

$$\overline{\boldsymbol{x}} = (\overline{x}, \overline{y}) = \tilde{\gamma}\Big((\tilde{x}, \tilde{y}) - \big(\tilde{x} - \tilde{\gamma}^{-1}\overline{x}, L(\tilde{x} - \tilde{\gamma}^{-1}\overline{x})\big)\Big) \in \boldsymbol{K}. \tag{6.25}$$

Therefore, $\boldsymbol{K}$ is closed.

Conversely, assume that $0 \in \operatorname{sri} \boldsymbol{K}$. Let $w \in K$. Then there exists $(\gamma, x, y) \in \mathbb{R}_{++} \times C \times D$ such that $w = \gamma(y - Lx)$. Hence $(0, w) = (\gamma(x - x), \gamma(y - Lx)) = \gamma((x, y) - (x, Lx)) \in \boldsymbol{K}$ and thus

$$\{0\} \times K \subset \boldsymbol{K}. \tag{6.26}$$

Since $\boldsymbol{K}$ is a linear subspace, there exists $(\gamma_1, x_1, y_1, z_1) \in \mathbb{R}_{++} \times C \times D \times \mathcal{H}$ such that $(0, -w) = \gamma_1((x_1, y_1) - (z_1, Lz_1))$. Hence $z_1 = x_1 \in C$ and thus $-w = \gamma_1(y_1 - Lx_1) \in K$. It follows that $-K \subset K$ and therefore that K is a linear subspace. Finally, let $\overline{w} \in \overline{K}$. Then there exist sequences $(\gamma_n)_{n\in\mathbb{N}}$ in $\mathbb{R}_{++}$, $(x_n)_{n\in\mathbb{N}}$ in C, and $(y_n)_{n\in\mathbb{N}}$ in D such that $\gamma_n(y_n - Lx_n) \to \overline{w}$. By (6.26), $(0, \gamma_n(y_n - Lx_n))_{n\in\mathbb{N}}$ lies in $\boldsymbol{K}$, which is closed. Thus $(0, \overline{w}) \in \boldsymbol{K}$ and we deduce the existence of $(\tilde{\gamma}, \tilde{x}, \tilde{y}, \tilde{z}) \in \mathbb{R}_{++} \times C \times D \times \mathcal{H}$ such that $(0, \overline{w}) = \tilde{\gamma}(\tilde{x} - \tilde{z}, \tilde{y} - L\tilde{z})$. It follows that $\tilde{z} = \tilde{x} \in C$ and hence that $\overline{w} = \tilde{\gamma}(\tilde{y} - L\tilde{x}) \in K$.

(ii): We have

$$\begin{aligned}
\operatorname{cone}\big(B - L(A)\big) &\subset \operatorname{cone}\big(D - L(C)\big) \\
&\subset \overline{\operatorname{span}}\big(D - L(C)\big) \\
&\subset \overline{\operatorname{span}}\big(\overline{\operatorname{conv}}\, B - L(\overline{\operatorname{conv}}\, A)\big) \\
&\subset \overline{\operatorname{span}}\big(B - L(A)\big) \\
&= \operatorname{cone}\big(B - L(A)\big).
\end{aligned} \tag{6.27}$$

Hence $\operatorname{cone}(D - L(C)) = \overline{\operatorname{span}}\,(D - L(C))$, i.e., $0 \in \operatorname{sri}(D - L(C))$. The last assertion follows from (i). ☐

6.3 Polar and Dual Cones

Definition 6.22 Let C be a subset of $\mathcal{H}$. The *polar cone* of C is

$$C^{\ominus} = \big\{u \in \mathcal{H} \;\big|\; \sup \langle C \mid u\rangle \leqslant 0\big\}, \tag{6.28}$$

and the *dual cone* of C is $C^{\oplus} = -C^{\ominus}$. If C is a nonempty convex cone, then C is *self-dual* if $C = C^{\oplus}$.

The next two results are immediate consequences of Definition 6.22 and (2.2).

Proposition 6.23 *Let C be a linear subspace of $\mathcal{H}$. Then $C^{\ominus} = C^{\oplus} = C^{\perp}$.*

Proposition 6.24 *Let C be a subset of $\mathcal{H}$. Then the following hold:*

(i) *Let $D \subset C$. Then $C^{\ominus} \subset D^{\ominus}$ and $C^{\oplus} \subset D^{\oplus}$.*
(ii) *$C^{\ominus}$ and $C^{\oplus}$ are nonempty closed convex cones.*
(iii) *$C^{\ominus} = (\operatorname{cone} C)^{\ominus} = (\operatorname{conv} C)^{\ominus} = \overline{C}^{\ominus}$.*
(iv) *$C^{\ominus} \cap C^{\oplus} = C^{\perp}$.*
(v) *Suppose that $\overline{\operatorname{cone}}\, C = -\overline{\operatorname{cone}}\, C$. Then $C^{\ominus} = C^{\oplus} = C^{\perp}$.*

Example 6.25 Let I be a nonempty set. Then $\ell^2_+(I)$ is self-dual. In particular, the positive orthant $\mathbb{R}^N_+$ in $\mathbb{R}^N$ is self-dual.

Proof. Let $(e_i)_{i\in I}$ be the standard unit vectors of $\ell^2_+(I)$. By Example 6.7, $\ell^2_+(I) = \overline{\operatorname{cone}}\operatorname{conv}\{e_i\}_{i\in I}$. Hence, using Proposition 6.24(iii), $(\ell^2_+(I))^{\oplus} = \{e_i\}^{\oplus}_{i\in I} = \big\{x \in \ell^2(I) \;\big|\; (\forall i \in I)\ \ \langle x \mid e_i\rangle \geqslant 0\big\} = \ell^2_+(I)$. ☐

Example 6.26 (Fejér) The convex cone of positive semidefinite symmetric matrices in $\mathbb{S}^N$ is self-dual.

Proof. See, e.g., [205, Corollary 7.5.4]. ☐

Proposition 6.27 *Let K_1 and K_2 be nonempty cones in $\mathcal{H}$. Then*

$$(K_1 + K_2)^{\ominus} = K_1^{\ominus} \cap K_2^{\ominus}. \tag{6.29}$$

If K_1 and K_2 are linear subspaces, then $(K_1 + K_2)^{\perp} = K_1^{\perp} \cap K_2^{\perp}$.

Proof. Fix $x_1 \in K_1$ and $x_2 \in K_2$. First, let $u \in (K_1 + K_2)^{\ominus}$. Then, for every $\lambda_1 \in \mathbb{R}_{++}$ and $\lambda_2 \in \mathbb{R}_{++}$, $\lambda_1 x_1 + \lambda_2 x_2 \in \lambda_1 K_1 + \lambda_2 K_2 = K_1 + K_2$ and therefore $\langle \lambda_1 x_1 + \lambda_2 x_2 \mid u\rangle \leqslant 0$. Setting $\lambda_1 = 1$ and letting $\lambda_2 \downarrow 0$ yields $u \in K_1^{\ominus}$. Likewise, setting $\lambda_2 = 1$ and letting $\lambda_1 \downarrow 0$ yields $u \in K_2^{\ominus}$. Thus, $(K_1 + K_2)^{\ominus} \subset K_1^{\ominus} \cap K_2^{\ominus}$. Conversely, let $u \in K_1^{\ominus} \cap K_2^{\ominus}$. Then $\langle x_1 \mid u\rangle \leqslant 0$ and $\langle x_2 \mid u\rangle \leqslant 0$; hence $\langle x_1 + x_2 \mid u\rangle \leqslant 0$. Thus $u \in (K_1+K_2)^{\ominus}$ and therefore $K_1^{\ominus} \cap K_2^{\ominus} \subset (K_1 + K_2)^{\ominus}$. Finally, the assertion concerning linear subspaces follows from Proposition 6.23. ☐

As just illustrated, the relationship between a cone and its polar cone is, in many respects, similar to that between a linear subspace and its orthogonal complement. The next three results partially generalize Corollary 3.24.

Proposition 6.28 *Let K be a nonempty closed convex cone in $\mathcal{H}$, let $x \in \mathcal{H}$, and let $p \in \mathcal{H}$. Then $p = P_K x \Leftrightarrow \big[p \in K,\ x - p \perp p,\ \text{and}\ x - p \in K^{\ominus}\big]$.*

Proof. Theorem 3.16 asserts that $p = P_K x$ if and only if

$$p \in K \quad \text{and} \quad (\forall y \in K) \quad \langle y - p \mid x - p \rangle \leqslant 0. \tag{6.30}$$

Suppose that (6.30) holds. Then $0 \in K$, $2p \in K$, and therefore $\langle -p \mid x - p \rangle \leqslant 0$ and $\langle 2p - p \mid x - p \rangle \leqslant 0$. Hence, $\langle p \mid x - p \rangle = 0$. In turn, $(\forall y \in K)$ $\langle y \mid x - p \rangle = \langle y - p \mid x - p \rangle + \langle p \mid x - p \rangle \leqslant 0$. Thus, $x - p \in K^{\ominus}$. Conversely, $(\forall y \in K)$ $\big[\langle p \mid x - p \rangle = 0 \text{ and } \langle y \mid x - p \rangle \leqslant 0\big] \Rightarrow \langle y - p \mid x - p \rangle \leqslant 0$. ☐

Example 6.29 Let I be a nonempty set, suppose that $\mathcal{H} = \ell^2(I)$, set $K = \ell^2_+(I)$, and let $x = (\xi_i)_{i \in I} \in \mathcal{H}$. Then $P_K x = (\max\{\xi_i, 0\})_{i \in I}$.

Proof. This is a direct application of Proposition 6.28 where, by Example 6.25, $K^{\ominus} = -K^{\oplus} = -K$. ☐

Theorem 6.30 (Moreau) *Let K be a nonempty closed convex cone in $\mathcal{H}$ and let $x \in \mathcal{H}$. Then the following hold:*

(i) $x = P_K x + P_{K^{\ominus}} x$.
(ii) $P_K x \perp P_{K^{\ominus}} x$.
(iii) $\|x\|^2 = d_K^2(x) + d_{K^{\ominus}}^2(x)$.

Proof. (i): Set $q = x - P_K x$. By Proposition 6.28, $q \in K^{\ominus}$, $x - q = P_K x \perp x - P_K x = q$, and $x - q = P_K x \in K \subset K^{\ominus\ominus}$. Appealing once more to Proposition 6.28, we conclude that $q = P_{K^{\ominus}} x$.

(ii): It follows from Proposition 6.28 and (i) that $P_K x \perp x - P_K x = P_{K^{\ominus}} x$.

(iii): Using (i) and (ii), we obtain $\|x\|^2 = \|P_{K^{\ominus}} x + P_K x\|^2 = \|P_{K^{\ominus}} x\|^2 + \|P_K x\|^2 = \|x - P_K x\|^2 + \|x - P_{K^{\ominus}} x\|^2 = d_K^2(x) + d_{K^{\ominus}}^2(x)$. ☐

There is only one way to split a vector into the sum of a vector in a closed linear subspace V and a vector in $V^{\ominus} = V^{\perp}$. For general convex cones, this is no longer true (for instance, in $\mathcal{H} = \mathbb{R}$, $0 = x - x$ for every $x \in K = \mathbb{R}_+ = -\mathbb{R}_-$). However, the decomposition provided by Theorem 6.30(i) is unique in several respects.

Corollary 6.31 *Let K be a nonempty closed convex cone in $\mathcal{H}$ and let $x \in \mathcal{H}$. Suppose that $y \in K \smallsetminus \{P_K x\}$ and $z \in K^{\ominus} \smallsetminus \{P_{K^{\ominus}} x\}$ satisfy $x = y + z$. Then the following hold:*

(i) $\|P_K x\| < \|y\|$ *and* $\|P_{K^{\ominus}} x\| < \|z\|$.
(ii) $\langle y \mid z \rangle < \langle P_K x \mid P_{K^{\ominus}} x \rangle = 0$.
(iii) $\|P_K x - P_{K^{\ominus}} x\| < \|y - z\|$.

Proof. (i): We deduce from Theorem 3.16 that $\|P_K x\| = \|x - P_{K^{\ominus}} x\| < \|x - z\| = \|y\|$ and that $\|P_{K^{\ominus}} x\| = \|x - P_K x\| < \|x - y\| = \|z\|$.

(ii): Since $y \in K$ and $z \in K^{\ominus}$, we have $\langle y \mid z \rangle \leqslant 0$. However, if $\langle y \mid z \rangle = 0$, then (i) and Theorem 6.30 yield $\|x\|^2 = \|y+z\|^2 = \|y\|^2 + \|z\|^2 > \|P_K x\|^2 + \|P_{K^{\ominus}} x\|^2 = \|x\|^2$, which is impossible.

(iii): By (i) and (ii), $\|y - z\|^2 = \|y\|^2 - 2\langle y \mid z \rangle + \|z\|^2 > \|P_K x\|^2 - 2\langle P_K x \mid P_{K^{\ominus}} x \rangle + \|P_{K^{\ominus}} x\|^2 = \|P_K x - P_{K^{\ominus}} x\|^2$. ☐

The following fact will be used repeatedly.

Proposition 6.32 *Let K be a nonempty closed convex cone in $\mathcal{H}$ and let $x \in \mathcal{H}$. Suppose that $\langle x \mid x - P_K x \rangle \leqslant 0$. Then $x \in K$.*

Proof. By Proposition 6.28, $\langle P_K x \mid x - P_K x \rangle = 0$. Therefore, $\|x - P_K x\|^2 = \langle x \mid x - P_K x \rangle - \langle P_K x \mid x - P_K x \rangle \leqslant 0$. Thus, $x = P_K x \in K$. ☐

Proposition 6.33 *Let C be a nonempty convex subset of $\mathcal{H}$. Then*

$$C^{\ominus\ominus} = \overline{\operatorname{cone}}\, C. \tag{6.31}$$

Proof. Set $K = \overline{\operatorname{cone}}\, C$. Since $C \subset C^{\ominus\ominus}$, Proposition 6.24(ii) yields $K \subset \overline{\operatorname{cone}}\, C^{\ominus\ominus} = C^{\ominus\ominus}$. Conversely, let $x \in C^{\ominus\ominus}$. By Proposition 6.24(iii) and Proposition 6.28, $x \in C^{\ominus\ominus} = K^{\ominus\ominus}$ and $x - P_K x \in K^{\ominus}$. Therefore $\langle x \mid x - P_K x \rangle \leqslant 0$ and, by Proposition 6.32, $x \in K$. Thus $C^{\ominus\ominus} \subset K$. ☐

Corollary 6.34 *Let K be a nonempty closed convex cone in $\mathcal{H}$. Then*

$$K^{\ominus\ominus} = K. \tag{6.32}$$

Proposition 6.35 *Let K_1 and K_2 be two nonempty convex cones in $\mathcal{H}$. Then*

$$(\overline{K_1} \cap \overline{K_2})^{\ominus} = \overline{K_1^{\ominus} + K_2^{\ominus}}. \tag{6.33}$$

Proof. By Proposition 6.27 and Proposition 6.33, $(K_1^{\ominus} + K_2^{\ominus})^{\ominus} = \overline{K_1} \cap \overline{K_2}$. Taking polars yields the result. ☐

Theorem 6.36 (Farkas) *Let $v \in \mathcal{H}$ and let $(u_i)_{i \in I}$ be a finite family in $\mathcal{H}$. Then $v \in \sum_{i \in I} \mathbb{R}_+ u_i \Leftrightarrow \bigcap_{i \in I} \{u_i\}^{\ominus} \subset \{v\}^{\ominus}$.*

Proof. Set $K = \sum_{i \in I} \mathbb{R}_+ u_i$. If $v \in K$, then clearly $\bigcap_{i \in I} \{u_i\}^{\ominus} \subset \{v\}^{\ominus}$. Conversely, assume that $\bigcap_{i \in I} \{u_i\}^{\ominus} \subset \{v\}^{\ominus}$. Proposition 6.8 and Proposition 6.28 yield $v - P_K v \in K^{\ominus} \subset \bigcap_{i \in I} \{u_i\}^{\ominus} \subset \{v\}^{\ominus}$, hence $\langle v \mid v - P_K v \rangle \leqslant 0$. In turn, Proposition 6.32 yields $v \in K$. ☐

The next result is a powerful generalization of Fact 2.25(iv)&(v), which correspond to the case when $U = \{0\}$ and $K = \{0\}$.

Proposition 6.37 *Let $\mathcal{K}$ be a real Hilbert space, let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, let U be a nonempty closed convex cone in $\mathcal{H}$, and let K be a nonempty closed convex cone in $\mathcal{K}$. Then the following hold:*

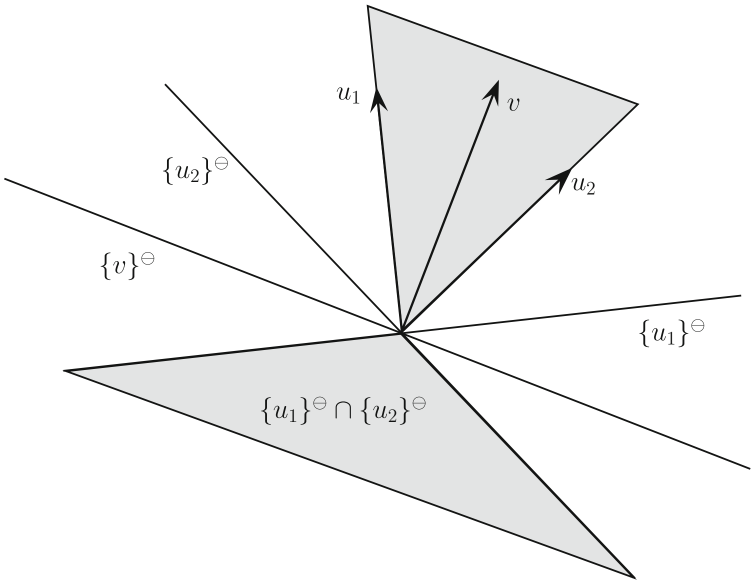

Fig. 6.1 Farkas's lemma: v lies in the cone generated by u_1 and u_2 if and only if the half-space $\{v\}^{\ominus}$ contains the cone $\{u_1\}^{\ominus} \cap \{u_2\}^{\ominus}$.

(i) $(L^{-1}(K))^{\ominus} = \overline{L^*(K^{\ominus})}$.
(ii) $(L^*)^{-1}(U) = (L(U^{\ominus}))^{\ominus}$.

Proof. (i): Set $C = L^{-1}(K)$ and $D = \overline{L^*(K^{\ominus})}$, and let $v \in K^{\ominus}$. Then $\sup \langle C \mid L^*v\rangle = \sup \langle L(C) \mid v\rangle \leqslant \sup \langle K \mid v\rangle = 0$ and hence $L^*v \in C^{\ominus}$. Therefore $L^*(K^{\ominus}) \subset C^{\ominus}$ and thus $D \subset C^{\ominus}$. Conversely, take $u \in C^{\ominus}$ and set $p = P_D u$, which is well defined since D is a nonempty closed convex cone. By Proposition 6.28, $u - p \in D^{\ominus}$. Therefore $\sup \langle K^{\ominus} \mid L(u-p)\rangle = \sup \langle L^*(K^{\ominus}) \mid u-p\rangle \leqslant 0$ and, using Corollary 6.34, we obtain $L(u-p) \in K^{\ominus\ominus} = K$. Thus, $u - p \in C$ and, in turn, $\langle u \mid u-p\rangle \leqslant 0$. Thus, Proposition 6.32 yields $u \in D$.

(ii): Since $L^* \in \mathcal{B}(\mathcal{K}, \mathcal{H})$, $(L^*)^{-1}(U)$ is a nonempty closed convex cone. Hence, Corollary 6.34, (i), and Proposition 6.24(iii) yield $(L^*)^{-1}(U) = ((L^*)^{-1}(U))^{\ominus\ominus} = \overline{L^{**}(U^{\ominus})}^{\ominus} = (L(U^{\ominus}))^{\ominus}$. ☐

6.4 Tangent and Normal Cones

Definition 6.38 Let C be a nonempty convex subset of $\mathcal{H}$ and let $x \in \mathcal{H}$. The *tangent cone* to C at x is

$$T_C x = \begin{cases} \overline{\operatorname{cone}}\,(C - x) = \overline{\bigcup_{\lambda \in \mathbb{R}_{++}} \lambda(C-x)}, & \text{if } x \in C;\\ \varnothing, & \text{otherwise,} \end{cases} \tag{6.34}$$

and the *normal cone* to C at x is

$$N_C x = \begin{cases} (C-x)^{\ominus} = \big\{u \in \mathcal{H} \;\big|\; \sup \langle C - x \mid u\rangle \leqslant 0\big\}, & \text{if } x \in C;\\ \varnothing, & \text{otherwise.}\end{cases} \tag{6.35}$$

Example 6.39 Let $C = B(0;1)$ and let $x \in C$. Then

$$T_C x = \begin{cases} \big\{y \in \mathcal{H} \;\big|\; \langle y \mid x\rangle \leqslant 0\big\}, & \text{if } \|x\| = 1;\\ \mathcal{H}, & \text{if } \|x\| < 1,\end{cases} \tag{6.36}$$

and

$$N_C x = \begin{cases} \mathbb{R}_+ x, & \text{if } \|x\| = 1;\\ \{0\}, & \text{if } \|x\| < 1.\end{cases} \tag{6.37}$$

Example 6.40 Let K be a nonempty convex cone in $\mathcal{H}$ and let $x \in K$. Then

$$T_K x = \overline{K + \mathbb{R}x} \quad \text{and} \quad N_K x = K^{\ominus} \cap \{x\}^{\perp}. \tag{6.38}$$

Proof. Proposition 6.3(i) yields $K + \mathbb{R}_+ x = K$, which implies that $K + \mathbb{R}x = K - \mathbb{R}_{++}x = \bigcup_{\lambda\in\mathbb{R}_{++}} K - \lambda x = \bigcup_{\lambda\in\mathbb{R}_{++}} \lambda(K - x)$. Taking closures, we obtain $T_K x = \overline{K + \mathbb{R}x}$. Let us now show that $N_K x = K^{\ominus} \cap \{x\}^{\perp}$. If $u \in K^{\ominus} \cap \{x\}^{\perp}$, then $(\forall y \in K)$ $\langle y - x \mid u\rangle = \langle y \mid u\rangle \leqslant 0$ and thus $u \in N_K x$. Conversely, take $u \in N_K x$. Then $(\forall y \in K)$ $\langle y - x \mid u\rangle \leqslant 0$. Since $\{x/2, 2x\} \subset K$, it follows that $\langle x \mid u\rangle = 0$ and hence that $u \in K^{\ominus}$. $\square$

Example 6.41 Let I be a nonempty set, suppose that $\mathcal{H} = \ell^2(I)$, set $K = \ell^2_+(I)$, and let $x = (\xi_i)_{i\in I} \in K$. Then

$$N_K x = \big\{(\nu_i)_{i\in I} \in \mathcal{H} \;\big|\; (\forall i \in I)\;\; \nu_i \leqslant 0 = \xi_i \nu_i\big\}. \tag{6.39}$$

Proof. Example 6.25 and Example 6.40 imply that $N_K x = \ell^2_-(I) \cap \{x\}^{\perp}$. Let $u = (\nu_i)_{i\in I} \in \ell^2_-(I)$. Since for every $i \in I$, $\nu_i \leqslant 0 \leqslant \xi_i$, we see that $\sum_{i\in I} \xi_i\nu_i = \langle x \mid u\rangle = 0$ if and only if $(\forall i \in I)$ $\xi_i\nu_i = 0$. $\square$

Example 6.42 Let N be a strictly positive integer, set $I = \{1, \ldots, N\}$, let $x = (\xi_i)_{i\in I} \in \mathbb{R}^N$, and let $y = (\eta_i)_{i\in I} \in \mathbb{R}^N$. Then the following hold:

(i) Suppose that $x \in \mathbb{R}^N_+$. Then $y \in N_{\mathbb{R}^N_+} x \Leftrightarrow (\forall i \in I) \begin{cases} \eta_i \leqslant 0, & \text{if } \xi_i = 0;\\ \eta_i = 0, & \text{if } \xi_i > 0.\end{cases}$

(ii) Suppose that $x \in \mathbb{R}^N_-$. Then $y \in N_{\mathbb{R}^N_-} x \Leftrightarrow (\forall i \in I) \begin{cases} \eta_i \geqslant 0, & \text{if } \xi_i = 0;\\ \eta_i = 0, & \text{if } \xi_i < 0.\end{cases}$

Example 6.43 Let C be an affine subspace of $\mathcal{H}$, let $V = C - C$ be its parallel linear subspace, and let $x \in \mathcal{H}$. Then

$$T_C x = \begin{cases} \overline{V}, & \text{if } x \in C;\\ \varnothing, & \text{otherwise,}\end{cases} \quad \text{and} \quad N_C x = \begin{cases} V^{\perp}, & \text{if } x \in C;\\ \varnothing, & \text{otherwise.}\end{cases} \tag{6.40}$$

Proposition 6.44 *Let C be a nonempty convex subset of $\mathcal{H}$ and let $x \in C$. Then the following hold:*

(i) $T_C^{\ominus} x = N_C x$ *and* $N_C^{\ominus} x = T_C x$.
(ii) $x \in \operatorname{core} C \Rightarrow T_C x = \mathcal{H} \Leftrightarrow N_C x = \{0\}$.

Proof. (i): Since $(C - x) \subset T_C x$, Proposition 6.24(i) yields $T_C^{\ominus} x \subset N_C x$. Now let $u \in N_C x$. Then $(\forall \lambda \in \mathbb{R}_{++})$ $\sup \langle \lambda(C - x) \mid u \rangle \leqslant 0$. Hence, $\sup \langle \operatorname{cone}(C - x) \mid u \rangle \leqslant 0$ and therefore $\sup \langle T_C x \mid u \rangle \leqslant 0$, i.e., $u \in T_C^{\ominus} x$. Altogether, $T_C^{\ominus} x = N_C x$. Furthermore, since $T_C x$ is a nonempty closed convex cone, Corollary 6.34 yields $T_C x = T_C^{\ominus\ominus} x = N_C^{\ominus} x$.

(ii): If $x \in \operatorname{core} C$, then $\operatorname{cone}(C - x) = \mathcal{H}$. In turn, $T_C x = \overline{\operatorname{cone}}\,(C - x) = \mathcal{H}$, and by (i), $N_C x = T_C^{\ominus} x = \mathcal{H}^{\ominus} = \{0\}$. Finally, it follows from (i) that $N_C x = \{0\} \Rightarrow T_C x = \{0\}^{\ominus} = \mathcal{H}$. ☐

Interior points of convex sets can be characterized via tangent and normal cones.

Proposition 6.45 *Let C be a convex subset of $\mathcal{H}$ such that $\operatorname{int} C \neq \varnothing$ and let $x \in C$. Then $x \in \operatorname{int} C \Leftrightarrow T_C x = \mathcal{H} \Leftrightarrow N_C x = \{0\}$.*

Proof. Set $D = C - x$. Then $0 \in D$ and $\operatorname{int} D = \operatorname{int} C - x \neq \varnothing$. In view of Proposition 6.17, $0 \in \operatorname{int} D \Leftrightarrow \overline{\operatorname{cone}}\, D = \mathcal{H}$, i.e., $x \in \operatorname{int} C \Leftrightarrow T_C x = \mathcal{H}$. The last equivalence is from Proposition 6.44. ☐

Corollary 6.46 *Suppose that $\mathcal{H}$ is finite-dimensional, let C be a nonempty convex subset of $\mathcal{H}$, and let $x \in C$. Then the following are equivalent:*

(i) $x \in \operatorname{int} C$.
(ii) $T_C x = \mathcal{H}$.
(iii) $N_C x = \{0\}$.

Proof. (i)⇒(ii)⇔(iii): (6.11) and Proposition 6.44(ii).

(iii)⇒(i): Set $U = \operatorname{aff} C$ and $V = U - U = U - x$. Then $C - x \subset U - x = V$ and it follows from Proposition 6.24(i) that $V^{\perp} = V^{\ominus} \subset (C - x)^{\ominus} = N_C x$. Since $N_C x = \{0\}$, we obtain $V^{\perp} = 0$ and thus $V = \mathcal{H}$. Hence $\operatorname{aff} C = \mathcal{H}$ and therefore $\operatorname{int} C = \operatorname{ri} C \neq \varnothing$ by Fact 6.14(i). The result follows from Proposition 6.45. ☐

We conclude this section with a characterization of projections onto closed convex sets.

Proposition 6.47 *Let C be a nonempty closed convex subset of $\mathcal{H}$, and let x and p be points in $\mathcal{H}$. Then $p = P_C x \Leftrightarrow x - p \in N_C p$.*

Proof. This follows at once from (3.10) and (6.35). ☐

6.5 Recession and Barrier Cones

Definition 6.48 Let C be a nonempty convex subset of $\mathcal{H}$. The *recession cone* of C is

$$\operatorname{rec} C = \big\{x \in \mathcal{H} \;\big|\; x + C \subset C\big\}, \tag{6.41}$$

and the *barrier cone* of C is

$$\operatorname{bar} C = \big\{u \in \mathcal{H} \;\big|\; \sup \langle C \mid u\rangle < +\infty\big\}. \tag{6.42}$$

Proposition 6.49 *Let C be a nonempty convex subset of $\mathcal{H}$. Then the following hold:*

(i) $\operatorname{rec} C$ *is a convex cone and* $0 \in \operatorname{rec} C$.
(ii) $\operatorname{bar} C$ *is a convex cone and* $C^{\ominus} \subset \operatorname{bar} C$.
(iii) *Suppose that C is bounded. Then* $\operatorname{bar} C = \mathcal{H}$.
(iv) *Suppose that C is a cone. Then* $\operatorname{bar} C = C^{\ominus}$.
(v) *Suppose that C is closed. Then* $(\operatorname{bar} C)^{\ominus} = \operatorname{rec} C$.

Proof. (i): It is readily verified that $0 \in \operatorname{rec} C$, that $\operatorname{rec} C + \operatorname{rec} C \subset \operatorname{rec} C$, and that $\operatorname{rec} C$ is convex. Hence, the result follows from Proposition 6.3(ii).

(ii): Clear from (6.28) and (6.42).

(iii): By Cauchy–Schwarz, $(\forall u \in \mathcal{H})\ \sup \langle C \mid u\rangle \leqslant \|u\| \sup \|C\| < +\infty$.

(iv): Take $u \in \operatorname{bar} C$. Since C is a cone, $\sup \langle C \mid u\rangle$ cannot be strictly positive, and hence $u \in C^{\ominus}$. Thus $\operatorname{bar} C \subset C^{\ominus}$, while $C^{\ominus} \subset \operatorname{bar} C$ by (ii).

(v): Take $x \in \operatorname{rec} C$. Then, for every $u \in \operatorname{bar} C$, $\langle x \mid u\rangle + \sup \langle C \mid u\rangle = \sup \langle x + C \mid u\rangle \leqslant \sup \langle C \mid u\rangle < +\infty$, which implies that $\langle x \mid u\rangle \leqslant 0$ and hence that $x \in (\operatorname{bar} C)^{\ominus}$. Thus, $\operatorname{rec} C \subset (\operatorname{bar} C)^{\ominus}$. Conversely, take $x \in (\operatorname{bar} C)^{\ominus}$ and $y \in C$, and set $p = P_C(x + y)$. By Proposition 6.47 and (ii), $x + y - p \in N_C p = (C - p)^{\ominus} \subset \operatorname{bar}(C - p) = \operatorname{bar} C$. Hence, since $x \in (\operatorname{bar} C)^{\ominus}$, we obtain $\langle x + y - p \mid x\rangle \leqslant 0$, and (3.10) yields $\|x + y - p\|^2 = \langle x + y - p \mid x\rangle + \langle x + y - p \mid y - p\rangle \leqslant 0$. Hence $x + y = p \in C$ and $x \in \operatorname{rec} C$. Thus, $(\operatorname{bar} C)^{\ominus} \subset \operatorname{rec} C$. ☐

Corollary 6.50 *Let K be a nonempty closed convex cone in $\mathcal{H}$. Then* $\operatorname{rec} K = K$.

Proof. It follows from Proposition 6.49(v), Proposition 6.49(iv), and Corollary 6.34 that $\operatorname{rec} K = (\operatorname{bar} K)^{\ominus} = K^{\ominus\ominus} = K$. ☐

The next result makes it clear why the recession cone of C is sometimes denoted by 0^+C.

Proposition 6.51 *Let C be a nonempty closed convex subset of $\mathcal{H}$ and let $x \in \mathcal{H}$. Then the following are equivalent:*

(i) $x \in \operatorname{rec} C$.
(ii) *There exist sequences $(x_n)_{n\in\mathbb{N}}$ in C and $(\alpha_n)_{n\in\mathbb{N}}$ in $]0,1]$ such that $\alpha_n \to 0$ and $\alpha_n x_n \to x$.*

(iii) *There exist sequences* $(x_n)_{n\in\mathbb{N}}$ *in* C *and* $(\alpha_n)_{n\in\mathbb{N}}$ *in* $]0,1]$ *such that* $\alpha_n \to 0$ *and* $\alpha_n x_n \rightharpoonup x$.

Proof. Take $y \in C$.

(i)⇒(ii): Proposition 6.49(i) yields $(\forall n \in \mathbb{N})$ $(n+1)x \in \operatorname{rec} C$. Now define $(\forall n \in \mathbb{N})$ $x_n = (n+1)x + y \in C$ and $\alpha_n = 1/(n+1)$. Then $\alpha_n \to 0$ and $\alpha_n x_n \to x$.

(ii)⇒(iii): Clear.

(iii)⇒(i): The sequence $(\alpha_n x_n + (1-\alpha_n)y)_{n\in\mathbb{N}}$ lies in C and Corollary 3.35 therefore implies that its weak limit $x + y$ belongs to C. It follows that $x \in \operatorname{rec} C$. ☐

Corollary 6.52 *Suppose that* $\mathcal{H}$ *is finite-dimensional. Let* C *be a nonempty closed convex subset of* $\mathcal{H}$. *Then* C *is bounded if and only if* $\operatorname{rec} C = \{0\}$.

Proof. If C is bounded, then clearly $\operatorname{rec} C = \{0\}$. Now assume that C is unbounded. Then there exist a sequence $(x_n)_{n\in\mathbb{N}}$ in C and a vector $y \in \mathcal{H}$ such that $\|x_n\| \to +\infty$ and $x_n/\|x_n\| \to y$. Hence $\|y\| = 1$, and thus Proposition 6.51 implies that $y \in (\operatorname{rec} C) \smallsetminus \{0\}$. ☐

Corollary 6.53 *Let* C *be a nonempty closed convex subset of* $\mathcal{H}$ *and suppose that* $0 \notin C$. *Then* $(\operatorname{cone} C) \cup (\operatorname{rec} C) = \overline{\operatorname{cone}}\, C$.

Proof. It is clear that $\operatorname{cone} C \subset \overline{\operatorname{cone}}\, C$. Now take $x \in \operatorname{rec} C$. By Proposition 6.51, x is the weak limit of a sequence $(\alpha_n x_n)_{n\in\mathbb{N}}$, where $(\alpha_n)_{n\in\mathbb{N}}$ lies in $]0,1]$, $\alpha_n \to 0$, and $(x_n)_{n\in\mathbb{N}}$ lies in C. Since $(\alpha_n x_n)_{n\in\mathbb{N}}$ lies in $\operatorname{cone} C$, x belongs to the weak closure of $\operatorname{cone} C$, which is $\overline{\operatorname{cone}}\, C$ by Theorem 3.34 and Proposition 6.2. Thus,

$$(\operatorname{cone} C) \cup (\operatorname{rec} C) \subset \overline{\operatorname{cone}}\, C. \tag{6.43}$$

Conversely, take $x \in \overline{\operatorname{cone}}\, C$. By Proposition 6.2(ii), there exist sequences $(\alpha_n)_{n\in\mathbb{N}}$ in $\mathbb{R}_{++}$ and $(x_n)_{n\in\mathbb{N}}$ in C such that $\alpha_n x_n \to x$. After passing to subsequences if necessary, we assume that $\alpha_n \to \alpha \in [0,+\infty]$. If $\alpha = +\infty$, then, since $\|\alpha_n x_n\| \to \|x\|$, it follows that $x_n \to 0$ and, in turn, that $0 \in C$, which violates our hypothesis. Hence $\alpha \in \mathbb{R}_+$. If $\alpha = 0$, then $x \in \operatorname{rec} C$ by Proposition 6.51. Otherwise, $\alpha \in \mathbb{R}_{++}$, in which case $x_n = (\alpha_n x_n)/\alpha_n \to x/\alpha \in C$, and hence $x \in \alpha C \subset \operatorname{cone} C$. ☐

Exercises

Exercise 6.1 Find a convex subset C of $\mathbb{R}^2$ such that $\overline{\operatorname{cone}}\, C \neq \bigcup_{\lambda\in\mathbb{R}_+} \lambda C$.

Exercise 6.2 Prove Proposition 6.3.

Exercise 6.3 Prove Proposition 6.4.

Exercise 6.4 Let $\alpha \in \mathbb{R}_{++}$ and set $K_\alpha = \big\{(x,\xi) \in \mathcal{H} \times \mathbb{R} \bigm| \|x\| \leqslant \alpha\xi\big\}$. Show that $K_\alpha^\oplus = K_{1/\alpha}$ and conclude that K_1 is self-dual.

Exercise 6.5 Let C be a convex subset of $\mathcal{H}$ and let $x \in \mathcal{H}$ be such that $\operatorname{cone}(C - x)$ is a linear subspace. Show that $x \in C$.

Exercise 6.6 Let C be a cone in $\mathcal{H}$, let $\mathcal{K}$ be a real Hilbert space, let D be a cone in $\mathcal{K}$, and let $L\colon \mathcal{H} \to \mathcal{K}$ be positively homogeneous. Show that $L(C)$ is a cone in $\mathcal{K}$ and that $L^{-1}(D)$ is a cone in $\mathcal{H}$.

Exercise 6.7 Suppose that $\mathcal{H}$ is finite-dimensional, and let C and D be nonempty convex subsets of $\mathcal{H}$. Show that the implication $C \subset D \Rightarrow \operatorname{ri} C \subset \operatorname{ri} D$ is false.

Exercise 6.8 Suppose that $\mathcal{H} = \ell^2(\mathbb{N})$, and set $C = \operatorname{lev}_{\leqslant 1} f$, where $f\colon \mathcal{H} \to \,]-\infty, +\infty] : (\xi_k)_{k\in\mathbb{N}} \mapsto \sum_{k\in\mathbb{N}} |\xi_k|$. Show that $0 \in (\operatorname{ri} C) \smallsetminus (\operatorname{sri} C)$.

Exercise 6.9 In connection with Proposition 6.8, find an infinite set C in $\mathbb{R}^2$ such that the convex cone $\sum_{c\in C} \mathbb{R}_+ c$ is not closed.

Exercise 6.10 Show that the conclusion of Proposition 6.16 fails if the assumption that $\operatorname{int} C \neq \varnothing$ is omitted.

Exercise 6.11 Let C be a subset of $\mathcal{H}$. Show that $-(C^\ominus) = (-C)^\ominus$, which justifies writing simply $-C^\ominus$ for these sets, and that $C^{\ominus\ominus} = C^{\oplus\oplus}$.

Exercise 6.12 Let K be a nonempty closed convex cone in $\mathcal{H}$. Then K is *acute* if $K \subset K^\oplus$, and K is *obtuse* if $K^\oplus \subset K$ (hence, a cone is self-dual if and only if it is both acute and obtuse). Prove that K is obtuse if and only if $K^\oplus$ is acute.

Exercise 6.13 Let K be a nonempty closed convex cone in $\mathcal{H}$. Prove that K is a linear subspace if and only if $K \cap K^\oplus = \{0\}$.

Exercise 6.14 Let K be a nonempty closed convex solid cone in $\mathcal{H}$. Show that $K^\ominus$ is pointed.

Exercise 6.15 Let K be a nonempty closed convex pointed cone in $\mathcal{H}$. Show the following:

(i) If $\mathcal{H}$ is finite-dimensional, then $K^\ominus$ is solid.
(ii) In general, $K^\ominus$ fails to be solid.

Exercise 6.16 Let N be a strictly positive integer, set $I = \{1, \ldots, N\}$, and set

$$K = \big\{(\xi_i)_{i\in I} \in \mathbb{R}^N \bigm| \xi_1 \geqslant \xi_2 \geqslant \cdots \geqslant \xi_N \geqslant 0\big\}. \tag{6.44}$$

Show that K is a nonempty pointed closed convex cone. Use Proposition 6.37 to show that

$$K^\ominus = \big\{(\zeta_i)_{i\in I} \in \mathbb{R}^N \bigm| \zeta_1 \leqslant 0, \zeta_1 + \zeta_2 \leqslant 0, \ldots, \zeta_1 + \cdots + \zeta_N \leqslant 0\big\}. \tag{6.45}$$

Furthermore, use Exercise 6.15(i) to show that K is solid.

Exercise 6.17 Let C be a nonempty convex subset of $\mathcal{H}$, and let $x \in C$. Show that $x \in \operatorname{qri} C \Leftrightarrow N_C x$ is a linear subspace.

Exercise 6.18 Let C and D be nonempty convex subsets of $\mathcal{H}$, and let $(x, y) \in C \times D$. Show that $N_C x \cap N_D y = N_{C+D}(x + y)$.

Exercise 6.19 Let C and D be convex subsets of $\mathcal{H}$, and suppose that $z \in C \cap D$. Show that the following are equivalent:

(i) C and D are separated.
(ii) $N_C z \cap (-N_D z) \neq \{0\}$.
(iii) $N_{C-D}\, 0 \neq \{0\}$.

Exercise 6.20 Let C be a convex subset of $\mathcal{H}$. Show that $\operatorname{span} C$ is closed if and only if $0 \in \operatorname{sri}(C - C)$. Furthermore, provide an example in which $\operatorname{span} C$ is not closed even though C is.

Exercise 6.21 Using Definition 6.38 directly, provide a proof of Example 6.39 and of Example 6.43.

Exercise 6.22 Find a closed convex subset C and a point $x \in C$ such that $T_C x = \mathcal{H}$ but $\operatorname{int} C = \varnothing$. Compare with Proposition 6.45.

Exercise 6.23 Suppose that $\mathcal{H} = \mathbb{R}^2$ and set $C = \{(\xi_1, \xi_2) \in \mathcal{H} \mid \xi_1^2 \leqslant \xi_2\}$. Determine $\operatorname{bar} C$ and observe that $\operatorname{bar} C$ is not closed.

Exercise 6.24 Let $v \in \mathcal{H}$ and let U be a nonempty finite subset of $\mathcal{H}$. Using Farkas's lemma (Theorem 6.36), show that $v \in \operatorname{span} U$ if and only if $\bigcap_{u \in U} \{u\}^\perp \subset \{v\}^\perp$.

Exercise 6.25 Let C be a nonempty closed convex subset of $\mathcal{H}$. Show that $\operatorname{ran}(\operatorname{Id} - P_C)$ is a cone and that $\overline{\operatorname{ran}}\,(\operatorname{Id} - P_C) \subset (\operatorname{rec} C)^\ominus$.

Exercise 6.26 Let C be a nonempty closed convex subset of $\mathcal{H}$ and let $u \in \mathcal{H}$. Show that $u \in (\operatorname{rec} C)^\ominus$ if and only if

$$(\forall x \in \mathcal{H}) \quad \lim_{\lambda \to +\infty} \frac{P_C(x + \lambda u)}{\lambda} = 0. \tag{6.46}$$

Exercise 6.27 Let C be a nonempty closed convex subset of $\mathcal{H}$. Use Exercise 6.25 and Exercise 6.26 to show that $\overline{\operatorname{ran}}\,(\operatorname{Id} - P_C) = (\operatorname{rec} C)^\ominus$.

Exercise 6.28 Suppose that $\mathcal{H}$ is infinite-dimensional and separable, and let $(e_n)_{n \in \mathbb{N}}$ be an orthonormal basis of $\mathcal{H}$. Set

$$C = \{x \in \mathcal{H} \mid (\forall n \in \mathbb{N}) \;\; |\langle x \mid e_n \rangle| \leqslant n\}. \tag{6.47}$$

Show that C is an unbounded closed convex set, and that $\operatorname{rec} C = \{0\}$. Compare with Corollary 6.52.

Exercise 6.29 Show that the conclusion of Corollary 6.53 fails if the assumption $0 \notin C$ is replaced by $0 \in C$.

Exercise 6.30 Consider Exercise 2.7 and its notation. Let C be a nonempty closed convex subset of $\mathcal{H}$ and denote its closure in $\operatorname{csm} \mathcal{H}$ by $\operatorname{csm} C$. Show that $\operatorname{csm} C = C \cup (\operatorname{rec}(C) \smallsetminus \{0\})$.

Chapter 7
Support Functions and Polar Sets

In this chapter, we develop basic results concerning support points, including the Bishop–Phelps theorem and the representation of a nonempty closed convex set as the intersection of the closed half-spaces containing it. Polar sets are also studied.

7.1 Support Points

Definition 7.1 Let C be a nonempty subset of $\mathcal{H}$, let $x \in C$, and suppose that $u \in \mathcal{H} \smallsetminus \{0\}$. If

$$\sup \langle C \mid u \rangle \leqslant \langle x \mid u \rangle, \tag{7.1}$$

then $\{y \in \mathcal{H} \mid \langle y \mid u \rangle = \langle x \mid u \rangle\}$ is a *supporting hyperplane* of C at x, and x is a *support point* of C with *normal vector* u. The set of support points of C is denoted by $\operatorname{spts} C$ and the closure of $\operatorname{spts} C$ by $\overline{\operatorname{spts}}\, C$.

Proposition 7.2 *Let C and D be nonempty subsets of $\mathcal{H}$ such that $C \subset D$. Then $C \cap \operatorname{spts} D \subset \operatorname{spts} C = C \cap \operatorname{spts} \overline{C}$.*

Proof. Let $x \in C \cap \operatorname{spts} D$. Then $x \in C$ and there exists $u \in \mathcal{H} \smallsetminus \{0\}$ such that $\sup \langle C \mid u \rangle \leqslant \sup \langle D \mid u \rangle \leqslant \langle x \mid u \rangle$. Hence, $x \in \operatorname{spts} C$. This verifies the announced inclusion, and the equality is clear from the definition. ☐

Proposition 7.3 *Let C be a nonempty convex subset of $\mathcal{H}$. Then*

$$\operatorname{spts} C = \{x \in C \mid N_C x \smallsetminus \{0\} \neq \varnothing\} = N_C^{-1}(\mathcal{H} \smallsetminus \{0\}). \tag{7.2}$$

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-48311-5_31

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_7

Proof. Let $x \in C$. Then $x \in \operatorname{spts} C \Leftrightarrow (\exists\, u \in \mathcal{H} \smallsetminus \{0\}) \sup \langle C - x \mid u \rangle \leqslant 0 \Leftrightarrow (\exists\, u \in \mathcal{H} \smallsetminus \{0\})\ u \in (C-x)^{\ominus} = N_C x$. ☐

Theorem 7.4 (Bishop–Phelps) *Let C be a nonempty closed convex subset of $\mathcal{H}$. Then* $\operatorname{spts} C = P_C(\mathcal{H} \smallsetminus C)$ *and* $\overline{\operatorname{spts}}\, C = \operatorname{bdry} C$.

Proof. We assume that $C \neq \mathcal{H}$. Fix $\varepsilon \in \mathbb{R}_{++}$ and let x be a support point of C with normal vector u. Then $\sup \langle C - x \mid (x + \varepsilon u) - x \rangle \leqslant 0$ and Theorem 3.16 implies that $x = P_C(x + \varepsilon u)$. Since $u \neq 0$, we note that $x + \varepsilon u \notin C$. Hence $\operatorname{spts} C \subset P_C(\mathcal{H} \smallsetminus C)$ and $x \in \operatorname{bdry} C$. Thus $\overline{\operatorname{spts}}\, C \subset \operatorname{bdry} C$. Next, assume that $P_C y = x$, for some $y \in \mathcal{H} \smallsetminus C$. Proposition 6.47 asserts that $0 \neq y - x \in N_C x$; hence $x \in \operatorname{spts} C$ by Proposition 7.3. Thus,

$$\operatorname{spts} C = P_C(\mathcal{H} \smallsetminus C). \tag{7.3}$$

Now take $z \in \operatorname{bdry} C$. Then there exists $y \in \mathcal{H} \smallsetminus C$ such that $\|z - y\| \leqslant \varepsilon$. Set $p = P_C y$. Then $p \in \operatorname{spts} C$ and Proposition 4.16 yields $\|p - z\| = \|P_C y - P_C z\| \leqslant \|y - z\| \leqslant \varepsilon$. Therefore, $z \in \overline{\operatorname{spts}}\, C$ and hence $\operatorname{bdry} C \subset \overline{\operatorname{spts}}\, C$. ☐

Proposition 7.5 *Let C be a convex subset of $\mathcal{H}$ such that* $\operatorname{int} C \neq \varnothing$. *Then* $\operatorname{bdry} C \subset \operatorname{spts} \overline{C}$ *and* $C \cap \operatorname{bdry} C \subset \operatorname{spts} C$.

Proof. We assume that $C \neq \mathcal{H}$. Set $D = \overline{C}$ and let $x \in \operatorname{bdry} C \subset D$. Then Proposition 3.45(iii) yields $x \in D \smallsetminus (\operatorname{int} D)$, and Proposition 6.45 guarantees the existence of a vector $u \in N_D x \smallsetminus \{0\}$. Hence, it follows from Proposition 7.3 that $x \in \operatorname{spts} D = \operatorname{spts} \overline{C}$. Therefore, $\operatorname{bdry} C \subset \operatorname{spts} \overline{C}$ and, furthermore, Proposition 7.2 yields $C \cap \operatorname{bdry} C \subset C \cap \operatorname{spts} \overline{C} = \operatorname{spts} C$. ☐

Corollary 7.6 *Let C be a nonempty closed convex subset of $\mathcal{H}$ and suppose that one of the following holds:*

(i) $\operatorname{int} C \neq \varnothing$.
(ii) *C is a closed affine subspace.*
(iii) *$\mathcal{H}$ is finite-dimensional.*

Then $\operatorname{spts} C = \operatorname{bdry} C$.

Proof. (i): Combine Theorem 7.4 and Proposition 7.5.

(ii): Assume that $C \neq \mathcal{H}$, let $V = C - C$ be the closed linear subspace parallel to C, let $x \in \operatorname{bdry} C$, and let $u \in V^{\perp} \smallsetminus \{0\}$. Then $C = x + V$, $P_C(x+u) = P_{x+V}(x+u) = x + P_V u = x$ by Proposition 3.19, and Theorem 7.4 therefore yields $x \in \operatorname{spts} C$.

(iii): In view of (i), assume that $\operatorname{int} C = \varnothing$. Let $D = \operatorname{aff} C$ and let $x \in \operatorname{bdry} C$. Then $D \neq \mathcal{H}$, since Proposition 6.12 asserts that $\operatorname{core} C = \operatorname{int} C = \varnothing$. On the other hand, since $\mathcal{H}$ is finite-dimensional, D is closed. Thus, D is a proper closed affine subspace of $\mathcal{H}$ and therefore $x \in \operatorname{bdry} D$. Altogether, (ii) and Proposition 7.2 imply that $x \in C \cap \operatorname{spts} D \subset \operatorname{spts} C$. ☐

A boundary point of a closed convex set C need not be a support point of C, even if C is a closed convex cone.

Example 7.7 Suppose that $\mathcal{H} = \ell^2(\mathbb{N})$ and set $C = \ell^2_+(\mathbb{N})$, which is a nonempty closed convex cone with empty interior (see Example 6.7). For every $x = (\xi_k)_{k\in\mathbb{N}} \in \mathcal{H}$, we have $P_C x = (\max\{\xi_k, 0\})_{k\in\mathbb{N}}$ by Example 6.29. Consequently, by Theorem 7.4, $x = (\xi_k)_{k\in\mathbb{N}} \in C$ is a support point of C if and only if $\big\{k \in \mathbb{N} \bigm| \xi_k = 0\big\} \neq \varnothing$. Therefore, $(1/2^k)_{k\in\mathbb{N}} \in \operatorname{bdry} C \smallsetminus \operatorname{spts} C$.

7.2 Support Functions

Definition 7.8 Let C be a subset of $\mathcal{H}$. The *support function of* C is

$$\sigma_C \colon \mathcal{H} \to [-\infty, +\infty] : u \mapsto \sup \langle C \mid u \rangle \,. \tag{7.4}$$

Example 7.9 Let Ω be a nonempty subset of $\mathbb{R}$, set $\underline{\omega} = \inf \Omega$, and set $\overline{\omega} = \sup \Omega$. Then

$$(\forall \xi \in \mathbb{R}) \quad \sigma_\Omega(\xi) = \begin{cases} \underline{\omega}\xi, & \text{if } \ \xi < 0; \\ 0, & \text{if } \ \xi = 0; \\ \overline{\omega}\xi, & \text{if } \ \xi > 0. \end{cases} \tag{7.5}$$

Example 7.10 Suppose that $\mathcal{H}$ is finite-dimensional, let $|||\cdot|||$ be a norm on $\mathcal{H}$, and let $C = \big\{x \in \mathcal{H} \bigm| |||x||| \leqslant 1\big\}$ be the associated closed unit ball. Then $\sigma_C = |||\cdot|||_*$ is the *dual norm* of $|||\cdot|||$.

Let C be a nonempty subset of $\mathcal{H}$, and suppose that $u \in \mathcal{H} \smallsetminus \{0\}$. If $\sigma_C(u) < +\infty$, then (7.4) implies that $\big\{x \in \mathcal{H} \bigm| \langle x \mid u \rangle \leqslant \sigma_C(u)\big\}$ is the smallest closed half-space with outer normal u that contains C (see Figure 7.1). Now suppose that, for some $x \in C$, we have $\sigma_C(u) = \langle x \mid u \rangle$. Then x is a support point of C and $\big\{x \in \mathcal{H} \bigm| \langle x \mid u \rangle = \sigma_C(u)\big\}$ is a supporting hyperplane of C at x.

Proposition 7.11 *Let C be a subset of $\mathcal{H}$ and set*

$$(\forall u \in \mathcal{H}) \quad H_u = \big\{x \in \mathcal{H} \bigm| \langle x \mid u \rangle \leqslant \sigma_C(u)\big\}. \tag{7.6}$$

Then $\overline{\operatorname{conv}}\, C = \bigcap_{u\in\mathcal{H}} H_u$.

Proof. We assume that $C \neq \varnothing$ and set $D = \bigcap_{u\in\mathcal{H}} H_u$. Since $(H_u)_{u\in\mathcal{H}}$ is a family of closed convex sets each of which contains C, we deduce that D is closed and convex and that $\overline{\operatorname{conv}}\, C \subset D$. Conversely, take $x \in D$ and set $p = P_{\overline{\operatorname{conv}}\, C} x$. By (3.10), $\sigma_{\overline{\operatorname{conv}}\, C}(x - p) \leqslant \langle p \mid x - p \rangle$. On the other hand, since $x \in D \subset H_{x-p}$, we have $\langle x \mid x - p \rangle \leqslant \sigma_C(x - p)$. Thus, $\|x - p\|^2 = \langle x \mid x - p \rangle - \langle p \mid x - p \rangle \leqslant \sigma_C(x - p) - \sigma_{\overline{\operatorname{conv}}\, C}(x - p) \leqslant 0$, and we conclude that $x = p \in \overline{\operatorname{conv}}\, C$. ☐

Corollary 7.12 *Let C be a closed convex subset of $\mathcal{H}$. Then C is the intersection of all the closed half-spaces of $\mathcal{H}$ which contain C.*

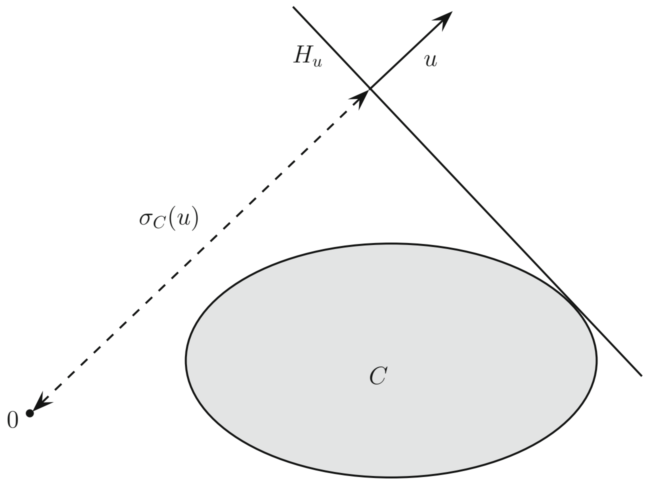

Fig. 7.1 Support function of C evaluated at a vector u such that $\|u\| = 1$. The half-space $H_u = \{x \in \mathcal{H} \mid \langle x \mid u\rangle \leqslant \sigma_C(u)\}$ contains C, and it follows from Example 3.23 that the distance from 0 to the hyperplane $\{x \in \mathcal{H} \mid \langle x \mid u\rangle = \sigma_C(u)\}$ is $\sigma_C(u)$.

Proposition 7.13 *Let C be a subset of $\mathcal{H}$. Then $\sigma_C = \sigma_{\operatorname{conv} C} = \sigma_{\overline{\operatorname{conv}}\, C}$.*

Proof. Assume that $C \neq \varnothing$ and fix $u \in \mathcal{H}$. Since, by Proposition 3.46, $C \subset \operatorname{conv} C \subset \overline{\operatorname{conv} C} = \overline{\operatorname{conv}}\, C$, we have $\sigma_C(u) \leqslant \sigma_{\operatorname{conv} C}(u) \leqslant \sigma_{\overline{\operatorname{conv}}\, C}(u)$. Now, let $x \in \operatorname{conv} C$. By Proposition 3.4, we suppose that $x = \sum_{i\in I} \alpha_i x_i$ for some finite families $(x_i)_{i\in I}$ in C and $(\alpha_i)_{i\in I}$ in $]0,1]$ such that $\sum_{i\in I} \alpha_i = 1$. Then $\langle x \mid u\rangle = \sum_{i\in I} \alpha_i \langle x_i \mid u\rangle \leqslant \sum_{i\in I} \alpha_i \sigma_C(u) = \sigma_C(u)$. Therefore $\sigma_{\operatorname{conv} C}(u) = \sup_{z\in \operatorname{conv} C} \langle z \mid u\rangle \leqslant \sigma_C(u)$. Finally, let $x \in \overline{\operatorname{conv} C} = \overline{\operatorname{conv}}\, C$, say $x_n \to x$ for some sequence $(x_n)_{n\in\mathbb{N}}$ in $\operatorname{conv} C$. Then $\langle x \mid u\rangle = \lim \langle x_n \mid u\rangle \leqslant \sigma_{\operatorname{conv} C}(u)$. Hence, $\sigma_{\overline{\operatorname{conv}}\, C}(u) = \sup_{z\in \overline{\operatorname{conv}}\, C} \langle z \mid u\rangle \leqslant \sigma_{\operatorname{conv} C}(u)$. ☐

7.3 Polar Sets

Definition 7.14 Let C be a subset of $\mathcal{H}$. The *polar set* of C is

$$C^{\odot} = \operatorname{lev}_{\leqslant 1} \sigma_C = \{u \in \mathcal{H} \mid (\forall x \in C) \;\; \langle x \mid u\rangle \leqslant 1\}. \tag{7.7}$$

Example 7.15 Let K be a cone in $\mathcal{H}$. Then $K^{\odot} = K^{\ominus}$ (Exercise 7.10). In particular, if K is a linear subspace, then Proposition 6.23 yields $K^{\odot} = K^{\ominus} = K^{\perp}$.

Proposition 7.16 *Let C and D be subsets of $\mathcal{H}$. Then the following hold:*

(i) $C^{\perp} \subset C^{\ominus} \subset C^{\odot}$.
(ii) $0 \in C^{\odot}$, *and* $C^{\odot}$ *is closed and convex.*
(iii) $C \cup \{0\} \subset C^{\odot\odot}$.
(iv) *Suppose that* $C \subset D$. *Then* $D^{\odot} \subset C^{\odot}$ *and* $C^{\odot\odot} \subset D^{\odot\odot}$.
(v) $C^{\odot\odot\odot} = C^{\odot}$.
(vi) $(\overline{\operatorname{conv}}\, C)^{\odot} = C^{\odot}$.

Proof. (i)–(iv): Immediate consequences of (2.2), (6.28), and (7.7).

(v): By (iii), $C \subset C^{\odot\odot}$ and $C^{\odot} \subset (C^{\odot})^{\odot\odot} = C^{\odot\odot\odot}$. Hence, using (iv), we obtain $C^{\odot\odot\odot} = (C^{\odot\odot})^{\odot} \subset C^{\odot} \subset C^{\odot\odot\odot}$.

(vi): By Proposition 7.13, $\sigma_C = \sigma_{\overline{\operatorname{conv}}\, C}$. ☐

Remark 7.17 In view of Example 7.15 and Proposition 7.16, we obtain the following properties for an arbitrary cone K in $\mathcal{H}$: $K \subset K^{\ominus\ominus}$, $K^{\ominus\ominus\ominus} = K^{\ominus}$, and $(\overline{\operatorname{conv}}\, K)^{\ominus} = K^{\ominus}$. In particular, we retrieve the following well-known facts for a linear subspace V of $\mathcal{H}$: $V \subset V^{\perp\perp}$, $V^{\perp\perp\perp} = V^{\perp}$, and $\overline{V}^{\perp} = V^{\perp}$.

Theorem 7.18 *Let C be a subset of $\mathcal{H}$. Then* $C^{\odot\odot} = \overline{\operatorname{conv}}\,(C \cup \{0\})$.

Proof. By Proposition 7.16(ii)&(iii), $\overline{\operatorname{conv}}\,(C \cup \{0\}) \subset \overline{\operatorname{conv}}\, C^{\odot\odot} = C^{\odot\odot}$. Conversely, suppose that $x \in \operatorname{lev}_{\leqslant 1} \sigma_{C^{\odot}} \smallsetminus \overline{\operatorname{conv}}\,(C \cup \{0\})$. It follows from Theorem 3.50 that there exists $u \in \mathcal{H} \smallsetminus \{0\}$ such that

$$\langle x \mid u \rangle > \sigma_{\overline{\operatorname{conv}}\,(C \cup \{0\})}(u) \geqslant \max\{\sigma_C(u), 0\}. \tag{7.8}$$

After scaling u if necessary, we assume that $\langle x \mid u \rangle > 1 \geqslant \sigma_C(u)$. Hence, $u \in C^{\odot}$ and therefore $1 < \langle u \mid x \rangle \leqslant \sigma_{C^{\odot}}(x) \leqslant 1$, which is impossible. ☐

Corollary 7.19 *Let C be a subset of $\mathcal{H}$. Then the following hold:*

(i) *C is closed, convex, and contains the origin if and only if* $C^{\odot\odot} = C$.
(ii) *C is a nonempty closed convex cone if and only if* $C^{\ominus\ominus} = C$.
(iii) *C is a closed linear subspace if and only if* $C^{\perp\perp} = C$.

Proof. (i): Combine Proposition 7.16(ii) and Theorem 7.18.

(ii): Combine (i), Proposition 6.24(ii), and Example 7.15.

(iii): Combine (ii) and Proposition 6.23. ☐

Exercises

Exercise 7.1 Let C be a subset of $\mathcal{H}$, let $\gamma \in \mathbb{R}_{++}$, and let $\rho \in \mathbb{R} \smallsetminus \{0\}$.

(i) Show that $\sigma_C \circ \rho\mathrm{Id} = \sigma_{\rho C}$.
(ii) Show that $\sigma_C \circ \gamma\mathrm{Id} = \gamma\sigma_C$.
(iii) Provide an example in which $\sigma_C \circ \rho\mathrm{Id} \neq \rho\sigma_C$.

Exercise 7.2 Let C and D be subsets of $\mathcal{H}$. Show that $\sigma_{C+D} = \sigma_C + \sigma_D$.

Exercise 7.3 Suppose that $\mathcal{H} = \mathbb{R}$ and let $-\infty < \alpha < \beta < +\infty$. Prove Example 7.9 and determine the support function σ_C in each of the following cases:

(i) $C = \{\alpha\}$.
(ii) $C = [\alpha, \beta]$.
(iii) $C = \left]-\infty, \beta\right]$.
(iv) $C = \left[\alpha, +\infty\right[$.
(v) $C = \mathbb{R}$.

Exercise 7.4 Let C be a subset of $\mathcal{H}$ and let $\gamma \in \mathbb{R}_{++}$. Show that $(\gamma C)^{\odot} = \gamma^{-1} C^{\odot}$.

Exercise 7.5 Let C and D be subsets of $\mathcal{H}$. Show that $(C \cup D)^{\odot} = C^{\odot} \cap D^{\odot}$.

Exercise 7.6 Provide two subsets C and D of $\mathbb{R}$ such that $(C \cap D)^{\odot} \neq C^{\odot} \cup D^{\odot}$.

Exercise 7.7 Set $C = B(0; 1)$. Show that $C^{\odot} = C$.

Exercise 7.8 Let C be a subset of $\mathcal{H}$ and suppose that $C^{\odot} = C$. Show that $C = B(0; 1)$.

Exercise 7.9 Suppose that $\mathcal{H} = \mathbb{R}^N$, let $p \in [1, +\infty]$, and set $C_p = \big\{x \in \mathcal{H} \;\big|\; \|x\|_p \leqslant 1\big\}$, where $\|\cdot\|_p$ is defined in (9.36). Compute $C_p^{\odot}$.

Exercise 7.10 Let C be a subset of $\mathcal{H}$. Show that, if C is a cone, then $C^{\ominus} = C^{\odot}$. Show that this identity fails if C is not a cone.

Chapter 8
Convex Functions

Convex functions, which lie at the heart of modern optimization, are introduced in this chapter. We study operations that preserve convexity and the interplay between various continuity properties.

8.1 Basic Properties and Examples

Definition 8.1 Let $f\colon \mathcal{H} \to [-\infty,+\infty]$. Then f is *convex* if its epigraph $\operatorname{epi} f = \big\{(x,\xi) \in \mathcal{H} \times \mathbb{R} \;\big|\; f(x) \leqslant \xi\big\}$ is a convex subset of $\mathcal{H} \times \mathbb{R}$. Moreover, f is *concave* if $-f$ is convex.

Proposition 8.2 *Let $f\colon \mathcal{H} \to [-\infty,+\infty]$ be convex. Then its domain $\operatorname{dom} f = \big\{x \in \mathcal{H} \;\big|\; f(x) < +\infty\big\}$ is convex.*

Proof. Set $L\colon \mathcal{H}\times\mathbb{R} \to \mathcal{H}\colon (x,\xi) \mapsto x$. Then L is linear and $\operatorname{dom} f = L(\operatorname{epi} f)$. It therefore follows from Proposition 3.5 that $\operatorname{dom} f$ is convex. $\square$

Example 8.3 Let C be a subset of $\mathcal{H}$. Then $\operatorname{epi}\iota_C = C \times \mathbb{R}_+$, and hence ι_C is a convex function if and only if C is a convex set.

Proposition 8.4 *Let $f\colon \mathcal{H} \to [-\infty,+\infty]$. Then f is convex if and only if*

$$(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)(\forall \alpha \in \,]0,1[\,)$$
$$f\big(\alpha x + (1-\alpha)y\big) \leqslant \alpha f(x) + (1-\alpha) f(y). \quad (8.1)$$

Proof. Note that $f \equiv +\infty \Leftrightarrow \operatorname{epi} f = \varnothing \Leftrightarrow \operatorname{dom} f = \varnothing$, in which case f is convex and (8.1) holds. So we assume that $\operatorname{dom} f \neq \varnothing$, and take $(x,\xi) \in$

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-48311-5_31

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_8

$\operatorname{epi} f$, $(y,\eta) \in \operatorname{epi} f$, and $\alpha \in]0,1[$. First, suppose that f is convex. Then $\alpha(x,\xi) + (1-\alpha)(y,\eta) \in \operatorname{epi} f$ and therefore

$$f\big(\alpha x + (1-\alpha)y\big) \leqslant \alpha\xi + (1-\alpha)\eta. \tag{8.2}$$

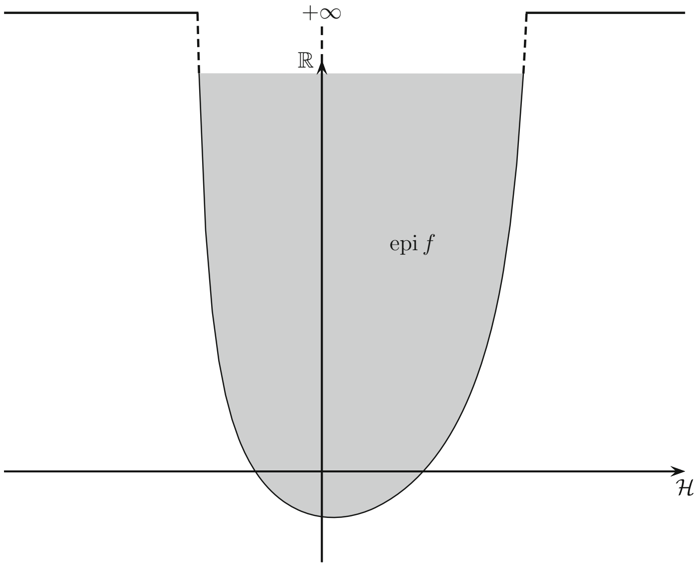

Fig. 8.1 A function $f\colon \mathcal{H} \to [-\infty,+\infty]$ is convex if $\operatorname{epi} f$ is convex in $\mathcal{H} \times \mathbb{R}$.

Letting $\xi \downarrow f(x)$ and $\eta \downarrow f(y)$ in (8.2), we obtain (8.1). Now assume that f satisfies (8.1). Then $f\big(\alpha x + (1-\alpha)y\big) \leqslant \alpha f(x) + (1-\alpha)f(y) \leqslant \alpha\xi + (1-\alpha)\eta$, and therefore $\alpha(x,\xi) + (1-\alpha)(y,\eta) \in \operatorname{epi} f$. □

Corollary 8.5 *Let $f\colon \mathcal{H} \to [-\infty,+\infty]$ be convex. Then, for every $\xi \in \mathbb{R}$, $\operatorname{lev}_{\leqslant\xi} f$ is convex.*

Proposition 8.6 *Let $(\mathcal{H}_i)_{i\in I}$ be a family of real Hilbert spaces, set $\mathcal{H} = \bigoplus_{i\in I}\mathcal{H}_i$, and let, for every i in I, $f_i\colon \mathcal{H}_i \to]-\infty,+\infty]$ be convex. Suppose that I is finite or that $\inf_{i\in I} f_i \geqslant 0$. Then $\bigoplus_{i\in I} f_i$ is convex.*

Proof. This is a consequence of Proposition 8.4. □

Definition 8.7 Let $f\colon \mathcal{H} \to]-\infty,+\infty]$ be a proper function. Then f is *strictly convex* if

$$(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)(\forall \alpha \in]0,1[)$$
$$x \neq y \quad\Rightarrow\quad f\big(\alpha x + (1-\alpha)y\big) < \alpha f(x) + (1-\alpha)f(y). \tag{8.3}$$

Now let C be a nonempty subset of $\operatorname{dom} f$. Then f is *convex on* C if

$$(\forall x \in C)(\forall y \in C)(\forall \alpha \in]0,1[\,) \qquad f\big(\alpha x + (1-\alpha)y\big) \leqslant \alpha f(x) + (1-\alpha) f(y), \tag{8.4}$$

and f is *strictly convex on* C if

$$(\forall x \in C)(\forall y \in C)(\forall \alpha \in]0,1[\,) \qquad x \neq y \;\;\Rightarrow\;\; f\big(\alpha x + (1-\alpha)y\big) < \alpha f(x) + (1-\alpha) f(y). \tag{8.5}$$

Remark 8.8 Let $f\colon \mathcal{H} \to]-\infty,+\infty]$ be proper and let C be a subset of $\operatorname{dom} f$. Then $f + \iota_C$ is convex if and only if C is convex and f is convex on C (see Exercise 8.2).

Example 8.9 The function $\|\cdot\|$ is convex. If $\mathcal{H} \neq \{0\}$, then $\|\cdot\|$ is not strictly convex.

Proof. Convexity is clear. Now take $x \in \mathcal{H} \smallsetminus \{0\}$ and $\alpha \in]0,1[$. Then $\|\alpha x + (1-\alpha)0\| = \alpha\|x\| + (1-\alpha)\|0\|$. Hence $\|\cdot\|$ is not strictly convex. ☐

Example 8.10 The function $\|\cdot\|^2$ is strictly convex.

Proof. This follows from Corollary 2.15. ☐

Proposition 8.11 *Let $f\colon \mathcal{H} \to [-\infty,+\infty]$. Then f is convex if and only if, for all finite families $(\alpha_i)_{i\in I}$ in $]0,1[$ such that $\sum_{i\in I}\alpha_i = 1$ and $(x_i)_{i\in I}$ in* $\operatorname{dom} f$, *we have*

$$f\Bigg(\sum_{i\in I}\alpha_i x_i\Bigg) \leqslant \sum_{i\in I}\alpha_i f(x_i). \tag{8.6}$$

Proof. Assume that f is convex and fix finite families $(x_i)_{i\in I}$ in $\operatorname{dom} f$ and $(\alpha_i)_{i\in I}$ in $]0,1[$ such that $\sum_{i\in I}\alpha_i = 1$. Take $(\forall i \in I)$ $(x_i,\xi_i) \in \operatorname{epi} f$. Then $\sum_{i\in I}\alpha_i(x_i,\xi_i) \in \operatorname{epi} f$ and therefore $f(\sum_{i\in I}\alpha_i x_i) \leqslant \sum_{i\in I}\alpha_i\xi_i$. To obtain (8.6), let $(\forall i \in I)$ $\xi_i \downarrow f(x_i)$. The converse implication follows from Proposition 8.4. ☐

Corollary 8.12 *Let $f\colon \mathcal{H} \to]-\infty,+\infty]$. Then the following are equivalent:*

(i) *f is convex.*
(ii) *For all finite families $(\alpha_i)_{i\in I}$ in $]0,1[$ such that $\sum_{i\in I}\alpha_i = 1$ and $(x_i)_{i\in I}$ in* $\operatorname{dom} f$, *we have $f(\sum_{i\in I}\alpha_i x_i) \leqslant \sum_{i\in I}\alpha_i f(x_i)$.*
(iii) $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})(\forall \alpha \in]0,1[\,)$ $f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha) f(y)$.

Proposition 8.13 *Let $f\colon \mathcal{H} \to]-\infty,+\infty]$ be proper. Then f is strictly convex if and only if for all finite families $(\alpha_i)_{i\in I}$ in $]0,1[$ such that $\sum_{i\in I}\alpha_i = 1$ and $(x_i)_{i\in I}$ in* $\operatorname{dom} f$, *we have $f(\sum_{i\in I}\alpha_i x_i) \leqslant \sum_{i\in I}\alpha_i f(x_i)$, and equality holds if and only if $\{x_i\}_{i\in I}$ is a singleton.*

Proof. Assume first that f is strictly convex. We prove the corresponding implication by induction on m, the number of elements in I. The result is clear for $m = 2$. We assume now that $m \geqslant 3$, that $I = \{1, \dots, m\}$, and that the result is true for families containing $m-1$ or fewer points, and we set $\mu = f(\sum_{i \in I} \alpha_i x_i) = \sum_{i \in I} \alpha_i f(x_i)$. Then

$$\mu \leqslant (1-\alpha_m) f\left(\sum_{i=1}^{m-1} \frac{\alpha_i}{1-\alpha_m} x_i\right) + \alpha_m f(x_m) \tag{8.7}$$

$$\leqslant (1-\alpha_m) \sum_{i=1}^{m-1} \frac{\alpha_i}{1-\alpha_m} f(x_i) + \alpha_m f(x_m) \tag{8.8}$$

$$= \mu. \tag{8.9}$$

Hence, the inequalities (8.7) and (8.8) are actually equalities, and the induction hypothesis yields $(1-\alpha_m)^{-1}(\sum_{i=1}^{m-1} \alpha_i x_i) = x_m$ and $x_1 = \cdots = x_{m-1}$. Therefore, $x_1 = \cdots = x_m$, as required. The reverse implication is clear by considering the case in which I contains exactly two elements. ☐

We now state a simple convexity condition for functions on the real line (for an extension, see Proposition 17.7).

Proposition 8.14 *Let $\phi\colon \mathbb{R} \to \left]-\infty, +\infty\right]$ be a proper function that is differentiable on a nonempty open interval I in* $\operatorname{dom} \phi$*. Then the following hold:*

(i) *Suppose that ϕ' is increasing on I. Then ϕ is convex on I.*
(ii) *Suppose that ϕ' is strictly increasing on I. Then ϕ is strictly convex on I.*

Proof. Fix x and y in I, and $\alpha \in \left]0,1\right[$. Set $\psi\colon \mathbb{R} \to \left]-\infty, +\infty\right] : z \mapsto \alpha\phi(x) + (1-\alpha)\phi(z) - \phi(\alpha x + (1-\alpha)z)$. Now let $z \in I$. Then

$$\psi'(z) = (1-\alpha)\big(\phi'(z) - \phi'(\alpha x + (1-\alpha)z)\big) \tag{8.10}$$

and $\psi'(x) = 0$.

(i): It follows from (8.10) that $\psi'(z) \leqslant 0$ if $z < x$, and that $\psi'(z) \geqslant 0$ if $z > x$. Thus, ψ achieves its infimum on I at x. In particular, $\psi(y) \geqslant \psi(x) = 0$, and the convexity of ϕ on I follows from Proposition 8.4.

(ii): It follows from (8.10) that $\psi'(z) < 0$ if $z < x$, and $\psi'(z) > 0$ if $z > x$. Thus, x is the unique minimizer of ψ on I. Hence, if $y \neq x$, then $\psi(y) > \psi(x) = 0$, and $\phi + \iota_I$ is strictly convex by (8.3). ☐

Example 8.15 Let $\psi\colon \mathbb{R} \to \mathbb{R}$ be increasing, let $\alpha \in \mathbb{R}$, and set

$$\phi\colon \mathbb{R} \to \left]-\infty, +\infty\right] : x \mapsto \begin{cases} \displaystyle\int_\alpha^x \psi(t)\, dt, & \text{if } x \geqslant \alpha; \\ +\infty, & \text{otherwise.} \end{cases} \tag{8.11}$$

Then ϕ is convex. If ψ is strictly increasing, then ϕ is strictly convex.

Proof. By Proposition 8.14, $\varphi\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto \int_\alpha^x \psi(t)\,dt$ is convex since $\varphi' = \psi$. Hence, the convexity of $\phi = \varphi + \iota_{[\alpha,+\infty[}$ follows from Proposition 8.4. The variant in which ψ is strictly convex is proved similarly. ☐

8.2 Convexity-Preserving Operations

In this section we describe some basic operations that preserve convexity. Further results will be provided in Proposition 12.11 and Proposition 12.36(ii) (see also Exercise 8.14).

Proposition 8.16 *Let $(f_i)_{i\in I}$ be a family of convex functions from $\mathcal{H}$ to $[-\infty,+\infty]$. Then $\sup_{i\in I} f_i$ is convex.*

Proof. By Lemma 1.6(i), $\operatorname{epi}\big(\sup_{i\in I} f_i\big) = \bigcap_{i\in I} \operatorname{epi} f_i$, which is convex as an intersection of convex sets by Example 3.2(iv). ☐

Proposition 8.17 *The set of convex functions from $\mathcal{H}$ to $]-\infty,+\infty]$ is closed under addition and multiplication by strictly positive real numbers.*

Proof. This is an easy consequence of Proposition 8.4. ☐

Proposition 8.18 *Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of convex functions from $\mathcal{H}$ to $]-\infty,+\infty]$ such that $(f_n)_{n\in\mathbb{N}}$ is pointwise convergent. Then $\lim f_n$ is convex.*

Proof. Set $f = \lim f_n$, and take x and y in $\operatorname{dom} f$, and $\alpha \in]0,1[$. By Corollary 8.12, $(\forall n \in \mathbb{N})$ $f_n(\alpha x + (1-\alpha)y) \leqslant \alpha f_n(x) + (1-\alpha) f_n(y)$. Passing to the limit, we obtain $f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha) f(y)$, and using the same result, we conclude that f is convex. ☐

Example 8.19 Let $(z_n)_{n\in\mathbb{N}}$ be a bounded sequence in $\mathcal{H}$. Then $f\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \overline{\lim}\,\|x - z_n\|$ is convex and Lipschitz continuous with constant 1.

Proof. Set $(\forall n \in \mathbb{N})$ $f_n = \sup\big\{\|\cdot - z_k\| \;\big|\; k \in \mathbb{N},\ k \geqslant n\big\}$. Then it follows from Example 8.9 and Proposition 8.16 that each f_n is convex. In addition, $(f_n)_{n\in\mathbb{N}}$ decreases pointwise to $f = \inf_{n\in\mathbb{N}} f_n$. Hence f is convex by Proposition 8.18. Now take x and y in $\mathcal{H}$. Then $(\forall n \in \mathbb{N})$ $\|x - z_n\| \leqslant \|x - y\| + \|y - z_n\|$. Thus

$$\overline{\lim}\,\|x - z_n\| \leqslant \|x - y\| + \overline{\lim}\,\|y - z_n\| \tag{8.12}$$

and hence $f(x) - f(y) \leqslant \|x - y\|$. Analogously, $f(y) - f(x) \leqslant \|y - x\|$. Altogether, $|f(x) - f(y)| \leqslant \|x - y\|$. ☐

Proposition 8.20 *Let $\mathcal{K}$ be a real Hilbert space, let $L\colon \mathcal{H} \to \mathcal{K}$ be affine, and let $g\colon \mathcal{K} \to]-\infty,+\infty]$ be convex. Then $g \circ L$ is convex.*

Proof. This is an immediate consequence of Proposition 8.4. ☐

Proposition 8.21 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ and $\phi\colon \mathbb{R} \to \left]-\infty,+\infty\right]$ be convex functions. Set $C = \operatorname{conv}(\mathbb{R} \cap \operatorname{ran} f)$ and extend ϕ to $\widetilde{\phi}\colon \left]-\infty,+\infty\right] \to \left]-\infty,+\infty\right]$ by setting $\widetilde{\phi}(+\infty) = +\infty$. Suppose that $C \subset \operatorname{dom}\phi$ and that ϕ is increasing on C. Then the composition $\widetilde{\phi} \circ f$ is convex.*

Proof. Note that $\operatorname{dom}(\widetilde{\phi} \circ f) = f^{-1}(\operatorname{dom}\phi) \subset \operatorname{dom} f$ and that $\widetilde{\phi} \circ f$ coincides with $\phi \circ f$ on $\operatorname{dom}(\widetilde{\phi} \circ f)$. Take x and y in $\operatorname{dom}(\widetilde{\phi} \circ f)$, and $\alpha \in \left]0,1\right[$. Using the convexity of f, we obtain $f(\alpha x+(1-\alpha)y) \leqslant \alpha f(x)+(1-\alpha)f(y) < +\infty$. Since ϕ is increasing and convex on C, we deduce that $(\phi\circ f)(\alpha x+(1-\alpha)y) \leqslant \phi(\alpha f(x) + (1-\alpha)f(y)) \leqslant \alpha(\phi \circ f)(x) + (1-\alpha)(\phi \circ f)(y)$. Hence, the result follows from Proposition 8.4. ☐

Example 8.22 Let $\psi\colon \mathbb{R} \to \mathbb{R}$ be a continuous increasing function such that $\psi(0) \geqslant 0$, and set

$$f\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \int_0^{\|x\|} \psi(t)\, dt. \tag{8.13}$$

Then the following hold:

(i) f is convex.
(ii) Suppose that ψ is strictly increasing. Then f is strictly convex.

Proof. Let $\phi\colon \mathbb{R} \to \left]-\infty,+\infty\right]$ be the function obtained by setting $\alpha = 0$ in (8.11). Then, since $\psi \geqslant 0$ on $\mathbb{R}_+$, ϕ is increasing on $\operatorname{dom}\phi = \mathbb{R}_+$ and $f = \phi \circ \|\cdot\|$.

(i): It follows from Example 8.15 that ϕ is convex, and in turn, from Example 8.9 and Proposition 8.21, that f is convex.

(ii): Suppose that x and y are two distinct points in $\mathcal{H}$ and let $\alpha \in \left]0,1\right[$. First, assume that $\|x\| \neq \|y\|$. Since ϕ is increasing and, by Example 8.15, strictly convex, we have $f(\alpha x + (1-\alpha)y) = \phi(\|\alpha x + (1-\alpha)y\|) \leqslant \phi(\alpha\|x\| + (1-\alpha)\|y\|) < \alpha\phi(\|x\|) + (1-\alpha)\phi(\|y\|) = \alpha f(x) + (1-\alpha)f(y)$. Now assume that $\|x\| = \|y\|$. Then Corollary 2.15 asserts that $\|\alpha x+(1-\alpha)y\|^2 = \alpha\|x\|^2 + (1-\alpha)\|y\|^2 - \alpha(1-\alpha)\|x-y\|^2 = \|x\|^2 - \alpha(1-\alpha)\|x-y\|^2 < \|x\|^2$. Hence, since ϕ is strictly increasing,

$$\begin{aligned} f(\alpha x + (1-\alpha)y) &= \phi(\|\alpha x + (1-\alpha)y\|) \\ &< \phi(\|x\|) \\ &= \alpha\phi(\|x\|) + (1-\alpha)\phi(\|y\|) \\ &= \alpha f(x) + (1-\alpha)f(y), \end{aligned} \tag{8.14}$$

which completes the proof. ☐

Example 8.23 Let $p \in \left]1,+\infty\right[$. Then $\|\cdot\|^p$ is strictly convex and

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|x+y\|^p \leqslant 2^{p-1}\big(\|x\|^p + \|y\|^p\big). \tag{8.15}$$

Proof. Set $\psi(t) = t$ if $t < 0$, and $\psi(t) = pt^{p-1}$ if $t \geqslant 0$. Then ψ is continuous, strictly increasing, and $\psi(0) = 0$. Hence, we deduce from Example 8.22(ii) that $\|\cdot\|^p$ is strictly convex. As a result, for every x and y in $\mathcal{H}$, $\|(x+y)/2\|^p \leqslant (\|x\|^p + \|y\|^p)/2$ and (8.15) follows. □

The next proposition describes a convex integral function in the setting of Example 2.6.

Proposition 8.24 *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $(\mathsf{H}, \langle\cdot \mid \cdot\rangle_{\mathsf{H}})$ be a separable real Hilbert space. Suppose that $\mathcal{H} = L^2((\Omega, \mathcal{F}, \mu); \mathsf{H})$, and let $\varphi\colon \mathsf{H} \to \left]-\infty, +\infty\right]$ be a measurable convex function such that*

$$(\forall x \in \mathcal{H})(\exists\, \varrho \in L^1((\Omega, \mathcal{F}, \mu); \mathbb{R})) \quad \varphi \circ x \geqslant \varrho \quad \mu\text{-a.e.} \tag{8.16}$$

Set

$$f\colon \mathcal{H} \to \left]-\infty, +\infty\right] : x \mapsto \int_\Omega \varphi\big(x(\omega)\big)\mu(d\omega). \tag{8.17}$$

Then $\operatorname{dom} f = \big\{x \in \mathcal{H} \bigm| \varphi \circ x \in L^1((\Omega, \mathcal{F}, \mu); \mathbb{R})\big\}$ *and f is convex.*

Proof. For every $x \in \mathcal{H}$, $\varphi \circ x$ is measurable and, in view of (8.16), $\int_\Omega (\varphi \circ x) d\mu$ is well defined and it never takes on the value $-\infty$. Hence, f is well defined and we obtain the above expression for its domain. Now take x and y in $\mathcal{H}$ and $\alpha \in \left]0, 1\right[$. It follows from Corollary 8.12 that, for μ-almost every $\omega \in \Omega$, we have $\varphi(\alpha x(\omega) + (1-\alpha)y(\omega)) \leqslant \alpha\varphi(x(\omega)) + (1-\alpha)\varphi(y(\omega))$. Upon integrating these inequalities over Ω with respect to μ, we obtain $f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha) f(y)$, hence the convexity of f. □

Proposition 8.25 *Let $\varphi\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ be convex, set $C = \{1\} \times \operatorname{epi} \varphi$, and let*

$$f\colon \mathbb{R} \times \mathcal{H} \to \left]-\infty, +\infty\right] : (\xi, x) \mapsto \begin{cases} \xi\varphi(x/\xi), & \text{if } \xi > 0;\\ +\infty, & \text{otherwise,} \end{cases} \tag{8.18}$$

be the perspective function *of φ. Then* $\operatorname{epi} f = \operatorname{cone} C$ *and f is convex.*

Proof. It follows from the convexity of φ that C is a convex subset of $\mathbb{R} \times \mathcal{H} \times \mathbb{R}$. Furthermore,

$$\begin{aligned} \operatorname{epi} f &= \big\{(\xi, x, \zeta) \in \mathbb{R}_{++} \times \mathcal{H} \times \mathbb{R} \bigm| \xi\varphi(x/\xi) \leqslant \zeta\big\}\\ &= \big\{(\xi, x, \zeta) \in \mathbb{R}_{++} \times \mathcal{H} \times \mathbb{R} \bigm| (x/\xi, \zeta/\xi) \in \operatorname{epi} \varphi\big\}\\ &= \big\{\xi(1, y, \eta) \bigm| \xi \in \mathbb{R}_{++} \text{ and } (y, \eta) \in \operatorname{epi} \varphi\big\}\\ &= \mathbb{R}_{++} C\\ &= \operatorname{cone} C. \end{aligned} \tag{8.19}$$

In view of Proposition 6.2(iii), $\operatorname{epi} f$ is therefore a convex set. □

When restricted to the probability simplex, the following divergences are positive.

Example 8.26 Suppose that $\mathcal{H} = \mathbb{R}^N$, set $I = \{1, \ldots, N\}$, and let $\phi\colon \mathbb{R} \to]-\infty, +\infty]$ be convex. Define the *Csiszár ϕ-divergence* between $x = (\xi_i)_{i\in I} \in \mathcal{H}$ and $y = (\eta_i)_{i\in I} \in \mathcal{H}$ as

$$d_\phi(x, y) = \begin{cases} \sum_{i\in I} \eta_i \phi(\xi_i/\eta_i), & \text{if } y \in \mathbb{R}^N_{++}; \\ +\infty, & \text{otherwise.} \end{cases} \tag{8.20}$$

Then $d_\phi\colon \mathcal{H} \times \mathcal{H} \to]-\infty, +\infty]$ is convex.

Proof. Combine Proposition 8.25 and Proposition 8.6. □

Example 8.27 Suppose that $\mathcal{H} = \mathbb{R}^N$ and set $I = \{1, \ldots, N\}$. For every $x = (\xi_i)_{i\in I} \in \mathcal{H}$ and every $y = (\eta_i)_{i\in I} \in \mathcal{H}$, let

$$f(x, y) = \begin{cases} \sum_{i\in I} \xi_i \ln(\xi_i/\eta_i), & \text{if } x \in \mathbb{R}^N_{++} \text{ and } y \in \mathbb{R}^N_{++}; \\ +\infty, & \text{otherwise} \end{cases} \tag{8.21}$$

be the *Kullback-Leibler divergence* between x and y. Then $f\colon \mathcal{H} \times \mathcal{H} \to]-\infty, +\infty]$ is convex.

Proof. Apply Example 8.26 to the function

$$\phi\colon \mathbb{R} \to]-\infty, +\infty] : t \mapsto \begin{cases} t \ln t, & \text{if } t > 0; \\ +\infty, & \text{otherwise,} \end{cases} \tag{8.22}$$

which is (strictly) convex by Proposition 8.14(ii). □

Example 8.28 Suppose that $\mathcal{H} = \mathbb{R}^N$ and set $I = \{1, \ldots, N\}$. For every $x = (\xi_i)_{i\in I} \in \mathcal{H}$ and every $y = (\eta_i)_{i\in I} \in \mathcal{H}$, let

$$f(x, y) = \begin{cases} \sum_{i\in I} (\xi_i - \eta_i)(\ln \xi_i - \ln \eta_i), & \text{if } x \in \mathbb{R}^N_{++} \text{ and } y \in \mathbb{R}^N_{++}; \\ +\infty, & \text{otherwise} \end{cases} \tag{8.23}$$

be the *Jeffreys divergence* between x and y. Then $f\colon \mathcal{H} \times \mathcal{H} \to]-\infty, +\infty]$ is convex.

Proof. Apply Example 8.26 to the function

$$\phi\colon \mathbb{R} \to]-\infty, +\infty] : t \mapsto \begin{cases} (t-1) \ln t, & \text{if } t > 0; \\ +\infty, & \text{otherwise,} \end{cases} \tag{8.24}$$

which is (strictly) convex by Proposition 8.14(ii). □

Example 8.29 Suppose that $\mathcal{H} = \mathbb{R}^N$ and set $I = \{1, \ldots, N\}$. For every $x = (\xi_i)_{i\in I} \in \mathcal{H}$ and every $y = (\eta_i)_{i\in I} \in \mathcal{H}$, let

$$f(x,y) = \begin{cases} \displaystyle\sum_{i\in I} \frac{|\xi_i - \eta_i|^2}{\eta_i}, & \text{if } y \in \mathbb{R}^N_{++}; \\ +\infty, & \text{otherwise} \end{cases} \tag{8.25}$$

be the *Pearson divergence* between x and y. Then $f\colon \mathcal{H}\times\mathcal{H} \to \left]-\infty,+\infty\right]$ is convex.

Proof. Apply Example 8.26 to the function

$$\phi\colon \mathbb{R} \to \left]-\infty,+\infty\right] : t \mapsto |t-1|^2, \tag{8.26}$$

which is (strictly) convex by Proposition 8.14(ii). ☐

Example 8.30 Suppose that $\mathcal{H} = \mathbb{R}^N$ and set $I = \{1,\ldots,N\}$. For every $x = (\xi_i)_{i\in I} \in \mathcal{H}$ and every $y = (\eta_i)_{i\in I} \in \mathcal{H}$, let

$$f(x,y) = \begin{cases} \displaystyle\sum_{i\in I} \left|\sqrt{\xi_i} - \sqrt{\eta_i}\right|^2, & \text{if } x \in \mathbb{R}^N_{++} \text{ and } y \in \mathbb{R}^N_{++}; \\ +\infty, & \text{otherwise} \end{cases} \tag{8.27}$$

be the *Hellinger divergence* between x and y. Then $f\colon \mathcal{H}\times\mathcal{H} \to \left]-\infty,+\infty\right]$ is convex.

Proof. Apply Example 8.26 to the *Hellinger entropy* function

$$\phi\colon \mathbb{R} \to \left]-\infty,+\infty\right] : t \mapsto \begin{cases} \left|\sqrt{t}-1\right|^2, & \text{if } t > 0; \\ +\infty, & \text{otherwise,} \end{cases} \tag{8.28}$$

which is (strictly) convex by Proposition 8.14(ii). ☐

Example 8.31 Suppose that $\mathcal{H} = \mathbb{R}^N$, set $I = \{1,\ldots,N\}$, and let $\alpha \in \left]1,+\infty\right[$. For every $x = (\xi_i)_{i\in I} \in \mathcal{H}$ and every $y = (\eta_i)_{i\in I} \in \mathcal{H}$, let

$$f(x,y) = \begin{cases} \displaystyle\sum_{i\in I} \xi_i^\alpha \eta_i^{1-\alpha}, & \text{if } x \in \mathbb{R}^N_{++} \text{ and } y \in \mathbb{R}^N_{++}; \\ +\infty, & \text{otherwise} \end{cases} \tag{8.29}$$

be the *Rényi divergence* between x and y. Then $f\colon \mathcal{H}\times\mathcal{H} \to \left]-\infty,+\infty\right]$ is convex.

Proof. Apply Example 8.26 to the function

$$\phi\colon \mathbb{R} \to \left]-\infty,+\infty\right] : t \mapsto \begin{cases} t^\alpha, & \text{if } t > 0; \\ +\infty, & \text{otherwise,} \end{cases} \tag{8.30}$$

which is (strictly) convex by Proposition 8.14(ii) or Example 8.23. ☐

Example 8.32 Let C be a convex subset of $\mathcal{H}$ and set

$$f\colon \mathbb{R}\times\mathcal{H}\to\,]-\infty,+\infty] : (\xi,x)\mapsto \begin{cases}\xi, & \text{if } \xi>0 \text{ and } x\in\xi C;\\ +\infty, & \text{otherwise.}\end{cases} \tag{8.31}$$

Then f is convex.

Proof. Apply Proposition 8.25 to $1+\iota_C$. □

Example 8.33 Let $\phi\colon\mathbb{R}\to\,]-\infty,+\infty]$ be convex and let

$$\phi^{\diamond}\colon \mathbb{R}\to\,]-\infty,+\infty] : \xi\mapsto \begin{cases}\xi\phi(1/\xi), & \text{if } \xi>0;\\ +\infty, & \text{otherwise}\end{cases} \tag{8.32}$$

be the *adjoint* of ϕ. Then $\phi^{\diamond}$ is convex.

Proof. Let f be the perspective function of ϕ, as defined in (8.18). Since f is convex by Proposition 8.25, so is $\phi^{\diamond}=f(\cdot,1)$. □

Example 8.34 The function

$$\phi\colon \mathbb{R}\to\,]-\infty,+\infty] : \xi\mapsto \begin{cases}1/\xi, & \text{if } \xi>0;\\ +\infty, & \text{otherwise}\end{cases} \tag{8.33}$$

is convex.

Proof. This follows from Example 8.33 and Example 8.10 since $\phi=(|\cdot|^2)^{\diamond}$. □

Proposition 8.35 *Let $\mathcal{K}$ be a real Hilbert space and let $F\colon\mathcal{H}\times\mathcal{K}\to\,]-\infty,+\infty]$ be convex. Then the* marginal function

$$f\colon\mathcal{H}\to[-\infty,+\infty] : x\mapsto \inf F(x,\mathcal{K}) \tag{8.34}$$

is convex.

Proof. Take x_1 and x_2 in $\operatorname{dom} f$, and $\alpha\in\,]0,1[$. Furthermore, let $\xi_1\in\,]f(x_1),+\infty[$ and $\xi_2\in\,]f(x_2),+\infty[$. Then there exist y_1 and y_2 in $\mathcal{K}$ such that $F(x_1,y_1)<\xi_1$ and $F(x_2,y_2)<\xi_2$. In turn, the convexity of F gives

$$\begin{aligned} f\big(\alpha x_1+(1-\alpha)x_2\big) &\leqslant F\big(\alpha x_1+(1-\alpha)x_2,\alpha y_1+(1-\alpha)y_2\big)\\ &\leqslant \alpha F(x_1,y_1)+(1-\alpha)F(x_2,y_2)\\ &< \alpha\xi_1+(1-\alpha)\xi_2. \end{aligned} \tag{8.35}$$

Thus, letting $\xi_1\downarrow f(x_1)$ and $\xi_2\downarrow f(x_2)$, we obtain $f(\alpha x_1+(1-\alpha)x_2)\leqslant \alpha f(x_1)+(1-\alpha)f(x_2)$. □

Example 8.36 Let C be a convex subset of $\mathcal{H}$ such that $0 \in C$. The *Minkowski gauge* of C, i.e., the function

$$m_C \colon \mathcal{H} \to \left]-\infty, +\infty\right] : x \mapsto \inf \big\{ \xi \in \mathbb{R}_{++} \bigm| x \in \xi C \big\}, \tag{8.36}$$

is convex. Furthermore, $m_C(0) = 0$,

$$(\forall x \in \mathcal{H})(\forall \lambda \in \mathbb{R}_{++}) \quad m_C(\lambda x) = \lambda m_C(x) = m_{(1/\lambda)C}(x), \tag{8.37}$$

and

$$(\forall x \in \operatorname{dom} m_C)(\forall \lambda \in \left]m_C(x), +\infty\right[) \quad x \in \lambda C. \tag{8.38}$$

Proof. The convexity of m_C follows from Example 8.32 and Proposition 8.35. Since $0 \in C$, it is clear that $m_C(0) = 0$. Now take $x \in \mathcal{H}$ and $\lambda \in \mathbb{R}_{++}$. Then

$$\begin{aligned} \big\{ \xi \in \mathbb{R}_{++} \bigm| \lambda x \in \xi C \big\} &= \lambda \big\{ \xi/\lambda \in \mathbb{R}_{++} \bigm| x \in (\xi/\lambda) C \big\} \\ &= \big\{ \xi \in \mathbb{R}_{++} \bigm| x \in \xi\big((1/\lambda)C\big) \big\}, \end{aligned} \tag{8.39}$$

and (8.37) follows by taking the infimum. Finally, suppose that $x \in \operatorname{dom} m_C$, $\lambda > m_C(x)$, and $x \notin \lambda C$. Now let $\mu \in \left]0, \lambda\right]$. Since $0 \in C$ and C is convex, we have $\mu C \subset \lambda C$ and therefore $x \notin \mu C$. Consequently, $m_C(x) \geqslant \lambda$, which is a contradiction. ☐

8.3 Topological Properties

We start with a technical fact.

Proposition 8.37 *Let $f \colon \mathcal{H} \to \left]-\infty, +\infty\right]$ be proper and convex, let $x_0 \in \operatorname{dom} f$, and let $\rho \in \mathbb{R}_{++}$. Then the following hold:*

(i) *Suppose that $\eta = \sup f(B(x_0; \rho)) < +\infty$ and let $\alpha \in \left]0, 1\right[$. Then*

$$\big(\forall x \in B(x_0; \alpha\rho)\big) \quad |f(x) - f(x_0)| \leqslant \alpha\big(\eta - f(x_0)\big). \tag{8.40}$$

(ii) *Suppose that $\delta = \operatorname{diam} f(B(x_0; 2\rho)) < +\infty$. Then f is Lipschitz continuous relative to $B(x_0; \rho)$ with constant δ/ρ.*

Proof. (i): Take $x \in B(x_0; \alpha\rho)$. The convexity of f yields

$$\begin{aligned} f(x) - f(x_0) &= f\big((1-\alpha)x_0 + \alpha\big(x - (1-\alpha)x_0\big)/\alpha\big) - f(x_0) \\ &\leqslant \alpha\big(f\big(x_0 + (x - x_0)/\alpha\big) - f(x_0)\big) \\ &\leqslant \alpha\big(\eta - f(x_0)\big). \end{aligned} \tag{8.41}$$

Likewise,

$$\begin{aligned} f(x_0) - f(x) &= f\left(\frac{x}{1+\alpha} + \frac{\alpha}{1+\alpha}\,\frac{(1+\alpha)x_0 - x}{\alpha}\right) - f(x) \\ &\leqslant \frac{\alpha}{1+\alpha}\big(f\big(x_0 + (x_0 - x)/\alpha\big) - f(x)\big) \\ &\leqslant \frac{\alpha}{1+\alpha}\Big(\big(\eta - f(x_0)\big) + \big(f(x_0) - f(x)\big)\Big), \end{aligned} \tag{8.42}$$

which after rearranging implies that $f(x_0) - f(x) \leqslant \alpha(\eta - f(x_0))$. Altogether, $|f(x) - f(x_0)| \leqslant \alpha(\eta - f(x_0))$.

(ii): Take distinct points x and y in $B(x_0; \rho)$ and set

$$z = x + \left(\frac{1}{\alpha} - 1\right)(x - y), \quad \text{where} \;\; \alpha = \frac{\|x - y\|}{\|x - y\| + \rho} < \frac{\|x - y\|}{\rho}. \tag{8.43}$$

Then $x = \alpha z + (1-\alpha)y$ and $\|z - x_0\| \leqslant \|z - x\| + \|x - x_0\| = \rho + \|x - x_0\| \leqslant 2\rho$. Therefore, y and z belong to $B(x_0; 2\rho)$ and hence, by convexity of f,

$$\begin{aligned} f(x) &= f\big(\alpha z + (1-\alpha)y\big) \\ &\leqslant f(y) + \alpha\big(f(z) - f(y)\big) \\ &\leqslant f(y) + \alpha\delta \\ &\leqslant f(y) + (\delta/\rho)\|x - y\|. \end{aligned} \tag{8.44}$$

Thus $f(x) - f(y) \leqslant (\delta/\rho)\|x - y\|$. Interchanging the roles of x and y, we conclude that $|f(x) - f(y)| \leqslant (\delta/\rho)\|x - y\|$. ☐

The following theorem captures the main continuity properties of convex functions.

Theorem 8.38 *Let $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ be proper and convex, and let $x_0 \in \operatorname{dom} f$. Then the following are equivalent:*

(i) *f is locally Lipschitz continuous near x_0.*
(ii) *f is continuous at x_0.*
(iii) *f is bounded on a neighborhood of x_0.*
(iv) *f is bounded above on a neighborhood of x_0.*

Moreover, if one of these conditions holds, then f is locally Lipschitz continuous on $\operatorname{int}\operatorname{dom} f$.

Proof. (i)⇒(ii)⇒(iii)⇒(iv): Clear.

(iv)⇒(ii): Take $\rho \in \mathbb{R}_{++}$ such that $\eta = \sup f(B(x_0; \rho)) < +\infty$, $\alpha \in \left]0, 1\right[$, and $x \in B(x_0; \alpha\rho)$. Proposition 8.37(i) implies that $|f(x) - f(x_0)| \leqslant \alpha(\eta - f(x_0))$. Therefore, f is continuous at x_0.

(iii)⇒(i): An immediate consequence of Proposition 8.37(ii).

We have shown that items (i)–(iv) are equivalent. Now assume that (iv) is satisfied, say $\eta = \sup f(B(x_0; \rho)) < +\infty$ for some $\rho \in \mathbb{R}_{++}$. Then f is locally

Lipschitz continuous near x_0. Take $x \in \operatorname{int}\operatorname{dom} f \smallsetminus \{x_0\}$ and $\gamma \in \mathbb{R}_{++}$ such that $B(x;\gamma) \subset \operatorname{dom} f$. Now set

$$y = x_0 + \frac{1}{1-\alpha}(x - x_0), \quad \text{where} \quad \alpha = \frac{\gamma}{\gamma + \|x - x_0\|} \in \left]0,1\right[. \tag{8.45}$$

Then $y \in B(x;\gamma)$. Now take $z \in B(x;\alpha\rho)$ and set $w = x_0 + (z - x)/\alpha = (z - (1-\alpha)y)/\alpha$. Then $w \in B(x_0;\rho)$ and $z = \alpha w + (1-\alpha)y$. Consequently,

$$f(z) \leqslant \alpha f(w) + (1-\alpha) f(y) \leqslant \alpha\eta + (1-\alpha) f(y), \tag{8.46}$$

and f is therefore bounded above on $B(x;\alpha\rho)$. By the already established equivalence between (i) and (iv) (applied to the point x), we conclude that f is locally Lipschitz continuous near x. ☐

Corollary 8.39 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex, and suppose that one of the following holds:*

(i) *f is bounded above on some neighborhood.*
(ii) *f is lower semicontinuous.*
(iii) *$\mathcal{H}$ is finite-dimensional.*

Then $\operatorname{cont} f = \operatorname{int}\operatorname{dom} f$.

Proof. Since the inclusion $\operatorname{cont} f \subset \operatorname{int}\operatorname{dom} f$ always holds, we assume that $\operatorname{int}\operatorname{dom} f \neq \varnothing$.

(i): Clear from Theorem 8.38.

(ii): Define a sequence $(C_n)_{n\in\mathbb{N}}$ of closed convex subsets of $\mathcal{H}$ by $(\forall n \in \mathbb{N})$ $C_n = \operatorname{lev}_{\leqslant n} f$. Then $\operatorname{dom} f = \bigcup_{n\in\mathbb{N}} C_n$ and, by Lemma 1.44(i), $\varnothing \neq \overline{\operatorname{int}\operatorname{dom} f} = \overline{\operatorname{int}\bigcup_{n\in\mathbb{N}} C_n} = \overline{\bigcup_{n\in\mathbb{N}} \operatorname{int} C_n}$. Hence there exists $n \in \mathbb{N}$ such that $\operatorname{int} C_n \neq \varnothing$, say $B(x;\rho) \subset C_n$, where $x \in C_n$ and $\rho \in \mathbb{R}_{++}$. Thus $\sup f(B(x;\rho)) \leqslant n$ and we apply (i).

(iii): Let $x \in \operatorname{int}\operatorname{dom} f$. Since $\mathcal{H}$ is finite-dimensional, there exist a finite family $(y_i)_{i\in I}$ in $\operatorname{dom} f$ and $\rho \in \mathbb{R}_{++}$ such that $B(x;\rho) \subset \operatorname{conv}\{y_i\}_{i\in I} \subset \operatorname{int}\operatorname{dom} f$ (see Exercise 3.20). Consequently, by Proposition 8.11, $\sup f(B(x;\rho)) \leqslant \sup f(\operatorname{conv}\{y_i\}_{i\in I}) \leqslant \sup \operatorname{conv}\{f(y_i)\}_{i\in I} = \max_{i\in I} f(y_i) < +\infty$. The conclusion follows from (i). ☐

Corollary 8.40 *Suppose that $\mathcal{H}$ is finite-dimensional and let $f\colon \mathcal{H} \to \mathbb{R}$ be convex. Then f is continuous.*

Corollary 8.41 *Suppose that $\mathcal{H}$ is finite-dimensional, let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex, and let C be a nonempty closed bounded subset of* $\operatorname{ri}\operatorname{dom} f$*. Then f is Lipschitz continuous relative to C.*

Proof. Let $z \in \operatorname{dom} f$. We work in the Euclidean space $\operatorname{span}(z - \operatorname{dom} f)$ and therefore assume, without loss of generality, that $C \subset \operatorname{int}\operatorname{dom} f$. By Corollary 8.39(iii), f is continuous on $\operatorname{int}\operatorname{dom} f$. Hence, by Theorem 8.38, for every $x \in C$, there exists an open ball B_x with center x such that f is β_x-Lipschitz continuous relative to B_x for some $\beta_x \in \mathbb{R}_{++}$. Thus, $C \subset \bigcup_{x \in C} B_x$, and since C is compact, it follows that there exists a finite subset D of C such that $C \subset \bigcup_{x \in D} B_x$. We therefore conclude that f is Lipschitz continuous relative to C with constant $\max_{x \in D} \beta_x$. ☐

Example 8.42 Suppose that $\mathcal{H}$ is infinite-dimensional and let $f : \mathcal{H} \to \mathbb{R}$ be a linear functional that is everywhere discontinuous; see Example 2.27 for a concrete construction. Set $C = \operatorname{lev}_{\leqslant 1} f$ and $H = \{x \in \mathcal{H} \mid f(x) = 0\}$. Then the following hold:

(i) f and $-f$ are convex, $\operatorname{int} \operatorname{dom} f = \operatorname{int} \operatorname{dom}(-f) = \mathcal{H}$, and $\operatorname{cont} f = \operatorname{cont}(-f) = \varnothing$.
(ii) f is neither lower semicontinuous nor upper semicontinuous.
(iii) C is convex, $0 \in \operatorname{core} C$, and $\operatorname{int} C = \varnothing$.
(iv) f is unbounded above and below on every nonempty open subset of $\mathcal{H}$.
(v) H is a hyperplane that is dense in $\mathcal{H}$.

Proof. (i): Since f is linear and $\operatorname{dom} f = \mathcal{H}$, it follows that f and $-f$ are convex with $\operatorname{int} \operatorname{dom} f = \operatorname{int} \operatorname{dom}(-f) = \operatorname{int} \mathcal{H} = \mathcal{H}$. The assumption that f is everywhere discontinuous implies that $\operatorname{cont} f = \operatorname{cont}(-f) = \varnothing$.

(ii): If f were lower semicontinuous, then (i) and Corollary 8.39(ii) would imply that $\varnothing = \operatorname{cont} f = \operatorname{int} \operatorname{dom} f = \mathcal{H}$, which is absurd. Hence f is not lower semicontinuous. Analogously, we deduce that $-f$ is not lower semicontinuous, i.e., that f is not upper semicontinuous.

(iii): The convexity of C follows from (i) and Corollary 8.5. Assume that $x_0 \in \operatorname{int} C$. Since f is bounded above on C by 1, it follows that f is bounded above on a neighborhood of x_0. Then Corollary 8.39(i) implies that $\operatorname{cont} f = \operatorname{int} \operatorname{dom} f$, which contradicts (i). Hence $\operatorname{int} C = \varnothing$. Now take $x \in \mathcal{H}$. If $x \in C$, then $[0, x] \subset C$ by linearity of f. Likewise, if $x \in \mathcal{H} \smallsetminus C$, then $[0, x/f(x)] \subset C$. Thus $\operatorname{cone} C = \mathcal{H}$ and therefore $0 \in \operatorname{core} C$.

(iv)&(v): Take $x_0 \in \mathcal{H}$ and $\varepsilon \in \mathbb{R}_{++}$. Corollary 8.39(i) and item (i) imply that f is unbounded both above and below on $B(x_0; \varepsilon)$. In particular, there exist points y_- and y_+ in $B(x_0; \varepsilon)$ such that $f(y_-) < 0 < f(y_+)$. Now set $y_0 = (f(y_+)y_- - f(y_-)y_+)/(f(y_+) - f(y_-))$. Then, by linearity of f, $y_0 \in [y_-, y_+] \subset B(x_0; \varepsilon)$ and $f(y_0) = 0$. ☐

Example 8.43 Suppose that C is a convex subset of $\mathcal{H}$ such that $0 \in \operatorname{int} C$. Then $\operatorname{dom} m_C = \mathcal{H}$ and m_C is continuous.

Proof. Let $x \in \mathcal{H}$. Then there exists $\varepsilon \in \mathbb{R}_{++}$ such that $\varepsilon x \in C$, which implies that $m_C(x) \leqslant 1/\varepsilon$. Thus $\operatorname{dom} m_C = \mathcal{H}$. Since C is a neighborhood of 0 and $m_C(C) \subset [0, 1]$, Corollary 8.39(i) implies that m_C is continuous. ☐

Example 8.44 Let $\rho \in \mathbb{R}_{++}$ and let

$$f\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto \begin{cases} \rho|x| - \dfrac{\rho^2}{2}, & \text{if } |x| > \rho; \\ \dfrac{|x|^2}{2}, & \text{if } |x| \leqslant \rho \end{cases} \tag{8.47}$$

be the *Huber function*. Then f is continuous and convex.

Proof. The convexity of f follows from Proposition 8.14(i) and its continuity from Corollary 8.40. □

Proposition 8.45 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex. Suppose that f is bounded above on a neighborhood of $x \in \operatorname{dom} f$. Then $\operatorname{int}\operatorname{epi} f \neq \varnothing$.*

Proof. Using Theorem 8.38, we obtain $\delta \in \mathbb{R}_{++}$ and $\beta \in \mathbb{R}_+$ such that

$$\big(\forall y \in B(x;\delta)\big) \quad |f(x) - f(y)| \leqslant \beta\|x-y\| \leqslant \beta\delta. \tag{8.48}$$

Now fix $\rho \in \left]2\beta\delta,+\infty\right[$, set $\gamma = \min\{\delta, \rho/2\} > 0$, and take $(y,\eta) \in \mathcal{H} \times \mathbb{R}$ such that

$$\|(y,\eta) - (x, f(x)+\rho)\|^2 \leqslant \gamma^2. \tag{8.49}$$

Then $\|y-x\| \leqslant \gamma \leqslant \delta$ and $|\eta - (f(x)+\rho)| \leqslant \gamma \leqslant \rho/2$. It follows from (8.48) that $y \in \operatorname{dom} f$ and

$$f(y) < f(x) + \rho/2 = f(x) + \rho - \rho/2 \leqslant f(x) + \rho - \gamma \leqslant \eta. \tag{8.50}$$

Thus $(y,\eta) \in \operatorname{epi} f$ and hence $B((x, f(x)+\rho);\gamma) \subset \operatorname{epi} f$. We conclude that $\operatorname{int}\operatorname{epi} f \neq \varnothing$. □

Proposition 8.46 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$. Then the following hold:*

(i) *Suppose that f is upper semicontinuous at a point $x_0 \in \mathcal{H}$ such that $f(x_0) < 0$. Then $x_0 \in \operatorname{int}\operatorname{lev}_{\leqslant 0} f$.*
(ii) *Suppose that f is convex and that there exists a point $x_0 \in \mathcal{H}$ such that $f(x_0) < 0$. Then $\operatorname{int}\operatorname{lev}_{\leqslant 0} f \subset \operatorname{lev}_{<0} f$.*

Proof. (i): If $x_0 \notin \operatorname{int}\operatorname{lev}_{\leqslant 0} f$, then there exists a sequence $(y_n)_{n\in\mathbb{N}}$ in $\mathcal{H} \smallsetminus \operatorname{lev}_{\leqslant 0} f = \operatorname{lev}_{>0} f$ such that $y_n \to x_0$. In turn, the upper semicontinuity of f at x_0 yields $f(x_0) \geqslant \overline{\lim} f(y_n) \geqslant 0 > f(x_0)$, which is absurd.

(ii): Fix $x \in \operatorname{int}\operatorname{lev}_{\leqslant 0} f$. Then $(\exists\, \rho \in \mathbb{R}_{++})$ $B(x;\rho) \subset \operatorname{lev}_{\leqslant 0} f$. We must show that $f(x) < 0$. We assume that $x \neq x_0$ since, if $x = x_0$, there is nothing to prove. Let $\delta \in \left]0, \rho/\|x - x_0\|\right]$. Set $y = x_0 + (1+\delta)(x - x_0) = x + \delta(x - x_0) \in \operatorname{lev}_{\leqslant 0} f$ and observe that $x = \alpha x_0 + (1-\alpha)y$, where $\alpha = \delta/(1+\delta) \in \left]0,1\right[$. Hence $f(x) \leqslant \alpha f(x_0) + (1-\alpha) f(y) < 0$. □

Corollary 8.47 *Let* $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ *be a convex function such that* $\operatorname{lev}_{<0} f \neq \varnothing$*. Suppose that one of the following holds:*

(i) f *is upper semicontinuous on* $\operatorname{lev}_{<0} f$*.*
(ii) f *is lower semicontinuous and* $\operatorname{dom} f$ *is open.*
(iii) $\mathcal{H}$ *is finite-dimensional and* $\operatorname{dom} f$ *is open.*

Then $\operatorname{int} \operatorname{lev}_{\leqslant 0} f = \operatorname{lev}_{<0} f$*.*

Proof. (i): This follows from Proposition 8.46.

(ii)&(iii): By Corollary 8.39, f is continuous on $\operatorname{int} \operatorname{dom} f = \operatorname{dom} f$ and hence upper semicontinuous on $\operatorname{lev}_{<0} f$. Thus, the claim follows from (i). ☐

Exercises

Exercise 8.1 Let $f\colon \mathbb{R} \to \left]-\infty,+\infty\right]$ be convex, let x_0, x_β, and x_1 be points in $\operatorname{dom} f$ such that $x_0 < x_1$ and $x_\beta = (1-\beta)x_0 + \beta x_1$, where $\beta \in \left]0,1\right[$, and suppose that the points $(x_0, f(x_0))$, $(x_\beta, f(x_\beta))$, and $(x_1, f(x_1))$ lie on a line. Show that $(\forall \alpha \in [0,1])\ f((1-\alpha)x_0 + \alpha x_1) = (1-\alpha)f(x_0) + \alpha f(x_1)$.

Exercise 8.2 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be a proper function and let C be a nonempty subset of $\operatorname{dom} f$. Show that $f + \iota_C$ is convex if and only if C is convex and f is convex on C.

Exercise 8.3 Provide examples of a function $f\colon \mathbb{R} \to \mathbb{R}$ and a nonempty subset C of $\mathbb{R}$ illustrating each of the following:

(i) f is not convex, C is convex, and f is convex on C.
(ii) f is not convex, C is not convex, and f is convex on C.

Exercise 8.4 Provide an example of a proper function $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ and a nonempty nonconvex subset C of $\operatorname{dom} f$ such that f is convex on C but $f + \iota_C$ is not convex.

Exercise 8.5 Provide an example of a function $f\colon \mathcal{H} \to \mathbb{R}$ and a nonempty nonconvex subset C of $\operatorname{dom} f$ such that f is convex on C but there exist distinct points x_1, x_2, and x_3 in C and real numbers α_1, α_2, and α_3 in $\left]0,1\right[$ such that $f(\sum_{i=1}^3 \alpha_i x_i) > \sum_{i=1}^3 \alpha_i f(x_i)$.

Exercise 8.6 Let $f\colon \mathbb{R} \to \left]-\infty,+\infty\right]$ be convex with $\operatorname{int} \operatorname{dom} f \neq \varnothing$. Show that f is strictly convex if and only if it is strictly convex on $\operatorname{int} \operatorname{dom} f$.

Exercise 8.7 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$. Show that f is convex if and only if

$$\begin{aligned}(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f) \quad & x \neq y \\ & \Rightarrow \quad (\forall z \in \left]x,y\right[) \quad \frac{f(z)-f(x)}{\|z-x\|} + \frac{f(z)-f(y)}{\|z-y\|} \leqslant 0. \quad (8.51)\end{aligned}$$

Exercise 8.8 Let $f\colon \mathbb{R} \to \,]-\infty,+\infty]$ be proper and convex, let $x_0 \in \operatorname{dom} f$, and let $\alpha \in \,]0,1[$. Set $g\colon \operatorname{dom} f \to \mathbb{R}\colon x \mapsto \alpha f(x) - f(x_0 + \alpha(x - x_0))$. Show that g is decreasing on $]-\infty, x_0] \cap \operatorname{dom} f$ and increasing on $[x_0, +\infty[\cap \operatorname{dom} f$. In addition show that, if f is strictly convex, then g is strictly decreasing on $]-\infty, x_0] \cap \operatorname{dom} f$ and strictly increasing on $[x_0, +\infty[\cap \operatorname{dom} f$.

Exercise 8.9 Show that the converse of Corollary 8.5 is false by providing an example of a function that is not convex and the lower level sets of which are all convex.

Exercise 8.10 (Arithmetic mean–geometric mean inequality) Let $(x_i)_{1 \leqslant i \leqslant m}$ be a finite family in $\mathbb{R}_+$. Show that

$$\sqrt[m]{x_1 \cdots x_m} \leqslant \frac{x_1 + \cdots + x_m}{m}, \tag{8.52}$$

and that equality occurs in (8.52) if and only if $x_1 = \cdots = x_m$.

Exercise 8.11 Let $\phi\colon \mathbb{R} \to \,]-\infty,+\infty]$ be a convex function such that $\operatorname{dom} \phi \subset \mathbb{R}_{++}$. Show that $\phi^{\diamond\diamond} = \phi$, where $\phi^{\diamond\diamond} = (\phi^{\diamond})^{\diamond}$ is the biadjoint of ϕ (see Example 8.33).

Exercise 8.12 Let $p \in [1,+\infty[$. Denoting the adjoint of ϕ by $\phi^{\diamond}$ (see Example 8.33), show that $\phi^{\diamond} = \phi$ in each of the following cases:

(i) $\phi\colon \mathbb{R} \to \,]-\infty,+\infty] : x \mapsto \begin{cases} |x^{1/p} - 1|^p, & \text{if } x > 0;\\ +\infty, & \text{otherwise.}\end{cases}$

(For $p = 2$, this is the *Hellinger entropy* function.)

(ii) $\phi\colon \mathbb{R} \to \,]-\infty,+\infty] : x \mapsto \begin{cases} -|x^{1/p} + 1|^p, & \text{if } x > 0;\\ +\infty, & \text{otherwise.}\end{cases}$

(iii) $\phi\colon \mathbb{R} \to \,]-\infty,+\infty] : x \mapsto \begin{cases} -\sqrt{x}, & \text{if } x > 0;\\ +\infty, & \text{otherwise.}\end{cases}$

Exercise 8.13 Let $(\mathcal{H}_i)_{i\in I}$ be a finite family of real Hilbert spaces, and let, for every i in I, $f_i\colon \mathcal{H}_i \to \,]-\infty,+\infty]$ be proper. Show that $\bigoplus_{i\in I} f_i$ is convex if and only if the functions $(f_i)_{i\in I}$ are. Furthermore, prove the corresponding statement for strict convexity.

Exercise 8.14 Let $f\colon \mathcal{H} \to \,]-\infty,+\infty]$ and $g\colon \mathcal{H} \to \,]-\infty,+\infty]$, and assume that

$$h\colon \mathcal{H} \to \,]-\infty,+\infty] : x \mapsto \inf_{y\in\mathcal{H}} \max\bigl\{f(y), g(x-y)\bigr\}. \tag{8.53}$$

Prove the following:

(i) $\operatorname{dom} h = \operatorname{dom} f + \operatorname{dom} g$.
(ii) $\inf h(\mathcal{H}) = \max\{\inf f(\mathcal{H}), \inf g(\mathcal{H})\}$.

(iii) $(\forall \eta \in \mathbb{R})$ $\operatorname{lev}_{<\eta} h = \operatorname{lev}_{<\eta} f + \operatorname{lev}_{<\eta} g$.
(iv) h is convex if f and g are.

Exercise 8.15 Determine the Minkowski gauge of the closed unit ball.

Exercise 8.16 Set $C = \big\{(\xi_1, \xi_2) \in \mathbb{R}^2 \;\big|\; \max\{|\xi_1|, |\xi_2|\} \leqslant 1\big\}$. Show that $m_C \colon \mathbb{R}^2 \to \mathbb{R} \colon (\xi_1, \xi_2) \mapsto \max\{|\xi_1|, |\xi_2|\}$ and that $m_{(1/2)C} = 2m_C$.

Exercise 8.17 Let C and D be convex subsets of $\mathcal{H}$ such that $0 \in C \subset D$. Show that $m_C \geqslant m_D$.

Exercise 8.18 Set $C = \big\{(\xi_1, \xi_2) \in \mathbb{R}^2 \;\big|\; \max\{|\xi_1|, |\xi_2|\} \leqslant 1\big\}$, $D = B(0; 1)$, and $Q = \frac{1}{2}C + \frac{1}{2}D$. Show that $0 \in \operatorname{int} Q$, that Q is closed and convex, that $m_Q \colon \mathbb{R}^2 \to \mathbb{R}$ is convex and continuous, that

$$(\forall (\xi_1, \xi_2) \in \mathbb{R}^2) \quad \max\{|\xi_1|, |\xi_2|\} \leqslant m_Q(\xi_1, \xi_2) \leqslant 2\max\{|\xi_1|, |\xi_2|\}, \tag{8.54}$$

that

$$(\forall \xi_1 \in \mathbb{R}_{++})(\forall \xi_2 \in \left[0, \tfrac{1}{2}\xi_1\right[) \quad m_Q(\xi_1, \xi_2) = \xi_1, \tag{8.55}$$

and that

$$(\forall (\xi_1, \xi_2) \in \mathbb{R}^2_+)(\forall (\eta_1, \eta_2) \in \mathbb{R}^2_+) \qquad \left.\begin{aligned} \xi_1 &\leqslant \eta_1 \\ \xi_2 &\leqslant \eta_2 \end{aligned}\right\} \;\Rightarrow\; m_Q(\xi_1, \xi_2) \leqslant m_Q(\eta_1, \eta_2). \tag{8.56}$$

Exercise 8.19 Suppose that $\mathcal{H}$ is finite-dimensional and let C be a convex subset of $\mathcal{H}$ such that $0 \in \operatorname{ri} C$. Show that m_C is lower semicontinuous.

Exercise 8.20 Let f and C be as in Example 8.42. Determine m_C.

Exercise 8.21 Suppose that $\mathcal{H}$ is infinite-dimensional. Provide an example of a convex subset C of $\mathcal{H}$ such that $0 \in C$ but m_C is not lower semicontinuous.

Exercise 8.22 Show that the conclusion of Corollary 8.47(ii) fails if $\operatorname{dom} f$ is not open.

Chapter 9
Lower Semicontinuous Convex Functions

The theory of convex functions is most powerful in the presence of lower semicontinuity. A key property of lower semicontinuous convex functions is the existence of a continuous affine minorant, which we establish in this chapter by projecting onto the epigraph of the function.

9.1 Lower Semicontinuous Convex Functions

We start by observing that various types of lower semicontinuity coincide for convex functions.

Theorem 9.1 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be convex. Then the following are equivalent:*

(i) *f is weakly sequentially lower semicontinuous.*
(ii) *f is sequentially lower semicontinuous.*
(iii) *f is lower semicontinuous.*
(iv) *f is weakly lower semicontinuous.*

Proof. The set $\operatorname{epi} f$ is convex by Definition 8.1. Hence, the equivalences follow from Lemma 1.24, Lemma 1.36, and Theorem 3.34. □

Definition 9.2 The set of lower semicontinuous convex functions from $\mathcal{H}$ to $[-\infty,+\infty]$ is denoted by $\Gamma(\mathcal{H})$.

The set $\Gamma(\mathcal{H})$ is closed under several important operations. For instance, it is straightforward to verify that $\Gamma(\mathcal{H})$ is closed under multiplication by strictly positive real numbers.

Proposition 9.3 *Let $(f_i)_{i\in I}$ be a family in $\Gamma(\mathcal{H})$. Then $\sup_{i\in I} f_i \in \Gamma(\mathcal{H})$.*

Proof. Combine Lemma 1.26 and Proposition 8.16. □

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-48311-5_31

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_9

Corollary 9.4 *Let $(f_i)_{i\in I}$ be a family in $\Gamma(\mathcal{H})$. Suppose that one of the following holds:*

(i) *I is finite and $-\infty \notin \bigcup_{i\in I} f_i(\mathcal{H})$.*
(ii) *$\inf_{i\in I} f_i \geqslant 0$.*

Then $\sum_{i\in I} f_i \in \Gamma(\mathcal{H})$.

Proof. (i): A consequence of Lemma 1.27 and Proposition 8.17.

(ii): Let $\mathcal{I}$ be the class of nonempty finite subsets of I and set $(\forall J \in \mathcal{I})$ $g_J = \sum_{i\in J} f_i$. Then it follows from (i) that $(\forall J \in \mathcal{I})$ $g_J \in \Gamma(\mathcal{H})$. However, (2.4) yields $\sum_{i\in I} f_i = \sup_{J\in\mathcal{I}} g_J$. In view of Proposition 9.3, the proof is complete. $\square$

Proposition 9.5 *Let $\mathcal{K}$ be a real Hilbert space, let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$, and let $f \in \Gamma(\mathcal{K})$. Then $f \circ L \in \Gamma(\mathcal{H})$.*

Proof. This is a consequence of Proposition 8.20. $\square$

Proposition 9.6 *Let $f \in \Gamma(\mathcal{H})$ and suppose that $-\infty \in f(\mathcal{H})$. Then f is nowhere real-valued, i.e., $f(\mathcal{H}) \subset \{-\infty, +\infty\}$.*

Proof. Let $x \in \mathcal{H}$ be such that $f(x) = -\infty$, let $y \in \mathcal{H}$, and let $\alpha \in \,]0,1[$. If $f(y) \neq +\infty$, then Proposition 8.4 yields $f(\alpha x + (1-\alpha)y) = -\infty$. In turn, since f is lower semicontinuous, $f(y) \leqslant \underline{\lim}_{\alpha\downarrow 0} f(\alpha x + (1-\alpha)y) = -\infty$, i.e., $f(y) = -\infty$. $\square$

The function $x \mapsto -\infty$ belongs to $\Gamma(\mathcal{H})$, which makes the following notion well defined.

Definition 9.7 Let $f\colon \mathcal{H} \to [-\infty, +\infty]$. Then

$$\breve{f} = \sup\big\{g \in \Gamma(\mathcal{H}) \;\big|\; g \leqslant f\big\} \tag{9.1}$$

is the *lower semicontinuous convex envelope* of f.

Proposition 9.8 *Let $f\colon \mathcal{H} \to [-\infty, +\infty]$. Then the following hold:*

(i) *$\breve{f}$ is the largest lower semicontinuous convex function majorized by f.*
(ii) *$(\forall x \in \mathcal{H})$ $\breve{f}(x) = \underline{\lim}_{y\to x} \breve{f}(y)$.*
(iii) *$\operatorname{epi} \breve{f}$ is closed and convex.*
(iv) *$\operatorname{conv} \operatorname{dom} f \subset \operatorname{dom} \breve{f} \subset \overline{\operatorname{conv}} \operatorname{dom} f$.*

Proof. (i): This is a consequence of (9.1) and Proposition 9.3.

(ii): This follows from (i) and Lemma 1.32(iv).

(iii): Combine (i), Lemma 1.24, and Definition 8.1.

(iv): By (i), $\breve{f} \leqslant f$ and $\breve{f}$ is convex. Hence, Proposition 8.2 yields

$$\operatorname{conv} \operatorname{dom} f \subset \operatorname{conv} \operatorname{dom} \breve{f} = \operatorname{dom} \breve{f}. \tag{9.2}$$

Now set $C = \overline{\operatorname{conv}}\, \operatorname{dom} f$ and

$$g\colon \mathcal{H} \to [-\infty,+\infty] : x \mapsto \begin{cases} \breve{f}(x), & \text{if } x \in C; \\ +\infty, & \text{if } x \notin C. \end{cases} \tag{9.3}$$

Using (iii), we note that $\operatorname{epi} g = (\operatorname{epi} \breve{f}) \cap (C \times \mathbb{R})$ is closed and convex. It follows from Lemma 1.24 and Definition 8.1 that

$$g \in \Gamma(\mathcal{H}). \tag{9.4}$$

Now fix $x \in \mathcal{H}$. If $x \in C$, then $g(x) = \breve{f}(x) \leqslant f(x)$; otherwise, $x \notin \operatorname{dom} f \subset C$ and therefore $g(x) = f(x) = +\infty$. Altogether, $g \leqslant f$ and, in view of (9.4), we obtain $g \leqslant \breve{f}$. Thus, $\operatorname{dom} \breve{f} \subset \operatorname{dom} g \subset C = \overline{\operatorname{conv}}\, \operatorname{dom} f$. ☐

Theorem 9.9 *Let* $f\colon \mathcal{H} \to [-\infty,+\infty]$. *Then* $\operatorname{epi} \breve{f} = \overline{\operatorname{conv}}\, \operatorname{epi} f$.

Proof. Set $E = \overline{\operatorname{conv}}\, \operatorname{epi} f$. Since $\breve{f} \leqslant f$, we have $\operatorname{epi} f \subset \operatorname{epi} \breve{f}$. Hence $E \subset \overline{\operatorname{conv}}\, \operatorname{epi} \breve{f} = \operatorname{epi} \breve{f}$ by Proposition 9.8(iii). It remains to show that $\operatorname{epi} \breve{f} \subset E$. We assume that $f \not\equiv +\infty$, since otherwise $\breve{f} = f$ and the conclusion is clear. Let us proceed by contradiction and assume that there exists

$$(x,\xi) \in \operatorname{epi} \breve{f} \smallsetminus E. \tag{9.5}$$

Since E is a nonempty closed convex subset of $\mathcal{H} \times \mathbb{R}$, Theorem 3.16 implies that the projection (p,π) of (x,ξ) onto E satisfies

$$\big(\forall (y,\eta) \in E\big) \quad \langle y-p \mid x-p\rangle + (\eta-\pi)(\xi-\pi) \leqslant 0. \tag{9.6}$$

Letting $\eta \uparrow +\infty$ in (9.6), we deduce that $\xi \leqslant \pi$. Let us first assume that $\xi = \pi$. Then (9.6) yields $(\forall y \in \overline{\operatorname{conv}}\, \operatorname{dom} f)$ $\langle y-p \mid x-p\rangle \leqslant 0$. Consequently, since Proposition 9.8(iv) asserts that $x \in \operatorname{dom} \breve{f} \subset \overline{\operatorname{conv}}\, \operatorname{dom} f$, we obtain $\|x-p\|^2 = 0$ and, in turn, $(p,\pi) = (x,\xi)$, which is impossible since $(x,\xi) \notin E$ by (9.5). Therefore, we must have

$$\xi < \pi. \tag{9.7}$$

Setting $u = (x-p)/(\pi-\xi)$ and letting $\eta = f(y)$ in (9.6), we get

$$(\forall y \in \operatorname{dom} f) \quad \langle y-p \mid u\rangle + \pi \leqslant f(y). \tag{9.8}$$

Consequently, f is minorized by the lower semicontinuous convex function $g\colon y \mapsto \langle y-p \mid u\rangle + \pi$, and it follows that $g \leqslant \breve{f}$. In particular, since $(x,\xi) \in \operatorname{epi} \breve{f}$, we have

$$\pi \leqslant \frac{\|x-p\|^2}{\pi-\xi} + \pi = g(x) \leqslant \breve{f}(x) \leqslant \xi, \tag{9.9}$$

which contradicts (9.7). We conclude that $\operatorname{epi} \breve{f} \subset E$. ☐

Corollary 9.10 *Let* $f\colon \mathcal{H} \to [-\infty, +\infty]$ *be convex. Then* $\bar{f} = \breve{f}$.

Proof. Combine Lemma 1.32(vi) and Theorem 9.9. ☐

Corollary 9.11 *Let* $f\colon \mathcal{H} \to [-\infty, +\infty]$ *be convex and such that* $\operatorname{lev}_{<0} f \neq \varnothing$. *Then* $\operatorname{lev}_{<0} f \subset \operatorname{lev}_{\leqslant 0} f \subset \operatorname{lev}_{\leqslant 0} \breve{f}$ *and* $\overline{\operatorname{lev}_{<0} f} = \overline{\operatorname{lev}_{\leqslant 0} f} = \operatorname{lev}_{\leqslant 0} \breve{f}$.

Proof. Take $x \in \mathcal{H}$. Then $f(x) < 0 \Rightarrow f(x) \leqslant 0 \Rightarrow \breve{f}(x) \leqslant 0$, which shows the inclusions. Now assume that $x \in \operatorname{lev}_{\leqslant 0} \breve{f}$. Since f is convex, Theorem 9.9 yields $\operatorname{epi} \breve{f} = \overline{\operatorname{epi}} f$. Hence there exists a sequence $(x_n, \xi_n)_{n\in\mathbb{N}}$ in $\operatorname{epi} f$ that converges to $(x, \breve{f}(x))$. Fix $z \in \operatorname{lev}_{<0} f$ and take a sequence $(\zeta_n)_{n\in\mathbb{N}}$ in $\mathbb{R}_{--}$ such that $(z, \zeta_n)_{n\in\mathbb{N}}$ lies in $\operatorname{epi} f$ and $\zeta_n \to f(z)$. Furthermore, fix a sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ in $]0, 1[$ such that $\varepsilon_n \to 0$. Define a sequence $(\alpha_n)_{n\in\mathbb{N}}$ in $\mathbb{R}_{++}$ by

$$(\forall n \in \mathbb{N}) \quad \alpha_n = \varepsilon_n + \begin{cases} 0, & \text{if } \xi_n \leqslant 0; \\ \xi_n/(\xi_n - \zeta_n), & \text{otherwise.} \end{cases} \tag{9.10}$$

Observe that $(\alpha_n)_{n\in\mathbb{N}}$ lies eventually in $]0, 1[$ and $\alpha_n \to 0$. Hence, eventually, $(\alpha_n(z, \zeta_n) + (1 - \alpha_n)(x_n, \xi_n))_{n\in\mathbb{N}}$ lies in $\operatorname{epi} f$ and therefore

$$\begin{aligned} f(\alpha_n z + (1 - \alpha_n)x_n) &\leqslant \alpha_n\zeta_n + (1 - \alpha_n)\xi_n \\ &= \xi_n + \varepsilon_n(\zeta_n - \xi_n) - \max\{\xi_n, 0\} \\ &< 0. \end{aligned} \tag{9.11}$$

Therefore the sequence $(\alpha_n z + (1 - \alpha_n)x_n)_{n\in\mathbb{N}}$, which converges to x, lies eventually in $\operatorname{lev}_{<0} f$. The result follows. ☐

9.2 Proper Lower Semicontinuous Convex Functions

As illustrated in Proposition 9.6, nonproper lower semicontinuous convex functions are of limited use. By contrast, proper lower semicontinuous convex functions will play a central role in this book.

Definition 9.12 The set of proper lower semicontinuous convex functions from $\mathcal{H}$ to $]-\infty, +\infty]$ is denoted by $\Gamma_0(\mathcal{H})$.

Example 9.13 Let $(e_i)_{i\in I}$ be a family in $\mathcal{H}$ and let $(\phi_i)_{i\in I}$ be a family in $\Gamma_0(\mathbb{R})$ such that $(\forall i \in I)\ \phi_i \geqslant \phi_i(0) = 0$. Set $f\colon \mathcal{H} \to]-\infty, +\infty] : x \mapsto \sum_{i\in I} \phi_i(\langle x \mid e_i\rangle)$. Then $f \in \Gamma_0(\mathcal{H})$.

Proof. Set $(\forall i \in I)\ f_i\colon \mathcal{H} \to]-\infty, +\infty] : x \mapsto \phi_i(\langle x \mid e_i\rangle)$. Then $f = \sum_{i\in I} f_i$ and $(\forall i \in I)\ 0 \leqslant f_i \in \Gamma_0(\mathcal{H})$. Thus, it follows from Corollary 9.4(ii) that $f \in \Gamma(\mathcal{H})$. Finally, since $f(0) = 0$, f is proper. ☐

Proposition 9.14 *Let $f \in \Gamma_0(\mathcal{H})$, let $x \in \mathcal{H}$, and let $y \in \operatorname{dom} f$. For every $\alpha \in \,]0,1[$, set $x_\alpha = (1-\alpha)x + \alpha y$. Then $\lim_{\alpha \downarrow 0} f(x_\alpha) = f(x)$.*

Proof. Using the lower semicontinuity and the convexity of f, we obtain $f(x) \leqslant \underline{\lim}_{\alpha\downarrow 0} f(x_\alpha) \leqslant \overline{\lim}_{\alpha\downarrow 0} f(x_\alpha) \leqslant \overline{\lim}_{\alpha\downarrow 0}(1-\alpha)f(x) + \alpha f(y) = f(x)$. Therefore, $\lim_{\alpha\downarrow 0} f(x_\alpha) = f(x)$. ☐

Corollary 9.15 *Let $f \in \Gamma_0(\mathbb{R})$. Then $f|_{\overline{\operatorname{dom} f}}$ is continuous.*

The conclusion of Corollary 9.15 fails in general Hilbert spaces and even in the Euclidean plane (apply Example 9.43 below with $\mathcal{H} = \mathbb{R}$ and $p = 2$).

Proposition 9.16 *Let f and g be functions in $\Gamma_0(\mathcal{H} \oplus \mathbb{R})$ such that $(\operatorname{dom} f) \cup (\operatorname{dom} g) \subset \mathcal{H} \times \mathbb{R}_+$. Assume that $f|_{\mathcal{H}\times\mathbb{R}_{++}} = g|_{\mathcal{H}\times\mathbb{R}_{++}} \not\equiv +\infty$. Then $f = g$.*

Proof. Let $(y,\eta) \in \operatorname{dom} f \cap \operatorname{dom} g \cap (\mathcal{H} \times \mathbb{R}_{++})$ and let $x \in \mathcal{H}$. Then, for every $\alpha \in \,]0,1[$,

$$(1-\alpha)(x,0) + \alpha(y,\eta) = \big((1-\alpha)x + \alpha y, \alpha\eta\big) \in \mathcal{H} \times \mathbb{R}_{++}. \tag{9.12}$$

Taking the limit as $\alpha \downarrow 0$, we obtain

$$f(x,0) \leftarrow f\big((1-\alpha)x + \alpha y, \alpha\eta\big) = g\big((1-\alpha)x + \alpha y, \alpha\eta\big) \to g(x,0) \tag{9.13}$$

by Proposition 9.14. Hence $f|_{\mathcal{H}\times\{0\}} = g|_{\mathcal{H}\times\{0\}}$. Finally, the assumption on the domain implies that $f|_{\mathcal{H}\times\mathbb{R}_{--}} = g|_{\mathcal{H}\times\mathbb{R}_{--}}$. ☐

We conclude this section with an extension of Fact 6.13.

Fact 9.17 (See [329, Corollary 13.2]) Let f and g be in $\Gamma_0(\mathcal{H})$. Then

$$\operatorname{int}(\operatorname{dom} f - \operatorname{dom} g) = \operatorname{core}(\operatorname{dom} f - \operatorname{dom} g). \tag{9.14}$$

9.3 Affine Minorization

A key property of functions in $\Gamma_0(\mathcal{H})$ is that they possess continuous affine minorants. To see this, we require the following two results.

Proposition 9.18 *Let $f \in \Gamma_0(\mathcal{H})$, let $(x,\xi) \in \mathcal{H} \times \mathbb{R}$, and let $(p,\pi) \in \mathcal{H} \times \mathbb{R}$. Then $(p,\pi) = P_{\operatorname{epi} f}(x,\xi)$ if and only if*

$$\max\{\xi, f(p)\} \leqslant \pi \tag{9.15}$$

and

$$(\forall y \in \operatorname{dom} f) \quad \langle y - p \mid x - p \rangle + \big(f(y) - \pi\big)\big(\xi - \pi\big) \leqslant 0. \tag{9.16}$$

Proof. Since $f \in \Gamma_0(\mathcal{H})$, the set $\operatorname{epi} f$ is nonempty, closed, and convex. Now set $(p,\pi) = P_{\operatorname{epi} f}(x,\xi)$. Then Theorem 3.16 implies that (p,π) is characterized by $(p,\pi) \in \operatorname{epi} f$ and $(\forall (y,\eta) \in \operatorname{epi} f)$ $\langle y-p \mid x-p\rangle + (\eta-\pi)(\xi-\pi) \leqslant 0$, which is equivalent to $f(p) \leqslant \pi$ and $(\forall y \in \operatorname{dom} f)(\forall \lambda \in \mathbb{R}_+)$ $\langle y-p \mid x-p\rangle + (f(y)+\lambda-\pi)(\xi-\pi) \leqslant 0$. By letting $\lambda \uparrow +\infty$, we deduce that $\xi \leqslant \pi$. The characterization follows. ☐

Proposition 9.19 *Let $f \in \Gamma_0(\mathcal{H})$, let $x \in \operatorname{dom} f$, let $\xi \in \left]-\infty, f(x)\right[$, and let $(p,\pi) \in \mathcal{H} \times \mathbb{R}$. Then $(p,\pi) = P_{\operatorname{epi} f}(x,\xi)$ if and only if*

$$\xi < f(p) = \pi \tag{9.17}$$

and

$$(\forall y \in \operatorname{dom} f) \quad \langle y-p \mid x-p\rangle \leqslant \big(f(y)-f(p)\big)\big(f(p)-\xi\big). \tag{9.18}$$

Proof. Suppose first that $(p,\pi) = P_{\operatorname{epi} f}(x,\xi)$. Since $p \in \operatorname{dom} f$, (9.16) yields

$$\big(f(p)-\pi\big)\big(\xi-\pi\big) \leqslant 0. \tag{9.19}$$

To establish that $\xi < f(p)$, we argue by contradiction. Suppose that $f(p) \leqslant \xi$. Then $f(p)-\pi \leqslant \xi-\pi$ and hence, since $\xi-\pi \leqslant 0$ by (9.15), we obtain $(f(p)-\pi)(\xi-\pi) \geqslant (\xi-\pi)^2$. In view of (9.19), we deduce that $\xi = \pi$. In turn, since $x \in \operatorname{dom} f$, (9.16) implies that $\langle x-p \mid x-p\rangle \leqslant 0$. Thus $x = p$ and hence $(p,\pi) = (x,\xi)$. This is impossible, since $(p,\pi) \in \operatorname{epi} f$ and $(x,\xi) \notin \operatorname{epi} f$. Thus,

$$\xi < f(p), \tag{9.20}$$

and (9.15) implies that $\xi < \pi$ and $f(p) \leqslant \pi$. Hence, (9.19) yields $f(p) = \pi$ and (9.17) holds. Combining (9.17) and Proposition 9.18, we obtain (9.18).

Conversely, if (9.17) and (9.18) hold, then Proposition 9.18 implies directly that $(p,\pi) = P_{\operatorname{epi} f}(x,\xi)$. ☐

Theorem 9.20 *Let $f \in \Gamma_0(\mathcal{H})$. Then f possesses a continuous affine minorant. More precisely,*

$$(\exists\, p \in \operatorname{dom} f)(\exists\, u \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \langle y-p \mid u\rangle + f(p) \leqslant f(y). \tag{9.21}$$

Proof. Fix $x \in \operatorname{dom} f$ and $\xi \in \left]-\infty, f(x)\right[$, and set $(p,\pi) = P_{\operatorname{epi} f}(x,\xi)$. Then, by (9.17), $f(p) > \xi$. Now set $u = (x-p)/(f(p)-\xi)$ and $g\colon \mathcal{H} \to \mathbb{R}\colon y \mapsto \langle y-p \mid u\rangle + f(p)$. Then (9.18) yields $g \leqslant f$. ☐

Corollary 9.21 *Let $f \in \Gamma_0(\mathcal{H})$. Then f is bounded below on every nonempty bounded subset of $\mathcal{H}$.*

Proof. Let C be a nonempty bounded subset of $\mathcal{H}$ and set $\beta = \sup_{x\in C} \|x\|$. Theorem 9.20 asserts that f has a continuous affine minorant, say $\langle \cdot \mid u\rangle + \eta$. Then, by Cauchy–Schwarz, $(\forall x \in C)$ $f(x) \geqslant \langle x \mid u\rangle + \eta \geqslant -\|x\|\,\|u\| + \eta \geqslant -\beta\|u\| + \eta > -\infty$. ☐

Example 9.22 Suppose that $\mathcal{H}$ is infinite-dimensional and let $f\colon \mathcal{H} \to \mathbb{R}$ be a discontinuous linear functional (see Example 2.27 and Example 8.42). Then f has no continuous affine minorant.

Proof. Assume that the conclusion is false, i.e., that there exist $u \in \mathcal{H}$ and $\eta \in \mathbb{R}$ such that $(\forall x \in \mathcal{H})\ \langle x \mid u\rangle + \eta \leqslant f(x)$. Then, since f is odd, $(\forall x \in \mathcal{H})$ $f(x) \leqslant \langle x \mid u\rangle - \eta \leqslant \|x\|\,\|u\| - \eta$. Consequently, $\sup f(B(0;1)) \leqslant \|u\| - \eta$ and therefore f is bounded above on a neighborhood of 0. This contradicts Corollary 8.39(i) since f is nowhere continuous. ☐

Theorem 9.23 *Let $f \in \Gamma_0(\mathcal{H})$ and let $x \in \operatorname{int}\operatorname{dom} f$. Then there exists a continuous affine minorant a of f such that $a(x) = f(x)$. In other words, $(\exists\, u \in \mathcal{H})(\forall y \in \mathcal{H})\ \langle y - x \mid u\rangle + f(x) \leqslant f(y)$.*

Proof. In view of Corollary 8.39, $x \in \operatorname{cont} f$. Hence, it follows from Theorem 8.38 and Proposition 8.45 that $\operatorname{int}\operatorname{epi} f \neq \varnothing$. In turn, Proposition 7.5 implies that $(x, f(x)) \in \operatorname{spts}(\operatorname{epi} f)$, and we therefore derive from Theorem 7.4 that there exists $(z, \zeta) \in (\mathcal{H} \times \mathbb{R}) \smallsetminus (\operatorname{epi} f)$ such that $(x, f(x)) = P_{\operatorname{epi} f}(z, \zeta)$. In view of Proposition 3.21 and since $x \in \operatorname{int}\operatorname{dom} f$, we assume that $z \in \operatorname{int}\operatorname{dom} f$. Thus, by Proposition 9.18, $\max\{\zeta, f(x)\} \leqslant f(x)$, i.e.,

$$f(x) \geqslant \zeta \tag{9.22}$$

and

$$(\forall y \in \operatorname{dom} f) \quad \langle y - x \mid z - x\rangle + \big(f(y) - f(x)\big)\big(\zeta - f(x)\big) \leqslant 0. \tag{9.23}$$

If $f(x) = \zeta$, then the above inequality evaluated at $y = z$ yields $z = x$, which contradicts the fact that $(z, \zeta) \neq (x, f(x))$. Hence $f(x) > \zeta$. Now set $u = (z - x)/(f(x) - \zeta)$. Then (9.23) becomes $(\forall y \in \operatorname{dom} f)\ \langle y - x \mid u\rangle + f(x) - f(y) \leqslant 0$, and the result follows. ☐

Proposition 9.24 (Jensen's inequality) *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space such that $\mu(\Omega) \in \mathbb{R}_{++}$, let $\phi \in \Gamma_0(\mathbb{R})$, and let $x\colon \Omega \to \mathbb{R}$ be a measurable function such that $\mu(\Omega)^{-1} \int_\Omega x(\omega)\mu(d\omega) \in \operatorname{int}\operatorname{dom} \phi$. Then*

$$\phi\bigg(\frac{1}{\mu(\Omega)} \int_\Omega x(\omega)\mu(d\omega)\bigg) \leqslant \frac{1}{\mu(\Omega)} \int_\Omega \phi(x(\omega))\mu(d\omega). \tag{9.24}$$

Proof. Since ϕ is lower semicontinuous, it is measurable, and so is therefore $\phi \circ x$. Now set $\xi = \mu(\Omega)^{-1} \int_\Omega x\, d\mu$. It follows from Theorem 9.23 that there exists $\alpha \in \mathbb{R}$ such that $(\forall \eta \in \mathbb{R})\ \alpha(\eta - \xi) + \phi(\xi) \leqslant \phi(\eta)$. Thus, for μ-almost every $\omega \in \Omega$, $\alpha(x(\omega) - \xi) + \phi(\xi) \leqslant \phi(x(\omega))$. Integrating these inequalities over Ω with respect to μ yields $\phi(\xi)\mu(\Omega) \leqslant \int_\Omega \phi(x(\omega))\mu(d\omega)$. ☐

Example 9.25 Let $(\Omega, \mathcal{F}, \mu)$ be a measure space such that $\mu(\Omega) \in \mathbb{R}_{++}$, let $(\mathsf{H}, \|\cdot\|_{\mathsf{H}})$ be a separable real Hilbert space, and take p and q in $\mathbb{R}_{++}$ such that $p < q$. Then the following hold:

(i) Let $x \in L^p((\Omega, \mathcal{F}, \mu); \mathsf{H})$. Then

$$\left(\int_\Omega \|x(\omega)\|_{\mathsf{H}}^p \mu(d\omega) \right)^{\frac{1}{p}} \leqslant \mu(\Omega)^{\frac{1}{p}-\frac{1}{q}} \left(\int_\Omega \|x(\omega)\|_{\mathsf{H}}^q \mu(d\omega) \right)^{\frac{1}{q}}. \tag{9.25}$$

(ii) $L^q((\Omega, \mathcal{F}, \mu); \mathsf{H}) \subset L^p((\Omega, \mathcal{F}, \mu); \mathsf{H})$.

Proof. (i): Set $\phi = |\cdot|^{q/p}$. Then it follows from Example 8.23 that ϕ is convex. Now set $y\colon \omega \mapsto \|x(\omega)\|_{\mathsf{H}}^p$. Since y is integrable, $\mu(\Omega)^{-1} \int_\Omega y \, d\mu \in \mathbb{R} = \operatorname{dom} \phi$, and Proposition 9.24 applied to y yields (9.25).

(ii): An immediate consequence of (i). ☐

Example 9.26 Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ and take p and q in $\mathbb{R}_{++}$ such that $p < q$. Suppose that $\mathsf{E}|X|^p < +\infty$. Then $\mathsf{E}^{1/p}|X|^p \leqslant \mathsf{E}^{1/q}|X|^q$.

Proof. Set $\mu = \mathsf{P}$ and $\mathsf{H} = \mathbb{R}$ in Example 9.25(i) (see Example 2.9). ☐

9.4 Recession Function

Proposition 9.27 *Let $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ be proper and convex, let $x \in \operatorname{dom} f$, let $y \in \mathcal{H}$, and set $\phi\colon \mathbb{R}_{++} \to \left]-\infty, +\infty\right] : \alpha \mapsto (f(x+\alpha y) - f(x))/\alpha$. Then ϕ is increasing.*

Proof. Fix α and β in $\mathbb{R}_{++}$ such that $\alpha < \beta$, and set $\lambda = \alpha/\beta$ and $z = x + \beta y$. If $f(z) = +\infty$, then certainly $\phi(\alpha) \leqslant \phi(\beta) = +\infty$. Otherwise, by (8.1), $f(x+\alpha y) = f(\lambda z + (1-\lambda)x) \leqslant \lambda f(z) + (1-\lambda) f(x) = f(x) + \lambda(f(z) - f(x))$; hence $\phi(\alpha) \leqslant \phi(\beta)$. ☐

Definition 9.28 Let $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ be proper and convex. The *recession function* of f at $y \in \mathcal{H}$ is

$$(\operatorname{rec} f)(y) = \sup_{x \in \operatorname{dom} f} \big(f(x+y) - f(x) \big). \tag{9.26}$$

The next result establishes a connection between recession functions and recession cones and provides an a posteriori motivation for the former.

Proposition 9.29 *Let $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ be proper and convex. Then* $\operatorname{epi} \operatorname{rec} f = \operatorname{rec} \operatorname{epi} f$.

Proof. Let $(y, \eta) \in \mathcal{H} \times \mathbb{R}$. Then it follows from (9.26) and (6.41) that

$$\begin{aligned} (y,\eta) \in \operatorname{epi} \operatorname{rec} f &\Leftrightarrow (\operatorname{rec} f)(y) \leqslant \eta \\ &\Leftrightarrow (\forall x \in \operatorname{dom} f) \quad f(x+y) - f(x) \leqslant \eta \\ &\Leftrightarrow (\forall (x,\xi) \in \operatorname{epi} f) \quad f(x+y) \leqslant \xi + \eta \end{aligned}$$

$$\begin{aligned}
&\Leftrightarrow (\forall (x,\xi) \in \operatorname{epi} f) \quad (x+y,\xi+\eta) \in \operatorname{epi} f \\
&\Leftrightarrow (y,\eta) \in \operatorname{rec} \operatorname{epi} f,
\end{aligned} \tag{9.27}$$

and the proof is complete. ☐

Proposition 9.30 *Let $f \in \Gamma_0(\mathcal{H})$, let $x \in \operatorname{dom} f$, and let $y \in \mathcal{H}$. Then the following hold:*

(i) $\operatorname{rec} f \in \Gamma_0(\mathcal{H})$.
(ii) $(\operatorname{rec} f)(y) = \lim_{\alpha\to+\infty}(f(x+\alpha y) - f(x))/\alpha$.
(iii) $(\operatorname{rec} f)(y) = \lim_{\alpha\to+\infty} f(x+\alpha y)/\alpha$.
(iv) $(\operatorname{rec} f)(y) = \sup_{\alpha\in\mathbb{R}_{++}}(f(x+\alpha y) - f(x))/\alpha$.
(v) *Suppose that* $\inf f(\mathcal{H}) > -\infty$. *Then* $\operatorname{rec} f \geqslant 0$.
(vi) *Let $g \in \Gamma_0(\mathcal{H})$ be such that* $\operatorname{dom} f \cap \operatorname{dom} g \neq \varnothing$. *Then* $\operatorname{rec}(f+g) = (\operatorname{rec} f) + (\operatorname{rec} g)$.
(vii) *Let $\mathcal{K}$ be a real Hilbert space, let $g \in \Gamma_0(\mathcal{K})$, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$ be such that* $\operatorname{ran} L \cap \operatorname{dom} g \neq \varnothing$. *Then* $\operatorname{rec}(g \circ L) = (\operatorname{rec} g) \circ L$.

Proof. (i): In view of (9.26), $\operatorname{rec} f$ is the supremum of a family of lower semicontinuous functions, and it is therefore likewise by Proposition 9.3. In addition, $(\operatorname{rec} f)(0) = 0$.

(ii): By Proposition 9.27, $\sigma = \lim_{\alpha\to+\infty}(f(x+\alpha y) - f(x))/\alpha$ is well defined. Using successively the lower semicontinuity of f and the convexity of f, we obtain, for every $z \in \operatorname{dom} f$,

$$\begin{aligned}
f(z+y) - f(z) &\leqslant \varliminf_{\alpha\to+\infty} f\big((1-1/\alpha)z + (1/\alpha)(x+\alpha y)\big) - f(z) \\
&\leqslant \varliminf_{\alpha\to+\infty} (1-1/\alpha)f(z) + (1/\alpha)f(x+\alpha y) - f(z) \\
&= \varliminf_{\alpha\to+\infty} \left(\frac{f(x+\alpha y) - f(x)}{\alpha} + \frac{f(x) - f(z)}{\alpha}\right) \\
&= \sigma.
\end{aligned} \tag{9.28}$$

Thus, (9.26) yields $(\operatorname{rec} f)(y) \leqslant \sigma$. Next, to show that $\sigma \leqslant (\operatorname{rec} f)(y)$, we assume that $(\operatorname{rec} f)(y) < +\infty$. Let α be a strictly positive rational number, say $\alpha = p/q$, where p and q are strictly positive integers. In view of (9.26), we have

$$f(x+py) - f(x) = \sum_{n=1}^{p} \Big(f\big(x+ny\big) - f\big(x+(n-1)y\big)\Big) \leqslant p(\operatorname{rec} f)(y). \tag{9.29}$$

Hence, by convexity of f,

$$\begin{aligned}
f(x+\alpha y) - f(x) &= f\big((1-1/q)x + (1/q)(x+py)\big) - f(x) \\
&\leqslant (1-1/q)f(x) + (1/q)f(x+py) - f(x) \\
&= \frac{f(x+py) - f(x)}{q} \\
&\leqslant \alpha(\operatorname{rec} f)(y).
\end{aligned} \tag{9.30}$$

Since f is lower semicontinuous, we deduce by density of $\mathbb{Q}$ in $\mathbb{R}$ that $(\forall \alpha \in \mathbb{R}_{++})$ $f(x+\alpha y) - f(x) \leqslant \alpha(\operatorname{rec} f)(y)$. In turn, (9.26) yields $\sigma \leqslant (\operatorname{rec} f)(y)$.

(iii): Clear in view of (ii).

(iv): This follows from (ii) and Proposition 9.27.

(v): Set $\mu = \inf f(\mathcal{H})$. Then $(\operatorname{rec} f)(y) = \lim_{\alpha\to+\infty}(f(x+\alpha y) - f(x))/\alpha = \lim_{\alpha\to+\infty} f(x+\alpha y)/\alpha \geqslant \lim_{\alpha\to+\infty} \mu/\alpha = 0$.

(vi): Since $f + g \in \Gamma_0(\mathcal{H})$, this follows from (iii).

(vii): Since $\operatorname{dom}(g \circ L) \neq \varnothing$, it follows from Lemma 1.28 and Proposition 8.20 that $g \circ L \in \Gamma_0(\mathcal{H})$. Now let $y \in \mathcal{H}$, let $x \in \operatorname{dom}(g\circ L)$, and set $z = Lx$. Then $z \in \operatorname{dom} g$ and (iii) yields $(\operatorname{rec}(g \circ L))(y) = \sup_{\alpha\in\mathbb{R}_{++}} \alpha^{-1}(g \circ L)(x+\alpha y) = \sup_{\alpha\in\mathbb{R}_{++}} \alpha^{-1} g(z+\alpha Ly) = (\operatorname{rec} g)(Ly)$. ☐

Example 9.31 Let f be a function in $\Gamma_0(\mathcal{H})$ that is positively homogeneous in the sense that $(\forall x \in \mathcal{H})(\forall \alpha \in \mathbb{R}_{++})$ $f(\alpha x) = \alpha f(x)$. Then $\operatorname{rec} f = f$.

Proof. Set $x = 0$ and let $y \in \mathcal{H}$. Since f is lower semicontinuous, $x \in \operatorname{dom} f$. Therefore, applying Proposition 9.30(iii) with $x = 0$ yields $(\operatorname{rec} f)(y) = \lim_{\alpha\to+\infty} f(\alpha y)/\alpha = f(y)$. ☐

Example 9.32 Let f be a function in $\Gamma_0(\mathcal{H})$ that is supercoercive in the sense that $\lim_{\|x\|\to+\infty} f(x)/\|x\| = +\infty$. Then $\operatorname{rec} f = \iota_{\{0\}}$.

Proof. Let $x \in \operatorname{dom} f$ and let $y \in \mathcal{H}$. If $y = 0$, then $(\operatorname{rec} f)(y) = 0$. Otherwise, by Proposition 9.30(iii),

$$(\operatorname{rec} f)(y) = \|y\| \lim_{\alpha\to+\infty} \frac{f(x+\alpha y)}{\|x+\alpha y\|}\frac{\|x+\alpha y\|}{\|\alpha y\|} = +\infty. \tag{9.31}$$

Altogether, $\operatorname{rec} f = \iota_{\{0\}}$. ☐

9.5 Construction of Functions in $\Gamma_0(\mathcal{H})$

We start with a basic tool for constructing functions in $\Gamma_0(\mathcal{H})$.

Proposition 9.33 *Let $g\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be a proper convex function such that $\operatorname{dom} g$ is open and g is continuous on $\operatorname{dom} g$. Set*

$$f\colon \mathcal{H} \to \left]-\infty,+\infty\right] : x \mapsto \begin{cases} g(x), & \text{if } x \in \operatorname{dom} g;\\ \varliminf_{y\to x} g(y), & \text{if } x \in \operatorname{bdry}\operatorname{dom} g;\\ +\infty, & \text{if } x \in \mathcal{H} \smallsetminus \overline{\operatorname{dom}}\, g. \end{cases} \tag{9.32}$$

Then $f = \breve{g}$ and $f \in \Gamma_0(\mathcal{H})$.

Proof. Set $C = \operatorname{dom} g$. To show that $f = \breve{g}$ we shall repeatedly utilize Proposition 9.8. Note that, since $g \geqslant \breve{g}$, we have $C \subset \operatorname{dom}\breve{g} \subset \overline{C}$. Let $x \in \mathcal{H}$ and assume first that $x \in C$. Then $+\infty > g(x) \geqslant \breve{g}(x)$. By Theorem 9.9, there exists a sequence $(x_n, \xi_n)_{n\in\mathbb{N}}$ in $\operatorname{epi} g$ such that $(x_n, \xi_n) \to (x, \breve{g}(x))$. Hence $\breve{g}(x) = \lim \xi_n = \underline{\lim}\, \xi_n \geqslant \underline{\lim}\, g(x_n) \geqslant \underline{\lim}\, \breve{g}(x_n) \geqslant \breve{g}(x)$ and so $f(x) = g(x) = \lim g(x_n) = \underline{\lim}\, g(x_n) = \breve{g}(x)$. Consequently, $f = \breve{g}$ on C. If $x \in \mathcal{H} \smallsetminus \overline{C}$, then $f(x) = +\infty = \breve{g}(x)$ and thus $f = \breve{g}$ on $\mathcal{H} \smallsetminus \overline{C}$. If $x \in (\operatorname{bdry} C) \smallsetminus (\operatorname{dom} \breve{g})$, then $+\infty \geqslant f(x) = \underline{\lim}_{y\to x}\, g(y) \geqslant \underline{\lim}_{y\to x}\, \breve{g}(y) = \breve{g}(x) = +\infty$ and thus $f(x) = \breve{g}(x) = +\infty$. Finally, we assume that $x \in (\operatorname{bdry} C) \cap (\operatorname{dom} \breve{g})$. Using Theorem 9.9 again, we see that there exists a sequence $(x_n, \xi_n)_{n\in\mathbb{N}}$ in $\operatorname{epi} g$ such that $(x_n, \xi_n) \to (x, \breve{g}(x))$. Hence $f(x) = \underline{\lim}_{y\to x}\, g(y) \geqslant \underline{\lim}_{y\to x}\, \breve{g}(y) = \breve{g}(x) = \lim \xi_n = \underline{\lim}\, \xi_n \geqslant \underline{\lim}\, g(x_n) \geqslant \underline{\lim}_{y\to x}\, g(y) = f(x)$ and therefore $f(x) = \breve{g}(x)$. We have verified that $f = \breve{g}$. It follows that f is lower semicontinuous and convex. Since f is real-valued on C, Proposition 9.6 implies that f is also proper. ☐

The following result concerns the construction of strictly convex functions in $\Gamma_0(\mathbb{R})$.

Proposition 9.34 *Let $g\colon \mathbb{R} \to \left]-\infty, +\infty\right]$ be strictly convex and proper, and suppose that* $\operatorname{dom} g = \left]\alpha, \beta\right[$, *where α and β are in $[-\infty, +\infty]$ and $\alpha < \beta$. Set*

$$f\colon \mathbb{R} \to \left]-\infty, +\infty\right] : x \mapsto \begin{cases} g(x), & \text{if } x \in \left]\alpha, \beta\right[;\\ \lim\limits_{y\downarrow\alpha} g(y), & \text{if } x = \alpha;\\ \lim\limits_{y\uparrow\beta} g(y), & \text{if } x = \beta;\\ +\infty, & \text{otherwise.} \end{cases} \tag{9.33}$$

Then f is strictly convex, $f = \breve{g}$, and $f \in \Gamma_0(\mathbb{R})$.

Proof. Proposition 9.14, Corollary 8.39(iii), and Proposition 9.33 imply that f is convex and that $f = \breve{g} \in \Gamma_0(\mathbb{R})$. To verify strict convexity, suppose that x and y are distinct points in $\operatorname{dom} f$, take $\gamma \in \left]0, 1\right[$, and suppose that $f(\gamma x + (1-\gamma)y) = \gamma f(x) + (1-\gamma) f(y)$. By Exercise 8.1, $(\forall \lambda \in \left]0,1\right[)\ f(\lambda x + (1-\lambda)y) = \lambda f(x) + (1-\lambda) f(y)$. Since $\left]x, y\right[\subset \left]\alpha, \beta\right[$ and $f = g$ on $\left]\alpha, \beta\right[$, this contradicts the strict convexity of g. ☐

The next two examples are consequences of Proposition 9.34 and Proposition 8.14(ii).

Example 9.35 (Entropy) The negative *Boltzmann–Shannon entropy*

$$\mathbb{R} \to \left]-\infty, +\infty\right] : x \mapsto \begin{cases} x\ln(x) - x, & \text{if } x > 0;\\ 0, & \text{if } x = 0;\\ +\infty, & \text{if } x < 0 \end{cases} \tag{9.34}$$

is strictly convex and belongs to $\Gamma_0(\mathbb{R})$.

Example 9.36 The following are strictly convex functions in $\Gamma_0(\mathbb{R})$:

(i) $x \mapsto \exp(x)$.

(ii) $x \mapsto |x|^p$, where $p \in \,]1,+\infty[$.

(iii) $x \mapsto \begin{cases} 1/x^p, & \text{if } x > 0; \\ +\infty, & \text{otherwise,} \end{cases}$ where $p \in [1,+\infty[$.

(iv) $x \mapsto \begin{cases} -x^p, & \text{if } x \geqslant 0; \\ +\infty, & \text{otherwise,} \end{cases}$ where $p \in \,]0,1[$.

(v) $x \mapsto \begin{cases} 1/\sqrt{1-x^2}, & \text{if } |x| < 1; \\ +\infty, & \text{otherwise.} \end{cases}$

(vi) $x \mapsto \begin{cases} -\sqrt{1-x^2}, & \text{if } |x| \leqslant 1; \\ +\infty, & \text{otherwise.} \end{cases}$

(vii) $x \mapsto \begin{cases} x\ln(x) + (1-x)\ln(1-x), & \text{if } x \in \,]0,1[\,; \\ 0, & \text{if } x \in \{0,1\}; \\ +\infty, & \text{otherwise} \end{cases}$

(negative *Fermi–Dirac entropy*).

(viii) $x \mapsto \begin{cases} -\ln(x), & \text{if } x > 0; \\ +\infty, & \text{otherwise} \end{cases}$ (negative *Burg entropy*).

Remark 9.37 By utilizing direct sum constructions (see Proposition 8.6 and Exercise 8.13), we can construct a (strictly) convex function in $\Gamma_0(\mathbb{R}^N)$ from (strictly) convex functions in $\Gamma_0(\mathbb{R})$.

Proposition 9.38 (Hölder's inequality) *Let $(\Omega,\mathcal{F},\mu)$ be a measure space, let $p \in \,]1,+\infty[$, and set $p^* = p/(p-1)$. Furthermore, let $x \in L^p(\Omega,\mathcal{F},\mu)$ and $y \in L^{p^*}(\Omega,\mathcal{F},\mu)$. Then*

$$\int_\Omega |x(\omega)y(\omega)|\mu(d\omega) \leqslant \left(\int_\Omega |x(\omega)|^p\mu(d\omega)\right)^{1/p} \left(\int_\Omega |x(\omega)|^{p^*}\mu(d\omega)\right)^{1/p^*}. \tag{9.35}$$

Proof. Set $\alpha = 1/p$, $\xi = (\int_\Omega |x(\omega)|^p\mu(d\omega))^{1/p}$, and $\eta = (\int_\Omega |x(\omega)|^{p^*}\mu(d\omega))^{1/p^*}$. If $0 \in \{\xi,\eta\}$, the result is clear. Suppose that $0 \notin \{\xi,\eta\}$. We derive from Example 9.36(viii) that, for every σ and every τ in $\mathbb{R}_{++}$, $\ln(\sigma^\alpha\tau^{1-\alpha}) = \alpha\ln(\sigma) + (1-\alpha)\ln(\tau) \leqslant \ln(\alpha\sigma + (1-\alpha)\tau)$. Hence $\sigma^\alpha\tau^{1-\alpha} \leqslant \alpha\sigma + (1-\alpha)\tau$, which remains true if $0 \in \{\sigma,\tau\}$. Given $\omega \in \Omega$, the above inequality applied to $\sigma = |x(\omega)|^p/\xi^p$ and $\tau = |y(\omega)|^{p^*}/\eta^{p^*}$ yields $|x(\omega)y(\omega)| \leqslant \xi\eta(\alpha|x(\omega)|^p/\xi^p + (1-\alpha)|y(\omega)|^{p^*}/\eta^{p^*})$. By integrating these inequalities over Ω with respect to μ, we obtain $\int_\Omega |x(\omega)y(\omega)|\mu(d\omega) \leqslant \xi\eta(\alpha + (1-\alpha)) = \xi\eta$. ☐

Example 9.39 (Hölder's inequality for sums) Suppose that $\mathcal{H} = \mathbb{R}^N$, set $I = \{1, \dots, N\}$, and let $p \in [1, +\infty]$. Set $p^* = p/(p-1)$ if $p \in\,]1, +\infty[$; $p^* = 1$ if $p = +\infty$; $p^* = +\infty$ if $p = 1$. Define the ℓ^p norm by

$$\big(\forall x = (\xi_i)_{i\in I} \in \mathcal{H}\big) \quad \|x\|_p = \begin{cases} \left(\displaystyle\sum_{i\in I} |\xi_i|^p\right)^{1/p}, & \text{if } p < +\infty; \\ \displaystyle\max_{i\in I} |\xi_i|, & \text{if } p = +\infty, \end{cases} \tag{9.36}$$

and let $x = (\xi_i)_{i\in I}$ and $y = (\eta_i)_{i\in I}$ be in $\mathcal{H}$. Then

$$|\langle x \mid y\rangle| \leqslant \sum_{i\in I} |\xi_i \eta_i| \leqslant \|x\|_p \, \|y\|_{p^*}. \tag{9.37}$$

Proof. The first inequality is clear, and so are the cases when $p \in \{1, +\infty\}$. On the other hand, if $p \notin \{1, +\infty\}$, then the second inequality is the special case of Proposition 9.38 in which $\Omega = I$, $\mathcal{F} = 2^\Omega$, and μ is the counting measure, i.e., for every $C \subset \Omega$, $\mu(C)$ is the cardinality of C. ☐

We now turn our attention to the construction of proper lower semicontinuous convex integral functions (see Example 2.6 for notation).

Proposition 9.40 *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, let $(\mathsf{H}, \langle \cdot \mid \cdot\rangle_{\mathsf{H}})$ be a separable real Hilbert space, and let $\varphi \in \Gamma_0(\mathsf{H})$. Suppose that $\mathcal{H} = L^2((\Omega, \mathcal{F}, \mu); \mathsf{H})$ and that one of the following holds:*

(i) $\mu(\Omega) < +\infty$.
(ii) $\varphi \geqslant \varphi(0) = 0$.

Set

$$\begin{aligned} f\colon\ \mathcal{H} &\to\,]-\infty, +\infty] \\ x &\mapsto \begin{cases} \displaystyle\int_\Omega \varphi\big(x(\omega)\big)\mu(d\omega), & \textit{if } \varphi \circ x \in L^1\big((\Omega, \mathcal{F}, \mu); \mathbb{R}\big); \\ +\infty, & \textit{otherwise.} \end{cases} \end{aligned} \tag{9.38}$$

Then $f \in \Gamma_0(\mathcal{H})$.

Proof. We first observe that, since φ is lower semicontinuous, it is measurable, and so is therefore $\varphi \circ x$ for every $x \in \mathcal{H}$. Let us now show that $f \in \Gamma_0(\mathcal{H})$.

(i): By Theorem 9.20, there exists a continuous affine function $\psi\colon \mathsf{H} \to \mathbb{R}$ such that $\varphi \geqslant \psi$, say $\psi = \langle \cdot \mid \mathsf{u}\rangle_{\mathsf{H}} + \eta$ for some $\mathsf{u} \in \mathsf{H}$ and $\eta \in \mathbb{R}$. Let us set $u\colon \Omega \to \mathsf{H}\colon \omega \mapsto \mathsf{u}$. Then $u \in \mathcal{H}$ since $\int_\Omega \|u(\omega)\|_{\mathsf{H}}^2 \mu(d\omega) = \|\mathsf{u}\|_{\mathsf{H}}^2 \mu(\Omega) < +\infty$. Moreover, for every $x \in \mathcal{H}$, $\varphi \circ x \geqslant \psi \circ x$ and

$$\int_\Omega \psi\big(x(\omega)\big)\mu(d\omega) = \int_\Omega \langle x(\omega) \mid \mathsf{u}\rangle_{\mathsf{H}}\, \mu(d\omega) + \eta\mu(\Omega) = \langle x \mid u\rangle + \eta\mu(\Omega) \in \mathbb{R}. \tag{9.39}$$

Thus, Proposition 8.24 asserts that f is well defined and convex, with $\operatorname{dom} f = \big\{x \in \mathcal{H} \;\big|\; \varphi \circ x \in L^1((\Omega, \mathcal{F}, \mu); \mathbb{R})\big\}$. It also follows from (9.38) and (9.39) that

$$(\forall x \in \operatorname{dom} f) \quad f(x) = \int_\Omega (\varphi - \psi)\big(x(\omega)\big)\mu(d\omega) + \langle x \mid u\rangle + \eta\mu(\Omega). \tag{9.40}$$

Now take $\mathsf{z} \in \operatorname{dom} \varphi$ and set $z \colon \Omega \to \mathsf{H} \colon \omega \mapsto \mathsf{z}$. Then $z \in \mathcal{H}$ and $\int_\Omega |\varphi \circ z| d\mu = |\varphi(\mathsf{z})|\mu(\Omega) < +\infty$. Hence, $\varphi \circ z \in L^1((\Omega, \mathcal{F}, \mu); \mathbb{R})$. This shows that f is proper. Next, to show that f is lower semicontinuous, let us fix $\xi \in \mathbb{R}$ and a sequence $(x_n)_{n\in\mathbb{N}}$ in $\operatorname{lev}_{\leqslant\xi} f$ that converges to some $x \in \mathcal{H}$. In view of Lemma 1.24, it suffices to show that $f(x) \leqslant \xi$. Since $\|x_n(\cdot) - x(\cdot)\|_{\mathsf{H}} \to 0$ in $L^2((\Omega, \mathcal{F}, \mu); \mathbb{R})$, there exists a subsequence $(x_{k_n})_{n\in\mathbb{N}}$ such that $x_{k_n}(\omega) \overset{\mathsf{H}}{\to} x(\omega)$ for μ-almost every $\omega \in \Omega$ [6, Theorem 2.5.1 & Theorem 2.5.3]. Now set $\phi = (\varphi - \psi) \circ x$ and $(\forall n \in \mathbb{N})$ $\phi_n = (\varphi - \psi) \circ x_{k_n}$. Since $\varphi - \psi$ is lower semicontinuous, we have

$$\phi(\omega) = (\varphi - \psi)\big(x(\omega)\big) \leqslant \underline{\lim}(\varphi - \psi)\big(x_{k_n}(\omega)\big) = \underline{\lim}\, \phi_n(\omega) \quad \mu\text{-a.e. on } \Omega. \tag{9.41}$$

On the other hand, since $\inf_{n\in\mathbb{N}} \phi_n \geqslant 0$, Fatou's lemma [6, Lemma 1.6.8] yields $\int_\Omega \underline{\lim}\, \phi_n d\mu \leqslant \underline{\lim} \int_\Omega \phi_n d\mu$. Hence, we derive from (9.40) and (9.41) that

$$\begin{aligned} f(x) &= \int_\Omega \phi d\mu + \langle x \mid u\rangle + \eta\mu(\Omega) \\ &\leqslant \int_\Omega \underline{\lim}\, \phi_n d\mu + \langle x \mid u\rangle + \eta\mu(\Omega) \\ &\leqslant \underline{\lim} \int_\Omega \phi_n d\mu + \lim \langle x_{k_n} \mid u\rangle + \eta\mu(\Omega) \\ &= \underline{\lim} \int_\Omega (\varphi \circ x_{k_n}) d\mu \\ &= \underline{\lim}\, f(x_{k_n}) \\ &\leqslant \xi. \end{aligned} \tag{9.42}$$

(ii): Since (8.16) holds with $\varrho = 0$, it follows from Proposition 8.24 that f is a well-defined convex function. In addition, since $\varphi(0) = 0$, (9.38) yields $f(0) = 0$. Thus, f is proper. Finally, to prove that f is lower semicontinuous, we follow the same procedure as above with $\psi = 0$. ☐

Example 9.41 (Boltzmann–Shannon entropy) Let $(\Omega, \mathcal{F}, \mu)$ be a finite measure space and suppose that $\mathcal{H} = L^2(\Omega, \mathcal{F}, \mu)$ (see Example 2.7). Using the convention $0 \ln(0) = 0$, set

$$f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$$
$$x \mapsto \begin{cases} \displaystyle\int_\Omega \big(x(\omega)\ln(x(\omega)) - x(\omega)\big)\mu(d\omega), & \text{if } x \geqslant 0 \ \ \mu\text{-a.e. on } \Omega; \\ +\infty, & \text{otherwise.} \end{cases} \tag{9.43}$$

Then $f \in \Gamma_0(\mathcal{H})$. In particular, this is true in the following cases:

(i) Entropy of a random variable: $\mathcal{H} = L^2(\Omega, \mathcal{F}, \mathsf{P})$, where $(\Omega, \mathcal{F}, \mathsf{P})$ is a probability space (see Example 2.9), and

$$f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$$
$$X \mapsto \begin{cases} \mathsf{E}\big(X\ln(X) - X\big), & \text{if } X \geqslant 0 \text{ a.s.}; \\ +\infty, & \text{otherwise.} \end{cases} \tag{9.44}$$

(ii) Discrete entropy: $\mathcal{H} = \mathbb{R}^N$ and

$$f\colon \qquad \mathcal{H} \qquad \to \left]-\infty,+\infty\right]$$
$$(\xi_k)_{1\leqslant k\leqslant N} \mapsto \begin{cases} \displaystyle\sum_{k=1}^N \xi_k\ln(\xi_k) - \xi_k, & \text{if } \displaystyle\min_{1\leqslant k\leqslant N} \xi_k \geqslant 0; \\ +\infty, & \text{otherwise.} \end{cases} \tag{9.45}$$

Proof. Denote by φ the function defined in (9.34). Then Example 9.35 asserts that $\varphi \in \Gamma_0(\mathbb{R})$. First, take $x \in \mathcal{H}$ such that $x \geqslant 0$ μ-a.e., and set $C = \big\{\omega \in \Omega \bigm| 0 \leqslant x(\omega) < 1\big\}$ and $D = \big\{\omega \in \Omega \bigm| x(\omega) \geqslant 1\big\}$. Since, for every $\xi \in \mathbb{R}_+$, $|\varphi(\xi)| = |\xi\ln(\xi) - \xi| \leqslant 1_{[0,1[}(\xi) + \xi^2 1_{[1,+\infty[}(\xi)$, we have

$$\begin{aligned} \int_\Omega \big|\varphi\big(x(\omega)\big)\big|\mu(d\omega) &= \int_C \big|\varphi\big(x(\omega)\big)\big|\mu(d\omega) + \int_D \big|\varphi\big(x(\omega)\big)\big|\mu(d\omega) \\ &\leqslant \mu(C) + \int_D |x(\omega)|^2\mu(d\omega) \\ &\leqslant \mu(\Omega) + \|x\|^2 \\ &< +\infty, \end{aligned} \tag{9.46}$$

and therefore $\varphi \circ x \in L^1\big((\Omega, \mathcal{F}, \mu); \mathbb{R}\big)$. Now take $x \in \mathcal{H}$ and set $A = \big\{\omega \in \Omega \bigm| x(\omega) \geqslant 0\big\}$ and $B = \big\{\omega \in \Omega \bigm| x(\omega) < 0\big\}$. Then

$$\begin{aligned} \int_\Omega \varphi(x(\omega))\mu(d\omega) &= \int_A \varphi(x(\omega))\mu(d\omega) + \int_B \varphi(x(\omega))\mu(d\omega) \\ &= \begin{cases} \displaystyle\int_\Omega x(\omega)\big(\ln(x(\omega)) - 1\big)\mu(d\omega), & \text{if } x \geqslant 0 \ \ \mu\text{-a.e. on } \Omega; \\ +\infty, & \text{otherwise} \end{cases} \\ &= f(x). \end{aligned} \tag{9.47}$$

Altogether, it follows from Proposition 9.40(i) with $\mathsf{H} = \mathbb{R}$ that $f \in \Gamma_0(\mathcal{H})$.

(i): This is the special case when μ is a probability measure.

(ii): Special case when $\Omega = \{1, \dots, N\}$, $\mathcal{F} = 2^\Omega$, and μ is the counting measure. ☐

Next, we consider the construction of the lower semicontinuous convex envelope of a perspective function.

Proposition 9.42 *Let* $\varphi \in \Gamma_0(\mathcal{H})$, *let*

$$f\colon \mathbb{R} \times \mathcal{H} \to \left]-\infty, +\infty\right] : (\xi, x) \mapsto \begin{cases} \xi\varphi(x/\xi), & \text{if } \xi > 0;\\ +\infty, & \text{otherwise} \end{cases} \tag{9.48}$$

be its perspective function, and set

$$g\colon \mathbb{R} \times \mathcal{H} \to \left]-\infty, +\infty\right] : (\xi, x) \mapsto \begin{cases} \xi\varphi(x/\xi), & \text{if } \xi > 0;\\ (\operatorname{rec}\varphi)(x), & \text{if } \xi = 0;\\ +\infty, & \text{if } \xi < 0. \end{cases} \tag{9.49}$$

Then $\breve{f} = g \in \Gamma_0(\mathbb{R} \oplus \mathcal{H})$.

Proof. Set $C = \{1\} \times \operatorname{epi}\varphi$. By Proposition 8.25, $\operatorname{epi} f = \operatorname{cone} C$. Because $(0,0) \notin C$, it follows from Theorem 9.9, Corollary 6.53, Proposition 9.29, and Lemma 1.6(ii) that $\operatorname{epi}\breve{f} = \overline{\operatorname{epi}}\, f = \overline{\operatorname{cone}}\, C = (\operatorname{cone} C) \cup (\operatorname{rec} C) = (\operatorname{epi} f) \cup (\{0\} \times \operatorname{rec}\operatorname{epi}\varphi) = (\operatorname{epi} f) \cup (\{0\} \times \operatorname{epi}\operatorname{rec}\varphi) = (\operatorname{epi} f) \cup \operatorname{epi}(\iota_{\{0\}} \oplus \operatorname{rec}\varphi) = \operatorname{epi}\min\{f, \iota_{\{0\}} \oplus \operatorname{rec}\varphi\} = \operatorname{epi} g$. ☐

Example 9.43 Suppose that $\mathcal{H} \neq \{0\}$ and let $p \in \left]1, +\infty\right[$. Then the function

$$g\colon \mathbb{R} \times \mathcal{H} \to \left]-\infty, +\infty\right] : (\xi, x) \mapsto \begin{cases} \|x\|^p/\xi^{p-1}, & \text{if } \xi > 0;\\ 0, & \text{if } (\xi, x) = (0,0);\\ +\infty, & \text{otherwise} \end{cases} \tag{9.50}$$

belongs to $\Gamma_0(\mathbb{R} \oplus \mathcal{H})$ and $g|_{\operatorname{dom} g}$ is not continuous at $(0,0)$.

Proof. Set $\varphi = \|\cdot\|^p$. Then it follows from Example 8.23 and Example 9.32 that $\operatorname{rec}\varphi = \iota_{\{0\}}$. Hence (9.50) coincides with (9.49) and the first claim is therefore an application of Proposition 9.42. Now set $x = (0,0)$, let $u \in \mathcal{H}$ be such that $\|u\| = 1$, fix a sequence $(\alpha_n)_{n\in\mathbb{N}}$ in $\mathbb{R}_{++}$ such that $\alpha_n \downarrow 0$, and set $(\forall n \in \mathbb{N})\ x_n = (\alpha_n^{p/(p-1)}, \alpha_n u)$. Then $(x_n)_{n\in\mathbb{N}}$ lies in $\operatorname{dom} g$ and $x_n \to x$, but $\lim g(x_n) = 1 \neq 0 = g(x)$. ☐

Corollary 9.44 *Let* $\mathcal{K}$ *be a real Hilbert space, let* $\varphi \in \Gamma_0(\mathcal{K})$, *let* $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, *let* $r \in \mathcal{K}$, *let* $u \in \mathcal{H}$, *let* $\rho \in \mathbb{R}$, *and set*

$$f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$$
$$x \mapsto \begin{cases} (\langle x \mid u\rangle - \rho)\varphi\bigg(\dfrac{Lx-r}{\langle x \mid u\rangle - \rho}\bigg), & \text{if } \langle x \mid u\rangle > \rho;\\ (\operatorname{rec}\varphi)\big(Lx-r\big), & \text{if } \langle x \mid u\rangle = \rho;\\ +\infty, & \text{if } \langle x \mid u\rangle < \rho. \end{cases} \tag{9.51}$$

Suppose that there exists $z \in \mathcal{H}$ such that $Lz \in r + (\langle z \mid u\rangle - \rho)\operatorname{dom}\varphi$ and $\langle z \mid u\rangle \geqslant \rho$. Then $f \in \Gamma_0(\mathcal{H})$.

Proof. Let g be the lower semicontinuous convex envelope of the perspective function of φ constructed as in (9.49), and let $T\colon \mathcal{H} \to \mathbb{R}\oplus\mathcal{K}\colon x \mapsto (\langle x \mid u\rangle - \rho, Lx - r)$. Then T is a continuous affine operator, while $g \in \Gamma_0(\mathbb{R}\oplus\mathcal{K})$ by Proposition 9.42. Therefore, it follows from Lemma 1.28 and Proposition 8.20 that $f = g\circ T$ is lower semicontinuous and convex. Finally, $z \in \operatorname{dom} f$ since $(\operatorname{rec}\varphi)(0) = 0$. ☐

Example 9.45 Let $\mathcal{K}$ be a real Hilbert space, let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$, let $r \in \mathcal{K}$, let $u \in \mathcal{H}$, let $\rho \in \mathbb{R}$, let $p \in \left]1,+\infty\right[$, and set

$$f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$$
$$x \mapsto \begin{cases} \dfrac{\|Lx-r\|^p}{(\langle x \mid u\rangle - \rho)^{p-1}}, & \text{if } \langle x \mid u\rangle > \rho;\\ 0 & \text{if } \langle x \mid u\rangle = \rho \text{ and } Lx = r;\\ +\infty, & \text{otherwise.} \end{cases} \tag{9.52}$$

Then $f \in \Gamma_0(\mathcal{H})$.

Proof. It follows from Example 9.32 that this is an application of Corollary 9.44 with $\varphi = \|\cdot\|^p$. ☐

We provide a lower semicontinuous version of the Csiszár ϕ-divergence of Example 8.26.

Example 9.46 Suppose that $\mathcal{H} = \mathbb{R}^N$, let $\phi \in \Gamma_0(\mathbb{R})$, and set $I = \{1,\ldots,N\}$. For every $x = (\xi_i)_{i\in I} \in \mathcal{H}$ and $y = (\eta_i)_{i\in I} \in \mathcal{H}$, set $I_-(y) = \big\{i \in I \;\big|\; \eta_i < 0\big\}$, $I_0(y) = \big\{i \in I \;\big|\; \eta_i = 0\big\}$, $I_+(y) = \big\{i \in I \;\big|\; \eta_i > 0\big\}$, and

$$d_\phi(x,y) = \begin{cases} \displaystyle\sum_{i\in I_0(y)} (\operatorname{rec}\phi)(\xi_i) + \sum_{i\in I_+(y)} \eta_i\phi(\xi_i/\eta_i), & \text{if } I_-(y) = \varnothing;\\ +\infty, & \text{if } I_-(y) \neq \varnothing. \end{cases} \tag{9.53}$$

Then $d_\phi \in \Gamma_0(\mathcal{H}\oplus\mathcal{H})$.

Proof. Set

$$\psi\colon \mathbb{R}\times\mathbb{R}\to\,]-\infty,+\infty] : (\eta,\xi)\mapsto \begin{cases} \eta\phi(\xi/\eta), & \text{if } \eta>0;\\ (\operatorname{rec}\phi)(\xi), & \text{if } \eta=0;\\ +\infty, & \text{if } \eta<0. \end{cases} \tag{9.54}$$

Then Proposition 9.42 asserts that $\psi\in\Gamma_0(\mathbb{R}^2)$. Therefore, $d_\phi=\bigoplus_{i\in I}\psi\in\Gamma_0(\mathcal{H}\oplus\mathcal{H})$. ☐

Remark 9.47 In Example 9.46 suppose additionally that $\phi(t)/t\to+\infty$ when $|t|\to+\infty$, and set $I_0(x,y)=\big\{i\in I \;\big|\; \xi_i\neq 0 \text{ and } \eta_i=0\big\}$. Then it follows from Example 9.32 that (9.53) reduces to

$$d_\phi(x,y)=\begin{cases} \displaystyle\sum_{i\in I_+(y)}\eta_i\phi(\xi_i/\eta_i), & \text{if } I_-(y)\cup I_0(x,y)=\varnothing;\\ +\infty, & \text{if } I_-(y)\cup I_0(x,y)\neq\varnothing. \end{cases} \tag{9.55}$$

In particular, this formula provides lower semicontinuous versions of the Csiszár ϕ-divergences described in Examples 8.27–8.31.

Example 9.48 Suppose that $\mathcal{H}=\mathbb{R}^N$, and set $I=\{1,\ldots,N\}$. For every $x=(\xi_i)_{i\in I}\in\mathcal{H}$ and every $y=(\eta_i)_{i\in I}\in\mathcal{H}$, set $I_-(y)=\big\{i\in I \;\big|\; \eta_i<0\big\}$, $J(x,y)=\big\{i\in I \;\big|\; (\xi_i\neq 0 \text{ and } \eta_i=0) \text{ or } (\xi_i<0 \text{ and } \eta_i>0)\big\}$, $I_+(y)=\big\{i\in I \;\big|\; \eta_i>0\big\}$, and

$$f(x,y)=\begin{cases} \displaystyle\sum_{i\in I_+(x)\cap I_+(y)}\xi_i\ln(\xi_i/\eta_i), & \text{if } I_-(y)\cup J(x,y)=\varnothing;\\ +\infty, & \text{otherwise}. \end{cases} \tag{9.56}$$

Then f is the lower semicontinuous hull of the Kullback-Leibler divergence of Example 8.27.

Exercises

Exercise 9.1 Let $f\colon\mathcal{H}\to\,]-\infty,+\infty]$ be lower semicontinuous and *midpoint convex* in the sense that

$$(\forall x\in\operatorname{dom}f)(\forall y\in\operatorname{dom}f)\quad f\left(\frac{x+y}{2}\right)\leqslant\frac{f(x)+f(y)}{2}. \tag{9.57}$$

Show that f is convex.

Exercise 9.2 Let $f\colon\mathcal{H}\to\,]-\infty,+\infty]$ be midpoint convex. Show that f need not be convex.

Exercise 9.3 Provide a family of continuous linear functions the supremum of which is neither continuous nor linear.

Exercise 9.4 Let $f \in \Gamma_0(\mathcal{H})$. Show that $\mathbb{R} \cap \operatorname{ran} f$ is convex, and provide an example in which $\operatorname{ran} f$ is not convex.

Exercise 9.5 Provide an example of a convex function $f : \mathcal{H} \to [-\infty, +\infty]$ such that $\operatorname{ran} f = \{-\infty, 0, +\infty\}$. Compare with Proposition 9.6.

Exercise 9.6 Set $\mathcal{C} = \big\{ C \subset \mathcal{H} \bigm| C \text{ is nonempty, closed, and convex} \big\}$ and set

$$(\forall C \in \mathcal{C}) \quad \Upsilon_C : \mathcal{H} \to [-\infty, +\infty] : x \mapsto \begin{cases} -\infty, & \text{if } x \in C; \\ +\infty, & \text{otherwise.} \end{cases} \tag{9.58}$$

Prove that $\mathcal{C} \to \big\{ f \in \Gamma(\mathcal{H}) \bigm| -\infty \in f(\mathcal{H}) \big\} : C \mapsto \Upsilon_C$ is a bijection.

Exercise 9.7 Let $f : \mathcal{H} \to \left]-\infty, +\infty\right]$ be convex. Show that f is continuous if and only if it is lower semicontinuous and $\operatorname{cont} f = \operatorname{dom} f$.

Exercise 9.8 Let $f : \mathcal{H} \to \left]-\infty, +\infty\right]$ be convex and set $\mu = \inf f(\mathcal{H})$. Prove the following statements:

(i) $f \in \Gamma(\mathcal{H}) \Leftrightarrow (\forall \xi \in \left]\mu, +\infty\right[)\ \operatorname{lev}_{\leqslant \xi} f = \overline{\operatorname{lev}_{< \xi} f}$.
(ii) $\operatorname{cont} f = \operatorname{dom} f \Leftrightarrow (\forall \xi \in \left]\mu, +\infty\right[)\ \operatorname{lev}_{< \xi} f = \operatorname{int} \operatorname{lev}_{\leqslant \xi} f$.
(iii) f is continuous $\Leftrightarrow (\forall \xi \in \left]\mu, +\infty\right[)\ \operatorname{lev}_{= \xi} f = \operatorname{bdry} \operatorname{lev}_{\leqslant \xi} f$.

Exercise 9.9 Let $(e_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$, let $(\omega_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}_+$, and let $(p_n)_{n\in\mathbb{N}}$ be a sequence in $\left[1, +\infty\right[$. Set $f : \mathcal{H} \to \left]-\infty, +\infty\right] : x \mapsto \sum_{n\in\mathbb{N}} \omega_n \, |\langle x \mid e_n \rangle|^{p_n}$. Show that $f \in \Gamma_0(\mathcal{H})$.

Exercise 9.10 Use Proposition 8.14(ii) and Proposition 9.34 to verify Example 9.35 and Example 9.36.

Exercise 9.11 Let $(\mathcal{H}_i)_{i\in I}$ be a finite family of real Hilbert spaces, set $\mathcal{H} = \bigoplus_{i\in I} \mathcal{H}_i$, and let, for every i in I, $f_i : \mathcal{H}_i \to \left]-\infty, +\infty\right]$ be proper and convex. Show that $\operatorname{rec} \bigoplus_{i\in I} f_i = \bigoplus_{i\in I} \operatorname{rec} f_i$.

Exercise 9.12 Use Proposition 9.33 to show that the function

$$f : \mathbb{R}^2 \to \left]-\infty, +\infty\right] : (\xi, \eta) \mapsto \begin{cases} \eta^2/\xi, & \text{if } \xi > 0; \\ 0, & \text{if } (\xi, \eta) = (0, 0); \\ +\infty, & \text{otherwise} \end{cases} \tag{9.59}$$

belongs to $\Gamma_0(\mathbb{R}^2)$.

Exercise 9.13 Let N be a strictly positive integer and set $I = \{1, \dots, N\}$. Show that the function

$$f\colon \mathbb{R}^N \to \left]-\infty, +\infty\right] : (\xi_i)_{i\in I} \mapsto \begin{cases} \dfrac{\sum_{i\in I} \xi_i^2}{\sum_{i\in I} \xi_i}, & \text{if } \sum_{i\in I} \xi_i > 0; \\ 0, & \text{if } (\forall i \in I)\ \xi_i = 0; \\ +\infty, & \text{otherwise} \end{cases} \tag{9.60}$$

is in $\Gamma_0(\mathbb{R}^N)$.

Chapter 10
Convex Functions: Variants

In this chapter we present variants of the notion of convexity for functions. The most important are the weaker notion of quasiconvexity and the stronger notions of uniform and strong convexity.

10.1 Between Linearity and Convexity

Definition 10.1 A function $f\colon \mathcal{H} \to [-\infty, +\infty]$ is

(i) *positively homogeneous* if $(\forall x \in \mathcal{H})(\forall \lambda \in \mathbb{R}_{++})\ f(\lambda x) = \lambda f(x)$;
(ii) *subadditive* if $(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)\ f(x+y) \leqslant f(x) + f(y)$;
(iii) *sublinear* if it is positively homogeneous and subadditive.

The proofs of the following results are left as Exercise 10.1 and Exercise 10.2, respectively.

Proposition 10.2 *Let* $f\colon \mathcal{H} \to [-\infty, +\infty]$. *Then* f *is positively homogeneous if and only if* $\operatorname{epi} f$ *is a cone; in this case,* $\operatorname{dom} f$ *is also a cone.*

Proposition 10.3 *Let* $f\colon \mathcal{H} \to [-\infty, +\infty]$ *be positively homogeneous. Then* f *is sublinear if and only if it is convex.*

Clearly, linearity implies sublinearity, which in turn implies convexity. However, as we now illustrate, neither implication is reversible.

Example 10.4 Suppose that $\mathcal{H} \neq \{0\}$. Then the following hold:

(i) $\|\cdot\|$ is sublinear, but not linear.
(ii) $\|\cdot\|^2$ is convex, but not sublinear.

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-48311-5_31

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_10

Example 10.5 Let $\varphi\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be convex and let

$$f\colon \mathbb{R}\times\mathcal{H} \to \left]-\infty,+\infty\right] : (\xi,x) \mapsto \begin{cases} \xi\varphi(x/\xi), & \text{if } \xi > 0;\\ +\infty, & \text{otherwise} \end{cases} \tag{10.1}$$

be its perspective function. Then f is sublinear.

Proof. It is clear from (10.1) that f is positively homogeneous. Hence, the result follows from Proposition 8.25 and Proposition 10.3. ☐

Example 10.6 Let $f \in \Gamma_0(\mathcal{H})$. Then $\operatorname{rec} f$ is sublinear.

Proof. Let $y \in \mathcal{H}$ and let $\lambda \in \mathbb{R}_{++}$. Then Proposition 9.30(ii) yields $(\operatorname{rec} f)(\lambda y) = \lim_{\alpha\to+\infty}(f(x+\alpha\lambda y) - f(x))/\alpha = \lambda \lim_{\alpha\to+\infty}(f(x+\alpha\lambda y) - f(x))/(\alpha\lambda) = \lambda(\operatorname{rec} f)(y)$. Hence $\operatorname{rec} f$ is positively homogeneous (alternatively, combine Proposition 9.29 and Proposition 10.2). Sublinearity follows from Proposition 9.30(i) and Proposition 10.3. ☐

10.2 Uniform and Strong Convexity

We now introduce more restrictive versions of strict convexity.

Definition 10.7 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper, and let $\phi\colon \mathbb{R}_+ \to [0,+\infty]$ be an increasing function that vanishes only at 0. Then f is *uniformly convex* with modulus ϕ if

$$\begin{aligned}&(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)(\forall \alpha \in \left]0,1\right[)\\ &\quad f\big(\alpha x + (1-\alpha)y\big) + \alpha(1-\alpha)\phi\big(\|x-y\|\big) \leqslant \alpha f(x) + (1-\alpha)f(y).\end{aligned} \tag{10.2}$$

If (10.2) holds with $\phi = (\beta/2)|\cdot|^2$ for some $\beta \in \mathbb{R}_{++}$, then f is *strongly convex* with constant β. Now let C be a nonempty subset of $\operatorname{dom} f$. Then f is uniformly convex on C with modulus ϕ if

$$\begin{aligned}&(\forall x \in C)(\forall y \in C)(\forall \alpha \in \left]0,1\right[)\\ &\quad f\big(\alpha x + (1-\alpha)y\big) + \alpha(1-\alpha)\phi\big(\|x-y\|\big) \leqslant \alpha f(x) + (1-\alpha)f(y).\end{aligned} \tag{10.3}$$

Clearly, strong convexity implies uniformly convexity, uniformly convexity implies strict convexity, and strict convexity implies convexity. In the exercises, we provide examples that show that none of these implications is reversible.

Proposition 10.8 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and let $\beta \in \mathbb{R}_{++}$. Then f is strongly convex with constant β if and only if $f - (\beta/2)\|\cdot\|^2$ is convex.*

Proof. A direct consequence of Corollary 2.15. ☐

Example 10.9 Suppose that $\mathcal{H} \neq \{0\}$. Then the following hold:

(i) $\|\cdot\|$ is sublinear, but not strictly (hence not uniformly) convex.
(ii) $\|\cdot\|^2$ is strongly convex with constant 2, but not positively homogeneous.

Example 10.10 Let $\phi\colon \mathbb{R}_+ \to \mathbb{R}_+$ be increasing and let C be a nonempty bounded convex subset of $\mathcal{H}$. Set

$$f\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \int_0^{\|x\|} \phi(t)\,dt. \tag{10.4}$$

Then f is uniformly convex on C [362, Theorem 4.1(ii)]. In particular, taking $p \in \left]1,+\infty\right[$ and $\phi\colon t \mapsto pt^{p-1}$, we obtain that $\|\cdot\|^p + \iota_C$ is uniformly convex.

Definition 10.11 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex. The *exact modulus of convexity* of f is

$$\begin{aligned} \varphi\colon\ \mathbb{R}_+ &\to [0,+\infty] \\ t &\mapsto \inf_{\substack{x\in\operatorname{dom} f,\, y\in\operatorname{dom} f,\\ \|x-y\|=t,\, \alpha\in]0,1[}} \left(\frac{\alpha f(x)+(1-\alpha)f(y)-f(\alpha x+(1-\alpha)y)}{\alpha(1-\alpha)}\right). \end{aligned} \tag{10.5}$$

Proposition 10.12 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex, with exact modulus of convexity φ. Then $\varphi(0)=0$,*

$$(\forall t \in \mathbb{R}_+)(\forall \gamma \in \left[1,+\infty\right[) \quad \varphi(\gamma t) \geqslant \gamma^2 \varphi(t), \tag{10.6}$$

and φ is increasing.

Proof. We assume that $\mathcal{H} \neq \{0\}$. It is clear that $\varphi(\mathbb{R}_+) \subset [0,+\infty]$, that $\varphi(0)=0$, that (10.6) holds when $t=0$ or $\gamma=1$, and that (10.6) implies that φ is increasing. To show (10.6), we fix $t \in \mathbb{R}_{++}$ and $\gamma \in \left]1,+\infty\right[$ such that $\varphi(\gamma t) < +\infty$, and we verify that

$$\varphi(\gamma t) \geqslant \gamma^2 \varphi(t). \tag{10.7}$$

We consider two cases.

(a) $\gamma \in \left]1,2\right[$: Fix $\varepsilon \in \mathbb{R}_{++}$. Since $\varphi(\gamma t) < +\infty$, there exist $x \in \operatorname{dom} f$, $y \in \operatorname{dom} f$, and $\alpha \in \left]0,1/2\right]$ such that $\|x-y\| = \gamma t$ and

$$\varphi(\gamma t) + \varepsilon > \frac{\alpha f(x)+(1-\alpha)f(y)-f(\alpha x+(1-\alpha)y)}{\alpha(1-\alpha)}. \tag{10.8}$$

Now set $z_\alpha = \alpha x + (1-\alpha)y$, $\delta = 1/\gamma$, and $z_\delta = \delta x + (1-\delta)y$. Then $\|z_\delta - y\| = t$, $\gamma\alpha \in \left]0,1\right[$, and $z_\alpha = \gamma\alpha z_\delta + (1-\gamma\alpha)y$. We derive from (10.5) that

$$f(z_\delta) \leqslant \delta f(x) + (1-\delta)f(y) - \delta(1-\delta)\varphi(\gamma t) \tag{10.9}$$

and that

$$f(z_\alpha) \leqslant \gamma\alpha f(z_\delta) + (1-\gamma\alpha)f(y) - \gamma\alpha(1-\gamma\alpha)\varphi(t). \tag{10.10}$$

Furthermore, (10.8) is equivalent to

$$f(z_\alpha) > \alpha f(x) + (1-\alpha)f(y) - \alpha(1-\alpha)\varphi(\gamma t) - \varepsilon\alpha(1-\alpha). \tag{10.11}$$

Combining (10.9)–(10.11), we deduce that

$$\gamma^2\varphi(t) < \varphi(\gamma t) + \frac{\varepsilon\gamma(1-\alpha)}{1-\gamma\alpha}. \tag{10.12}$$

However, since $\alpha \in \left]0,1/2\right]$ and $\gamma \in \left]1,2\right[$, we have $(1-\alpha)/(1-\gamma\alpha) \leqslant 1/(2-\gamma)$. Thus, $\gamma^2\varphi(t) < \varphi(\gamma t) + \varepsilon\gamma/(2-\gamma)$ and, letting $\varepsilon \downarrow 0$, we obtain (10.7).

(b) $\gamma \in \left[2,+\infty\right[$: We write γ as a product of finitely many factors in $\left]1,2\right[$ and invoke (a) repeatedly to obtain (10.7). ☐

Corollary 10.13 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex, with exact modulus of convexity φ. Then f is uniformly convex if and only if φ vanishes only at 0; in this case, f is uniformly convex with modulus φ.*

Proof. Suppose that f is uniformly convex with modulus ϕ. Then

$$(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)(\forall \alpha \in \left]0,1\right[)$$
$$\phi\big(\|x-y\|\big) \leqslant \frac{\alpha f(x) + (1-\alpha)f(y) - f\big(\alpha x + (1-\alpha)y\big)}{\alpha(1-\alpha)} \tag{10.13}$$

and hence $\phi \leqslant \varphi$. Since ϕ vanishes only at 0, so does φ, since Proposition 10.12 asserts that $\varphi(0) = 0$. Conversely, if φ vanishes only at 0, then Proposition 10.12 implies that f is uniformly convex with modulus φ. ☐

Proposition 10.14 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex, with exact modulus of convexity φ, and set*

$$\psi\colon \mathbb{R}_+ \to [0,+\infty]$$
$$t \mapsto \inf_{\substack{x\in\operatorname{dom} f,\, y\in\operatorname{dom} f,\\ \|x-y\|=t}} \left(\tfrac{1}{2}f(x) + \tfrac{1}{2}f(y) - f\Big(\tfrac{1}{2}x + \tfrac{1}{2}y\Big)\right). \tag{10.14}$$

Then the following hold:

(i) $2\psi \leqslant \varphi \leqslant 4\psi$.
(ii) f *is uniformly convex if and only if* ψ *vanishes only at* 0.

Proof. (i): Let $t \in \mathbb{R}_+$. Since $\varphi(0) = \psi(0) = 0$, we assume that $\operatorname{dom} f$ is not a singleton and that $t > 0$. Take $\alpha \in \left]0, 1/2\right]$ and two points x_0 and y_0 in $\operatorname{dom} f$ such that $\|x_0 - y_0\| = t$ (if no such points exist then $\varphi(t) = \psi(t) = +\infty$ and hence (i) holds at t). By convexity and (10.14),

$$\begin{aligned} f\big(\alpha x_0 + (1-\alpha)y_0\big) &= f\Big(2\alpha\big(\tfrac{1}{2}x_0 + \tfrac{1}{2}y_0\big) + (1-2\alpha)y_0\Big) \\ &\leqslant 2\alpha f\big(\tfrac{1}{2}x_0 + \tfrac{1}{2}y_0\big) + (1-2\alpha)f(y_0) \\ &\leqslant 2\alpha\big(\tfrac{1}{2}f(x_0) + \tfrac{1}{2}f(y_0) - \psi(\|x_0 - y_0\|)\big) + (1-2\alpha)f(y_0) \\ &\leqslant \alpha f(x_0) + (1-\alpha)f(y_0) - 2\alpha(1-\alpha)\psi(t). \end{aligned} \tag{10.15}$$

Hence, Definition 10.11 yields $2\psi(t) \leqslant \varphi(t)$. On the other hand,

$$\varphi(t) \leqslant \inf_{\substack{x \in \operatorname{dom} f,\, y \in \operatorname{dom} f,\\ \|x-y\|=t}} \left(\frac{\tfrac{1}{2}f(x) + \tfrac{1}{2}f(y) - f\big(\tfrac{1}{2}x + \tfrac{1}{2}y\big)}{\tfrac{1}{4}}\right) = 4\psi(t). \tag{10.16}$$

Altogether, (i) is verified.

(ii): In view of (i), φ and ψ vanish at the same points. Hence (ii) follows from Corollary 10.13. ☐

Proposition 10.15 *Let* $g\colon \mathcal{H} \to \mathbb{R}_+$ *be uniformly convex, with exact modulus of convexity* φ*, and let* $p \in \left[1, +\infty\right[$*. Then* g^p *is uniformly convex, and its exact modulus of convexity* χ *satisfies*

$$\chi \geqslant 2^{1-2p} \min\big\{p2^{1-p}, 1 - 2^{-p}\big\}\varphi^p. \tag{10.17}$$

Proof. Let $t \in \mathbb{R}_+$, let $x \in \operatorname{dom} g^p = \operatorname{dom} g = \mathcal{H}$, and let $y \in \mathcal{H}$. We assume that $t > 0$ and that $\|x - y\| = t$. Now set $\alpha = \frac{1}{2}g(x) + \frac{1}{2}g(y)$, $\beta = g(\frac{1}{2}x + \frac{1}{2}y)$, and $\gamma = \varphi(t)/4$. Since Corollary 10.13 asserts that g is uniformly convex with modulus φ, we have

$$\alpha \geqslant \alpha - \beta \geqslant \gamma > 0. \tag{10.18}$$

If $\beta \leqslant \gamma/2$, then $\beta^p \leqslant \gamma^p 2^{-p}$ and $\alpha^p \geqslant \gamma^p$, so that $\alpha^p - \beta^p \geqslant \gamma^p(1 - 2^{-p})$. On the other hand, if $\beta > \gamma/2$, the mean value theorem yields $\alpha^p - \beta^p \geqslant p\beta^{p-1}(\alpha - \beta) > p(\gamma/2)^{p-1}\gamma = p2^{1-p}\gamma^p$. Altogether, we always have $\alpha^p - \beta^p \geqslant \gamma^p \min\{p2^{1-p}, 1 - 2^{-p}\}$. Thus, since $|\cdot|^p$ is convex by Example 8.23,

$$\begin{aligned} \tfrac{1}{2}g^p(x) + \tfrac{1}{2}g^p(y) - g^p\big(\tfrac{1}{2}x + \tfrac{1}{2}y\big) &\geqslant \big(\tfrac{1}{2}g(x) + \tfrac{1}{2}g(y)\big)^p - g^p\big(\tfrac{1}{2}x + \tfrac{1}{2}y\big) \\ &\geqslant 2^{-2p}\varphi^p(t) \min\big\{p2^{1-p}, 1 - 2^{-p}\big\}. \end{aligned} \tag{10.19}$$

Hence, (10.17) follows from Proposition 10.14(i). ☐

Example 10.16 Let $p \in [2, +\infty[$. Then $\|\cdot\|^p$ is uniformly convex.

Proof. By Example 10.9(ii), $\|\cdot\|^2$ is strongly convex with constant 2, hence uniformly convex with modulus $|\cdot|^2$. Since $p/2 \in [1, +\infty[$, Proposition 10.15 implies that $\|\cdot\|^p = (\|\cdot\|^2)^{p/2}$ is uniformly convex. ☐

Proposition 10.17 *Let $f\colon \mathcal{H} \to]-\infty, +\infty]$ be proper and convex, and let C be a nonempty compact convex subset of* $\operatorname{dom} f$ *such that f is strictly convex on C and $f|_C$ is continuous. Then f is uniformly convex on C.*

Proof. Set $g = f + \iota_C$, define ψ for g as in (10.14), and take $t \in \mathbb{R}_+$ such that $\psi(t) = 0$. Then there exist sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ in $\operatorname{dom} g = C$ such that $\|x_n - y_n\| \equiv t$ and

$$\tfrac{1}{2}g(x_n) + \tfrac{1}{2}g(y_n) - g\big(\tfrac{1}{2}x_n + \tfrac{1}{2}y_n\big) \to 0. \tag{10.20}$$

Invoking the compactness of C and after passing to subsequences if necessary, we assume that there exist $x \in C$ and $y \in C$ such that $x_n \to x$ and $y_n \to y$. Hence $\|x - y\| = t$, and since $g|_C = f|_C$ is continuous, (10.20) yields $\frac{1}{2}g(x) + \frac{1}{2}g(y) = g(\frac{1}{2}x + \frac{1}{2}y)$. In turn, the strict convexity of g forces $x = y$, i.e., $t = 0$. Thus, the result follows from Proposition 10.14(ii). ☐

Corollary 10.18 *Suppose that $\mathcal{H}$ is finite-dimensional, let $f\colon \mathcal{H} \to \mathbb{R}$ be strictly convex, and let C be a nonempty bounded closed convex subset of $\mathcal{H}$. Then f is uniformly convex on C.*

Proof. Combine Corollary 8.40 and Proposition 10.17. ☐

The following example illustrates the importance of the hypotheses in Proposition 10.17 and Corollary 10.18.

Example 10.19 Suppose that $\mathcal{H} = \mathbb{R}^2$ and set

$$f\colon \mathcal{H} \to]-\infty, +\infty] : (\xi, \eta) \mapsto \begin{cases} 0, & \text{if } \xi = \eta = 0; \\ \dfrac{\eta^2}{2\xi} + \eta^2, & \text{if } \xi > 0 \text{ and } \eta > 0; \\ +\infty, & \text{otherwise.} \end{cases} \tag{10.21}$$

Furthermore, fix $\rho \in \mathbb{R}_{++}$ and set $C = B(0; \rho) \cap \operatorname{dom} f$. Then f is strictly convex, C is a nonempty bounded convex subset of $\operatorname{dom} f$, f is strictly convex on C, and $f|_C$ is lower semicontinuous. However, f is not uniformly convex on C.

Proof. Set $g = f + \iota_C$. We verify here only the lack of uniform convexity of g since the other properties follow from those established in [313, p. 253]. For every $\eta \in]0, \rho[$, if we set $z_\eta = \big(\sqrt{\rho^2 - \eta^2}, \eta\big)$, then $\|z_\eta\| = \rho$, $z_\eta \in C$, and

$$\big(\forall \alpha \in]0,1[\,\big) \quad \frac{\alpha g(z_\eta) + (1-\alpha)g(0) - g\big(\alpha z_\eta + (1-\alpha)0\big)}{\alpha(1-\alpha)} = \eta^2. \tag{10.22}$$

Denoting the exact modulus of convexity of g by φ, we deduce that $0 \leqslant \varphi(\rho) \leqslant \inf_{\eta\in]0,\rho[} \eta^2 = 0$. Hence $\varphi(\rho) = 0$, and in view of Corollary 10.13, we conclude that g is not uniformly convex. ☐

10.3 Quasiconvexity

Definition 10.20 Let $f\colon \mathcal{H} \to [-\infty,+\infty]$. Then f is *quasiconvex* if its lower level sets $(\operatorname{lev}_{\leqslant\xi} f)_{\xi\in\mathbb{R}}$ are convex.

Example 10.21 Let $f\colon \mathcal{H} \to [-\infty,+\infty]$ be convex. Then f is quasiconvex.

Proof. A direct consequence of Corollary 8.5. ☐

Example 10.22 Let $f\colon \mathbb{R} \to [-\infty,+\infty]$ be increasing or decreasing. Then f is quasiconvex.

Example 10.23 Let $f\colon \mathcal{H} \to [-\infty,+\infty]$ be quasiconvex and let $\eta\in\mathbb{R}$. Then $\min\{f,\eta\}$ is quasiconvex.

Proof. Set $g = \min\{f,\eta\}$ and let $\xi\in\mathbb{R}$. If $\xi \geqslant \eta$, then $\operatorname{lev}_{\leqslant\xi} g = \mathcal{H}$ is convex. On the other hand, if $\xi < \eta$, then $\operatorname{lev}_{\leqslant\xi} g = \operatorname{lev}_{\leqslant\xi} f$ is also convex. ☐

Proposition 10.24 *Let $(f_i)_{i\in I}$ be a family of quasiconvex functions from $\mathcal{H}$ to $[-\infty,+\infty]$. Then $\sup_{i\in I} f_i$ is quasiconvex.*

Proof. Let $\xi\in\mathbb{R}$. Then $\operatorname{lev}_{\leqslant\xi} \sup_{i\in I} f_i = \bigcap_{i\in I} \operatorname{lev}_{\leqslant\xi} f_i$ is convex as an intersection of convex sets. ☐

Proposition 10.25 *Let $f\colon \mathcal{H} \to \,]-\infty,+\infty]$ be quasiconvex. Then the following are equivalent:*

(i) *f is weakly sequentially lower semicontinuous.*
(ii) *f is sequentially lower semicontinuous.*
(iii) *f is lower semicontinuous.*
(iv) *f is weakly lower semicontinuous.*

Proof. Since the sets $(\operatorname{lev}_{\leqslant\xi} f)_{\xi\in\mathbb{R}}$ are convex, the equivalences follow from Lemma 1.24, Lemma 1.36, and Theorem 3.34. ☐

The following proposition is clear from Definition 10.20.

Proposition 10.26 *Let $f\colon \mathcal{H} \to [-\infty,+\infty]$. Then f is quasiconvex if and only if*

$$(\forall x\in\operatorname{dom} f)(\forall y\in\operatorname{dom} f)(\forall\alpha\in\,]0,1[\,)\quad f\big(\alpha x+(1-\alpha)y\big) \leqslant \max\big\{f(x),f(y)\big\}. \tag{10.23}$$

We now turn our attention to strict versions of quasiconvexity suggested by (10.23).

Definition 10.27 Let $f\colon \mathcal{H} \to]-\infty,+\infty]$ be proper, let C be a nonempty subset of $\operatorname{dom} f$, and let $\phi\colon \mathbb{R}_+ \to [0,+\infty]$ be an increasing function that vanishes only at 0. Then f is

(i) *strictly quasiconvex* if

$$(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)(\forall \alpha \in]0,1[\,)\\ x \neq y \;\;\Rightarrow\;\; f\big(\alpha x + (1-\alpha)y\big) < \max\big\{f(x), f(y)\big\}; \quad (10.24)$$

(ii) *uniformly quasiconvex* with modulus ϕ if

$$(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)(\forall \alpha \in]0,1[\,)\\ f\big(\alpha x + (1-\alpha)y\big) + \alpha(1-\alpha)\phi\big(\|x-y\|\big) \leqslant \max\big\{f(x), f(y)\big\}; \quad (10.25)$$

(iii) *strictly quasiconvex on* C if

$$(\forall x \in C)(\forall y \in C)(\forall \alpha \in]0,1[\,)\\ x \neq y \;\;\Rightarrow\;\; f\big(\alpha x + (1-\alpha)y\big) < \max\big\{f(x), f(y)\big\}; \quad (10.26)$$

(iv) *uniformly quasiconvex on* C with modulus ϕ if

$$(\forall x \in C)(\forall y \in C)(\forall \alpha \in]0,1[\,)\\ f\big(\alpha x + (1-\alpha)y\big) + \alpha(1-\alpha)\phi\big(\|x-y\|\big) \leqslant \max\big\{f(x), f(y)\big\}. \quad (10.27)$$

Remark 10.28 Each type of quasiconvexity in Definition 10.27 is implied by its convex counterpart, and uniform quasiconvexity implies strict quasiconvexity. As examples in the remainder of this chapter show, these notions are all distinct.

Example 10.29 Let $f\colon \mathbb{R} \to]-\infty,+\infty]$ be proper, and strictly increasing or strictly decreasing on $\operatorname{dom} f$. Then it follows from (10.24) that f is strictly quasiconvex.

Example 10.30 Let $f\colon \mathcal{H} \to \mathbb{R}$ and let $\phi\colon \operatorname{ran} f \to \mathbb{R}$ be increasing. Then the following hold:

(i) Suppose that f is strictly quasiconvex and that ϕ is strictly increasing. Then $\phi \circ f$ is strictly quasiconvex.
(ii) Suppose that $\phi \circ f$ is strictly convex. Then f is strictly quasiconvex.

Proof. Assume that x and y are distinct points in $\mathcal{H} = \operatorname{dom} f = \operatorname{dom}(\phi \circ f)$, and let $\alpha \in]0,1[$.

(i): In view of (10.24), $f(\alpha x + (1-\alpha)y) < \max\{f(x), f(y)\}$. Therefore

$$\begin{aligned}(\phi \circ f)(\alpha x + (1-\alpha)y) &= \phi\big(f(\alpha x + (1-\alpha)y)\big)\\ &< \phi\big(\max\{f(x), f(y)\}\big)\\ &= \max\big\{(\phi \circ f)(x), (\phi \circ f)(y)\big\}. \end{aligned} \tag{10.28}$$

(ii): We derive from (8.3) that

$$\begin{aligned}\phi\big(f(\alpha x + (1-\alpha)y)\big) &= (\phi \circ f)(\alpha x + (1-\alpha)y)\\ &< \alpha(\phi \circ f)(x) + (1-\alpha)(\phi \circ f)(y)\\ &\leqslant \max\big\{(\phi \circ f)(x), (\phi \circ f)(y)\big\}\\ &= \phi\big(\max\{f(x), f(y)\}\big). \end{aligned} \tag{10.29}$$

Hence, $f(\alpha x + (1-\alpha)y) < \max\{f(x), f(y)\}$). ☐

Example 10.31 Let $p \in \mathbb{R}_{++}$. Then the following hold:

(i) $\|\cdot\|^p$ is strictly quasiconvex.
(ii) Suppose that $\mathcal{H} \neq \{0\}$ and that $p < 1$. Then $\|\cdot\|^p$ is not convex and not uniformly quasiconvex.

Proof. (i): Take $f = \|\cdot\|^2$ (which is strictly quasiconvex by Example 8.10) and $\phi\colon \mathbb{R}_+ \to \mathbb{R}\colon t \mapsto t^{p/2}$ in Example 10.30(i).

(ii): Let $z \in \mathcal{H}$ be such that $\|z\| = 1$. If we set $x = z$ and $y = 0$, and let $\alpha \in \,]0,1[$, then (8.1) turns into the false inequality $\alpha^p \leqslant \alpha$. Thus, by Proposition 8.4, f is not convex. Now assume that $\|\cdot\|^p$ is uniformly quasiconvex with modulus ϕ, let $t \in \mathbb{R}_{++}$, and let $s \in \,]t,+\infty[$. Setting $x = (s-t)z$, $y = (s+t)z$, and $\alpha = 1/2$, we estimate

$$\begin{aligned}\frac{\phi(2t)}{4} &\leqslant \max\big\{\|(s-t)z\|^p, \|(s+t)z\|^p\big\} - \big\|\tfrac{1}{2}(s-t)z + \tfrac{1}{2}(s+t)z\big\|^p\\ &= (s+t)^p - s^p. \end{aligned} \tag{10.30}$$

Letting $s \uparrow +\infty$, we deduce that $\phi(2t) = 0$, which is impossible. ☐

Example 10.32 Suppose that $\mathcal{H} = \mathbb{R}$ and set $f\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \nu x$, where $\nu \in \mathbb{R}_{++}$. Then f is convex, not strictly convex, but uniformly quasiconvex with modulus $\phi\colon t \mapsto \nu t$.

Proof. Take x and y in $\mathbb{R}$ such that $x < y$, and take $\alpha \in \,]0,1[$. Since $f(\alpha x + (1-\alpha)y) = \alpha f(x) + (1-\alpha)f(y)$, f is convex but not strictly so. The uniform quasiconvexity follows from the identity $f(\alpha x + (1-\alpha)y) = \nu y + \alpha\nu(x-y) = \max\{f(x), f(y)\} - \alpha\nu|x-y| \leqslant \max\{f(x), f(y)\} - \alpha(1-\alpha)\nu|x-y|$. ☐

Exercises

Exercise 10.1 Prove Proposition 10.2.

Exercise 10.2 Prove Proposition 10.3.

Exercise 10.3 Let $f: \mathcal{H} \to [-\infty, +\infty]$. Show that f is sublinear if and only if $\operatorname{epi} f$ is a convex cone.

Exercise 10.4 Check Example 10.9.

Exercise 10.5 Suppose that $\mathcal{H} \neq \{0\}$ and let $p \in \left]1, 2\right[$. Show that $\|\cdot\|^p$ is not uniformly convex.

Exercise 10.6 Show that

$$f: \mathbb{R} \times \mathcal{H} \to \left]-\infty, +\infty\right] : (\xi, x) \mapsto \begin{cases} \dfrac{\xi^4 + \xi^2\|x\|^2 + \|x\|^3}{\xi^2}, & \text{if } \xi > 0; \\ +\infty, & \text{otherwise} \end{cases} \tag{10.31}$$

is strongly convex.

Exercise 10.7 Let C be a nonempty convex subset of $\mathcal{H}$, let φ denote the exact modulus of convexity of ι_C, and let $t \in \mathbb{R}_+$. Show that $\varphi(t) = 0$ if $t < \operatorname{diam} C$ and that $\varphi(t) = +\infty$ if $t > \operatorname{diam} C$.

Exercise 10.8 Suppose that $\mathcal{H} = \mathbb{R}$ and set $f = |\cdot|^4$. Show that f is not strongly convex. Determine the function ψ defined in (10.14) explicitly and conclude from the same result that f is uniformly convex.

Exercise 10.9 Use Exercise 8.8 to show that the function ψ defined in (10.14) is increasing.

Exercise 10.10 Let $f \in \Gamma_0(\mathbb{R})$, and let C be a nonempty bounded closed interval in $\operatorname{dom} f$ such that f is strictly convex on C. Show that f is uniformly convex on C.

Exercise 10.11 Show that none of the following functions from $\Gamma_0(\mathbb{R})$ is uniformly convex:

(i) The negative Boltzmann–Shannon entropy (Example 9.35).

(ii) $x \mapsto \begin{cases} 1/x, & \text{if } x > 0; \\ +\infty, & \text{if } x \leqslant 0. \end{cases}$

(iii) The negative Burg entropy (Example 9.36(viii)).

Exercise 10.12 Set

$$f: \mathbb{R} \to \mathbb{R}: x \mapsto \begin{cases} 2x, & \text{if } x \leqslant 0; \\ x, & \text{if } x > 0. \end{cases} \tag{10.32}$$

Show that f is not convex but that it is uniformly quasiconvex.

Exercise 10.13 Suppose that $\mathcal{H} \neq \{0\}$ and set

$$f\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \frac{\|x\|}{\|x\|+1}. \tag{10.33}$$

Show that f is not convex, not uniformly quasiconvex, but strictly quasiconvex.

Exercise 10.14 Set

$$f\colon \mathbb{R} \to \left]-\infty,+\infty\right] : x \mapsto \begin{cases} \dfrac{x}{\lfloor x\rfloor+1} + \displaystyle\sum_{k=2}^{\lfloor x\rfloor+1} \frac{1}{k}, & \text{if } x \geqslant 0; \\ +\infty, & \text{if } x < 0. \end{cases} \tag{10.34}$$

Show that f is not convex, not uniformly quasiconvex, but strictly quasiconvex.

Exercise 10.15 Show that the sum of a quasiconvex function and a convex function may fail to be quasiconvex.

Exercise 10.16 Let $\mathcal{K}$ be a real Hilbert space and let $F\colon \mathcal{H}\times\mathcal{K} \to \left]-\infty,+\infty\right]$ be quasiconvex. Show that the marginal function $f\colon \mathcal{H} \to [-\infty,+\infty] : x \mapsto \inf F(x,\mathcal{K})$ is quasiconvex.

Exercise 10.17 Suppose that $\mathcal{H} \neq \{0\}$. Provide a function that is convex, but not strictly quasiconvex.

Exercise 10.18 Show that the converse of Example 10.30(ii) is false by providing a function $f\colon \mathbb{R} \to \mathbb{R}_+$ that is strictly quasiconvex, but such that f^2 is not strictly convex.

Chapter 11
Convex Minimization Problems

Convex optimization is one of the main areas of application of convex analysis. This chapter deals with the issues of existence and uniqueness in minimization problems and investigates properties of minimizing sequences.

11.1 Infima and Suprema

Proposition 11.1 *Let $f\colon \mathcal{H} \to [-\infty, +\infty]$ and let C be a nonempty subset of $\mathcal{H}$. Then the following hold:*

(i) *Suppose that f is lower semicontinuous. Then $\sup f(\overline{C}) = \sup f(C)$.*
(ii) *Suppose that f is convex. Then $\sup f(\operatorname{conv} C) = \sup f(C)$.*
(iii) *Let $u \in \mathcal{H}$. Then $\sup \langle \overline{\operatorname{conv}}\, C \mid u\rangle = \sup \langle C \mid u\rangle$ and $\inf \langle \overline{\operatorname{conv}}\, C \mid u\rangle = \inf \langle C \mid u\rangle$.*
(iv) *Suppose that $f \in \Gamma_0(\mathcal{H})$, that C is convex, and that $\operatorname{dom} f \cap \operatorname{int} C \neq \varnothing$. Then $\inf f(\overline{C}) = \inf f(C)$.*

Proof. (i): Since $C \subset \overline{C}$, we have $\sup f(C) \leqslant \sup f(\overline{C})$. Now take $x \in \overline{C}$. Then there exists a sequence $(x_n)_{n\in\mathbb{N}}$ in C such that $x_n \to x$. Thus $f(x) \leqslant \underline{\lim}\, f(x_n) \leqslant \sup f(C)$, and we conclude that $\sup f(\overline{C}) \leqslant \sup f(C)$.

(ii): Since $C \subset \operatorname{conv} C$, we have $\sup f(C) \leqslant \sup f(\operatorname{conv} C)$. Now take $x \in \operatorname{conv} C$, say $x = \sum_{i\in I} \alpha_i x_i$, where $(\alpha_i)_{i\in I}$ is a finite family in $]0,1[$ that satisfies $\sum_{i\in I} \alpha_i = 1$, and where $(x_i)_{i\in I}$ is a finite family in C. Then, since f is convex, Corollary 8.12 yields $f(x) = f\big(\sum_{i\in I} \alpha_i x_i\big) \leqslant \sum_{i\in I} \alpha_i f(x_i) \leqslant \sum_{i\in I} \alpha_i \sup f(C) = \sup f(C)$. Therefore, $\sup f(\operatorname{conv} C) \leqslant \sup f(C)$.

(iii): This follows from (i) and (ii), since $\langle \cdot \mid u\rangle$ and $-\langle \cdot \mid u\rangle$ are continuous and convex.

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-48311-5_31

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_11

(iv): It is clear that $\inf f(\overline{C}) \leqslant \inf f(C)$. Now take $x_0 \in \overline{C}$, $x_1 \in \operatorname{dom} f \cap \operatorname{int} C$, and set $(\forall \alpha \in]0,1[)\ x_\alpha = (1-\alpha)x_0 + \alpha x_1$. Using Proposition 9.14 and Proposition 3.44, we deduce that $f(x_0) = \lim_{\alpha \downarrow 0} f(x_\alpha) \geqslant \inf f(C)$. Therefore, $\inf f(\overline{C}) \geqslant \inf f(C)$. ☐

Example 11.2 Let C be a nonempty subset of $\mathcal{H}$. Then the support function σ_C is a sublinear function in $\Gamma_0(\mathcal{H})$, $\sigma_C = \sigma_{\overline{\operatorname{conv}}\, C}$, and $\operatorname{dom} \sigma_C = \operatorname{bar} \overline{\operatorname{conv}}\, C$. If C is bounded, then σ_C is real-valued and continuous on $\mathcal{H}$.

Proof. Definition 7.8 implies that $\sigma_C(0) = 0$ and that σ_C is the supremum of the family of (continuous, hence) lower semicontinuous and (linear, hence) convex functions $(\langle x \mid \cdot \rangle)_{x \in C}$. Therefore, σ_C is lower semicontinuous and convex by Proposition 9.3. On the other hand, it is clear from Definition 7.8 that σ_C is positively homogeneous. Altogether, σ_C is sublinear. Furthermore, Proposition 11.1(iii) (alternatively, Proposition 7.13) implies that $\sigma_C = \sigma_{\overline{\operatorname{conv}}\, C}$ and hence that $\operatorname{dom} \sigma_C = \operatorname{bar} \overline{\operatorname{conv}}\, C$ by (6.42). If C is bounded, then $\operatorname{bar} \overline{\operatorname{conv}}\, C = \mathcal{H}$ by Proposition 6.49(iii), and the continuity of σ_C is a consequence of Corollary 8.39(ii). ☐

11.2 Minimizers

Definition 11.3 Let $f\colon \mathcal{H} \to]-\infty,+\infty]$ be proper, and let $x \in \mathcal{H}$. Then x is a *(global) minimizer* of f if $f(x) = \inf f(\mathcal{H})$, i.e. (see Section 1.5), $f(x) = \min f(\mathcal{H}) \in \mathbb{R}$. The set of minimizers of f is denoted by $\operatorname{Argmin} f$. If $\operatorname{Argmin} f$ is a singleton, its unique element is denoted by $\operatorname{argmin}_{x\in\mathcal{H}} f(x)$. Now let C be a subset of $\mathcal{H}$ such that $C \cap \operatorname{dom} f \neq \varnothing$. A *minimizer of f over C* is a minimizer of $f + \iota_C$. The set of minimizers of f over C is denoted by $\operatorname{Argmin}_C f$. If there exists $\rho \in \mathbb{R}_{++}$ such that x is a minimizer of f over $B(x;\rho)$, then x is a *local minimizer* of f.

The following result underlines the fundamental importance of convexity in minimization problems.

Proposition 11.4 *Let $f\colon \mathcal{H} \to]-\infty,+\infty]$ be proper and convex. Then every local minimizer of f is a minimizer.*

Proof. Take $x \in \mathcal{H}$ and $\rho \in \mathbb{R}_{++}$ such that $f(x) = \min f(B(x;\rho))$. Fix $y \in \operatorname{dom} f \smallsetminus B(x;\rho)$, and set $\alpha = 1 - \rho/\|x-y\|$ and $z = \alpha x + (1-\alpha)y$. Then $\alpha \in]0,1[$ and $z \in B(x;\rho)$. In view of the convexity of f, we deduce that

$$f(x) \leqslant f(z) = f\big(\alpha x + (1-\alpha)y\big) \leqslant \alpha f(x) + (1-\alpha) f(y). \tag{11.1}$$

Therefore $f(x) \leqslant f(y)$. ☐

Proposition 11.5 *Let $f\colon \mathcal{H} \to]-\infty,+\infty]$ be proper and convex, and let C be a subset of $\mathcal{H}$. Suppose that x is a minimizer of f over C such that $x \in \operatorname{int} C$. Then x is a minimizer of f.*

Proof. There exists $\rho \in \mathbb{R}_{++}$ such that $B(x;\rho) \subset C$. Therefore, $f(x) = \inf f(B(x;\rho))$, and the conclusion follows from Proposition 11.4. ☐

Proposition 11.6 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and quasiconvex. Then* $\operatorname{Argmin} f$ *is convex.*

Proof. This follows from Definition 10.20. ☐

Proposition 11.7 *Let $\phi\colon \mathbb{R} \to \left]-\infty,+\infty\right]$ be proper, even, and convex. Then the following hold:*

(i) $0 \in \operatorname{Argmin} \phi$.
(ii) *ϕ is increasing on $\mathbb{R}_+$.*
(iii) *ϕ is decreasing on $\mathbb{R}_-$.*

Now suppose that, in addition, ϕ vanishes only at 0*. Then the following hold:*

(iv) *ϕ is strictly increasing on $\mathbb{R}_+$.*
(v) *ϕ is strictly decreasing on $\mathbb{R}_-$.*

Proof. (i): Let $\xi \in \mathbb{R}$. Then $\phi(0) = \phi((1/2)\xi + (1/2)(-\xi)) \leqslant (1/2)(\phi(\xi) + \phi(-\xi)) = \phi(\xi)$.

(ii): Take ξ and η in $\mathbb{R}_{++}$ such that $\xi < \eta$ and set $\alpha = \xi/\eta$. Then $\alpha \in \left]0,1\right[$ and (i) yields $\phi(0) \leqslant \phi(\eta)$. Hence, $\phi(\xi) = \phi(\alpha\eta + (1-\alpha)0) \leqslant \alpha\phi(\eta) + (1-\alpha)\phi(0) \leqslant \phi(\eta)$.

(iii): This follows from (ii) since ϕ is even.

(iv): Take ξ and η in $\mathbb{R}_{++}$ such that $\xi < \eta$ and set $\alpha = \xi/\eta$. Then $\alpha \in \left]0,1\right[$ and, since $\operatorname{Argmin} \phi = \{0\}$, $\phi(0) < \phi(\eta)$. Hence, $\phi(\xi) = \phi(\alpha\eta + (1-\alpha)0) \leqslant \alpha\phi(\eta) + (1-\alpha)\phi(0) < \phi(\eta)$.

(v): This follows from (iv) since ϕ is even. ☐

11.3 Uniqueness of Minimizers

Proposition 11.8 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be quasiconvex and let C be a convex subset of $\mathcal{H}$ such that $C \cap \operatorname{dom} f \neq \varnothing$. Suppose that one of the following holds:*

(i) *$f + \iota_C$ is strictly quasiconvex.*
(ii) *f is convex,* $(\operatorname{int} C) \cap \operatorname{Argmin} f = \varnothing$, *and C is* strictly convex, *i.e.,*

$$(\forall x \in C)(\forall y \in C) \quad x \neq y \quad \Rightarrow \quad \frac{x+y}{2} \in \operatorname{int} C. \tag{11.2}$$

Then f has at most one minimizer over C.

Proof. We assume that C is not a singleton. Set $\mu = \inf f(C)$ and suppose that there exist two distinct points x and y in $C \cap \operatorname{dom} f$ such that $f(x) = f(y) = \mu$. Since x and y lie in the convex set $C \cap \operatorname{lev}_{\leqslant \mu} f$, so does $z = (x+y)/2$. Therefore $f(z) = \mu$.

(i): It follows from the strict quasiconvexity of $f + \iota_C$ that $\mu = f(z) < \max\{f(x), f(y)\} = \mu$, which is impossible.

(ii): We have $z \in \operatorname{int} C$ and $f(z) = \inf f(C)$. Since f is convex, it follows from Proposition 11.5 that $f(z) = \inf f(\mathcal{H})$. Therefore, $z \in (\operatorname{int} C) \cap \operatorname{Argmin} f = \varnothing$, which is absurd. ☐

Corollary 11.9 *Let $f\colon \mathcal{H} \to \,]-\infty, +\infty]$ be proper and strictly convex. Then f has at most one minimizer.*

Proof. Since f is strictly quasiconvex, the result follows from Proposition 11.8(i) with $C = \mathcal{H}$. ☐

11.4 Existence of Minimizers

Theorem 11.10 *Let $f\colon \mathcal{H} \to \,]-\infty, +\infty]$ be lower semicontinuous and quasiconvex, and let C be a closed convex subset of $\mathcal{H}$ such that, for some $\xi \in \mathbb{R}$, $C \cap \operatorname{lev}_{\leqslant \xi} f$ is nonempty and bounded. Then f has a minimizer over C.*

Proof. Since f is lower semicontinuous and quasiconvex, it follows from Proposition 10.25 that f is weakly lower semicontinuous. On the other hand, since C and $\operatorname{lev}_{\leqslant \xi} f$ are closed and convex, the set $D = C \cap \operatorname{lev}_{\leqslant \xi} f$ is closed and convex, and, by assumption, bounded. Thus, D is nonempty and weakly compact by Theorem 3.37. Consequently, since minimizing f over C is equivalent to minimizing f over D, the claim follows from Lemma 2.30 and Theorem 1.29 in $\mathcal{H}^{\mathrm{weak}}$. ☐

Definition 11.11 Let $f\colon \mathcal{H} \to [-\infty, +\infty]$. Then f is *coercive* if

$$\lim_{\|x\| \to +\infty} f(x) = +\infty, \tag{11.3}$$

and *supercoercive* if

$$\lim_{\|x\| \to +\infty} \frac{f(x)}{\|x\|} = +\infty. \tag{11.4}$$

By convention, f is coercive and supercoercive if $\mathcal{H} = \{0\}$.

Proposition 11.12 *Let $f\colon \mathcal{H} \to [-\infty, +\infty]$. Then f is coercive if and only if its lower level sets $(\operatorname{lev}_{\leqslant \xi} f)_{\xi \in \mathbb{R}}$ are bounded.*

Proof. Suppose that, for some $\xi \in \mathbb{R}$, $\operatorname{lev}_{\leqslant \xi} f$ is unbounded. Then we can find a sequence $(x_n)_{n\in\mathbb{N}}$ in $\operatorname{lev}_{\leqslant \xi} f$ such that $\|x_n\| \to +\infty$. As a result, f is not coercive. Conversely, suppose that the lower level sets of f are bounded and take a sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ such that $\|x_n\| \to +\infty$. Then, for every $\xi \in \mathbb{R}_{++}$, we can find $N \in \mathbb{N}$ such that $\inf_{n \geqslant N} f(x_n) \geqslant \xi$. Therefore $f(x_n) \to +\infty$. ☐

Proposition 11.13 *Suppose that $\mathcal{H}$ is finite-dimensional and let $f \in \Gamma_0(\mathcal{H})$. Then f is coercive if and only if there exists $\xi \in \mathbb{R}$ such that* $\operatorname{lev}_{\leqslant \xi} f$ *is nonempty and bounded.*

Proof. If f is coercive, then all level sets $(\operatorname{lev}_{\leqslant \xi} f)_{\xi \in \mathbb{R}}$ are bounded by Proposition 11.12. Now suppose that $\operatorname{lev}_{\leqslant \xi} f$ is nonempty and bounded, and take $x \in \operatorname{lev}_{\leqslant \xi} f$. It is clear that all lower level sets at a lower height are bounded. Take $\eta \in \left]\xi, +\infty\right[$ and suppose that $\operatorname{lev}_{\leqslant \eta} f$ is unbounded. By Corollary 6.52, $\operatorname{rec} \operatorname{lev}_{\leqslant \eta} f \neq \{0\}$. Take $y \in \operatorname{rec} \operatorname{lev}_{\leqslant \eta} f$. Since $x \in \operatorname{lev}_{\leqslant \eta} f$, it follows that $(\forall \lambda \in \mathbb{R}_{++})$ $x + \lambda y \in \operatorname{lev}_{\leqslant \eta} f$. For every $\lambda \in \left]1, +\infty\right[$, we have $x + y = (1-\lambda^{-1})x + \lambda^{-1}(x+\lambda y)$ and hence $f(x+y) \leqslant (1-\lambda^{-1})f(x) + \lambda^{-1} f(x+\lambda y) \leqslant (1-\lambda^{-1})f(x) + \lambda^{-1}\eta$, which implies that $\lambda(f(x+y) - f(x)) \leqslant \eta - f(x)$ and, in turn, that $f(x+y) \leqslant f(x) \leqslant \xi$. We conclude that $x + \operatorname{rec} \operatorname{lev}_{\leqslant \eta} f \subset \operatorname{lev}_{\leqslant \xi} f$, which is impossible, since $\operatorname{lev}_{\leqslant \xi} f$ is bounded and $\operatorname{rec} \operatorname{lev}_{\leqslant \eta} f$ is unbounded. Therefore, all lower level sets $(\operatorname{lev}_{\leqslant \eta} f)_{\eta \in \mathbb{R}}$ are bounded and Proposition 11.12 implies that f is coercive. ☐

Proposition 11.14 *Let f be in $\Gamma_0(\mathcal{H})$, and let $g \colon \mathcal{H} \to \left]-\infty, +\infty\right]$ be supercoercive. Then $f + g$ is supercoercive.*

Proof. According to Theorem 9.20, f is minorized by a continuous affine functional, say $x \mapsto \langle x \mid u \rangle + \eta$, where $u \in \mathcal{H}$ and $\eta \in \mathbb{R}$. Thus, $(\forall x \in \mathcal{H})$ $f(x) + g(x) \geqslant \langle x \mid u \rangle + \eta + g(x) \geqslant -\|x\|\,\|u\| + \eta + g(x)$. We conclude that $(f(x) + g(x))/\|x\| \geqslant -\|u\| + (\eta + g(x))/\|x\| \to +\infty$ as $\|x\| \to +\infty$. ☐

Proposition 11.15 *Let $f \in \Gamma_0(\mathcal{H})$ and let C be a closed convex subset of $\mathcal{H}$ such that $C \cap \operatorname{dom} f \neq \varnothing$. Suppose that one of the following holds:*

(i) *f is coercive.*
(ii) *C is bounded.*

Then f has a minimizer over C.

Proof. Since $C \cap \operatorname{dom} f \neq \varnothing$, there exists $x \in \operatorname{dom} f$ such that $D = C \cap \operatorname{lev}_{\leqslant f(x)} f$ is nonempty, closed, and convex. Moreover, D is bounded since C or, by Proposition 11.12, $\operatorname{lev}_{\leqslant f(x)} f$ is. The result therefore follows from Theorem 11.10. ☐

Corollary 11.16 *Let f and g be in $\Gamma_0(\mathcal{H})$. Suppose that* $\operatorname{dom} f \cap \operatorname{dom} g \neq \varnothing$ *and that one of the following holds:*

(i) *f is supercoercive.*
(ii) *f is coercive and g is bounded below.*

Then $f + g$ is coercive and it has a minimizer over $\mathcal{H}$. If f or g is strictly convex, then $f + g$ has exactly one minimizer over $\mathcal{H}$.

Proof. It follows from Corollary 9.4 that $f + g \in \Gamma_0(\mathcal{H})$. Hence, in view of Proposition 11.15(i), it suffices to show that $f + g$ is coercive in both cases.

The uniqueness of the minimizer assertion will then follow from Corollary 11.9 by observing that $f+g$ is strictly convex.

(i): By Proposition 11.14, $f+g$ is supercoercive, hence coercive.

(ii): Set $\mu = \inf g(\mathcal{H}) > -\infty$. Then $(f+g)(x) \geqslant f(x)+\mu \to +\infty$ as $\|x\| \to +\infty$ by coercivity of f. ☐

Corollary 11.17 *Let $f \in \Gamma_0(\mathcal{H})$ be strongly convex. Then f is supercoercive and it has exactly one minimizer over $\mathcal{H}$.*

Proof. Set $q = (1/2)\|\cdot\|^2$. By Proposition 10.8, there exists $\beta \in \mathbb{R}_{++}$ such that $f-\beta q$ is convex. Hence, $f = \beta q + (f-\beta q)$ is the sum of the supercoercive function βq and $f-\beta q \in \Gamma_0(\mathcal{H})$. Therefore, f is supercoercive by Proposition 11.14. In view of Corollary 11.16, f has exactly one minimizer. ☐

Proposition 11.18 (Asymptotic center) *Let $(z_n)_{n\in\mathbb{N}}$ be a bounded sequence in $\mathcal{H}$, let C be a nonempty closed convex subset of $\mathcal{H}$, let $T\colon C \to C$ be nonexpansive, and set $f\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \varlimsup \|x-z_n\|^2$. Then the following hold:*

(i) *f is strongly convex with constant 2.*
(ii) *f is supercoercive.*
(iii) *$f+\iota_C$ is strongly convex and supercoercive; its unique minimizer, denoted by z_C, is called the* asymptotic center *of $(z_n)_{n\in\mathbb{N}}$ relative to C.*
(iv) *Suppose that $z \in \mathcal{H}$ and that $z_n \rightharpoonup z$. Then $(\forall x \in \mathcal{H})$ $f(x) = \|x-z\|^2 + f(z)$ and $z_C = P_C z$.*
(v) *Suppose that $(z_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to C. Then $P_C z_n \to z_C$.*
(vi) *Suppose that $(\forall n \in \mathbb{N})$ $z_{n+1} = Tz_n$. Then $z_C \in \operatorname{Fix} T$.*
(vii) *Suppose that $z_n - Tz_n \to 0$. Then $z_C \in \operatorname{Fix} T$.*

Proof. (i): Let $x \in \mathcal{H}$, let $y \in \mathcal{H}$, and let $\alpha \in \left]0,1\right[$. Corollary 2.15 yields $(\forall n \in \mathbb{N})$ $\|\alpha x + (1-\alpha)y - z_n\|^2 = \alpha\|x-z_n\|^2 + (1-\alpha)\|y-z_n\|^2 - \alpha(1-\alpha)\|x-y\|^2$. Now take the limit superior.

(ii)&(iii): By (i), f is strongly convex with constant 2, as is $f+\iota_C$. Hence, the claim follows from Corollary 11.17.

(iv): Let $x \in \mathcal{H}$. Then $(\forall n \in \mathbb{N})$ $\|x-z_n\|^2 = \|x-z\|^2 + \|z-z_n\|^2 + 2\langle x-z \mid z-z_n\rangle$. Consequently, $f(x) = \varlimsup \|x-z_n\|^2 = \|x-z\|^2 + f(z)$, and thus $P_C z$ minimizes $f+\iota_C$. Thus (iii) implies that $P_C z = z_C$.

(v): By Proposition 5.7, $\overline{y} = \lim P_C z_n$ is well defined. For every $n \in \mathbb{N}$ and every $y \in C$, $\|\overline{y}-z_n\| \leqslant \|\overline{y}-P_C z_n\| + \|P_C z_n - z_n\| \leqslant \|\overline{y}-P_C z_n\| + \|y-z_n\|$. Hence, $\varlimsup \|\overline{y}-z_n\| \leqslant \inf_{y\in C} \varlimsup \|y-z_n\|$ and $\overline{y}$ is thus a minimizer of $f+\iota_C$. By (iii), $\overline{y} = z_C$.

(vi): Observe that $z_C \in C$ and that $Tz_C \in C$. For every $n \in \mathbb{N}$, $\|Tz_C - z_{n+1}\| = \|Tz_C - Tz_n\| \leqslant \|z_C - z_n\|$. Thus, taking the limit superior, we obtain $(f+\iota_C)(Tz_C) \leqslant (f+\iota_C)(z_C)$. By (iii), $Tz_C = z_C$.

(vii): For every $n \in \mathbb{N}$, $\|Tz_C - z_n\| \leqslant \|Tz_C - Tz_n\| + \|Tz_n - z_n\| \leqslant \|z_C - z_n\| + \|Tz_n - z_n\|$. Hence $\varlimsup \|Tz_C - z_n\| \leqslant \varlimsup \|z_C - z_n\|$, and thus $(f+\iota_C)(Tz_C) \leqslant (f+\iota_C)(z_C)$. Again by (iii), $Tz_C = z_C$. ☐

Corollary 11.19 *Let C be a nonempty closed convex subset of $\mathcal{H}$ and let $T\colon C \to C$ be nonexpansive. For every $z_0 \in C$, set $(\forall n \in \mathbb{N})$ $z_{n+1} = Tz_n$. Then the following are equivalent:*

(i) $\operatorname{Fix} T \neq \varnothing$.
(ii) *For every $z_0 \in C$, $(z_n)_{n\in\mathbb{N}}$ is bounded.*
(iii) *For some $z_0 \in C$, $(z_n)_{n\in\mathbb{N}}$ is bounded.*

Proof. (i)⇒(ii): Combine Example 5.3 with Proposition 5.4(i).
(ii)⇒(iii): Clear.
(iii)⇒(i): This follows from Proposition 11.18(vi). ☐

11.5 Minimizing Sequences

Minimizing sequences were introduced in Definition 1.8. In this section we investigate some of their properties.

Proposition 11.20 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be a proper coercive function. Then every minimizing sequence of f is bounded.*

Proof. This follows at once from Definition 1.8 and Proposition 11.12. ☐

Proposition 11.21 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be a proper lower semicontinuous quasiconvex function and let $(x_n)_{n\in\mathbb{N}}$ be a minimizing sequence of f that converges weakly to some point $x \in \mathcal{H}$. Then $f(x) = \min f(\mathcal{H})$.*

Proof. It follows from Proposition 10.25 that f is weakly sequentially lower semicontinuous. Hence $\inf f(\mathcal{H}) \leqslant f(x) \leqslant \underline{\lim} f(x_n) = \inf f(\mathcal{H})$. ☐

Remark 11.22 Let $f \in \Gamma_0(\mathcal{H})$ and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\operatorname{dom} f$.

(i) Suppose that $(x_n)_{n\in\mathbb{N}}$ converges strongly to a minimizer x of f and that $\mathcal{H} = \mathbb{R}$ or $x \in \operatorname{int}\operatorname{dom} f$. Then $(x_n)_{n\in\mathbb{N}}$ is a minimizing sequence of f. Indeed, it follows from Corollary 9.15 in the former case, and from Corollary 8.39 in the latter, that $f(x_n) \to f(x)$.
(ii) Suppose that $(x_n)_{n\in\mathbb{N}}$ converges strongly to a minimizer x of f. Then $(x_n)_{n\in\mathbb{N}}$ is not necessarily a minimizing sequence of f (see the construction in the proof of Example 9.43).
(iii) Suppose that $\mathcal{H}$ is infinite-dimensional and that $(x_n)_{n\in\mathbb{N}}$ converges weakly to a minimizer x of f. Then $(x_n)_{n\in\mathbb{N}}$ is not necessarily a minimizing sequence of f, even if $x \in \operatorname{int}\operatorname{dom} f$. For instance, suppose that $(x_n)_{n\in\mathbb{N}}$ is an orthonormal sequence in $\mathcal{H}$ and set $f = \|\cdot\|$. Then, as seen in Example 2.32, $x_n \rightharpoonup 0$ and $f(0) = 0 = \inf f(\mathcal{H})$, while $f(x_n) \equiv 1$.

The next examples illustrate various behaviors of minimizing sequences.

Example 11.23 Suppose that $\mathcal{H} = \mathbb{R}^2$. It follows from Example 9.43 that the function

$$f\colon \mathcal{H} \to \left]-\infty,+\infty\right] : (\xi_1,\xi_2) \mapsto \begin{cases} \xi_2^2/\xi_1, & \text{if } \xi_1 > 0;\\ 0, & \text{if } \xi_1 = \xi_2 = 0;\\ +\infty, & \text{otherwise} \end{cases} \tag{11.5}$$

belongs to $\Gamma_0(\mathcal{H})$. Moreover, $\operatorname{Argmin} f = \mathbb{R}_+ \times \{0\}$. Now let $p \in \left[1,+\infty\right[$ and set $(\forall n \in \mathbb{N})\ x_n = ((n+1)^{p+2}, n+1)$. Then $(x_n)_{n\in\mathbb{N}}$ is a minimizing sequence of f since $f(x_n) - \min f(\mathcal{H}) = 1/(n+1)^p \downarrow 0$. However, $d_{\operatorname{Argmin} f}(x_n) = n+1 \uparrow +\infty$. Thus, $f(x_n) - \min f(\mathcal{H}) = O(1/n^p)$ while $(\forall x \in \operatorname{Argmin} f)$ $\|x_n - x\| \uparrow +\infty$.

Example 11.24 Suppose that $\mathcal{H} = \mathbb{R}^2$ and consider the convex function

$$f\colon \mathcal{H} \to \mathbb{R}\colon (\xi_1,\xi_2) \mapsto \xi_1 + \|(\xi_1,\xi_2)\|. \tag{11.6}$$

Then $\operatorname{Argmin} f = \mathbb{R}_- \times \{0\}$. Now set $(\forall n \in \mathbb{N})\ x_n = (-n, 1)$. Then $(x_n)_{n\in\mathbb{N}}$ is a minimizing sequence of f since $f(x_n) - \min f(\mathcal{H}) = -n + \sqrt{n^2+1} \downarrow 0$. However, $(\forall n \in \mathbb{N})\ d_{\operatorname{Argmin} f}(x_n) = 1$.

Example 11.25 Suppose that $\mathcal{H} = \mathbb{R}^2$ and set

$$f\colon \mathcal{H} \to \mathbb{R}\colon (\xi_1,\xi_2) \mapsto \begin{cases} \xi_1 \ln(\xi_1/\xi_2) - \xi_1 + \xi_2, & \text{if } (\xi_1,\xi_2) \in \mathbb{R}^2_{++};\\ \xi_2, & \text{if } (\xi_1,\xi_2) \in \{0\} \times \mathbb{R}_+;\\ +\infty, & \text{otherwise.} \end{cases} \tag{11.7}$$

It follows from Example 9.48 that $f \in \Gamma_0(\mathcal{H})$. Moreover, $\min f(\mathcal{H}) = 0$ and $\operatorname{Argmin} f = \mathbb{R}_+(1,1)$. Let $(\varepsilon_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}_{++}$ such that $\varepsilon_n \downarrow 0$ and set $(\forall n \in \mathbb{N})\ x_n = (\varepsilon_n, \varepsilon_n)$, $y_n = (\varepsilon_n, \exp(-1/\varepsilon_n))$, and $z_n = (\varepsilon_n, \exp(-1/\varepsilon_n^2))$. Then $\lim x_n = \lim y_n = \lim z_n = (0,0) \in \operatorname{Argmin} f$ while $f(x_n) \equiv 0$, $f(y_n) \to 1$, and $f(z_n) \to +\infty$.

Example 11.26 Suppose that $x \in \mathcal{H} \smallsetminus \{0\}$, and set $f = \iota_{[-x,x]}$ and $(\forall n \in \mathbb{N})$ $x_n = (-1)^n x$. Then $(x_n)_{n\in\mathbb{N}}$ is a bounded minimizing sequence of f that is not weakly convergent.

Example 11.27 Suppose that $\mathcal{H}$ is infinite-dimensional and separable. Let $(\omega_k)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}_{++}$ such that $\omega_k \to 0$, let $(x_k)_{k\in\mathbb{N}}$ be an orthonormal basis of $\mathcal{H}$, and set

$$f\colon \mathcal{H} \to \left]-\infty,+\infty\right] : x \mapsto \sum_{k\in\mathbb{N}} \omega_k \left|\langle x \mid x_k\rangle\right|^2. \tag{11.8}$$

Then f is real-valued, continuous, and strictly convex, and 0 is its unique minimizer. However, f is not coercive. Moreover, $(x_n)_{n\in\mathbb{N}}$ is a minimizing sequence of f that converges weakly but not strongly to 0.

Proof. Note that f vanishes only at 0. Moreover, by Parseval, $(\forall x \in \mathcal{H})\ \|x\|^2 = \sum_{k\in\mathbb{N}} |\langle x \mid x_k\rangle|^2 \geqslant f(x)/\sup_{k\in\mathbb{N}} \omega_k$. Hence, f is real-valued. Furthermore, since the functions $(\omega_k |\langle \cdot \mid x_k\rangle|^2)_{k\in\mathbb{N}}$ are positive, convex, and continuous, it follows from Corollary 9.4 that f is convex and lower semicontinuous. Thus, Corollary 8.39(ii) implies that f is continuous.

To show that f is strictly convex, take x and y in $\mathcal{H}$ such that $x \neq y$, and fix $\alpha \in\,]0,1[$. Then there exists $m \in \mathbb{N}$ such that $\sqrt{\omega_m}\,\langle x \mid x_m\rangle \neq \sqrt{\omega_m}\,\langle y \mid x_m\rangle$. Since $|\cdot|^2$ is strictly convex (Example 8.10), we get

$$\omega_m |\langle \alpha x + (1-\alpha)y \mid x_m\rangle|^2 < \alpha\omega_m |\langle x \mid x_m\rangle|^2 + (1-\alpha)\omega_m |\langle y \mid x_m\rangle|^2 \tag{11.9}$$

and, for every $k \in \mathbb{N} \smallsetminus \{m\}$,

$$\omega_k |\langle \alpha x + (1-\alpha)y \mid x_k\rangle|^2 \leqslant \alpha\omega_k |\langle x \mid x_k\rangle|^2 + (1-\alpha)\omega_k |\langle y \mid x_k\rangle|^2. \tag{11.10}$$

Thus, $f(\alpha x+(1-\alpha)y) < \alpha f(x)+(1-\alpha)f(y)$. Now set $(\forall n \in \mathbb{N})\ y_n = x_n/\sqrt{\omega_n}$. Then $\|y_n\| = 1/\sqrt{\omega_n} \to +\infty$, but $f(y_n) \equiv 1$. Therefore, f is not coercive. Finally, $f(x_n) = \omega_n \to 0 = f(0) = \inf f(\mathcal{H})$ and, by Example 2.32, $x_n \rightharpoonup 0$ but $x_n \not\to 0$. ☐

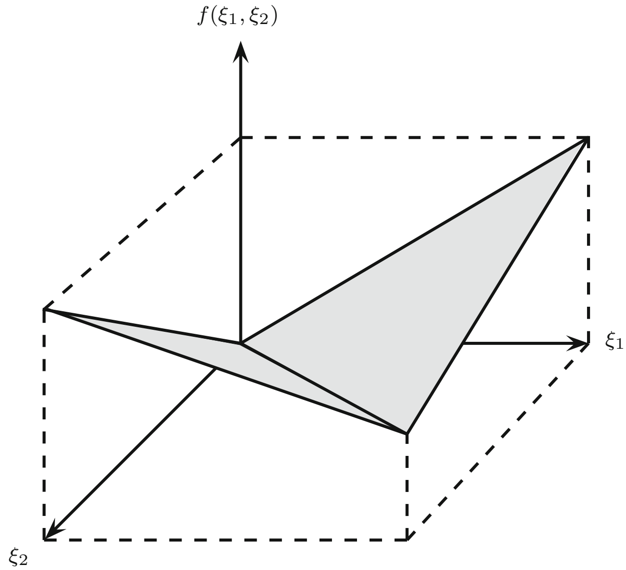

Fig. 11.1 Graph of the function in Example 11.28.

Example 11.28 Suppose that $\mathcal{H} = \mathbb{R}^2$. This example provides a coercive function $f \in \Gamma_0(\mathcal{H})$ with a unique minimizer $\overline{x}$ for which alternating minimizations produce a convergent sequence that is not a minimizing sequence and the limit of which is not $\overline{x}$. Suppose that $\mathcal{H} = \mathbb{R}^2$ and set (see Figure 11.1)

$$\begin{aligned} f\colon \quad \mathcal{H} \quad &\to \left]-\infty, +\infty\right] \\ (\xi_1, \xi_2) &\mapsto \max\{2\xi_1 - \xi_2, 2\xi_2 - \xi_1\} + \iota_{\mathbb{R}^2_+}(\xi_1, \xi_2) \\ &= \begin{cases} 2\xi_1 - \xi_2, & \text{if } \xi_1 \geqslant \xi_2 \geqslant 0; \\ 2\xi_2 - \xi_1, & \text{if } \xi_2 \geqslant \xi_1 \geqslant 0; \\ +\infty, & \text{otherwise.} \end{cases} \end{aligned} \tag{11.11}$$

It is clear that f is lower semicontinuous and convex as the sum of two such functions, and that it is coercive. Moreover, f admits $\overline{x} = (0, 0)$ as its unique minimizer and $\inf f(\mathcal{H}) = 0$. Given an initial point $x_0 \in \mathbb{R}_+ \times \mathbb{R}_{++}$, we define iteratively an alternating minimization sequence $(x_n)_{n\in\mathbb{N}}$ as follows: at iteration n, $x_n = (\xi_{1,n}, \xi_{2,n})$ is known and we construct $x_{n+1} = (\xi_{1,n+1}, \xi_{2,n+1})$ by first letting $\xi_{1,n+1}$ be the minimizer of $f(\cdot, \xi_{2,n})$ over $\mathbb{R}$ and then letting $\xi_{2,n+1}$ be the minimizer of $f(\xi_{1,n+1}, \cdot)$ over $\mathbb{R}$. In view of (11.11), for every integer $n \geqslant 1$, we obtain $x_n = (\xi_{2,0}, \xi_{2,0}) \neq \overline{x}$ and $f(x_n) = \xi_{2,0} \neq \inf f(\mathcal{H})$.

We now provide weak and strong convergence conditions for minimizing sequences.

Proposition 11.29 *Let $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ be proper, lower semicontinuous, and quasiconvex, and let $(x_n)_{n\in\mathbb{N}}$ be a minimizing sequence of f. Suppose that there exists $\xi \in \left]\inf f(\mathcal{H}), +\infty\right[$ such that $C = \operatorname{lev}_{\leqslant \xi} f$ is bounded. Then the following hold:*

(i) *The sequence $(x_n)_{n\in\mathbb{N}}$ has a weak sequential cluster point and every such point is a minimizer of f.*
(ii) *Suppose that $f + \iota_C$ is strictly quasiconvex. Then f possesses a unique minimizer x and $x_n \rightharpoonup x$.*
(iii) *Suppose that $f + \iota_C$ is uniformly quasiconvex. Then f possesses a unique minimizer x and $x_n \to x$.*

Proof. Without loss of generality, we assume that $(x_n)_{n\in\mathbb{N}}$ lies entirely in the bounded closed convex set C.

(i): The existence of a weak sequential cluster point is guaranteed by Lemma 2.45. The second assertion follows from Proposition 11.21.

(ii): Uniqueness follows from Proposition 11.8(i). In turn, the second assertion follows from (i) and Lemma 2.46.

(iii): Since $f + \iota_C$ is strictly quasiconvex, we derive from (ii) that f possesses a unique minimizer $x \in C$. Now fix $\alpha \in \left]0, 1\right[$ and let ϕ be the modulus of uniform quasiconvexity of $f + \iota_C$. Then it follows from (10.25) that, for every $n \in \mathbb{N}$,

$$
\begin{aligned}
f(x) + \alpha(1-\alpha)\phi\big(\|x_n - x\|\big) &= \inf f(\mathcal{H}) + \alpha(1-\alpha)\phi\big(\|x_n - x\|\big) \\
&\leqslant f\big(\alpha x_n + (1-\alpha)x\big) + \alpha(1-\alpha)\phi\big(\|x_n - x\|\big) \\
&\leqslant \max\{f(x_n), f(x)\} \\
&= f(x_n). \qquad (11.12)
\end{aligned}
$$

Consequently, since $f(x_n) \to f(x)$, we obtain $\phi(\|x_n - x\|) \to 0$ and, in turn, $\|x_n - x\| \to 0$. ☐

Corollary 11.30 *Let $f \in \Gamma_0(\mathcal{H})$ be coercive and let $(x_n)_{n\in\mathbb{N}}$ be a minimizing sequence of f. Then the following hold:*

(i) *The sequence $(x_n)_{n\in\mathbb{N}}$ has a weak sequential cluster point and every such point is a minimizer of f.*
(ii) *Suppose that f is strictly convex. Then f possesses a unique minimizer x and $x_n \rightharpoonup x$.*
(iii) *Suppose that f is uniformly convex on every nonempty bounded subset of* $\operatorname{dom} f$. *Then f possesses a unique minimizer x and $x_n \to x$.*

Another instance of strong convergence of minimizing sequences in convex variational problems is the following.

Proposition 11.31 *Let $f \in \Gamma_0(\mathcal{H})$ and let C be a bounded closed convex subset of $\mathcal{H}$ such that $C \cap \operatorname{dom} f \neq \varnothing$. Suppose that $C \cap \operatorname{Argmin} f = \varnothing$ and that C is uniformly convex, i.e., there exists an increasing function $\phi\colon [0, \operatorname{diam} C] \to \mathbb{R}_+$ that vanishes only at 0 such that*

$$
(\forall x \in C)(\forall y \in C) \quad B\Big(\frac{x+y}{2}; \phi\big(\|x-y\|\big)\Big) \subset C, \qquad (11.13)
$$

and let $(x_n)_{n\in\mathbb{N}}$ be a minimizing sequence of $f + \iota_C$. Then f has a unique minimizer x over C and $x_n \to x$.

Proof. We assume that $\mathcal{H} \neq \{0\}$ and that C is not a singleton. The existence of x follows from Proposition 11.15(ii) and its uniqueness from Proposition 11.8(ii). In turn, we deduce from Proposition 11.29(i) and Lemma 2.46 that $x_n \rightharpoonup x$. Moreover, since $C \cap \operatorname{Argmin} f = \varnothing$, it follows from Proposition 11.5 that $x \in \operatorname{bdry} C$. Hence, since (11.13) asserts that $\operatorname{int} C \neq \varnothing$, we derive from Corollary 7.6(i) that x is a support point of C. Denote by u an associated normal vector such that $\|u\| = 1$. Now let $n \in \mathbb{N}$ and set $z_n = (x_n + x)/2 + \phi(\|x_n - x\|)u$. Then it follows from (11.13) that $z_n \in C$ and from (7.1) that $\langle x_n - x \mid u\rangle/2 + \phi(\|x_n - x\|) = \langle z_n - x \mid u\rangle \leqslant 0$. Hence, $\phi(\|x_n - x\|) \leqslant \langle x - x_n \mid u\rangle/2 \to 0$ and therefore $x_n \to x$. ☐

Exercises

Exercise 11.1 Provide a function $f \in \Gamma_0(\mathcal{H})$ and a nonempty set $C \subset \operatorname{int}\operatorname{dom} f$ such that $\inf f(\overline{C}) < \inf f(C)$. Compare with Proposition 11.1(iv).

Exercise 11.2 Provide a function $f \in \Gamma_0(\mathcal{H})$ and a nonempty convex set $C \subset \operatorname{dom} f$ such that $\inf f(\overline{C}) < \inf f(C)$. Compare with Proposition 11.1(iv).

Exercise 11.3 Provide a convex function $f\colon \mathcal{H} \to \mathbb{R}$ and a nonempty convex set $C \subset \mathcal{H}$ such that $\inf f(\overline{C}) < \inf f(C)$. Compare with Proposition 11.1(iv).

Exercise 11.4 Let $f \in \Gamma_0(\mathcal{H})$, let $\beta \in \mathbb{R}_{++}$, and suppose that f is β-strongly convex. Let x be the unique minimizer of f and let $(x_n)_{n\in\mathbb{N}}$ be a minimizing sequence of f. Show that

$$(\forall n \in \mathbb{N}) \quad \|x_n - x\| \leqslant 2\sqrt{\frac{f(x_n) - f(x)}{\beta}}. \tag{11.14}$$

Exercise 11.5 Suppose that $\mathcal{H}$ is finite-dimensional, let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and lower semicontinuous, and let $(x_n)_{n\in\mathbb{N}}$ be a minimizing sequence of f. Suppose that there exists $\xi \in \left]\inf f(\mathcal{H}),+\infty\right[$ such that $\operatorname{lev}_{\leqslant \xi} f$ is nonempty and bounded. Show that $d_{\operatorname{Argmin} f}(x_n) \to 0$. Compare with Example 11.23 and Example 11.24.

Exercise 11.6 Show that Proposition 11.4 is false if f is merely quasiconvex.

Exercise 11.7 Find a function $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ that is proper and convex, a set $C \subset \mathcal{H}$, and a minimizer x of f over C such that $x \in \operatorname{bdry} C$ and $\operatorname{Argmin}(f + \iota_C) \cap \operatorname{Argmin} f = \varnothing$. Compare with Proposition 11.5.

Exercise 11.8 Find a convex function $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ and a convex set $C \subset \mathcal{H}$ such that $C \cap \operatorname{dom} f \neq \varnothing$, $C \cap \operatorname{Argmin} f = \varnothing$, and f has at least two minimizers over C. Compare with Proposition 11.8(ii).

Exercise 11.9 Let C be a convex subset of $\mathcal{H}$. Show that C is strictly convex if and only if $(\forall x \in C)(\forall y \in C)\ x \neq y \Rightarrow \left]x,y\right[\subset \operatorname{int} C$.

Exercise 11.10 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be a proper lower semicontinuous coercive quasiconvex function. Without using Theorem 11.10, derive from Proposition 3.20 and Proposition 11.12 that f has a minimizer.

Exercise 11.11 Show that the conclusion of Proposition 11.13 is false if f is not lower semicontinuous, even when $\mathcal{H} = \mathbb{R}^2$.

Exercise 11.12 Show that the conclusion of Proposition 11.13 is false if $\mathcal{H}$ is infinite-dimensional.

Exercise 11.13 Show that Proposition 11.14 is false if "supercoercive" is replaced by "coercive," even if $g \in \Gamma_0(\mathcal{H})$.

Exercise 11.14 Provide an alternative proof for Theorem 4.29 using Corollary 11.19.

Exercise 11.15 In Example 11.27, suppose that $(x_n)_{n \in \mathbb{N}}$ is an orthonormal sequence that is not an orthonormal basis. What is $\operatorname{Argmin} f$? Is f strictly convex?

Exercise 11.16 Let $f \in \Gamma_0(\mathcal{H})$ be bounded below, let $\beta \in \mathbb{R}_{++}$, let $y \in \operatorname{dom} f$, let $p \in \,]1, +\infty[$, and set $\alpha = f(y) - \inf f(\mathcal{H})$. Prove that there exists $z \in \mathcal{H}$ such that $\|z - y\| \leqslant \beta$ and

$$(\forall x \in \mathcal{H}) \quad f(z) + \frac{\alpha}{\beta^p}\|z - y\|^p \leqslant f(x) + \frac{\alpha}{\beta^p}\|x - y\|^p. \tag{11.15}$$

Exercise 11.17 Find a function $f \in \Gamma_0(\mathcal{H})$ such that $\operatorname{Argmin} f$ is a singleton and such that f has an unbounded minimizing sequence.

Exercise 11.18 Show that $B(0; \rho)$ is uniformly convex and that (11.13) holds with $\phi\colon [0, 2\rho] \to \mathbb{R}_+ : t \mapsto \rho - \sqrt{\rho^2 - (t/2)^2}$.

Chapter 12
Infimal Convolution

This chapter is devoted to a fundamental convexity-preserving operation for functions: the infimal convolution.

Special attention is given to the Moreau envelope and the proximity operator.

12.1 Definition and Basic Facts

Definition 12.1 Let f and g be functions from $\mathcal{H}$ to $]-\infty,+\infty]$. The *infimal convolution* (or *epi-sum*) of f and g is

$$f \,\Box\, g \colon \mathcal{H} \to [-\infty,+\infty] : x \mapsto \inf_{y\in\mathcal{H}} \big(f(y) + g(x-y)\big), \tag{12.1}$$

and it is *exact at a point* $x \in \mathcal{H}$ if $(f \,\Box\, g)(x) = \min_{y\in\mathcal{H}} f(y) + g(x-y)$, i.e. (see Definition 1.7),

$$(\exists\, y \in \mathcal{H}) \quad (f \,\Box\, g)(x) = f(y) + g(x-y) \in \,]-\infty,+\infty]\,; \tag{12.2}$$

$f \,\Box\, g$ is *exact* if it is exact at every point of its domain, in which case it is denoted by $f \,\boxdot\, g$.

Example 12.2 Let C be a subset of $\mathcal{H}$. Then it follows from (1.41) and (1.47) that $d_C = \iota_C \,\Box\, \|\cdot\|$. Moreover, Remark 3.11(i) asserts that, if C is nonempty and open, this infimal convolution is never exact on $\mathcal{H} \smallsetminus C$, though always real-valued.

Example 12.3 Let C and D be nonempty subsets of $\mathcal{H}$. Then $\iota_C \,\boxdot\, \iota_D = \iota_{C+D}$.

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-48311-5_31

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_12

Example 12.4 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ and let $y \in \mathcal{H}$. Then $\iota_{\{y\}} \boxdot f = \tau_y f$.

Definition 12.5 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ and let $u \in \mathcal{H}$. Then f possesses a *continuous affine minorant with slope* u if $f - \langle \cdot \mid u\rangle$ is bounded below.

Proposition 12.6 *Let f, g, and h be functions from $\mathcal{H}$ to $\left]-\infty,+\infty\right]$. Then the following hold:*

(i) *Suppose that f and g possess continuous affine minorants with slope $u \in \mathcal{H}$. Then $f \,\square\, g$ possesses a continuous affine minorant with slope u and $-\infty \notin (f \,\square\, g)(\mathcal{H})$.*
(ii) $\operatorname{dom}(f \,\square\, g) = \operatorname{dom} f + \operatorname{dom} g$.
(iii) $f \,\square\, g = g \,\square\, f$.
(iv) *Suppose that f, g, and h possess continuous affine minorants with the same slope. Then $f \,\square\, (g \,\square\, h) = (f \,\square\, g) \,\square\, h$.*

Proof. (i): By assumption, there exist $\eta \in \mathbb{R}$ and $\mu \in \mathbb{R}$ such that $f \geqslant \langle \cdot \mid u\rangle + \eta$ and $g \geqslant \langle \cdot \mid u\rangle + \mu$. Now fix $x \in \mathcal{H}$. Then, for every $y \in \mathcal{H}$, $f(y) + g(x-y) \geqslant \langle x \mid u\rangle + \eta + \mu$ and, therefore, $f \,\square\, g \geqslant \langle \cdot \mid u\rangle + \eta + \mu > -\infty$.

(ii)&(iii): Observe that (12.1) can be rewritten as

$$f \,\square\, g\colon \mathcal{H} \to [-\infty,+\infty] : x \mapsto \inf_{\substack{(u,v)\in\mathcal{H}\times\mathcal{H}\\ u+v=x}} \big(f(u) + g(v)\big). \tag{12.3}$$

(iv): It follows from (i) that $g \,\square\, h$ and $f \,\square\, g$ are functions from $\mathcal{H}$ to $\left]-\infty,+\infty\right]$. Therefore, the infimal convolutions $f \,\square\, (g \,\square\, h)$ and $(f \,\square\, g) \,\square\, h$ are well defined. Furthermore, using (12.3), we obtain

$$\begin{aligned}(\forall x \in \mathcal{H}) \quad \big(f \,\square\, (g \,\square\, h)\big)(x) &= \inf_{\substack{(u,v,w)\in\mathcal{H}\times\mathcal{H}\times\mathcal{H}\\ u+v+w=x}} \big(f(u) + g(v) + h(w)\big)\\ &= \big((f \,\square\, g) \,\square\, h\big)(x), \end{aligned} \tag{12.4}$$

as desired. ☐

Example 12.7 Set $f\colon \mathbb{R} \to \mathbb{R}\colon \xi \mapsto \xi$ and $g = -f$. Then $f \,\square\, g \equiv -\infty$. This shows that Proposition 12.6(i) fails if the assumption on the minorants is not satisfied.

Proposition 12.8 *Let f and g be functions from $\mathcal{H}$ to $\left]-\infty,+\infty\right]$. Then the following hold:*

(i) $\operatorname{epi} f + \operatorname{epi} g \subset \operatorname{epi}(f \,\square\, g)$.
(ii) *Suppose that $f \,\square\, g = f \boxdot g$. Then* $\operatorname{epi}(f \,\square\, g) = \operatorname{epi} f + \operatorname{epi} g$.

Proof. (i): Take $(x,\xi) \in \operatorname{epi} f$ and $(y,\eta) \in \operatorname{epi} g$. Then (12.3) yields

$$(f \,\square\, g)(x+y) \leqslant f(x) + g(y) \leqslant \xi + \eta. \tag{12.5}$$

Therefore $(x+y, \xi+\eta) \in \operatorname{epi}(f \,\square\, g)$.

(ii): Take $(x,\xi) \in \operatorname{epi}(f \,\Box\, g)$. In view of (i), it suffices to show that $(x,\xi) \in \operatorname{epi} f + \operatorname{epi} g$. By assumption, there exists $y \in \mathcal{H}$ such that $(f \,\Box\, g)(x) = f(y) + g(x-y) \leqslant \xi$. Therefore $(x-y, \xi - f(y)) \in \operatorname{epi} g$ and, in turn, $(x,\xi) = (y, f(y)) + (x-y, \xi - f(y)) \in \operatorname{epi} f + \operatorname{epi} g$. ☐

A central example of infimal convolution is obtained by using a power of the norm.

Proposition 12.9 *Let $f \colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper, let $p \in \left[1,+\infty\right[$, and set*

$$(\forall \gamma \in \mathbb{R}_{++}) \quad g_\gamma = f \,\Box\, \left(\frac{1}{\gamma p}\|\cdot\|^p\right). \tag{12.6}$$

Then the following hold for every $\gamma \in \mathbb{R}_{++}$ and every $x \in \mathcal{H}$:

(i) $\operatorname{dom} g_\gamma = \mathcal{H}$.
(ii) *Take $\mu \in \left]\gamma,+\infty\right[$. Then* $\inf f(\mathcal{H}) \leqslant g_\mu(x) \leqslant g_\gamma(x) \leqslant f(x)$.
(iii) $\inf g_\gamma(\mathcal{H}) = \inf f(\mathcal{H})$.
(iv) $g_\mu(x) \downarrow \inf f(\mathcal{H})$ *as* $\mu \uparrow +\infty$.
(v) g_γ *is bounded above on every ball in* $\mathcal{H}$.

Proof. Let $x \in \mathcal{H}$ and $\gamma \in \mathbb{R}_{++}$.

(i): By Proposition 12.6(ii), $\operatorname{dom} g_\gamma = \operatorname{dom} f + \operatorname{dom}(\|\cdot\|^p/(\gamma p)) = \operatorname{dom} f + \mathcal{H} = \mathcal{H}$.

(ii): It is clear that $g_\mu(x) \leqslant g_\gamma(x)$. On the other hand,

$$\inf f(\mathcal{H}) \leqslant g_\gamma(x) = \inf_{y \in \mathcal{H}} \left(f(y) + \frac{1}{\gamma p}\|x-y\|^p \right) \leqslant f(x). \tag{12.7}$$

(iii): This follows from (ii).

(iv): Let $y \in \mathcal{H}$. Then, for every $\mu \in \mathbb{R}_{++}$, $g_\mu(x) \leqslant f(y) + (\mu p)^{-1}\|x-y\|^p$ and therefore $\varlimsup_{\mu\uparrow+\infty} g_\mu(x) \leqslant f(y)$. Appealing to (ii) and taking the infimum over $y \in \mathcal{H}$, we obtain

$$\inf f(\mathcal{H}) \leqslant \varliminf_{\mu\uparrow+\infty} g_\mu(x) \leqslant \varlimsup_{\mu\uparrow+\infty} g_\mu(x) \leqslant \inf f(\mathcal{H}). \tag{12.8}$$

(v): Fix $z \in \operatorname{dom} f$ and $\rho \in \mathbb{R}_{++}$. Then

$$\begin{aligned}\big(\forall y \in B(x;\rho)\big) \quad g_\gamma(y) &\leqslant f(z) + \|y-z\|^p/(\gamma p) \\ &\leqslant f(z) + 2^{p-1}\big(\|y-x\|^p + \|x-z\|^p\big)/(\gamma p) && (12.9) \\ &\leqslant f(z) + 2^{p-1}\big(\rho^p + \|x-z\|^p\big)/(\gamma p) \\ &< +\infty, && (12.10)\end{aligned}$$

where (12.9) follows either from the triangle inequality or from (8.15), depending on whether $p = 1$ or $p > 1$. ☐

Remark 12.10 In (12.6), we may have $g_\gamma \equiv -\infty$, even if f is real-valued. For instance, suppose that $\mathcal{H}$ is infinite-dimensional and let f be a discontinuous linear functional (see Example 2.27 for a construction and Example 8.42 for properties). Since, for every $\rho \in \mathbb{R}_{++}$, f is unbounded below on $B(0;\rho)$, there exists a sequence $(y_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ such that $y_n \to 0$ and $f(y_n) \to -\infty$. Thus, for every $x \in \mathcal{H}$, $g_\gamma(x) \leqslant f(y_n) + (\gamma p)^{-1}\|x - y_n\|^p \to -\infty$. We conclude that $g_\gamma \equiv -\infty$.

12.2 Infimal Convolution of Convex Functions

Proposition 12.11 *Let f and g be convex functions from $\mathcal{H}$ to $]-\infty,+\infty]$. Then $f \,\Box\, g$ is convex.*

Proof. Take $F\colon \mathcal{H}\times\mathcal{H} \to]-\infty,+\infty] : (x,y) \mapsto f(y) + g(x-y)$ in Proposition 8.35. ☐

Corollary 12.12 *Let C be a convex subset of $\mathcal{H}$. Then d_C is convex.*

Proof. As seen in Example 12.2, $d_C = \iota_C \,\Box\, \|\cdot\|$. Since ι_C and $\|\cdot\|$ are convex (see Example 8.3 and Example 8.9), so is d_C by Proposition 12.11. ☐

The following examples show that the infimal convolution of two functions in $\Gamma_0(\mathcal{H})$ need not be exact or lower semicontinuous.

Example 12.13 Set $f\colon \mathbb{R} \to]-\infty,+\infty] : x \mapsto 1/x$ if $x > 0$; $+\infty$ otherwise, and set $g = f^\vee$. Then the following hold:

(i) $f \in \Gamma_0(\mathbb{R})$ and $g \in \Gamma_0(\mathbb{R})$.
(ii) $f \,\Box\, g \equiv 0$ and $f \,\Box\, g$ is nowhere exact.
(iii) Set $\varphi = \iota_C$ and $\psi = \iota_D$, where $C = \operatorname{epi} f$ and $D = \operatorname{epi} g$. It follows from (i) that C and D are nonempty closed convex subsets of $\mathbb{R}^2$. Therefore φ and ψ are in $\Gamma_0(\mathbb{R}^2)$. However, $C + D$ is the open upper half-plane in $\mathbb{R}^2$ and thus $\varphi \,\boxdot\, \psi = \iota_{C+D}$ is not lower semicontinuous.

We now present conditions under which the infimal convolution of two functions in $\Gamma_0(\mathcal{H})$ is exact and in $\Gamma_0(\mathcal{H})$ (see also Proposition 15.7).

Proposition 12.14 *Let f and g be in $\Gamma_0(\mathcal{H})$, and suppose that one of the following holds:*

(i) *f is supercoercive.*
(ii) *f is coercive and g is bounded below.*

Then $f \,\Box\, g = f \,\boxdot\, g \in \Gamma_0(\mathcal{H})$.

Proof. By Proposition 12.6(ii), $\operatorname{dom} f \,\Box\, g = \operatorname{dom} f + \operatorname{dom} g \neq \varnothing$. Now let $x \in \operatorname{dom} f \,\Box\, g$. Then $\operatorname{dom} f \cap \operatorname{dom} g(x - \cdot) \neq \varnothing$ and hence Corollary 11.16 implies that $f + g(x - \cdot)$ has a minimizer over $\mathcal{H}$. Thus, $(\forall x \in \operatorname{dom}(f \,\Box\, g))$ $(f \,\Box\, g)(x) = \min_{y \in \mathcal{H}} f(y) + g(x - y) \in \mathbb{R}$. Therefore, $f \,\Box\, g = f \,\boxdot\, g$ and $f \,\Box\, g$ is proper. In view of Proposition 12.11 and Theorem 9.1, to complete the proof it suffices to show that $f \,\Box\, g$ is sequentially lower semicontinuous. To this end, let $x \in \mathcal{H}$ and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that $x_n \to x$. We need to show that

$$(f \,\Box\, g)(x) \leqslant \underline{\lim}(f \,\Box\, g)(x_n). \tag{12.11}$$

After passing to a subsequence and relabeling, we assume that the sequence $((f \,\Box\, g)(x_n))_{n\in\mathbb{N}}$ converges, say $(f \,\Box\, g)(x_n) \to \mu \in [-\infty, +\infty[$. Let $(y_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that $(\forall n \in \mathbb{N})$ $(f \,\Box\, g)(x_n) = f(y_n) + g(x_n - y_n)$. We claim that

$$(y_n)_{n\in\mathbb{N}} \text{ is bounded.} \tag{12.12}$$

Assume that (12.12) is false. After passing to a subsequence and relabeling, we obtain $0 \neq \|y_n\| \to +\infty$. We now show that a contradiction ensues from each hypothesis.

(i): By Theorem 9.20, g possesses a continuous affine minorant, say $\langle \cdot \mid u\rangle + \eta$, where $u \in \mathcal{H}$ and $\eta \in \mathbb{R}$. Using the supercoercivity of f, we get

$$\begin{aligned}
+\infty > \mu &\\
&\leftarrow (f \,\Box\, g)(x_n)\\
&= f(y_n) + g(x_n - y_n)\\
&\geqslant f(y_n) + \langle x_n - y_n \mid u\rangle + \eta\\
&\geqslant \|y_n\|\left(\frac{f(y_n)}{\|y_n\|} - \|u\|\right) + \langle x_n \mid u\rangle + \eta\\
&\to +\infty,
\end{aligned} \tag{12.13}$$

which is impossible.

(ii): Since f is coercive, we have $f(y_n) \to +\infty$. Hence, $g(x_n - y_n) \to -\infty$ since $f(y_n) + g(x_n - y_n) \to \mu < +\infty$. However, this is impossible since g is bounded below.

Hence, (12.12) holds in both cases. After passing to a subsequence and relabeling, we assume that $(y_n)_{n\in\mathbb{N}}$ converges weakly to some point $y \in \mathcal{H}$. Then $x_n - y_n \rightharpoonup x - y$ and thus $\mu = \lim(f \,\Box\, g)(x_n) = \lim f(y_n) + g(x_n - y_n) \geqslant \underline{\lim}\, f(y_n) + \underline{\lim}\, g(x_n - y_n) \geqslant f(y) + g(x - y) \geqslant (f \,\Box\, g)(x)$. Therefore, (12.11) is verified and the proof is complete. $\square$

The next result examines the properties of the infimal convolution of a convex function with a power of the norm.

Proposition 12.15 *Let $f \in \Gamma_0(\mathcal{H})$, let $\gamma \in \mathbb{R}_{++}$, and let $p \in \left]1, +\infty\right[$. Then the infimal convolution*

$$f \,\square\, \left(\frac{1}{\gamma p}\|\cdot\|^p\right) : \mathcal{H} \to \left]-\infty, +\infty\right] : x \mapsto \inf_{y \in \mathcal{H}} \left(f(y) + \frac{1}{\gamma p}\|x - y\|^p\right) \quad (12.14)$$

is convex, real-valued, continuous, and exact. Moreover, for every $x \in \mathcal{H}$, the infimum in (12.14) is uniquely attained.

Proof. Let us define g_γ as in (12.6). We observe that $(\gamma p)^{-1}\|\cdot\|^p$ is supercoercive and, by Example 8.23, strictly convex. Proposition 12.14(i) implies that g_γ is proper, lower semicontinuous, convex, and exact. Combining this, Proposition 12.9(v), and Corollary 8.39(i), we obtain that g_γ is real-valued and continuous. The statement concerning the unique minimizer follows from Corollary 11.16(i). ☐

In the next two sections, we examine the cases $p = 1$ and $p = 2$ in (12.14) in more detail.

12.3 Pasch–Hausdorff Envelope

Definition 12.16 Let $f : \mathcal{H} \to \left]-\infty, +\infty\right]$ and let $\beta \in \mathbb{R}_+$. The *β-Pasch–Hausdorff envelope* of f is $f \,\square\, \big(\beta\|\cdot\|\big)$.

Observe that the Pasch–Hausdorff envelope enjoys all the properties listed in Proposition 12.9.

Proposition 12.17 *Let $f : \mathcal{H} \to \left]-\infty, +\infty\right]$ be proper, let $\beta \in \mathbb{R}_+$, and let g be the β-Pasch–Hausdorff envelope of f. Then exactly one of the following holds:*

(i) *f possesses a β-Lipschitz continuous minorant, and g is the largest β-Lipschitz continuous minorant of f.*
(ii) *f possesses no β-Lipschitz continuous minorant, and $g \equiv -\infty$.*

Proof. We first note that Proposition 12.9(i) implies that

$$\begin{aligned}(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad g(x) &= \inf_{z \in \mathcal{H}} \big(f(z) + \beta\|x - z\|\big) \\ &\leqslant \beta\|x - y\| + \inf_{z \in \mathcal{H}} \big(f(z) + \beta\|y - z\|\big) \\ &= \beta\|x - y\| + g(y) \\ &< +\infty. \end{aligned} \quad (12.15)$$

(i): Suppose that f possesses a β-Lipschitz continuous minorant $h\colon \mathcal{H} \to \mathbb{R}$. Then

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \begin{aligned} h(x) &\leqslant h(y) + \beta\|x-y\| \\ &\leqslant f(y) + \beta\|x-y\|. \end{aligned} \tag{12.16}$$

Taking the infimum over $y \in \mathcal{H}$, we obtain

$$(\forall x \in \mathcal{H}) \quad h(x) \leqslant g(x). \tag{12.17}$$

Consequently, $-\infty \notin g(\mathcal{H})$ and (12.15) implies that g is real-valued and β-Lipschitz continuous. On the other hand, it follows from Proposition 12.9(ii) that $g \leqslant f$.

(ii): Suppose that f possesses no β-Lipschitz continuous minorant and that g is real-valued at some point in $\mathcal{H}$. We derive from (12.15) that g is everywhere real-valued and β-Lipschitz continuous. If $\beta = 0$, then $g \equiv \inf f(\mathcal{H})$, which contradicts the assumption that f has no 0-Lipschitz continuous minorant. On the other hand, if $\beta > 0$, then Proposition 12.9(ii) implies that g is a minorant of f, and we once again reach a contradiction. ☐

We deduce at once the following involution property from Proposition 12.17(i).

Corollary 12.18 *Let $f\colon \mathcal{H} \to \mathbb{R}$ be β-Lipschitz continuous for some $\beta \in \mathbb{R}_{++}$. Then f is its own β-Pasch–Hausdorff envelope.*

Corollary 12.19 *Let C be a nonempty subset of $\mathcal{H}$, let $\beta \in \mathbb{R}_{++}$, and let $h\colon C \to \mathbb{R}$ be β–Lipschitz continuous. Set*

$$f\colon \mathcal{H} \to \left]-\infty,+\infty\right] : x \mapsto \begin{cases} h(x), & \text{if } x \in C; \\ +\infty, & \text{otherwise.} \end{cases} \tag{12.18}$$

Then $f \,\square\, (\beta\|\cdot\|)$ is a β–Lipschitz continuous extension of h.

Proof. Set $g = f \,\square\, (\beta\|\cdot\|)$. Then, for every $x \in C$,

$$g(x) = \inf_{y\in\mathcal{H}} \big(f(y)+\beta\|x-y\|\big) = \inf_{y\in C} \big(h(y)+\beta\|x-y\|\big) = h(x) > -\infty. \tag{12.19}$$

Hence, $g \not\equiv -\infty$ and Proposition 12.17 implies that g is the largest β-Lipschitz continuous minorant of f. In particular, $g|_C \leqslant f|_C = h$. On the other hand, (12.19) yields $g|_C \geqslant h$. Altogether, $g|_C = h$ and the proof is complete. ☐

12.4 Moreau Envelope and Proximity Operator

The most important instance of (12.6) is obtained when $p = 2$.

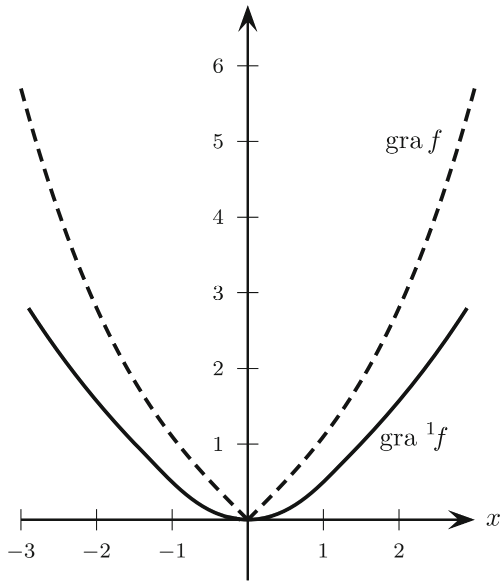

Fig. 12.1 Graphs of $f\colon x \mapsto |x| + 0.1|x|^3$ and of its Moreau envelope ^1f.

Definition 12.20 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ and let $\gamma \in \mathbb{R}_{++}$. The *Moreau envelope* of f of parameter γ is

$$^{\gamma}f = f \,\Box \left(\frac{1}{2\gamma}\|\cdot\|^2\right). \tag{12.20}$$

Example 12.21 Let C be a subset of $\mathcal{H}$ and let $\gamma \in \mathbb{R}_{++}$. Then $^{\gamma}\iota_C = (2\gamma)^{-1}d_C^2$.

As a special case of (12.6) with $p > 1$, the Moreau envelope inherits all the properties recorded in Proposition 12.9 and Proposition 12.15. In addition, it possesses specific properties that we now examine.

Proposition 12.22 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$, $\gamma \in \mathbb{R}_{++}$, and $\mu \in \mathbb{R}_{++}$. Then the following hold:*

(i) $^{\mu}(\gamma f) = \gamma\,(^{\gamma\mu}f)$.
(ii) $^{\gamma}(^{\mu}f) = {}^{(\gamma+\mu)}f$.

Proof. Fix $x \in \mathcal{H}$.
(i): We derive from (12.20) that

$$^{\mu}(\gamma f)(x) = \gamma \inf_{y\in\mathcal{H}} \left(f(y) + \frac{1}{2\gamma\mu}\|x-y\|^2\right) = \gamma\,(^{\gamma\mu}f)(x). \tag{12.21}$$

(ii): Set $\alpha = \mu/(\mu+\gamma)$. Then it follows from Corollary 2.15 that

$$\begin{aligned}
{}^{\gamma}({}^{\mu}f)(x) &= \inf_{z\in\mathcal{H}} \left(\inf_{y\in\mathcal{H}} \left(f(y) + \frac{1}{2\mu}\|z-y\|^2 \right) + \frac{1}{2\gamma}\|z-x\|^2 \right) \\
&= \inf_{y\in\mathcal{H}} \left(f(y) + \frac{1}{2\alpha\gamma} \inf_{z\in\mathcal{H}} \left(\alpha\|z-x\|^2 + (1-\alpha)\|z-y\|^2\right) \right) \\
&= \inf_{y\in\mathcal{H}} \Bigg(f(y) \\
&\qquad + \frac{1}{2\alpha\gamma} \inf_{z\in\mathcal{H}} \left(\|z-(\alpha x+(1-\alpha)y)\|^2 + \alpha(1-\alpha)\|x-y\|^2\right) \Bigg) \\
&= \inf_{y\in\mathcal{H}} \left(f(y) + \frac{1}{2(\gamma+\mu)}\|x-y\|^2 \right), \qquad (12.22)
\end{aligned}$$

which yields ${}^{\gamma}({}^{\mu}f)(x) = {}^{(\gamma+\mu)}f(x)$. ☐

The following definition concerns the case when $p = 2$ in Proposition 12.15.

Definition 12.23 Let $f \in \Gamma_0(\mathcal{H})$ and let $x \in \mathcal{H}$. Then $\operatorname{Prox}_f x$ is the unique point in $\mathcal{H}$ that satisfies

$$ {}^{1}f(x) = \min_{y\in\mathcal{H}} \left(f(y) + \frac{1}{2}\|x-y\|^2 \right) = f(\operatorname{Prox}_f x) + \frac{1}{2}\|x - \operatorname{Prox}_f x\|^2. \quad (12.23)$$

The operator $\operatorname{Prox}_f \colon \mathcal{H} \to \mathcal{H}$ is the *proximity operator*—or *proximal mapping*—of f.

Remark 12.24 Let $f \in \Gamma_0(\mathcal{H})$, let $\gamma \in \mathbb{R}_{++}$, and let $x \in \mathcal{H}$. Proposition 12.22(i) with $\mu = 1$ yields ${}^{1}(\gamma f) = \gamma({}^{\gamma}f)$. Hence we derive from (12.23) that

$$ {}^{\gamma}f(x) = f(\operatorname{Prox}_{\gamma f} x) + \frac{1}{2\gamma}\|x - \operatorname{Prox}_{\gamma f} x\|^2. \quad (12.24)$$

Example 12.25 Let C be a nonempty closed convex subset of $\mathcal{H}$. Then $\operatorname{Prox}_{\iota_C} = P_C$.

The next two results generalize (3.10) and Proposition 4.16, respectively.

Proposition 12.26 *Let $f \in \Gamma_0(\mathcal{H})$, and let x and p be in $\mathcal{H}$. Then*

$$ p = \operatorname{Prox}_f x \quad \Leftrightarrow \quad (\forall y \in \mathcal{H}) \quad \langle y - p \mid x - p\rangle + f(p) \leqslant f(y). \quad (12.25)$$

Proof. Let $y \in \mathcal{H}$. First, suppose that $p = \operatorname{Prox}_f x$ and set $(\forall \alpha \in \left]0,1\right[)$ $p_\alpha = \alpha y + (1-\alpha)p$. Then, for every $\alpha \in \left]0,1\right[$, (12.23) and the convexity of f yield

$$\begin{aligned} f(p) &\leqslant f(p_\alpha) + \frac{1}{2}\|x - p_\alpha\|^2 - \frac{1}{2}\|x - p\|^2 \\ &\leqslant \alpha f(y) + (1-\alpha) f(p) - \alpha \langle x - p \mid y - p \rangle + \frac{\alpha^2}{2}\|y - p\|^2 \end{aligned} \tag{12.26}$$

and hence $\langle y - p \mid x - p \rangle + f(p) \leqslant f(y) + (\alpha/2)\|y - p\|^2$. Letting $\alpha \downarrow 0$, we obtain the desired inequality. Conversely, suppose that $\langle y - p \mid x - p \rangle + f(p) \leqslant f(y)$. Then certainly $f(p) + (1/2)\|x - p\|^2 \leqslant f(y) + (1/2)\|x - p\|^2 + \langle x - p \mid p - y \rangle + (1/2)\|p - y\|^2 = f(y) + (1/2)\|x - y\|^2$ and we conclude that $p = \operatorname{Prox}_f x$. □

Proposition 12.27 *Let $f \in \Gamma_0(\mathcal{H})$, let $x \in \mathcal{H}$, and set $\phi\colon \mathbb{R}_{++} \to \mathbb{R}\colon \gamma \mapsto f(\operatorname{Prox}_{\gamma f} x)$. Then ϕ is decreasing, $\sup \phi(\mathbb{R}_{++}) \leqslant f(x)$, and $(\forall \gamma \in \mathbb{R}_{++})$ $\|x - \operatorname{Prox}_{\gamma f} x\|^2 + \gamma f(\operatorname{Prox}_{\gamma f} x) \leqslant \gamma f(x)$.*

Proof. Let $\gamma \in \mathbb{R}_{++}$ and $\mu \in \mathbb{R}_{++}$, and set $p = \operatorname{Prox}_{\gamma f} x$ and $q = \operatorname{Prox}_{\mu f} x$. By Proposition 12.26, $\langle q - p \mid x - p \rangle \leqslant -\gamma(f(p) - f(q))$ and $\langle p - q \mid x - q \rangle \leqslant \mu(f(p) - f(q))$. Adding these inequalities yields $\|p - q\|^2 \leqslant (\mu - \gamma)(f(p) - f(q))$, hence $(\mu - \gamma)(\phi(\gamma) - \phi(\mu)) \geqslant 0$. Finally, we derive from (12.25) that $\|x - \operatorname{Prox}_{\gamma f} x\|^2 + \gamma f(\operatorname{Prox}_{\gamma f} x) \leqslant \gamma f(x)$ and so $\phi(\gamma) = f(\operatorname{Prox}_{\gamma f} x) \leqslant f(x)$. □

Proposition 12.28 *Let $f \in \Gamma_0(\mathcal{H})$. Then Prox_f and $\operatorname{Id} - \operatorname{Prox}_f$ are firmly nonexpansive.*

Proof. Take x and y in $\mathcal{H}$, and set $p = \operatorname{Prox}_f x$ and $q = \operatorname{Prox}_f y$. Then Proposition 12.26 yields $\langle q - p \mid x - p \rangle + f(p) \leqslant f(q)$ and $\langle p - q \mid y - q \rangle + f(q) \leqslant f(p)$. Since p and q lie in $\operatorname{dom} f$, upon adding these two inequalities, we get $0 \leqslant \langle p - q \mid (x - p) - (y - q) \rangle$ and conclude via Proposition 4.4. □

Proposition 12.29 *Let $f \in \Gamma_0(\mathcal{H})$. Then*

$$\operatorname{Fix} \operatorname{Prox}_f = \operatorname{Argmin} f. \tag{12.27}$$

Proof. Let $x \in \mathcal{H}$. Then it follows from Proposition 12.26 that $x = \operatorname{Prox}_f x \Leftrightarrow (\forall y \in \mathcal{H})\ \langle y - x \mid x - x \rangle + f(x) \leqslant f(y) \Leftrightarrow (\forall y \in \mathcal{H})\ f(x) \leqslant f(y) \Leftrightarrow x \in \operatorname{Argmin} f$. □

It follows from Proposition 12.15 that the Moreau envelope of $f \in \Gamma_0(\mathcal{H})$ is convex, real-valued, and continuous. The next result states that it is actually Fréchet differentiable on $\mathcal{H}$.

Proposition 12.30 *Let $f \in \Gamma_0(\mathcal{H})$ and let $\gamma \in \mathbb{R}_{++}$. Then ${}^{\gamma}f\colon \mathcal{H} \to \mathbb{R}$ is Fréchet differentiable on $\mathcal{H}$, and its gradient*

$$\nabla({}^{\gamma}f) = \gamma^{-1}\big(\operatorname{Id} - \operatorname{Prox}_{\gamma f}\big) \tag{12.28}$$

is γ^{-1}-Lipschitz continuous.

Proof. Assume that x and y are distinct points in $\mathcal{H}$, and set $p = \operatorname{Prox}_{\gamma f} x$ and $q = \operatorname{Prox}_{\gamma f} y$. Using (12.24) and Proposition 12.26, we obtain

$$\begin{aligned} {}^{\gamma}f(y) - {}^{\gamma}f(x) &= f(q) - f(p) + \big(\|y-q\|^2 - \|x-p\|^2\big)/(2\gamma) \\ &\geqslant \big(2\langle q-p \mid x-p\rangle + \|y-q\|^2 - \|x-p\|^2\big)/(2\gamma) \\ &= \big(\|y-q-x+p\|^2 + 2\langle y-x \mid x-p\rangle\big)/(2\gamma) \\ &\geqslant \langle y-x \mid x-p\rangle/\gamma. \end{aligned} \tag{12.29}$$

Likewise, ${}^{\gamma}f(y) - {}^{\gamma}f(x) \leqslant \langle y-x \mid y-q\rangle/\gamma$. Combining the last two inequalities and using the firm nonexpansiveness of Prox_f (Proposition 12.28), we get

$$\begin{aligned} 0 &\leqslant {}^{\gamma}f(y) - {}^{\gamma}f(x) - \langle y-x \mid x-p\rangle/\gamma \\ &\leqslant \langle y-x \mid (y-q)-(x-p)\rangle/\gamma \\ &\leqslant (\|y-x\|^2 - \|q-p\|^2)/\gamma \\ &\leqslant \|y-x\|^2/\gamma. \end{aligned} \tag{12.30}$$

Thus, $\lim_{y\to x}\big({}^{\gamma}f(y) - {}^{\gamma}f(x) - \big\langle y-x \mid \gamma^{-1}(x-p)\big\rangle\big)/\|y-x\| = 0$. Finally, Lipschitz continuity follows from Proposition 12.28. ☐

Corollary 12.31 *Let C be a nonempty closed convex subset of $\mathcal{H}$. Then d_C^2 is Fréchet differentiable on $\mathcal{H}$ and*

$$\nabla d_C^2 = 2(\mathrm{Id} - P_C). \tag{12.31}$$

Proof. Apply Proposition 12.30 with $f = \iota_C$ and $\gamma = 1/2$, and use Example 12.21 and Example 12.25. ☐

Proposition 12.32 *Let K be a nonempty closed convex cone in $\mathcal{H}$, and set $q = (1/2)\|\cdot\|^2$. Then $\nabla(q\circ P_K) = \nabla((1/2)d_{K^\ominus}^2) = P_K$.*

Proof. Using Theorem 6.30(i) and Corollary 12.31, we obtain $\nabla(q\circ P_K) = \nabla(q\circ(\mathrm{Id} - P_{K^\ominus})) = \nabla((1/2)d_{K^\ominus}^2) = \mathrm{Id} - P_{K^\ominus} = P_K$. ☐

Proposition 12.33 *Let $f \in \Gamma_0(\mathcal{H})$ and let $x \in \mathcal{H}$. Then the nets $({}^{\gamma}f(x))_{\gamma\in\mathbb{R}_{++}}$ and $(f(\operatorname{Prox}_{\gamma f} x))_{\gamma\in\mathbb{R}_{++}}$ are decreasing and the following hold:*

(i) ${}^{\gamma}f(x) \downarrow \inf f(\mathcal{H})$ *and* $f(\operatorname{Prox}_{\gamma f} x) \downarrow \inf f(\mathcal{H})$ *as* $\gamma \uparrow +\infty$.
(ii) ${}^{\gamma}f(x) \uparrow f(x)$ *as* $\gamma \downarrow 0$.
(iii) *Suppose that $x \in \operatorname{dom} f$. Then $f(\operatorname{Prox}_{\gamma f} x) \uparrow f(x)$ and, furthermore, $\gamma^{-1}\|x - \operatorname{Prox}_{\gamma f} x\|^2 \to 0$ as $\gamma \downarrow 0$.*

Proof. The nets are decreasing by Proposition 12.9(ii) and Proposition 12.27.

(i): Since $\inf f(\mathcal{H}) \leqslant f(\operatorname{Prox}_{\gamma f} x) \leqslant f(\operatorname{Prox}_{\gamma f} x) + (2\gamma)^{-1}\|x - \operatorname{Prox}_{\gamma f} x\|^2 = {}^{\gamma}f(x)$ by (12.24), we obtain (i) from Proposition 12.9(iv).

(ii)&(iii): Observe that $f(\operatorname{Prox}_{\gamma f} x) \uparrow \sup_{\kappa\in\mathbb{R}_{++}} f(\operatorname{Prox}_{\kappa f} x) \leqslant f(x)$ as $\gamma \downarrow 0$ by Proposition 12.27. Now set $\mu = \sup_{\gamma\in\mathbb{R}_{++}} {}^{\gamma}f(x)$. It follows from

Proposition 12.9(ii) that ${}^{\gamma}f(x) \uparrow \mu \leqslant f(x)$ as $\gamma \downarrow 0$. Therefore, we assume that $\mu < +\infty$, and it is enough to show that $\lim_{\gamma\downarrow 0} {}^{\gamma}f(x) \geqslant \lim_{\gamma\downarrow 0} f(\operatorname{Prox}_{\gamma f} x) \geqslant f(x)$. We deduce from (12.24) that

$$(\forall \gamma \in \mathbb{R}_{++}) \quad \mu \geqslant {}^{\gamma}f(x) = f(\operatorname{Prox}_{\gamma f} x) + \frac{1}{2\gamma}\|x - \operatorname{Prox}_{\gamma f} x\|^2. \tag{12.32}$$

Now set $g = f + (1/2)\|x - \cdot\|^2$. Then (12.32) implies that $(\forall \gamma \in \,]0,1[)$ $\operatorname{Prox}_{\gamma f} x \in \operatorname{lev}_{\leqslant \mu} g$. Since g is coercive by Corollary 11.16(i), we derive from Proposition 11.12 that $\nu = \sup_{\gamma \in]0,1[} \|\operatorname{Prox}_{\gamma f} x\| < +\infty$. On the other hand, Theorem 9.20 asserts that there exist $u \in \mathcal{H}$ and $\eta \in \mathbb{R}$ such that $f \geqslant \langle \cdot \mid u\rangle + \eta$. Therefore, (12.32) yields

$$\begin{aligned}\big(\forall \gamma \in \,]0,1[\,\big) \quad \mu &\geqslant \langle \operatorname{Prox}_{\gamma f} x \mid u\rangle + \eta + \frac{1}{2\gamma}\|x - \operatorname{Prox}_{\gamma f} x\|^2 \\ &\geqslant -\nu\|u\| + \eta + \frac{1}{2\gamma}\|x - \operatorname{Prox}_{\gamma f} x\|^2,\end{aligned} \tag{12.33}$$

from which we deduce that

$$\|x - \operatorname{Prox}_{\gamma f} x\|^2 \leqslant 2\gamma\big(\mu + \nu\|u\| - \eta\big) \to 0 \quad \text{as} \quad \gamma \downarrow 0. \tag{12.34}$$

In turn, since f is lower semicontinuous, it follows from Proposition 12.27 that

$$\begin{aligned}\lim_{\gamma\downarrow 0} {}^{\gamma}f(x) &= \lim_{\gamma\downarrow 0}\left(f(\operatorname{Prox}_{\gamma f} x) + \frac{1}{2\gamma}\|x - \operatorname{Prox}_{\gamma f} x\|^2\right) \\ &\geqslant \underline{\lim}_{\gamma\downarrow 0} f(\operatorname{Prox}_{\gamma f} x) \\ &\geqslant f(x),\end{aligned} \tag{12.35}$$

which provides the desired inequalities. This establishes (ii). Finally, (iii) follows using (12.24). ☐

12.5 Infimal Postcomposition

Definition 12.34 Let $f\colon \mathcal{H} \to [-\infty, +\infty]$, let $\mathcal{K}$ be a real Hilbert space, and let $L\colon \mathcal{H} \to \mathcal{K}$. The *infimal postcomposition* of f by L is

$$L \triangleright f\colon \mathcal{K} \to [-\infty, +\infty] : y \mapsto \inf f\big(L^{-1}\{y\}\big) = \inf_{\substack{x \in \mathcal{H} \\ Lx = y}} f(x), \tag{12.36}$$

and it is *exact at a point* $y \in \mathcal{K}$ if $(L \triangleright f)(y) = \min_{x \in L^{-1}\{y\}} f(x)$, i.e.,

$$(\exists\, x \in \mathcal{H}) \quad Lx = y \quad \text{and} \quad (L \triangleright f)(y) = f(x) \in \,]-\infty, +\infty]\,; \tag{12.37}$$

$L \triangleright f$ is *exact* if it is exact at every point of its domain, in which case it is denoted by $L \mathbin{\blacktriangleright} f$.

Example 12.35 Set $\mathcal{K} = \mathbb{R}^N$, set $I = \{1, \ldots, N\}$, let m be a strictly positive integer, and set $K = \{1, \ldots, m\}$. Let $p \in [1, +\infty[$ and, for every $x = (\xi_i)_{i \in I} \in \mathcal{K}$, set $\|x\|_p = (\sum_{i \in I} |\xi_i|^p)^{1/p}$. Let $(I_k)_{k \in K}$ be nonempty subsets of I such that $\bigcup_{k \in K} I_k = I$ and set

$$\mathcal{H} = \big\{ \boldsymbol{x} = (x_k)_{k \in K} \;\big|\; (\forall k \in K)\; x_k = (\xi_{i,k})_{i \in I} \in \mathcal{K} \text{ and } (\forall i \in I \smallsetminus I_k)\; \xi_{i,k} = 0 \big\}. \tag{12.38}$$

Set $L \colon \mathcal{H} \to \mathcal{K} \colon \boldsymbol{x} \mapsto \sum_{k \in K} x_k$ and $f \colon \mathcal{H} \to \mathbb{R} \colon \boldsymbol{x} \mapsto \sum_{k \in K} \|x_k\|_p$. Then the *latent group lasso* (or *group lasso with overlap*) penalty is $\|\cdot\|_{\mathsf{lgl}} = L \rhd f$, that is,

$$(\forall y \in \mathcal{K}) \quad \|y\|_{\mathsf{lgl}} = \inf_{\substack{\boldsymbol{x} \in \mathcal{H} \\ \sum_{k \in K} x_k = y}} \sum_{k \in K} \|x_k\|_p. \tag{12.39}$$

If the sets $(I_k)_{k \in K}$ are pairwise disjoint, then $\|\cdot\|_{\mathsf{lgl}}$ reduces to the *group lasso* penalty.

Proposition 12.36 *Let $f \colon \mathcal{H} \to]-\infty, +\infty]$ and let $L \colon \mathcal{H} \to \mathcal{K}$, where $\mathcal{K}$ is a real Hilbert space. Then the following hold:*

(i) $\operatorname{dom}(L \rhd f) = L(\operatorname{dom} f)$.
(ii) *Suppose that f is convex and that L is affine. Then $L \rhd f$ is convex.*

Proof. (i): Take $y \in \mathcal{K}$. Then $y \in \operatorname{dom}(L \rhd f) \Leftrightarrow [\,(\exists x \in \mathcal{H})\; f(x) < +\infty$ and $Lx = y\,] \Leftrightarrow y \in L(\operatorname{dom} f)$.

(ii): The function $F \colon \mathcal{K} \times \mathcal{H} \to [-\infty, +\infty] : (y, x) \mapsto f(x) + \iota_{\operatorname{gra} L}(x, y)$ is convex and so is its marginal function $L \rhd f$ by Proposition 8.35. ☐

Proposition 12.37 *Let f and g be functions from $\mathcal{H}$ to $]-\infty, +\infty]$ and set $L \colon \mathcal{H} \times \mathcal{H} \to \mathcal{H} \colon (x, y) \mapsto x + y$. Then $f \,\square\, g = L \rhd (f \oplus g)$.*

Proof. A direct consequence of Definition 12.1 and Definition 12.34. ☐

Exercises

Exercise 12.1 Let $f \colon \mathcal{H} \to [-\infty, +\infty]$. Determine $f \,\square\, 0$ and $f \,\square\, \iota_{\{0\}}$.

Exercise 12.2 Let $f \colon \mathcal{H} \to]-\infty, +\infty]$ be subadditive and such that $f(0) = 0$. Show that $f \,\square\, f = f$.

Exercise 12.3 Let f and g be functions in $\Gamma_0(\mathcal{H})$. Show that the set $\operatorname{epi} f + \operatorname{epi} g$ is closed if and only if $f \,\square\, g$ is lower semicontinuous and exact on $\big\{ z \in \mathcal{H} \;\big|\; (f \,\square\, g)(z) > -\infty \big\}$.

Exercise 12.4 Provide continuous and convex functions f and g from $\mathbb{R}$ to $\mathbb{R}$ such that $\operatorname{epi} f + \operatorname{epi} g$ is strictly contained in $\operatorname{epi}(f \,\square\, g)$.

Exercise 12.5 Find functions f and g in $\Gamma_0(\mathbb{R})$ such that $f \mathbin{\Box} g = f \mathbin{\Box} 0$ while $g \neq 0$.

Exercise 12.6 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper, and let $g\colon \mathcal{H} \to \mathbb{R}$ be sublinear. Suppose that there exists $\beta \in \mathbb{R}_+$ such that g is β-Lipschitz, and that $f \mathbin{\Box} g$ is proper. Show that $f \mathbin{\Box} g$ is β-Lipschitz. Compare with Example 1.48.

Exercise 12.7 Let C be a subset of $\mathcal{H}$. Show that d_C is convex if and only if $\overline{C}$ is convex.

Exercise 12.8 Let K be a nonempty closed convex cone in $\mathcal{H}$, let $\phi\colon \mathbb{R} \to \mathbb{R}$ be an even convex function, and set $f\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \phi(\|P_K x\|)$. Show that f is convex. Is this property still true if K is an arbitrary nonempty closed convex set?

Exercise 12.9 Let $f \in \Gamma_0(\mathcal{H})$, let $\gamma \in \mathbb{R}_{++}$, let $\alpha \in \mathbb{R}_{++}$, and let $u \in \mathcal{H}$.

(i) Set $g = f^{\vee}$. Show that ${}^{\gamma}g = \big({}^{\gamma}f\big)^{\vee}$.
(ii) Set $g = f + (2\alpha)^{-1}\|\cdot\|^2$ and $\beta = (\alpha\gamma)/(\alpha+\gamma)$. Show that

$$ {}^{\gamma}g = \frac{1}{2(\alpha+\gamma)}\|\cdot\|^2 + \big({}^{\beta}f\big)\left(\frac{\alpha\cdot}{\alpha+\gamma}\right). \tag{12.40}$$

(iii) Set $g = f + \langle \cdot \mid u\rangle$. Show that ${}^{\gamma}g = \langle \cdot \mid u\rangle - (\gamma/2)\|u\|^2 + {}^{\gamma}f(\cdot - \gamma u)$.

Exercise 12.10 Compute the proximity operator and the Moreau envelope ${}^{1}f$ for $f \in \Gamma_0(\mathbb{R})$ in the following cases:

(i) $f = |\cdot|$.
(ii) $f = |\cdot|^3$.
(iii) $f = |\cdot| + |\cdot|^3$ (see Figure 12.1).

Exercise 12.11 Let $f \in \Gamma_0(\mathcal{H})$, and set $f_0 = \iota_{\{0\}}$ and $(\forall n \in \mathbb{N})$ $f_{n+1} = f \mathbin{\Box} f_n$. Show that

$$ \big(\forall n \in \mathbb{N} \smallsetminus \{0\}\big) \quad f_n = nf\Big(\frac{\cdot}{n}\Big). \tag{12.41}$$

Exercise 12.12 Let $\rho \in \mathbb{R}_{++}$ and let $p \in \left]1,+\infty\right[$. In connection with Example 8.44 and Proposition 12.15, show that $f = (\rho|\cdot|) \mathbin{\Box} ((1/p)|\cdot|^p)$ is the *Huber function of order* p, i.e.,

$$ f\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto \begin{cases} \rho|x| - \dfrac{p-1}{p}\rho^{\frac{p}{p-1}}, & \text{if } |x| > \rho^{1/(p-1)};\\ \dfrac{|x|^p}{p}, & \text{if } |x| \leqslant \rho^{1/(p-1)}. \end{cases} \tag{12.42}$$

Exercise 12.13 Suppose that $\mathcal{H} = \mathbb{R}$ and set $f\colon x \mapsto (1/2)|x|^2$. Show that the following operators are well defined and find an explicit expression for them:

(i) $P\colon \mathbb{R} \to \mathbb{R}\colon z \mapsto \underset{x\in\mathbb{R}}{\operatorname{argmin}} \ \big(f(x) + (1/2)|z-x|^2\big) = \operatorname{Prox}_f z$.

(ii) $Q\colon \mathbb{R} \to \mathbb{R}\colon z \mapsto \underset{x\in\mathbb{R}}{\operatorname{argmin}} \ \big(f(x) + |z-x|\big)$.

What happens if one iterates either P or Q?

Exercise 12.14 Let $\beta \in \mathbb{R}_{++}$. Show that the β-Pasch–Hausdorff envelope of a convex function from $\mathcal{H}$ to $\left]-\infty,+\infty\right]$ is convex. In addition, find a function from $\mathcal{H}$ to $\left]-\infty,+\infty\right]$ that has a nonconvex β-Pasch–Hausdorff envelope.

Exercise 12.15 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$, and let $g\colon \mathcal{H} \to \mathbb{R}$ be upper semicontinuous. Show that

$$f \,\square\, g = \bar{f} \,\square\, g, \tag{12.43}$$

where $\bar{f}$ is the lower semicontinuous envelope of f defined in (1.44).

Exercise 12.16 Let $A \in \mathcal{B}(\mathcal{H})$ be self-adjoint, strictly positive, and surjective, set $q_A\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \frac{1}{2}\langle x \mid Ax\rangle$, and let $b \in \mathcal{H}$. Show that

$$(\forall x \in \mathcal{H}) \quad q_A(x) - \langle x \mid b\rangle = q_A(A^{-1}b) - \big\langle A^{-1}b \mid b\big\rangle + q_A(x - A^{-1}b). \tag{12.44}$$

Deduce that the unique minimizer of the function $q_A - \langle \cdot \mid b\rangle$ is $A^{-1}b$, at which point the value is $-q_A(A^{-1}b) = -q_{A^{-1}}(b)$.

Exercise 12.17 Let A and B be self-adjoint, strictly positive, surjective operators in $\mathcal{B}(\mathcal{H})$ such that $A+B$ is surjective, and set $q_A\colon x \mapsto \frac{1}{2}\langle x \mid Ax\rangle$. Show that

$$q_A \,\boxdot\, q_B = q_{(A^{-1}+B^{-1})^{-1}}. \tag{12.45}$$

Exercise 12.18 For all $y \in \mathcal{H}$ and $\alpha \in \mathbb{R}_{++}$, set $q_{y,\alpha}\colon \mathcal{H} \to \left]-\infty,+\infty\right] : x \mapsto \frac{1}{2}\|\alpha^{-1}(x-y)\|^2$ if $\alpha > 0$; and $q_{y,\alpha} = \iota_{\{y\}}$ if $\alpha = 0$. Let y and z be in $\mathcal{H}$, and let α and β be in $\mathbb{R}$. Show that

$$q_{y,\alpha} \,\boxdot\, q_{z,\beta} = q_{y+z,\sqrt{\alpha^2+\beta^2}}. \tag{12.46}$$

Exercise 12.19 Show that the conclusion of Proposition 12.33(iii) fails if $x \notin \operatorname{dom} f$.

Exercise 12.20 Let f and g be functions from $\mathcal{H}$ to $\left]-\infty,+\infty\right]$. Define the *strict epigraph* of f by

$$\operatorname{epi}_< f = \big\{(x,\xi) \in \mathcal{H} \times \mathbb{R} \;\big|\; f(x) < \xi\big\}. \tag{12.47}$$

Show that $\operatorname{epi}_< f + \operatorname{epi}_< g = \operatorname{epi}_<(f \,\square\, g)$.

Exercise 12.21 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$, let $\mathcal{K}$ be a real Hilbert space, let $L\colon \mathcal{H} \to \mathcal{K}$, and use the same notation as in (12.47). Show that $\operatorname{epi}_<(L \rhd f) = (L \times \operatorname{Id})(\operatorname{epi}_< f) = \big\{(Lx,\xi) \;\big|\; (x,\xi) \in \operatorname{epi}_< f\big\} \subset \mathcal{K} \times \mathbb{R}$.

Chapter 13
Conjugation

Functional transforms make it possible to investigate problems from a different perspective and sometimes simplify their investigation. In convex analysis, the most suitable notion of a transform is the Legendre transform, which maps a function to its Fenchel conjugate. This transform is studied in detail in this chapter. In particular, it is shown that the conjugate of an infimal convolution is the sum of the conjugates. The key result of this chapter is the Fenchel–Moreau theorem, which states that the proper convex lower semicontinuous functions are precisely those functions that coincide with their biconjugates.

13.1 Definition and Examples

Definition 13.1 Let $f\colon \mathcal{H} \to [-\infty, +\infty]$. The *conjugate* (or *Fenchel conjugate*, or *Legendre transform*, or *Legendre–Fenchel transform*) of f is

$$f^*\colon \mathcal{H} \to [-\infty, +\infty] : u \mapsto \sup_{x\in\mathcal{H}} \big(\langle x \mid u\rangle - f(x)\big), \tag{13.1}$$

and the *biconjugate* of f is $f^{**} = (f^*)^*$.

Let us illustrate Definition 13.1 through a variety of examples (see also Figure 13.1).

Example 13.2 Suppose that $\mathcal{H} = \mathbb{R}$.

(i) Let $p \in\,]1, +\infty[$ and set $p^* = p/(p-1)$. Then

$$\left(\frac{|\cdot|^p}{p}\right)^* = \frac{|\cdot|^{p^*}}{p^*}. \tag{13.2}$$

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-48311-5_31

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_13

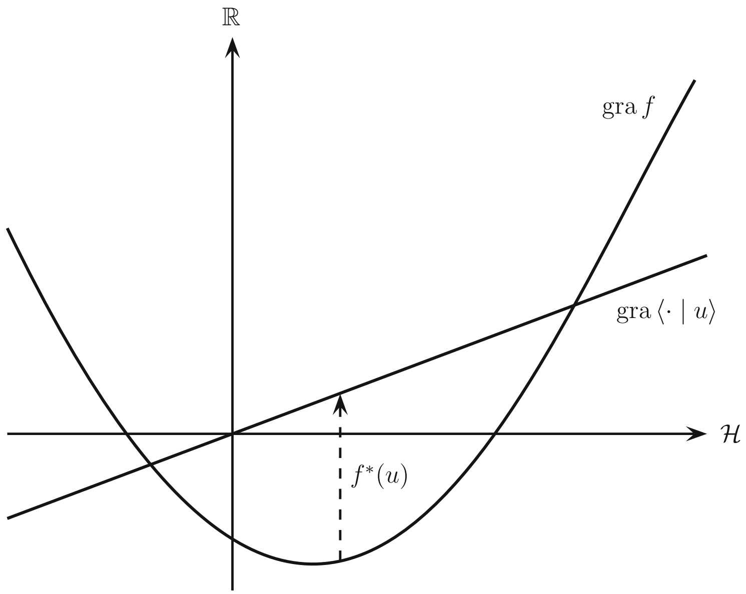

Fig. 13.1 $f^*(u)$ is the supremum of the signed vertical distance between the graph of f and that of the continuous linear functional $\langle \cdot \mid u \rangle$.

(ii) Let $f\colon x \mapsto \begin{cases} 1/x, & \text{if } x > 0; \\ +\infty, & \text{if } x \leqslant 0. \end{cases}$

Then $f^*\colon u \mapsto \begin{cases} -2\sqrt{-u}, & \text{if } u \leqslant 0; \\ +\infty, & \text{if } u > 0. \end{cases}$

(iii) Let $f\colon x \mapsto \begin{cases} -\ln(x), & \text{if } x > 0; \\ +\infty, & \text{if } x \leqslant 0 \end{cases}$

(negative *Burg entropy*). Then $f^*\colon u \mapsto \begin{cases} -\ln(-u) - 1, & \text{if } u < 0; \\ +\infty, & \text{if } u \geqslant 0. \end{cases}$

(iv) $\cosh^*\colon u \mapsto u \operatorname{arcsinh}(u) - \sqrt{u^2+1}$.

(v) $\exp^*\colon u \mapsto \begin{cases} u\ln(u) - u, & \text{if } u > 0; \\ 0, & \text{if } u = 0; \\ +\infty, & \text{if } u < 0 \end{cases}$

(negative *Boltzmann–Shannon entropy*).

(vi) Let $f\colon x \mapsto \begin{cases} x\ln(x) + (1-x)\ln(1-x), & \text{if } x \in \left]0,1\right[;\\ 0, & \text{if } x \in \{0,1\};\\ +\infty, & \text{otherwise}\end{cases}$

(negative *Fermi–Dirac entropy*). Then $f^*\colon u \mapsto \ln(1+e^u)$ and $f^{*\vee}$ is the *logistic loss* function.

(vii) Let $f\colon x \mapsto \begin{cases} x\ln(x) - (x+1)\ln(x+1), & \text{if } x > 0;\\ 0, & \text{if } x = 0;\\ +\infty, & \text{otherwise}\end{cases}$

(*Bose-Einstein entropy*). Then

$$f^*\colon u \mapsto \begin{cases} -\ln(1-e^u), & \text{if } u < 0;\\ +\infty, & \text{if } u \geqslant 0.\end{cases} \tag{13.3}$$

(viii) Let $f\colon x \mapsto \sqrt{1+x^2}$. Then $f^*\colon u \mapsto \begin{cases} -\sqrt{1-u^2}, & \text{if } |u| \leqslant 1;\\ +\infty, & \text{if } |u| > 1.\end{cases}$

Proof. Exercise 13.1. ☐

Example 13.3 Below are some direct applications of (13.1).

(i) Let $f = \iota_C$, where C is a subset of $\mathcal{H}$. Then (7.4) yields $f^* = \sigma_C$.
(ii) Let $f = \iota_K$, where K is a nonempty cone in $\mathcal{H}$. Then (i) yields $f^* = \sigma_K = \iota_{K^\ominus}$.
(iii) Let $f = \iota_V$, where V is a linear subspace of $\mathcal{H}$. Then (ii) yields $f^* = \iota_{V^\ominus} = \iota_{V^\perp}$.
(iv) Let $f = \iota_{B(0;1)}$. Then (i) yields $f^* = \sigma_{B(0;1)} = \sup_{\|x\|\leqslant 1} \langle x \mid \cdot\rangle = \|\cdot\|$.
(v) Let $f = \|\cdot\|$. Then $f^* = \iota_{B(0;1)}$.

Proof. (i)–(iv): Exercise 13.2.

(v): Let $u \in \mathcal{H}$. If $\|u\| \leqslant 1$, then Cauchy–Schwarz yields $0 = \langle 0 \mid u\rangle - \|0\| \leqslant \sup_{x\in\mathcal{H}}(\langle x \mid u\rangle - \|x\|) \leqslant \sup_{x\in\mathcal{H}}(\|x\|(\|u\|-1)) = 0$. Therefore $f^*(u) = 0$. On the other hand, if $\|u\| > 1$, then $\sup_{x\in\mathcal{H}}(\langle x \mid u\rangle - \|x\|) \geqslant \sup_{\lambda\in\mathbb{R}_{++}}(\langle \lambda u \mid u\rangle - \|\lambda u\|) = \|u\|(\|u\|-1)\sup_{\lambda\in\mathbb{R}_{++}} \lambda = +\infty$. Altogether, $f^* = \iota_{B(0;1)}$. ☐

Example 13.4 Let $\varphi\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper, let $\gamma \in \mathbb{R}_{++}$, and set $f = \varphi + \gamma^{-1}q$, where $q = (1/2)\|\cdot\|^2$. Then

$$f^* = \gamma q - {}^{\gamma}\varphi \circ \gamma\mathrm{Id} = \gamma q - (\varphi \,\square\, (\gamma^{-1}q)) \circ \gamma\mathrm{Id}. \tag{13.4}$$

Proof. Let $u \in \mathcal{H}$. Definition 12.20 yields

$$\begin{aligned} f^*(u) &= -\inf_{x\in\mathcal{H}} \big(f(x) - \langle x \mid u\rangle\big) \\ &= \frac{\gamma}{2}\|u\|^2 - \inf_{x\in\mathcal{H}}\left(\varphi(x) + \frac{1}{2\gamma}\|x-\gamma u\|^2\right) \\ &= \frac{\gamma}{2}\|u\|^2 - {}^{\gamma}\varphi(\gamma u), \end{aligned} \tag{13.5}$$

which provides (13.4). ☐

Example 13.5 Let C be a nonempty subset of $\mathcal{H}$ and let $f = \iota_C + \|\cdot\|^2/2$. Then $f^* = (\|\cdot\|^2 - d_C^2)/2$.

Proof. Set $\varphi = \iota_C$ and $\gamma = 1$ in Example 13.4. ☐

Example 13.6 Set $f = (1/2)\|\cdot\|^2$. Then $f^* = f$.

Proof. Set $C = \mathcal{H}$ in Example 13.5. ☐

Example 13.7 Let $\rho \in \mathbb{R}_{++}$ and let $f = \iota_{B(0;\rho)} + (1/2)\|\cdot\|^2$. Then

$$(\forall u \in \mathcal{H}) \quad f^*(u) = \begin{cases} \rho\|u\| - \dfrac{\rho^2}{2}, & \text{if } \|u\| > \rho; \\ \dfrac{\|u\|^2}{2}, & \text{if } \|u\| \leqslant \rho. \end{cases} \tag{13.6}$$

In particular, if $\mathcal{H} = \mathbb{R}$, we infer that $(\iota_{[-\rho,\rho]} + (1/2)|\cdot|^2)^*$ is the Huber function (see Example 8.44).

Proof. Set $C = B(0;\rho)$ in Example 13.5 and use Example 3.18. ☐

Example 13.8 Let $\phi\colon \mathbb{R} \to \left]-\infty,+\infty\right]$ be even. Then $(\phi\circ\|\cdot\|)^* = \phi^*\circ\|\cdot\|$.

Proof. If $\mathcal{H} = \{0\}$, then $(\phi\circ\|\cdot\|)^*(0) = -\phi(0) = (\phi^*\circ\|\cdot\|)(0)$. Now assume that $\mathcal{H} \neq \{0\}$. Then, for every $u \in \mathcal{H}$,

$$\begin{aligned} (\phi\circ\|\cdot\|)^*(u) &= \sup_{\rho\in\mathbb{R}_+}\ \sup_{\|x\|=1} \big(\langle \rho x \mid u\rangle - \phi(\|\rho x\|)\big) \\ &= \sup_{\rho\in\mathbb{R}_+} \big(\rho\|u\| - \phi(\rho)\big) \\ &= \sup_{\rho\in\mathbb{R}} \big(\rho\|u\| - \phi(\rho)\big) \\ &= \phi^*(\|u\|), \end{aligned} \tag{13.7}$$

as required. ☐

Example 13.9 Let $\varphi\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ and let f be its perspective function, defined in (8.18). Then $f^* = \iota_C$, where $C = \big\{(\mu,u) \in \mathbb{R}\times\mathcal{H} \;\big|\; \mu + \varphi^*(u) \leqslant 0\big\}$.

Proof. Let $\mu \in \mathbb{R}$ and $u \in \mathcal{H}$. It follows from (8.18) that

$$
\begin{aligned}
f^*(\mu,u) &= \sup_{\xi\in\mathbb{R}_{++}} \Big(\sup_{x\in\mathcal{H}} \xi\mu + \langle x \mid u\rangle - \xi\varphi(x/\xi) \Big)\\
&= \sup_{\xi\in\mathbb{R}_{++}} \xi\Big(\mu + \sup_{x\in\mathcal{H}} \big(\langle x/\xi \mid u\rangle - \varphi(x/\xi)\big) \Big)\\
&= \sup_{\xi\in\mathbb{R}_{++}} \xi\big(\mu + \varphi^*(u)\big)\\
&= \begin{cases} 0, & \text{if } \mu+\varphi^*(u)\leqslant 0;\\ +\infty, & \text{otherwise,}\end{cases}
\end{aligned}
\tag{13.8}
$$

which completes the proof. ☐

13.2 Basic Properties

Let us first record some direct consequences of Definition 13.1.

Proposition 13.10 *Let $f\colon \mathcal{H}\to[-\infty,+\infty]$. Then the following hold:*

(i) $f^*(0) = -\inf f(\mathcal{H})$.
(ii) $-\infty\in f^*(\mathcal{H}) \Leftrightarrow f\equiv+\infty \Leftrightarrow f^*\equiv-\infty$.
(iii) *Suppose that f^* is proper. Then f is proper.*
(iv) *Let $u\in\mathcal{H}$. Then*

$$f^*(u) = \sup_{x\in\operatorname{dom} f}\big(\langle x\mid u\rangle - f(x)\big) = \sup_{(x,\xi)\in\operatorname{epi} f}\big(\langle x\mid u\rangle - \xi\big).$$

(v) $f^* = \iota^*_{\operatorname{epi} f}(\,\cdot\,,-1) = \sigma_{\operatorname{epi} f}(\,\cdot\,,-1)$.
(vi) *Suppose that f is proper. Then $f^* = \iota^*_{\operatorname{gra} f}(\,\cdot\,,-1) = \sigma_{\operatorname{gra} f}(\,\cdot\,,-1)$.*

Proposition 13.11 *Let $f\colon\mathcal{H}\to\left]-\infty,+\infty\right]$ be proper, let $u\in\mathcal{H}$, and let $\mu\in\mathbb{R}$. Then*

$$\sigma_{\operatorname{epi} f}(u,-\mu) = \begin{cases} \mu f^*(u/\mu), & \text{if } \mu>0;\\ \sigma_{\operatorname{dom} f}(u), & \text{if } \mu=0;\\ +\infty, & \text{if } \mu<0.\end{cases} \tag{13.9}$$

Proof. If $\mu>0$, then $\sigma_{\operatorname{epi} f}(u,-\mu) = \mu\sigma_{\operatorname{epi} f}(u/\mu,-1) = \mu f^*(u/\mu)$ by Proposition 13.10(v). The cases when $\mu=0$ and $\mu<0$ are clear. ☐

Proposition 13.12 *Let $f\colon\mathcal{H}\to[-\infty,+\infty]$. Then the following hold:*

(i) *Let $(u,\mu)\in\mathcal{H}\times\mathbb{R}$. Then $(u,\mu)\in\operatorname{epi} f^* \Leftrightarrow \langle\cdot\mid u\rangle-\mu\leqslant f$.*
(ii) *$f^*\equiv+\infty$ if and only if f possesses no continuous affine minorant.*
(iii) *Suppose that $\operatorname{dom} f^*\neq\varnothing$. Then f is bounded below on every bounded subset of $\mathcal{H}$.*

Proof. (i): $(u,\mu) \in \operatorname{epi} f^* \Leftrightarrow f^*(u) \leqslant \mu \Leftrightarrow (\forall x \in \mathcal{H})\ \langle x \mid u\rangle - f(x) \leqslant \mu$.

(ii): By (i), there is a bijection between the points of the epigraph of f^* and the continuous affine minorants of f, and $\operatorname{epi} f^* = \varnothing \Leftrightarrow f^* \equiv +\infty$.

(iii): By (ii), if $\operatorname{dom} f^* \neq \varnothing$, then f possesses a continuous affine minorant, say $\langle \cdot \mid u\rangle + \mu$. Now let C be a bounded set in $\mathcal{H}$ and let $\beta = \sup_{x\in C} \|x\|$. Then, by Cauchy–Schwarz, $(\forall x \in C)\ f(x) \geqslant \langle x \mid u\rangle + \mu \geqslant -\|x\|\,\|u\| + \mu \geqslant -\beta\|u\| - \mu > -\infty$. ☐

Proposition 13.13 *Let $f\colon \mathcal{H} \to [-\infty,+\infty]$. Then $f^* \in \Gamma(\mathcal{H})$.*

Proof. We assume that $f \not\equiv +\infty$. By Proposition 13.10(iv), f^* is the supremum of the lower semicontinuous convex functions $\big(\langle x \mid \cdot\rangle - \xi\big)_{(x,\xi)\in\operatorname{epi} f}$. The result therefore follows from Proposition 9.3. ☐

Example 13.14 Let $\varphi\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and let $\gamma \in \mathbb{R}_{++}$. Then, by Proposition 13.13, $(\gamma/2)\|\cdot\|^2 - {}^{\gamma}\varphi(\gamma\cdot)$ is lower semicontinuous and convex as a conjugate function (see Example 13.4). Likewise, it follows from Example 13.5 that, for every nonempty subset C of $\mathcal{H}$, $\|\cdot\|^2 - d_C^2 \in \Gamma_0(\mathcal{H})$.

Proposition 13.15 (Fenchel–Young inequality) *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper. Then*

$$(\forall x \in \mathcal{H})(\forall u \in \mathcal{H}) \quad f(x) + f^*(u) \geqslant \langle x \mid u\rangle\,. \tag{13.10}$$

Proof. Fix x and u in $\mathcal{H}$. Since f is proper, it follows from Proposition 13.10(ii) that $-\infty \notin f^*(\mathcal{H})$. Thus, if $f(x) = +\infty$, the inequality trivially holds. On the other hand, if $f(x) < +\infty$, then (13.1) yields $f^*(u) \geqslant \langle x \mid u\rangle - f(x)$ and the inequality follows. ☐

Proposition 13.16 *Let f and g be functions from $\mathcal{H}$ to $[-\infty,+\infty]$. Then the following hold:*

(i) $f^{**} \leqslant f$.
(ii) $f \leqslant g \Rightarrow [\ f^* \geqslant g^*$ *and* $f^{**} \leqslant g^{**}\ ]$.
(iii) $f^{***} = f^*$.
(iv) $(\breve{f}\,)^* = f^*$.

Proof. (i)&(ii): These are direct consequences of (13.1).

(iii): It follows from (i) and (ii) that $f^{***} \geqslant f^*$. On the other hand, (i) applied to f^* yields $f^{***} \leqslant f^*$.

(iv): It follows from Proposition 13.13 and Definition 9.7 that $f^{**} \leqslant \breve{f} \leqslant f$. Hence, by (iii) and (ii), $f^* = f^{***} \geqslant (\breve{f}\,)^* \geqslant f^*$. ☐

Example 13.17 Set $f\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto -|x|$. Then $\breve{f} \equiv -\infty$ and $f^* \equiv +\infty$.

Example 13.18 Suppose that $\mathcal{H} \neq \{0\}$, let $\mathcal{S}(\mathcal{H})$ be the linear subspace of $\mathcal{B}(\mathcal{H})$ of self-adjoint operators, and define the *Löwner partial ordering* on $\mathcal{S}(\mathcal{H})$ by

$$(\forall A\in \mathcal{S}(\mathcal{H}))(\forall B\in \mathcal{S}(\mathcal{H}))\ \big[A\succcurlyeq B \ \Leftrightarrow\ (\forall x\in\mathcal{H})\ \langle Ax \mid x\rangle \geqslant \langle Bx \mid x\rangle\big]. \tag{13.11}$$

We also consider the associated strict ordering

$$(\forall A\in \mathcal{S}(\mathcal{H}))(\forall B\in \mathcal{S}(\mathcal{H}))\ \big[A\succ B \ \Leftrightarrow\ (\forall x\in\mathcal{H}\smallsetminus\{0\})\ \langle Ax \mid x\rangle > \langle Bx \mid x\rangle\big]. \tag{13.12}$$

Let $\alpha\in\mathbb{R}_{++}$, let $\beta\in\mathbb{R}_{++}$, and suppose that A and B are operators in $\mathcal{S}(\mathcal{H})$ such that $\alpha\mathrm{Id}\succcurlyeq A\succcurlyeq B\succcurlyeq \beta\mathrm{Id}$. Then the following hold:

(i) $\beta^{-1}\mathrm{Id}\succcurlyeq B^{-1}\succcurlyeq A^{-1}\succcurlyeq \alpha^{-1}\mathrm{Id}$.
(ii) $(\forall x\in\mathcal{H})\ \big\langle A^{-1}x \mid x\big\rangle \geqslant \|A\|^{-1}\|x\|^2$.
(iii) $\|A^{-1}\|\leqslant \beta^{-1}$.
(iv) A is nonexpansive $\Leftrightarrow$ $\mathrm{Id}\succcurlyeq A^*A$.
(v) A is strictly quasinonexpansive and $\operatorname{Fix} A=\{0\}$ $\Leftrightarrow$ A is strictly nonexpansive $\Leftrightarrow$ $\mathrm{Id}\succ A^*A$.

Proof. (i): It suffices to show that $B^{-1}\succcurlyeq A^{-1}$. Let us set $(\forall x\in\mathcal{H})$ $f(x)=(1/2)\langle Ax \mid x\rangle$ and $g(x)=(1/2)\langle Bx\mid x\rangle$. Then

$$\begin{aligned}(\forall u\in\mathcal{H})\quad f^*(u) &= \sup_{x\in\mathcal{H}}\Big(\langle x\mid u\rangle-\tfrac12\langle Ax\mid x\rangle\Big)\\ &= \tfrac12\big\langle A^{-1}u\mid u\big\rangle-\tfrac12\inf_{x\in\mathcal{H}}\Big(\big\langle A(x-A^{-1}u)\mid x-A^{-1}u\big\rangle\Big)\\ &= \tfrac12\big\langle A^{-1}u\mid u\big\rangle,\end{aligned} \tag{13.13}$$

and g^* is computed likewise. Since $f\geqslant g$, Proposition 13.16(ii) yields $g^*\geqslant f^*$, i.e., $(\forall u\in\mathcal{H})$ $(1/2)\big\langle B^{-1}u\mid u\big\rangle\geqslant(1/2)\big\langle A^{-1}u\mid u\big\rangle$.

(ii): Since $\|A\|\mathrm{Id}\succcurlyeq A$, (i) yields $A^{-1}\succcurlyeq\|A\|^{-1}\mathrm{Id}$.

(iii): We have $A^{-1}\in\mathcal{S}(\mathcal{H})$ and, by (i), $(\forall x\in\mathcal{H})$ $\|x\|^2/\beta\geqslant\big\langle A^{-1}x\mid x\big\rangle$. Taking the supremum over $B(0;1)$ yields $1/\beta\geqslant\|A^{-1}\|$.

(iv): A is nonexpansive $\Leftrightarrow$ $(\forall x\in\mathcal{H})$ $\|Ax\|^2\leqslant\|x\|^2$ $\Leftrightarrow$ $(\forall x\in\mathcal{H})$ $\langle A^*Ax\mid x\rangle\leqslant\langle\mathrm{Id}x\mid x\rangle$ $\Leftrightarrow$ $\mathrm{Id}\succcurlyeq A^*A$.

(v): A is strictly quasinonexpansive and $\operatorname{Fix}A=\{0\}$ $\Leftrightarrow$ $(\forall x\in\mathcal{H}\smallsetminus\{0\})$ $\|Ax\|^2<\|x\|^2$ $\Leftrightarrow$ A is strictly nonexpansive $\Leftrightarrow$ $(\forall x\in\mathcal{H}\smallsetminus\{0\})$ $\langle A^*Ax\mid x\rangle<\langle\mathrm{Id}x\mid x\rangle$ $\Leftrightarrow$ $\mathrm{Id}\succ A^*A$. ☐

Proposition 13.19 (Self-conjugacy) *Let $f\colon\mathcal{H}\to[-\infty,+\infty]$. Then*

$$f=f^*\quad\Leftrightarrow\quad f=(1/2)\|\cdot\|^2. \tag{13.14}$$

Proof. Set $q=(1/2)\|\cdot\|^2$. Then the identity $q^*=q$ is known from Example 13.6. Conversely, if $f=f^*$, then f is proper by virtue of Proposition 13.10(ii), and the Fenchel–Young inequality (Proposition 13.15) yields $(\forall x\in\mathcal{H})$ $2f(x)\geqslant\langle x\mid x\rangle$, i.e., $f\geqslant q$. Therefore, by Proposition 13.16(ii), $q=q^*\geqslant f^*=f$ and we conclude that $f=q$. ☐

Remark 13.20

(i) Suppose that $\mathcal{H} \neq \{0\}$. Since $(1/2)\|\cdot\|^2$ is the only self-conjugate function, a convex cone K in $\mathcal{H}$ cannot be self-polar since $K^{\ominus} = K \Leftrightarrow \iota_K^* = \iota_K$. In particular, we recover the fact that a linear subspace of $\mathcal{H}$ cannot be self-orthogonal.

(ii) Let $f = \iota_K$, where K is a self-dual closed convex cone in $\mathcal{H}$. Then

$$f^* = \sigma_K = \iota_{K^{\ominus}} = \iota_{K^{\oplus}}^{\vee} = f^{\vee}. \tag{13.15}$$

(iii) Another function that satisfies the identity $f^* = f^{\vee}$ is

$$f\colon \mathcal{H} = \mathbb{R} \to \left]-\infty, +\infty\right] : x \mapsto \begin{cases} -\ln(x) - \frac{1}{2}, & \text{if } x > 0; \\ +\infty, & \text{if } x \leqslant 0. \end{cases} \tag{13.16}$$

Proposition 13.21 *Let $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ be even. Then f^* is even.*

Proof. Let $u \in \mathcal{H}$. Then $f^*(-u) = \sup_{x \in \mathcal{H}}(\langle x \mid -u\rangle - f(x)) = \sup_{x \in \mathcal{H}}(\langle x \mid u\rangle - f(-x)) = \sup_{x \in \mathcal{H}}(\langle x \mid u\rangle - f(x)) = f^*(u)$. ☐

Proposition 13.22 *Let $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ be such that $f \geqslant f(0) = 0$. Then $f^* \geqslant f^*(0) = 0$.*

Proof. We have $f^*(0) = -\inf f(\mathcal{H}) = -f(0) = 0$. Moreover, for every $u \in \mathcal{H}$, (13.1) yields $f^*(u) \geqslant \langle 0 \mid u\rangle - f(0) = 0$. ☐

Proposition 13.23 *Let $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$. Then the following hold:*

(i) $(\forall \alpha \in \mathbb{R}_{++})\ (\alpha f)^* = \alpha f^*(\cdot/\alpha)$.
(ii) $(\forall \alpha \in \mathbb{R}_{++})\ (\alpha f(\cdot/\alpha))^* = \alpha f^*$.
(iii) $(\forall y \in \mathcal{H})(\forall v \in \mathcal{H})(\forall \beta \in \mathbb{R})\ \big(\tau_y f + \langle \cdot \mid v\rangle + \beta\big)^* = \tau_v f^* + \langle y \mid \cdot\rangle - \langle y \mid v\rangle - \beta$.
(iv) *Let $L \in \mathcal{B}(\mathcal{H})$ be bijective. Then $(f \circ L)^* = f^* \circ L^{*-1}$.*
(v) $f^{\vee *} = f^{*\vee}$.
(vi) *Let V be a closed linear subspace of $\mathcal{H}$ such that $\operatorname{dom} f \subset V$. Then $\big(f|_V\big)^* \circ P_V = f^* = f^* \circ P_V$.*

Proof. (i)–(iv): Straightforward from (13.1).

(v): Take $L = -\mathrm{Id}$ in (iv).

(vi): Let $u \in \mathcal{H}$. Then $f^*(u) = \sup_{x \in V}(\langle x \mid u\rangle - f(x)) = \sup_{x \in V}(\langle P_V x \mid u\rangle - f|_V(x)) = \sup_{x \in V}(\langle x \mid P_V u\rangle - f|_V(x)) = \big(f|_V\big)^*(P_V u)$. Consequently, we obtain $\big(f|_V\big)^*(P_V u) = \big(f|_V\big)^*(P_V P_V u) = f^*(P_V u)$. ☐

Proposition 13.24 *Let $\mathcal{K}$ be a real Hilbert space, and let f and g be functions from $\mathcal{H}$ to $\left]-\infty, +\infty\right]$. Then the following hold:*

(i) $(f \,\square\, g)^* = f^* + g^*$.
(ii) *Suppose that f and g are proper. Then $(f + g)^* \leqslant f^* \,\square\, g^*$.*

(iii) $(\forall\gamma \in \mathbb{R}_{++})$ $({}^{\gamma}f)^* = f^* + (\gamma/2)\|\cdot\|^2$.
(iv) *Let* $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$. *Then* $(L \triangleright f)^* = f^* \circ L^*$.
(v) *Let* $L \in \mathcal{B}(\mathcal{K},\mathcal{H})$. *Then* $(f \circ L)^* \leqslant L^* \triangleright f^*$.

Proof. (i): For every $u \in \mathcal{H}$, we have

$$\begin{aligned}(f \,\square\, g)^*(u) &= \sup_{x\in\mathcal{H}} \Big(\langle x \mid u\rangle - \inf_{y\in\mathcal{H}} \big(f(y) + g(x-y)\big)\Big) \\ &= \sup_{y\in\mathcal{H}} \Big(\langle y \mid u\rangle - f(y) + \sup_{x\in\mathcal{H}} \big(\langle x-y \mid u\rangle - g(x-y)\big)\Big) \\ &= f^*(u) + g^*(u). \end{aligned} \tag{13.17}$$

(ii): Items (i) and (ii) in Proposition 13.16 yield successively $f + g \geqslant f^{**} + g^{**}$ and $(f+g)^* \leqslant (f^{**}+g^{**})^*$. However, by (i) above, $(f^{**}+g^{**})^* = (f^* \,\square\, g^*)^{**} \leqslant f^* \,\square\, g^*$.

(iii): Take $g = \|\cdot\|^2/(2\gamma)$ in (i).

(iv): Let $v \in \mathcal{K}$. Then

$$\begin{aligned}(L \triangleright f)^*(v) &= \sup_{y\in\mathcal{K}} \big(\langle y \mid v\rangle - \inf_{Lx=y} f(x)\big) \\ &= \sup_{y\in\mathcal{K}} \sup_{x\in L^{-1}\{y\}} \big(\langle Lx \mid v\rangle - f(x)\big) \\ &= \sup_{x\in\mathcal{H}} \big(\langle x \mid L^*v\rangle - f(x)\big) \\ &= f^*(L^*v). \end{aligned} \tag{13.18}$$

(v): By (iv) and Proposition 13.16(i), $(L^* \triangleright f^*)^* = f^{**} \circ L^{**} = f^{**} \circ L \leqslant f \circ L$. Hence, by Proposition 13.16(ii), $L^* \triangleright f^* \geqslant (L^* \triangleright f^*)^{**} \geqslant (f \circ L)^*$. ☐

Corollary 13.25 *Let* $\mathcal{K}$ *be a real Hilbert space, let* $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$, *let* $g\colon \mathcal{K} \to \left]-\infty,+\infty\right]$, *and let* $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$. *Then the following hold:*

(i) *Suppose that* $-\infty \notin (L \triangleright f)(\mathcal{K})$. *Then* $((L \triangleright f) \,\square\, g)^* = (f^* \circ L^*) + g^*$.
(ii) *Suppose that* $\operatorname{dom} f \neq \varnothing$ *and* $\operatorname{dom} g \cap \operatorname{ran} L \neq \varnothing$. *Then* $(f + (g \circ L))^* \leqslant f^* \,\square\, (L^* \triangleright g^*)$.

Proof. (i): By Proposition 13.24(i)&(iv), $((L \triangleright f) \,\square\, g)^* = (L \triangleright f)^* + g^* = (f^* \circ L^*) + g^*$.

(ii): By Proposition 13.24(ii)&(v), $(f + (g \circ L))^* \leqslant f^* \,\square\, (g \circ L)^* \leqslant f^* \,\square\, (L^* \triangleright g^*)$. ☐

Example 13.26 Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\phi\colon \mathbb{R} \to \left]-\infty,+\infty\right]$ be increasing on $\mathbb{R}_+$ and even, and set $f = \phi \circ d_C$. Then $f^* = \sigma_C + \phi^* \circ \|\cdot\|$.

Proof. Since ϕ is increasing on $\mathbb{R}_+$, we have for every $x \in \mathcal{H}$ and every $y \in C$,

$$\inf_{z\in C} \phi\big(\|x-z\|\big) \leqslant \phi\big(\|x - P_C x\|\big) = \phi\Big(\inf_{z\in C}\|x-z\|\Big) \leqslant \phi\big(\|x-y\|\big). \quad (13.19)$$

Taking the infimum over $y \in C$ then yields $(\forall x \in \mathcal{H})\ \inf_{z\in C} \phi(\|x-z\|) = \phi(\inf_{z\in C}\|x-z\|)$. Thus, $f = \iota_C \,\square\, (\phi\circ\|\cdot\|)$. In turn, since ϕ is even, we derive from Proposition 13.24(i), Example 13.3(i), and Example 13.8 that

$$f^* = \big(\iota_C \,\square\, (\phi\circ\|\cdot\|)\big)^* = \iota_C^* + (\phi\circ\|\cdot\|)^* = \sigma_C + \phi^*\circ\|\cdot\|, \quad (13.20)$$

which completes the proof. ☐

Example 13.27 Let C be a nonempty closed convex subset of $\mathcal{H}$, let $p \in \left]1,+\infty\right[$, and set $p^* = p/(p-1)$.

(i) Setting $\phi = |\cdot|$ in Example 13.26 and using Example 13.3(v) yields $d_C^* = \sigma_C + \iota_{B(0;1)}$.
(ii) If V is a closed linear subspace of $\mathcal{H}$, then (i) and Example 13.3(iii) yield $d_V^* = \iota_{V^\perp\cap B(0;1)}$.
(iii) In view of Example 13.2(i), setting $\phi = (1/p)|\cdot|^p$ in Example 13.26 yields $((1/p)d_C^p)^* = \sigma_C + (1/p^*)\|\cdot\|^{p^*}$.
(iv) Setting $C = \{0\}$ in (iii) yields $((1/p)\|\cdot\|^p)^* = (1/p^*)\|\cdot\|^{p^*}$.

Here is a generalization of Proposition 13.16(ii).

Proposition 13.28 *Let $(f_i)_{i\in I}$ be a family of proper functions from $\mathcal{H}$ to $\left]-\infty,+\infty\right]$. Then the following hold:*

(i) $\big(\inf_{i\in I} f_i\big)^* = \sup_{i\in I} f_i^*$.
(ii) $\big(\sup_{i\in I} f_i\big)^* \leqslant \inf_{i\in I} f_i^*$.

Proof. (i): By definition,

$$\begin{aligned}(\forall u\in\mathcal{H})\quad \Big(\inf_{i\in I} f_i\Big)^*(u) &= \sup_{x\in\mathcal{H}}\Big(\langle x \mid u\rangle + \sup_{i\in I} -f_i(x)\Big)\\ &= \sup_{i\in I}\sup_{x\in\mathcal{H}}\big(\langle x\mid u\rangle - f_i(x)\big)\\ &= \sup_{i\in I} f_i^*(u). \end{aligned} \quad (13.21)$$

(ii): Set $g = \sup_{i\in I} f_i$. For every $i \in I$, $f_i \leqslant g$ and therefore $g^* \leqslant f_i^*$ by Proposition 13.16(ii). Hence, $g^* \leqslant \inf_{i\in I} f_i^*$. ☐

Proposition 13.29 *Let $f\colon \mathcal{H}\to\left]-\infty,+\infty\right]$ be proper, let $\gamma\in\mathbb{R}_{++}$, and set $q = (1/2)\|\cdot\|^2$. Then*

$$(\gamma f - q)^* = \gamma(\gamma q - f^*)^* - q \quad (13.22)$$

and $(\gamma q - f^*)^*$ is γ^{-1}-strongly convex.

Proof. Set $\varphi = \gamma q - f^*$. Then

$$\varphi = \gamma q - \sup_{u\in\mathcal{H}} \big(\langle \cdot \mid u\rangle - f(u)\big) = \inf_{u\in\mathcal{H}} \big(\gamma q - \langle \cdot \mid u\rangle + f(u)\big). \tag{13.23}$$

Therefore, we derive from Proposition 13.28(i) that

$$\begin{aligned}
\gamma\varphi^* &= \gamma \sup_{u\in\mathcal{H}} \big(\gamma q - \langle \cdot \mid u\rangle + f(u)\big)^* \\
&= \sup_{u\in\mathcal{H}} \big(\gamma\big(\gamma q - \langle \cdot \mid u\rangle\,\big)^* - \gamma f(u)\big) \\
&= \sup_{u\in\mathcal{H}} \big(q(\cdot + u) - \gamma f(u)\big) \\
&= q + \sup_{u\in\mathcal{H}} \big(\langle \cdot \mid u\rangle - (\gamma f - q)(u)\big) \\
&= q + (\gamma f - q)^*,
\end{aligned} \tag{13.24}$$

which yields (13.22). In turn, $(\gamma q - f^*)^* - \gamma^{-1}q = \gamma^{-1}(\gamma f - q)^*$ is convex by Proposition 13.13. Hence, the second claim follows from Proposition 10.8. ☐

Proposition 13.30 *Let $(\mathcal{H}_i)_{i\in I}$ be a finite family of real Hilbert spaces and, for every $i \in I$, let $f_i \colon \mathcal{H}_i \to \left]-\infty, +\infty\right]$. Then*

$$\Big(\bigoplus_{i\in I} f_i\Big)^* = \bigoplus_{i\in I} f_i^*. \tag{13.25}$$

Proof. Set $\boldsymbol{\mathcal{H}} = \bigoplus_{i\in I}\mathcal{H}_i$ and $\boldsymbol{f} = \bigoplus_{i\in I} f_i$, and let $\boldsymbol{u} = (u_i)_{i\in I} \in \boldsymbol{\mathcal{H}}$. We have

$$\begin{aligned}
\boldsymbol{f}^*(\boldsymbol{u}) &= \sup_{\boldsymbol{x}\in\boldsymbol{\mathcal{H}}} \big(\langle \boldsymbol{u} \mid \boldsymbol{x}\rangle - \boldsymbol{f}(\boldsymbol{x})\big) \\
&= \sup_{\boldsymbol{x}\in\boldsymbol{\mathcal{H}}} \Bigg(\sum_{i\in I} \langle u_i \mid x_i\rangle - \sum_{i\in I} f_i(x_i)\Bigg) \\
&= \sum_{i\in I} \sup_{x_i\in\mathcal{H}_i} \big(\langle u_i \mid x_i\rangle - f_i(x_i)\big) \\
&= \sum_{i\in I} f_i^*(u_i),
\end{aligned} \tag{13.26}$$

and we obtain the conclusion. ☐

Example 13.31 Suppose that $\mathcal{H} = \mathbb{R}^N$, set $I = \{1, \ldots, N\}$, let $p \in \left]1, +\infty\right[$, and set $p^* = p/(p-1)$. Then $((1/p)\|\cdot\|_p^p)^* = (1/p^*)\|\cdot\|_{p^*}^{p^*}$.

Proof. We derive from Example 13.2(i) that $((1/p)|\cdot|^p)^* = (1/p^*)|\cdot|^{p^*}$. Hence it follows from Proposition 13.30 that $((1/p)\|\cdot\|_p^p)^* = (\bigoplus_{i\in I}(1/p)|\cdot|^p)^* = \bigoplus_{i\in I}(1/p^*)|\cdot|^{p^*} = (1/p^*)\|\cdot\|_{p^*}^{p^*}$. ☐

Example 13.32 Suppose that $\mathcal{H} = \mathbb{R}^N$, set $I = \{1, \dots, N\}$, let $p \in [1, +\infty]$, and set $B_p = \{x \in \mathcal{H} \mid \|x\|_p \leqslant 1\}$. Set $p^* = p/(p-1)$ if $p \in]1, +\infty[$; $p^* = 1$ if $p = +\infty$; $p^* = +\infty$ if $p = 1$. Then $\|\cdot\|_p^* = \iota_{B_{p^*}}$.

Proof. Let $u \in \mathcal{H}$. First suppose that $p = 1$. We derive from Example 13.3(v) that $|\cdot|^* = \iota_{[-1,1]}$. Hence, Proposition 13.30 yields $\|\cdot\|_1^* = (\bigoplus_{i \in I} |\cdot|)^* = \bigoplus_{i \in I} |\cdot|^* = \bigoplus_{i \in I} \iota_{[-1,1]} = \iota_{B_\infty}$, as claimed. We now suppose that $p > 1$. Then Hölder's inequality (Example 9.39) asserts that $(\forall x \in \mathcal{H})\ \langle x \mid u \rangle \leqslant \|x\|_p \|u\|_{p^*}$. Hence, if $\|u\|_{p^*} \leqslant 1$, then $0 = \langle 0 \mid u \rangle - \|0\|_p \leqslant \sup_{x \in \mathcal{H}} (\langle x \mid u \rangle - \|x\|_p) \leqslant \sup_{x \in \mathcal{H}} (\|x\|_p (\|u\|_{p^*} - 1)) = 0$. Therefore $f^*(u) = 0$. Now assume that $\|u\|_{p^*} > 1$, and set $u = (\mu_i)_{i \in I}$ and $v = (\operatorname{sign}(\mu_i)|\mu_i|^{p^*-1})_{i \in I}$, where $0^0 = 1$. Then

$$\begin{aligned}
\sup_{x \in \mathcal{H}} \big(\langle x \mid u \rangle - \|x\|_p \big) &\geqslant \sup_{\lambda \in \mathbb{R}_{++}} \big(\langle \lambda v \mid u \rangle - \|\lambda v\|_p \big) \\
&= \big(\|u\|_{p^*}^{p^*} - \|u\|_{p^*}^{p^*/p} \big) \sup_{\lambda \in \mathbb{R}_{++}} \lambda \\
&= \|u\|_{p^*}^{p^*-1} \big(\|u\|_{p^*} - 1 \big) \sup_{\lambda \in \mathbb{R}_{++}} \lambda \\
&= +\infty.
\end{aligned} \tag{13.27}$$

Altogether, $f^* = \iota_{B_{p^*}}$. ☐

The remainder of this section concerns the conjugates of bivariate functions.

Proposition 13.33 *Let $\mathcal{K}$ be a real Hilbert space, let $F \colon \mathcal{H} \times \mathcal{K} \to]-\infty, +\infty]$ be a proper function, and let $f \colon \mathcal{H} \to [-\infty, +\infty] : x \mapsto \inf F(x, \mathcal{K})$. Then $f^* = F^*(\cdot, 0)$.*

Proof. Fix $u \in \mathcal{H}$. Then

$$\begin{aligned}
f^*(u) &= \sup_{x \in \mathcal{H}} \Big(\langle x \mid u \rangle - \inf_{y \in \mathcal{K}} F(x, y) \Big) \\
&= \sup_{x \in \mathcal{H}} \sup_{y \in \mathcal{K}} \Big(\langle x \mid u \rangle + \langle y \mid 0 \rangle - F(x, y) \Big) \\
&= \sup_{(x,y) \in \mathcal{H} \times \mathcal{K}} \big(\langle (x, y) \mid (u, 0) \rangle - F(x, y) \big) \\
&= F^*(u, 0),
\end{aligned} \tag{13.28}$$

which establishes the identity. ☐

Definition 13.34 Let $F \colon \mathcal{H} \times \mathcal{H} \to [-\infty, +\infty]$. Then F is *autoconjugate* if $F^* = F^\intercal$, where $F^\intercal \colon \mathcal{H} \times \mathcal{H} \colon (u, x) \mapsto F(x, u)$.

Proposition 13.35 *Let $F \in \Gamma(\mathcal{H} \oplus \mathcal{H})$. Then $F^{*\intercal} = F^{\intercal *}$.*

Proof. Exercise 13.10. ☐

Proposition 13.36 *Let $F \in \Gamma(\mathcal{H} \oplus \mathcal{H})$ be autoconjugate. Then $F \geqslant \langle \cdot \mid \cdot \rangle$ and $F^* \geqslant \langle \cdot \mid \cdot \rangle$.*

Proof. Take $(x, u) \in \mathcal{H} \times \mathcal{H}$. Then, the Fenchel–Young inequality (Proposition 13.15) yields $2F^*(u, x) = 2F(x, u) = F(x, u) + F^{\mathsf{T}}(u, x) = F(x, u) + F^*(u, x) \geqslant \langle (x, u) \mid (u, x) \rangle = 2 \langle x \mid u \rangle$ and the result follows. ☐

13.3 The Fenchel–Moreau Theorem

As seen in Proposition 13.16(i), a function $f : \mathcal{H} \to \left]-\infty, +\infty\right]$ is related to its biconjugate via the inequality $f^{**} \leqslant f$. In general $f^{**} \neq f$ since the left-hand side is always lower semicontinuous and convex by Proposition 13.13, while the right-hand side need not be. The next theorem characterizes those functions that coincide with their biconjugates.

Theorem 13.37 (Fenchel–Moreau) *Let $f : \mathcal{H} \to \left]-\infty, +\infty\right]$ be proper. Then f is lower semicontinuous and convex if and only if $f = f^{**}$. In this case, f^* is proper as well.*

Proof. If $f = f^{**}$, then f is lower semicontinuous and convex as a conjugate function by Proposition 13.13. Conversely, suppose that f is lower semicontinuous and convex. Fix $x \in \mathcal{H}$, take $\xi \in \left]-\infty, f(x)\right[$, and set $(p, \pi) = P_{\operatorname{epi} f}(x, \xi)$. Proposition 9.18 states that

$$\pi \geqslant \xi \quad \text{and} \quad (\forall y \in \operatorname{dom} f) \quad \langle y - p \mid x - p \rangle \leqslant \big(\pi - \xi\big)\big(f(y) - \pi\big). \tag{13.29}$$

If $\pi > \xi$, then upon setting $v = (x - p)/(\pi - \xi)$, we get from (13.29) that

$$\begin{aligned}(\forall y \in \operatorname{dom} f) \quad \langle y \mid v \rangle - f(y) &\leqslant \langle p \mid v \rangle - \pi \\ &= \langle x \mid v \rangle - (\pi - \xi)\|v\|^2 - \pi \\ &\leqslant \langle x \mid v \rangle - \pi.\end{aligned} \tag{13.30}$$

However, (13.30) implies that $f^*(v) \leqslant \langle x \mid v \rangle - \pi$ and, in turn, that $\pi \leqslant \langle x \mid v \rangle - f^*(v) \leqslant f^{**}(x)$. To sum up,

$$\pi > \xi \quad \Rightarrow \quad f^{**}(x) > \xi. \tag{13.31}$$

We now show that $f^{**}(x) = f(x)$. Since $\operatorname{dom} f \neq \varnothing$, we first consider the case $x \in \operatorname{dom} f$. Then (9.17) yields $\pi = f(p) > \xi$, and it follows from (13.31) and Proposition 13.16(i) that $f(x) \geqslant f^{**}(x) > \xi$. Since ξ can be chosen arbitrarily in $\left]-\infty, f(x)\right[$, we get $f^{**}(x) = f(x)$. Thus, f and f^{**} coincide on $\operatorname{dom} f \neq \varnothing$. Therefore $+\infty \not\equiv f^{**} \not\equiv -\infty$ and it follows from Proposition 13.10(ii) that $-\infty \notin f^*(\mathcal{H}) \neq \{+\infty\}$, i.e., that f^* is proper. Now suppose that $x \notin \operatorname{dom} f$.

If $\pi > \xi$, it follows from (13.31) that $f^{**}(x) > \xi$ and, since ξ can be any real number, we obtain $f^{**}(x) = +\infty = f(x)$. Otherwise, $\pi = \xi$ and, since $(x,\xi) \notin \operatorname{epi} f$ and $(p,\pi) \in \operatorname{epi} f$, we have $\|x-p\| > 0$. Now, fix $w \in \operatorname{dom} f^*$ and set $u = x - p$. Then it follows from (13.1) and (13.29) that

$$(\forall y \in \operatorname{dom} f) \quad \langle y \mid w\rangle - f(y) \leqslant f^*(w) \quad \text{and} \quad \langle y \mid u\rangle \leqslant \langle p \mid u\rangle . \tag{13.32}$$

Next, let $\lambda \in \mathbb{R}_{++}$. Then (13.32) yields

$$(\forall y \in \operatorname{dom} f) \quad \langle y \mid w + \lambda u\rangle - f(y) \leqslant f^*(w) + \langle p \mid \lambda u\rangle . \tag{13.33}$$

Hence, $f^*(w+\lambda u) \leqslant f^*(w)+\langle \lambda u \mid p\rangle = f^*(w)+\langle w + \lambda u \mid x\rangle-\langle w \mid x\rangle-\lambda\|u\|^2$. Consequently, $f^{**}(x) \geqslant \langle w + \lambda u \mid x\rangle - f^*(w+\lambda u) \geqslant \langle w \mid x\rangle+\lambda\|u\|^2 - f^*(w)$. Since λ can be arbitrarily large, we must have $f^{**}(x) = +\infty$. ☐

Corollary 13.38 *Let $f \in \Gamma_0(\mathcal{H})$. Then $f^* \in \Gamma_0(\mathcal{H})$ and $f^{**} = f$.*

Proof. Combine Theorem 13.37 and Proposition 13.13. ☐

Corollary 13.39 *Let $f \in \Gamma_0(\mathcal{H})$ and let $g\colon \mathcal{H} \to [-\infty,+\infty]$. Then $f \leqslant g \Leftrightarrow f^* \geqslant g^*$.*

Proof. Proposition 13.16(ii), Corollary 13.38, and Proposition 13.16(i) yield $f \leqslant g \Rightarrow f^* \geqslant g^* \Rightarrow f = f^{**} \leqslant g^{**} \leqslant g$. ☐

Corollary 13.40 *Let f and g be in $\Gamma_0(\mathcal{H})$. Then $f = g^* \Leftrightarrow g = f^*$.*

Example 13.41 Suppose that $\mathcal{H} = \mathbb{R}^N$, let $p \in [1,+\infty]$, set $B_p = \big\{x \in \mathcal{H} \;\big|\; \|x\|_p \leqslant 1\big\}$, and let $\|\cdot\|_{p,*}$ denote the dual norm of $\|\cdot\|_p$ (see Example 7.10). Set $p^* = p/(p-1)$ if $p \in\,]1,+\infty[$; $p^* = 1$ if $p = +\infty$; $p^* = +\infty$ if $p = 1$. Then $\|\cdot\|_{p,*} = \|\cdot\|_{p^*}$.

Proof. By Example 13.32 and Corollary 13.40, $\|\cdot\|_{p,*} = \iota^*_{B_p} = \|\cdot\|_{p^*}$. ☐

Corollary 13.42 *Let $f \in \Gamma_0(\mathcal{H})$. Then f is the supremum of its continuous affine minorants.*

Proof. As seen in Proposition 13.12(i), a continuous affine minorant of f assumes the form $\langle\cdot \mid u\rangle - \mu$, where $(u,\mu) \in \operatorname{epi} f^*$. On the other hand, it follows from Theorem 13.37 and Proposition 13.10(iv) that $(\forall x \in \mathcal{H})$ $f(x) = f^{**}(x) = \sup_{(u,\mu)\in \operatorname{epi} f^*}(\langle u \mid x\rangle - \mu)$. ☐

Example 13.43 Here are some consequences of Theorem 13.37.

(i) Let C be a nonempty subset of $\mathcal{H}$. Then it follows from Example 13.3(i) and Proposition 7.13 that $\iota_C^{**} = \sigma_C^* = \sigma^*_{\overline{\operatorname{conv}}\, C} = \iota^{**}_{\overline{\operatorname{conv}}\, C} = \iota_{\overline{\operatorname{conv}}\, C}$.

(ii) Let V be a closed linear subspace of $\mathcal{H}$. Then $\iota_V^{**} = \iota_V$. On the other hand, $\iota_V^* = \sigma_V = \iota_{V^\perp}$ and therefore $\iota_V^{**} = \sigma_V^* = \iota^*_{V^\perp} = \iota_{V^{\perp\perp}}$. Altogether, we recover the well-known identity $V^{\perp\perp} = V$.

(iii) More generally, let K be a nonempty closed convex cone in $\mathcal{H}$. Then $\iota_K^{**} = \iota_K$. On the other hand, $\iota_K^* = \sigma_K = \iota_{K^\ominus}$ and therefore $\iota_K^{**} = \sigma_K^* = \iota_{K^\ominus}^* = \iota_{K^{\ominus\ominus}}$. Thus, we recover Corollary 6.34, namely $K^{\ominus\ominus} = K$.

Proposition 13.44 *Let $f\colon \mathcal{H} \to [-\infty,+\infty]$ be convex and let $x \in \mathcal{H}$. Suppose that $f(x) \in \mathbb{R}$. Then the following are equivalent:*

(i) *f is lower semicontinuous at x.*
(ii) $f^{**}(x) = \breve{f}(x) = \bar{f}(x) = f(x)$.
(iii) $f^{**}(x) = f(x)$.

Moreover, each of the above implies that $f \geqslant \bar{f} = \breve{f} = f^{**} \in \Gamma_0(\mathcal{H})$.

Proof. In view of Proposition 13.13, Proposition 13.16(i), Proposition 9.8(i), and Corollary 9.10, we have $f^{**} \leqslant \breve{f} = \bar{f} \leqslant f$. Furthermore, $\breve{f}(x) = \bar{f}(x) = f(x) \in \mathbb{R}$ and therefore Proposition 9.6 implies that $\breve{f} \in \Gamma_0(\mathcal{H})$. Hence, we deduce from Proposition 13.16(iv) and Corollary 13.38 that $f^{**} = \breve{f}^{**} = \breve{f} = \bar{f} \leqslant f$. Thus, by Lemma 1.32(v), (i) $\Leftrightarrow f(x) = \bar{f}(x) \Leftrightarrow$ (iii) $\Leftrightarrow$ (ii). ☐

Proposition 13.45 *Let $f\colon \mathcal{H} \to \,]-\infty,+\infty]$. If f has a continuous affine minorant (equivalently, if* $\operatorname{dom} f^* \neq \varnothing$*), then $f^{**} = \breve{f}$; otherwise $f^{**} = -\infty$.*

Proof. If $f \equiv +\infty$, then $\breve{f} = f$, and Proposition 13.10(iv) yields $f^{**} \equiv +\infty$ and hence $f^{**} = \breve{f}$. Now suppose that $f \not\equiv +\infty$. As seen in Proposition 13.12(ii), if f possesses no continuous affine minorant, then $f^* \equiv +\infty$ and therefore $f^{**} \equiv -\infty$. Otherwise, there exists a continuous affine function $a\colon \mathcal{H} \to \mathbb{R}$ such that $a \leqslant f$. Hence, $a = \breve{a} \leqslant \breve{f} \leqslant f$, and $\breve{f}$ is therefore proper. Thus, we have $\breve{f} \in \Gamma_0(\mathcal{H})$, and it follows from Proposition 13.16(iv) and Theorem 13.37 that $f^{**} = \big(\breve{f}\,\big)^{**} = \breve{f}$. ☐

Proposition 13.46 *Let $f\colon \mathcal{H} \to \,]-\infty,+\infty]$ be a proper convex function that has a continuous affine minorant. Then the following hold:*

(i) $\operatorname{dom} f \subset \operatorname{dom} f^{**} \subset \overline{\operatorname{dom}}\, f$.
(ii) $\operatorname{epi} f^{**} = \overline{\operatorname{epi}}\ f$.
(iii) $(\forall x \in \mathcal{H})\ f^{**}(x) = \varliminf_{y \to x} f(y)$.

Proof. Proposition 13.44 yields $f^{**} = \breve{f} = \bar{f}$.

(i): Combine Proposition 8.2 and Proposition 9.8(iv).

(ii): Lemma 1.32(vi) yields $\operatorname{epi} f^{**} = \operatorname{epi} \bar{f} = \overline{\operatorname{epi}}\, f$.

(iii): Let $x \in \mathcal{H}$. Then Proposition 13.44 and Lemma 1.32(iv) yield $f^{**}(x) = \bar{f}(x) = \varliminf_{y \to x} f(y)$. ☐

Here is a sharpening of Proposition 13.28(ii).

Proposition 13.47 *Let $(f_i)_{i \in I}$ be a family of functions in $\Gamma_0(\mathcal{H})$ such that* $\sup_{i \in I} f_i \not\equiv +\infty$*. Then* $\big(\sup_{i \in I} f_i\big)^* = \big(\inf_{i \in I} f_i^*\big)\breve{}$.

Proof. Theorem 13.37 and Proposition 13.28(i) imply that $\sup_{i\in I} f_i = \sup_{i\in I} f_i^{**} = \big(\inf_{i\in I} f_i^*\big)^*$. Hence, $\big(\sup_{i\in I} f_i\big)^* = \big(\inf_{i\in I} f_i^*\big)^{**}$ and the claim follows from Proposition 13.45. ☐

Proposition 13.48 *Let $\mathcal{K}$ be a real Hilbert space, let $g \in \Gamma_0(\mathcal{K})$, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$ be such that* $\operatorname{dom} g \cap \operatorname{ran} L \neq \varnothing$. *Then* $(g \circ L)^* = (L^* \triangleright g^*)^{**} = (L^* \triangleright g^*)^{\smile}$.

Proof. Proposition 13.24(v) gives $(g \circ L)^* \leqslant L^* \triangleright g^*$. Thus, it follows from Proposition 13.16(ii), Proposition 13.24(iv), Corollary 13.38, and Proposition 13.16(i) that $(g\circ L)^{**} \geqslant (L^* \triangleright g^*)^* = g^{**} \circ L^{**} = g \circ L \geqslant (g \circ L)^{**}$. Consequently, $(g\circ L)^* = (g\circ L)^{***} = (L^* \triangleright g^*)^{**}$ by Proposition 13.16(iii). On the other hand, Proposition 13.24(iv) yields $\varnothing \neq \operatorname{dom}(g \circ L) = \operatorname{dom}(L^* \triangleright g^*)^*$. Therefore, by Proposition 13.45, $(L^* \triangleright g^*)^{**} = (L^* \triangleright g^*)^{\smile}$. ☐

The next result draws a connection between conjugate and recession functions.

Proposition 13.49 *Let $f \in \Gamma_0(\mathcal{H})$. Then*

$$\operatorname{rec}(f^*) = \sigma_{\operatorname{dom} f} \quad \textit{and} \quad \operatorname{rec} f = \sigma_{\operatorname{dom} f^*}. \tag{13.34}$$

Proof. Using Proposition 13.11, we set

$$\begin{aligned} F \colon \mathcal{H} \times \mathbb{R} &\to \left]-\infty,+\infty\right] \\ (u,\mu) &\mapsto \sigma_{\operatorname{epi} f}(u,-\mu) = \begin{cases} \mu f^*(u/\mu), & \text{if } \mu > 0;\\ \sigma_{\operatorname{dom} f}(u), & \text{if } \mu = 0;\\ +\infty, & \text{if } \mu < 0. \end{cases} \end{aligned} \tag{13.35}$$

Then $F \in \Gamma_0(\mathcal{H} \oplus \mathbb{R})$ and $\operatorname{dom} F \subset \mathcal{H} \times \mathbb{R}_+$. Furthermore, set

$$G \colon \mathcal{H} \times \mathbb{R} \to \left]-\infty,+\infty\right] : (u,\mu) \mapsto \begin{cases} \mu f^*(u/\mu), & \text{if } \mu > 0;\\ (\operatorname{rec}(f^*))(u), & \text{if } \mu = 0;\\ +\infty, & \text{if } \mu < 0. \end{cases} \tag{13.36}$$

By Proposition 9.42, $G \in \Gamma_0(\mathcal{H} \oplus \mathbb{R})$. Clearly, $(\operatorname{dom} F) \cup (\operatorname{dom} G) \subset \mathcal{H} \times \mathbb{R}_+$ and $F|_{\mathcal{H}\times\mathbb{R}_{++}} = G|_{\mathcal{H}\times\mathbb{R}_{++}}$. Since $\operatorname{dom} f^* \neq \varnothing$ by Corollary 13.38, we deduce that $\varnothing \neq (\operatorname{dom} f^*) \times \{1\} \subset \operatorname{dom} F \cap \operatorname{dom} G \cap (\mathcal{H} \times \mathbb{R}_{++})$. Therefore, $F = G$ by Proposition 9.16. It follows that $(\forall u \in \mathcal{H})$ $\sigma_{\operatorname{dom} f}(u) = F(u,0) = G(u,0) = (\operatorname{rec}(f^*))(u)$. Hence $\sigma_{\operatorname{dom} f} = \operatorname{rec}(f^*)$. Applying this identity to f^* and using Corollary 13.38, we obtain $\sigma_{\operatorname{dom} f^*} = \operatorname{rec} f$. ☐

We conclude this section with a result on the conjugation of convex integral functions.

Proposition 13.50 *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, let $(\mathsf{H}, \langle \cdot \mid \cdot \rangle_{\mathsf{H}})$ be a separable real Hilbert space, and let $\varphi \in \Gamma_0(\mathsf{H})$. Suppose that $\mathcal{H} = L^2((\Omega,\mathcal{F},\mu);\mathsf{H})$ and that one of the following holds:*

(a) $\mu(\Omega) < +\infty$.
(b) $\varphi \geqslant \varphi(0) = 0$.

Set

$$f\colon \mathcal{H} \to \left]-\infty, +\infty\right] \\ x \mapsto \begin{cases} \displaystyle\int_\Omega \varphi\big(x(\omega)\big)\mu(d\omega), & \text{if } \varphi \circ x \in L^1\big((\Omega, \mathcal{F}, \mu); \mathbb{R}\big); \\ +\infty, & \text{otherwise}, \end{cases} \tag{13.37}$$

and

$$g\colon \mathcal{H} \to \left]-\infty, +\infty\right] \\ u \mapsto \begin{cases} \displaystyle\int_\Omega \varphi^*\big(u(\omega)\big)\mu(d\omega), & \text{if } \varphi^* \circ u \in L^1\big((\Omega, \mathcal{F}, \mu); \mathbb{R}\big); \\ +\infty, & \text{otherwise}. \end{cases} \tag{13.38}$$

Then the following hold:

(i) *$f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$.*
(ii) *Suppose that H is separable, and that $(\Omega, \mathcal{F}, \mu)$ is complete (every subset of a set in $\mathcal{F}$ of μ-measure zero is in $\mathcal{F}$) and σ-finite (Ω is a countable union of sets in $\mathcal{F}$ of finite μ-measure). Then $f^* = g$.*

Proof. (i): We have shown in Proposition 9.40 that $f \in \Gamma_0(\mathcal{H})$. Likewise, since Corollary 13.38 implies that $\varphi^* \in \Gamma_0(\mathsf{H})$ and Proposition 13.22 implies that $\varphi^* \geqslant \varphi^*(0) = 0$, g is a well-defined function in $\Gamma_0(\mathcal{H})$.

(ii): This follows from (i) and [317] (see also [318, Theorem 21(a)]). ☐

Example 13.51 Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a complete probability space, let $\mathcal{H}$ be the space of square-integrable random variables (see Example 2.9), and set $f\colon \mathcal{H} \to \left]-\infty, +\infty\right] : X \mapsto (1/p)\mathsf{E}|X|^p$, where $p \in \left]1, +\infty\right[$. Then $f^*\colon \mathcal{H} \to \left]-\infty, +\infty\right] : X \mapsto (1/p^*)\mathsf{E}|X|^{p^*}$, where $p^* = p/(p-1)$.

Proof. This follows from Proposition 13.50 and Example 13.2(i). ☐

Exercises

Exercise 13.1 Prove Example 13.2.

Exercise 13.2 Prove items (i)–(iv) in Example 13.3.

Exercise 13.3 Let $f \in \Gamma_0(\mathcal{H})$ and let $\alpha \in \mathbb{R}_{++}$. Define $\alpha \star f = \alpha f \circ \alpha^{-1}\mathrm{Id}$. Prove the following:

(i) $\operatorname{epi}(\alpha \star f) = \alpha \operatorname{epi} f$.
(ii) $(\alpha f)^* = \alpha \star f^*$.
(iii) $(\alpha \star f)^* = \alpha f^*$.
(iv) $(\alpha^2 f \circ \alpha^{-1}\mathrm{Id})^* = \alpha^2 f^* \circ \alpha^{-1}\mathrm{Id}$.

The operation $(\alpha, f) \mapsto \alpha \star f$ is sometimes called *epi-multiplication*, it is the property dual to pointwise multiplication under conjugation.

Exercise 13.4 Set $f\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto (1/2)x^2 - |x| - \exp(-|x|) + 1$ and $V\colon x \mapsto W(-\exp(-x-1))$, where W is the Lambert W-function, i.e., the inverse of the bijection $[-1, +\infty[\, \to [-1/e, +\infty[\, : \xi \mapsto \xi \exp(\xi)$. Show that

$$f^*\colon u \mapsto -\tfrac{1}{2}V\big(|u|\big)^2 - V\big(|u|\big) + \tfrac{1}{2}u^2 + |u| - \tfrac{1}{2}. \tag{13.39}$$

Exercise 13.5 Suppose that $p \in \,]0,1[$ and set

$$f\colon \mathbb{R} \to \,]-\infty, +\infty] : x \mapsto \begin{cases} -\frac{1}{p}x^p, & \text{if } x \geqslant 0;\\ +\infty, & \text{otherwise.}\end{cases} \tag{13.40}$$

Show that

$$f^*\colon \mathbb{R} \to \,]-\infty, +\infty] : u \mapsto \begin{cases} -\frac{1}{p^*}|u|^{p^*}, & \text{if } u < 0;\\ +\infty, & \text{otherwise,}\end{cases} \tag{13.41}$$

where $\frac{1}{p} + \frac{1}{p^*} = 1$.

Exercise 13.6 Let f and g be functions from $\mathcal{H}$ to $]-\infty, +\infty]$ and let $\alpha \in \,]0,1[$. Show that $(\alpha f + (1-\alpha)g)^* \leqslant \alpha f^* + (1-\alpha)g^*$.

Exercise 13.7 Derive Proposition 13.24(i) from Proposition 13.33.

Exercise 13.8 Set $f\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \frac{1}{2}\overline{\lim}\,\|x - z_n\|^2$, where $(z_n)_{n\in\mathbb{N}}$ is a bounded sequence in $\mathcal{H}$ (see also Example 8.19 and Proposition 11.18) and define $g\colon \mathcal{H} \to \mathbb{R}\colon u \mapsto \frac{1}{2}\|u\|^2 + \underline{\lim}\,\langle z_n \mid u\rangle$. Show that $f^* \leqslant g$, and provide an example of a sequence $(z_n)_{n\in\mathbb{N}}$ for which $f^* = g$ and another one for which $f^* < g$.

Exercise 13.9 Prove Proposition 6.27 via Proposition 13.24(i).

Exercise 13.10 Prove Proposition 13.35.

Exercise 13.11 Let C and D be nonempty subsets of $\mathcal{H}$ such that D is closed and convex. Show that $C \subset D \Leftrightarrow \sigma_C \leqslant \sigma_D$.

Chapter 14
Further Conjugation Results

In this chapter, we exhibit several deeper results on conjugation. We first discuss Moreau's decomposition principle, whereby a vector is decomposed in terms of the proximity operator of a lower semicontinuous convex function and that of its conjugate. This powerful nonlinear principle extends the standard linear decomposition with respect to a closed linear subspace and its orthogonal complement. Basic results concerning the proximal average and positively homogeneous functions are also presented. Also discussed are the Moreau–Rockafellar theorem, which characterizes coercivity in terms of an interiority condition, and the Toland–Singer theorem, which provides an appealing formula for the conjugate of a difference.

14.1 Moreau's Decomposition

In this section, we take a closer look at the infimal convolution of a convex function in $\Gamma_0(\mathcal{H})$ and the function $q = (1/2)\|\cdot\|^2$.

Proposition 14.1 *Let $f \in \Gamma_0(\mathcal{H})$, let $\gamma \in \mathbb{R}_{++}$, and set $q = (1/2)\|\cdot\|^2$. Then*

$$(f + \gamma q)^* = f^* \mathbin{\boxdot} \big(\gamma^{-1} q\big) = {}^{\gamma}(f^*). \tag{14.1}$$

Proof. It follows from Corollary 13.38, Proposition 13.19, and Proposition 13.24(i) that $f + \gamma q = f^{**} + (\gamma^{-1}q)^* = (f^* \mathbin{\square} (\gamma^{-1}q))^*$. Since Corollary 13.38 and Proposition 12.15 imply that $f^* \mathbin{\square} (\gamma^{-1}q) = f^* \mathbin{\boxdot} (\gamma^{-1}q) \in \Gamma_0(\mathcal{H})$, we deduce from Theorem 13.37 and (12.20) that $(f + \gamma q)^* = (f^* \mathbin{\boxdot} (\gamma^{-1}q))^{**} = f^* \mathbin{\boxdot} (\gamma^{-1}q) = {}^{\gamma}(f^*)$. $\square$

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_14

The next proposition characterizes functions with strongly convex conjugates.

Proposition 14.2 *Let $f\colon \mathcal{H} \to \mathbb{R}$ be continuous and convex, let $\gamma \in \mathbb{R}_{++}$, and set $q = (1/2)\|\cdot\|^2$. Then the following are equivalent:*

(i) *$f^* - \gamma^{-1}q$ is convex, i.e., f^* is γ^{-1}-strongly convex.*
(ii) *$\gamma q - f$ is convex.*

Proof. (i)⇒(ii): Set $h = f^* - \gamma^{-1}q$. Since $f \in \Gamma_0(\mathcal{H})$ and h is convex, we have $h \in \Gamma_0(\mathcal{H})$ and, by Corollary 13.38, $h^* \in \Gamma_0(\mathcal{H})$. Hence, using Theorem 13.37 and Example 13.4, we obtain

$$f = f^{**} = (h + \gamma^{-1}q)^* = \gamma q - {}^{\gamma}h \circ \gamma \mathrm{Id}. \tag{14.2}$$

Hence, since it follows from Proposition 12.15 that ${}^{\gamma}h$ is convex, we deduce that $\gamma q - f = {}^{\gamma}h \circ \gamma\mathrm{Id}$ is convex.

(ii)⇒(i): Set $g = \gamma q - f$. Then $g \in \Gamma_0(\mathcal{H})$ and therefore Corollary 13.38 yields $g = g^{**}$. Thus, $f = \gamma q - g = \gamma q - (g^*)^*$. In turn, Proposition 13.29 yields $f^* = ((\gamma g^* - q)^* + q)/\gamma$. Thus, Proposition 13.13 implies that $f^* - q/\gamma = (\gamma g^* - q)^*/\gamma$ is convex. ☐

Theorem 14.3 *Let $f \in \Gamma_0(\mathcal{H})$ and let $\gamma \in \mathbb{R}_{++}$. Then the following hold:*

(i) *Set $q = (1/2)\|\cdot\|^2$. Then*

$$\gamma^{-1}q = f \boxdot (\gamma^{-1}q) + (f^* \boxdot (\gamma q)) \circ \gamma^{-1}\mathrm{Id} = {}^{\gamma}f + {}^{1/\gamma}(f^*) \circ \gamma^{-1}\mathrm{Id}. \tag{14.3}$$

(ii) $\mathrm{Id} = \mathrm{Prox}_{\gamma f} + \gamma\, \mathrm{Prox}_{f^*/\gamma} \circ \gamma^{-1}\mathrm{Id}$.
(iii) *Let $x \in \mathcal{H}$. Then*

$$f(\mathrm{Prox}_{\gamma f} x) + f^*(\mathrm{Prox}_{f^*/\gamma}(x/\gamma)) = \big\langle \mathrm{Prox}_{\gamma f} x \mid \mathrm{Prox}_{f^*/\gamma}(x/\gamma) \big\rangle. \tag{14.4}$$

Proof. (i): It follows from Example 13.4, Proposition 12.15, and Proposition 14.1 that

$$\begin{aligned}\gamma^{-1}q &= (f^* + \gamma q)^* + \big(f^* \boxdot (\gamma q)\big) \circ \gamma^{-1}\mathrm{Id}\\ &= f \boxdot \big(\gamma^{-1}q\big) + \big(f^* \boxdot (\gamma q)\big) \circ \gamma^{-1}\mathrm{Id}.\end{aligned} \tag{14.5}$$

(ii): Take the Fréchet derivative in (i) using Proposition 12.30.

(iii): Set $p = \mathrm{Prox}_{\gamma f} x$, $p^* = \mathrm{Prox}_{f^*/\gamma}(x/\gamma)$, and let $y \in \mathcal{H}$. Then (ii) and Proposition 12.26 yield $\langle y - p \mid p^* \rangle + f(p) \leqslant f(y)$. Therefore, it follows from Proposition 13.15 that $f(p) + f^*(p^*) = f(p) + \sup_{y \in \mathcal{H}}(\langle y \mid p^* \rangle - f(y)) \leqslant \langle p \mid p^* \rangle \leqslant f(p) + f^*(p^*)$. Hence, $f(p) + f^*(p^*) = \langle p \mid p^* \rangle$. ☐

The striking symmetry obtained when $\gamma = 1$ in Theorem 14.3 deserves to be noted.

Remark 14.4 Let $f \in \Gamma_0(\mathcal{H})$ and set $q = (1/2)\|\cdot\|^2$. Then Theorem 14.3 yields *Moreau's decomposition*

$$(f \,\Box\, q) + (f^* \,\Box\, q) = q \quad \text{and} \quad \operatorname{Prox}_f + \operatorname{Prox}_{f^*} = \operatorname{Id}. \tag{14.6}$$

Thus, using Proposition 12.30, we obtain

$$\operatorname{Prox}_f = \operatorname{Id} - \nabla(f \,\Box\, q) = \nabla(f^* \,\Box\, q). \tag{14.7}$$

If $f = \iota_K$ in (14.6), where K is a nonempty closed convex cone in $\mathcal{H}$, then we recover Moreau's conical decomposition (Theorem 6.30). In particular, if K is a closed linear subspace, we recover the identities $d_K^2 + d_{K^\perp}^2 = \|\cdot\|^2$ and $P_K + P_{K^\perp} = \operatorname{Id}$ already established in Corollary 3.24.

Example 14.5 Let $\rho \in \mathbb{R}_{++}$, set $f = \rho\|\cdot\|$, and let $x \in \mathcal{H}$. Then Example 13.3(v) and Proposition 13.23(i) yield $f^* = \iota_{B(0;\rho)}$. We therefore derive from (14.6), Example 12.25, and (3.13) that

$$\operatorname{Prox}_f x = \begin{cases} \big(1 - \rho/\|x\|\big)x, & \text{if } \|x\| > \rho; \\ 0, & \text{if } \|x\| \leqslant \rho. \end{cases} \tag{14.8}$$

In other words, Prox_f is the soft thresholder of Example 4.17. In turn, we derive from (12.23) that

$$^1 f(x) = f(\operatorname{Prox}_f x) + \frac{1}{2}\|x - \operatorname{Prox}_f x\|^2 = \begin{cases} \rho\|x\| - \dfrac{\rho^2}{2}, & \text{if } \|x\| > \rho; \\ \dfrac{\|x\|^2}{2}, & \text{if } \|x\| \leqslant \rho \end{cases} \tag{14.9}$$

is the generalized Huber function of Example 13.7. In particular, if $\mathcal{H} = \mathbb{R}$, it follows from Example 8.44 that the Moreau envelope of $\rho|\cdot|$ is the Huber function.

14.2 Proximal Average

Definition 14.6 Let f and g be in $\Gamma_0(\mathcal{H})$. The *proximal average* of f and g is

$$\begin{aligned} \operatorname{pav}(f, g)\colon \mathcal{H} &\to \left]-\infty, +\infty\right] \\ x &\mapsto \tfrac{1}{2} \inf_{\substack{(y,z) \in \mathcal{H}\times\mathcal{H} \\ y+z=2x}} \Big(f(y) + g(z) + \tfrac{1}{4}\|y - z\|^2 \Big). \end{aligned} \tag{14.10}$$

Proposition 14.7 *Set $L\colon \mathcal{H}\times\mathcal{H} \to \mathcal{H}\colon (y, z) \mapsto (y + z)/2$, let f and g be in $\Gamma_0(\mathcal{H})$, and set $F\colon \mathcal{H}\times\mathcal{H} \to \left]-\infty, +\infty\right] : (y, z) \mapsto \frac{1}{2}f(y) + \frac{1}{2}g(z) + \frac{1}{8}\|y - z\|^2$. Then the following hold:*

(i) $\operatorname{pav}(f,g) = \operatorname{pav}(g,f)$.
(ii) $\operatorname{pav}(f,g) = L \triangleright F$.
(iii) $\operatorname{dom}\operatorname{pav}(f,g) = \frac{1}{2}\operatorname{dom} f + \frac{1}{2}\operatorname{dom} g$.
(iv) $\operatorname{pav}(f,g)$ *is a proper convex function.*

Proof. (i): Clear from Definition 14.6.

(ii): Definition 12.34 and Definition 14.6 imply that $\operatorname{pav}(f,g) = L \triangleright F$. Take $x \in \operatorname{dom}\operatorname{pav}(f,g)$ and set $h\colon y \mapsto \|y\|^2 + (f(y) + g(2x-y) - 2\langle y \mid x\rangle + \|x\|^2)$. It suffices to show that h has a minimizer. Since h is a strongly convex function in $\Gamma_0(\mathcal{H})$, this follows from Corollary 11.17, which asserts that h has a unique minimizer.

(iii): This follows from (ii) and Proposition 12.36(i).

(iv): Since F is convex and L is linear, the convexity of $\operatorname{pav}(f,g)$ follows from (ii) and Proposition 12.36(ii). On the other hand, properness follows from (iii). ☐

Corollary 14.8 *Let f and g be in $\Gamma_0(\mathcal{H})$, and set $q = (1/2)\|\cdot\|^2$. Then the following hold:*

(i) $\operatorname{pav}(f,g) \in \Gamma_0(\mathcal{H})$.
(ii) $\big(\operatorname{pav}(f,g)\big)^* = \operatorname{pav}(f^*,g^*)$.
(iii) $\operatorname{pav}(f,g) \,\square\, q = \frac{1}{2}(f \,\square\, q) + \frac{1}{2}(g \,\square\, q)$.
(iv) $\operatorname{Prox}_{\operatorname{pav}(f,g)} = \frac{1}{2}\operatorname{Prox}_f + \frac{1}{2}\operatorname{Prox}_g$.

Proof. We define an operator Θ on $\Gamma_0(\mathcal{H}) \times \Gamma_0(\mathcal{H})$ by

$$\Theta\colon (f_1,f_2) \mapsto \big(\tfrac{1}{2}(f_1+q)\circ(2\mathrm{Id})\big) \,\square\, \big(\tfrac{1}{2}(f_2+q)\circ(2\mathrm{Id})\big). \tag{14.11}$$

Definition 14.6 and Lemma 2.12(ii) yield, for every $f_1 \in \Gamma_0(\mathcal{H})$, every $f_2 \in \Gamma_0(\mathcal{H})$, and every $x \in \mathcal{H}$,

$$\operatorname{pav}(f_1,f_2)(x) = \inf_{\substack{(y,z)\in\mathcal{H}\times\mathcal{H}\\ y+z=x}} \Big(\tfrac{1}{2}f_1(2y) + \tfrac{1}{2}f_2(2z) + 2q(y) + 2q(z)\Big) - q(x). \tag{14.12}$$

Hence

$$(\forall f_1 \in \Gamma_0(\mathcal{H}))(\forall f_2 \in \Gamma_0(\mathcal{H})) \quad \operatorname{pav}(f_1,f_2) = \Theta(f_1,f_2) - q. \tag{14.13}$$

Proposition 13.24(i), Proposition 13.23(ii), and Proposition 14.1 yield

$$\begin{aligned}\big(\Theta(f,g)\big)^* &= \big(\tfrac{1}{2}(f+q)\circ(2\mathrm{Id})\big)^* + \big(\tfrac{1}{2}(g+q)\circ(2\mathrm{Id})\big)^*\\ &= \tfrac{1}{2}(f+q)^* + \tfrac{1}{2}(g+q)^*\\ &= \tfrac{1}{2}(f^* \,\square\, q) + \tfrac{1}{2}(g^* \,\square\, q).\end{aligned} \tag{14.14}$$

In view of (14.13), Proposition 13.29, (14.14), (14.6), Proposition 14.1, Proposition 13.23(i), Proposition 13.24(i), and (14.11), we get

$$\begin{aligned}
\big(\operatorname{pav}(f,g)\big)^* &= \big(\Theta(f,g)-q\big)^* \\
&= \big(q-\big(\Theta(f,g)\big)^*\big)^* - q \\
&= \Big(\tfrac{1}{2}\big(q-(f^*\,\Box\, q)\big)+\tfrac{1}{2}\big(q-(g^*\,\Box\, q)\big)\Big)^* - q \\
&= \big(\tfrac{1}{2}(f\,\Box\, q)+\tfrac{1}{2}(g\,\Box\, q)\big)^* - q \\
&= \big(\tfrac{1}{2}(f\,\Box\, q)\big)^*\,\Box\,\big(\tfrac{1}{2}(g\,\Box\, q)\big)^* - q \\
&= \big(\tfrac{1}{2}(f^*+q)\circ(2\mathrm{Id})\big)\,\Box\,\big(\tfrac{1}{2}(g^*+q)\circ(2\mathrm{Id})\big) - q \\
&= \Theta(f^*,g^*)-q. \qquad (14.15)
\end{aligned}$$

On the other hand, $\operatorname{pav}(f^*,g^*) = \Theta(f^*,g^*)-q$ by (14.13). Combining this with (14.15), we obtain (ii). In turn, $(\operatorname{pav}(f,g))^{**} = (\operatorname{pav}(f^*,g^*))^* = \operatorname{pav}(f^{**},g^{**}) = \operatorname{pav}(f,g)$, which implies (i) by virtue of Proposition 14.7(iv) and Proposition 13.13. Furthermore, using (ii), Proposition 14.1, (14.13), and (14.14), we deduce that

$$\begin{aligned}
\operatorname{pav}(f^*,g^*)\,\Box\, q &= \big(\operatorname{pav}(f,g)\big)^*\,\Box\, q \\
&= \big(\operatorname{pav}(f,g)+q\big)^* \\
&= \big(\Theta(f,g)\big)^* \\
&= \tfrac{1}{2}(f^*\,\Box\, q)+\tfrac{1}{2}(g^*\,\Box\, q). \qquad (14.16)
\end{aligned}$$

Hence, upon replacing in (14.16) f by f^* and g by g^*, we obtain (iii). Finally, (14.16) and (14.7) yield

$$\operatorname{Prox}_{\operatorname{pav}(f,g)} = \nabla\big((\operatorname{pav}(f,g))^*\,\Box\, q\big) = \nabla\big(\tfrac{1}{2}(f^*\,\Box\, q)+\tfrac{1}{2}(g^*\,\Box\, q)\big), \qquad (14.17)$$

which implies (iv). ☐

The proofs of the following results are left as Exercise 14.3 and Exercise 14.4.

Proposition 14.9 *Let f and g be in $\Gamma_0(\mathcal{H})$. Then*

$$\Big(\tfrac{1}{2}f^*+\tfrac{1}{2}g^*\Big)^* \leqslant \operatorname{pav}(f,g) \leqslant \tfrac{1}{2}f+\tfrac{1}{2}g. \qquad (14.18)$$

Proposition 14.10 *Let F and G be in $\Gamma_0(\mathcal{H}\oplus\mathcal{H})$. Then* $(\operatorname{pav}(F,G))^{\intercal} = \operatorname{pav}(F^{\intercal},G^{\intercal})$.

14.3 Positively Homogeneous Functions

Proposition 14.11 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ and set*

$$C = \big\{u \in \mathcal{H} \;\big|\; (\forall x \in \mathcal{H}) \;\; \langle x \mid u\rangle \leqslant f(x)\big\}. \tag{14.19}$$

Then the following are equivalent:

(i) *f is positively homogeneous and $f \in \Gamma_0(\mathcal{H})$.*
(ii) *$f = \sigma_C$ and C is nonempty, closed, and convex.*
(iii) *f is the support function of a nonempty closed convex subset of $\mathcal{H}$.*

Proof. (i)⇒(ii): Let $u \in \mathcal{H}$. We deduce from Proposition 9.14 that, given $y \in \operatorname{dom} f$, $f(0) = \lim_{\alpha\downarrow 0} f((1-\alpha)0 + \alpha y) = \lim_{\alpha\downarrow 0} \alpha f(y) = 0$. Thus, if $u \in C$, we obtain $f^*(u) = \sup_{x\in\mathcal{H}} \langle x \mid u\rangle - f(x) \leqslant 0 = \langle 0 \mid u\rangle - f(0) \leqslant f^*(u)$ and, in turn, $f^*(u) = 0$. On the other hand, if $u \notin C$, then there exists $z \in \mathcal{H}$ such that $\langle z \mid u\rangle - f(z) > 0$. Consequently, $(\forall \lambda \in \mathbb{R}_{++})$ $f^*(u) \geqslant \langle \lambda z \mid u\rangle - f(\lambda z) = \lambda(\langle z \mid u\rangle - f(z))$. Since λ can be arbitrarily large, we conclude that $f^*(u) = +\infty$. Altogether, $f^* = \iota_C$. However, since $f \in \Gamma_0(\mathcal{H})$, Corollary 13.38 yields $\iota_C = f^* \in \Gamma_0(\mathcal{H})$, which shows that C is a nonempty closed convex set. On the other hand, we deduce from Theorem 13.37 that $f = f^{**} = \iota_C^* = \sigma_C$.

(ii)⇒(iii): Clear.

(iii)⇒(i): Example 11.2. ☐

The next result establishes a connection between the polar set and the Minkowski gauge.

Proposition 14.12 *Let C be a convex subset of $\mathcal{H}$ such that $0 \in C$. Then $m_C^* = \iota_{C^\odot}$.*

Proof. Fix $u \in \mathcal{H}$. First, suppose that $u \in C^\odot$, let $x \in \operatorname{dom} m_C$, and let $\lambda \in \left]m_C(x),+\infty\right[$. Then, by (8.38), there exists $y \in C$ such that $x = \lambda y$ and, in turn, such that $\langle x \mid u\rangle = \lambda \langle y \mid u\rangle \leqslant \lambda$. Taking the limit as $\lambda \downarrow m_C(x)$ yields $\langle x \mid u\rangle \leqslant m_C(x)$, and we deduce from Proposition 13.10(iv) that $m_C^*(u) \leqslant 0$. On the other hand, $m_C^*(u) \geqslant \langle 0 \mid u\rangle - m_C(0) = 0$. Altogether, m_C^* and $\iota_{C^\odot}$ coincide on $C^\odot$. Now, suppose that $u \notin C^\odot$. Then there exists $x \in C$ such that $\langle x \mid u\rangle > 1 \geqslant m_C(x)$ and, using (8.37), we deduce that $m_C^*(u) \geqslant \sup_{\lambda\in\mathbb{R}_{++}} \langle \lambda x \mid u\rangle - m_C(\lambda x) = \sup_{\lambda\in\mathbb{R}_{++}} \lambda(\langle x \mid u\rangle - m_C(x)) = +\infty$. Therefore, m_C^* and $\iota_{C^\odot}$ coincide also on $\mathcal{H} \smallsetminus C^\odot$. ☐

Corollary 14.13 *Let C be a closed convex subset of $\mathcal{H}$ such that $0 \in C$. Then the following hold:*

(i) *$C = \operatorname{lev}_{\leqslant 1} m_C$.*
(ii) *Suppose that $0 \in \operatorname{int} C$. Then $\operatorname{int} C = \operatorname{lev}_{<1} m_C$.*

Proof. (i): It is clear that $C \subset \operatorname{lev}_{\leqslant 1} m_C$. Now assume the existence of a vector $x \in (\operatorname{lev}_{\leqslant 1} m_C) \smallsetminus C$. Theorem 3.50 provides $u \in \mathcal{H} \smallsetminus \{0\}$ such that $\langle x \mid u\rangle > \sigma_C(u) \geqslant 0$ since $0 \in C$. Hence, after scaling u if necessary, we assume that $\langle x \mid u\rangle > 1 \geqslant \sigma_C(u)$ so that $u \in C^{\odot}$. Using Proposition 14.12 and Proposition 13.16(i), we obtain the contradiction $1 < \langle u \mid x\rangle \leqslant \sigma_{C^{\odot}}(x) = \iota^*_{C^{\odot}}(x) = m_C^{**}(x) \leqslant m_C(x) \leqslant 1$.

(ii): Example 8.43 states that m_C is continuous on $\mathcal{H}$. In view of (i) and Corollary 8.47(ii) applied to $m_C - 1$, we deduce that $\operatorname{int} C = \operatorname{int}(\operatorname{lev}_{\leqslant 1} m_C) = \operatorname{lev}_{<1} m_C$. ☐

14.4 Coercivity

We now investigate the interplay between coercivity and conjugation.

Proposition 14.14 *Let $f \in \Gamma_0(\mathcal{H})$, let $\alpha \in \mathbb{R}_{++}$, and consider the following properties:*

(i) $\varliminf_{\|x\|\to+\infty} f(x)/\|x\| > \alpha$.
(ii) $(\exists\, \beta \in \mathbb{R})\ f \geqslant \alpha\|\cdot\| + \beta$.
(iii) $(\exists\, \gamma \in \mathbb{R})\ f^*\big|_{B(0;\alpha)} \leqslant \gamma$.
(iv) $\varliminf_{\|x\|\to+\infty} f(x)/\|x\| \geqslant \alpha$.

Then (i)⇒(ii)⇔(iii)⇒(iv).

Proof. (i)⇒(ii): There exists $\rho \in \mathbb{R}_{++}$ such that

$$\big(\forall x \in \mathcal{H} \smallsetminus B(0;\rho)\big) \quad f(x) \geqslant \alpha\|x\|. \tag{14.20}$$

Now set $\mu = \inf f(B(0;\rho))$. Then we deduce from Proposition 13.12(iii) that $\mu > -\infty$. Thus $(\forall x \in B(0;\rho))\ \alpha\|x\| \leqslant \alpha\rho \leqslant (\alpha\rho - \mu) + f(x)$. Hence,

$$\big(\forall x \in B(0;\rho)\big) \quad f(x) \geqslant \alpha\|x\| + (\mu - \alpha\rho). \tag{14.21}$$

Altogether, (ii) holds with $\beta = \min\{0, \mu - \alpha\rho\}$.

(ii)⇔(iii): Corollary 13.39 and Example 13.3(v) yield $\alpha\|\cdot\| + \beta \leqslant f \Leftrightarrow f^* \leqslant (\alpha\|\cdot\| + \beta)^* \Leftrightarrow f^* \leqslant \iota_{B(0;\alpha)} - \beta$.

(ii)⇒(iv): $\varliminf_{\|x\|\to+\infty} f(x)/\|x\| \geqslant \varliminf_{\|x\|\to+\infty} (\alpha + \beta/\|x\|) = \alpha$. ☐

Proposition 14.14 yields at once the following result.

Proposition 14.15 *Let $f \in \Gamma_0(\mathcal{H})$ and consider the following properties:*

(i) *f is supercoercive.*
(ii) *f^* is bounded on every bounded subset of $\mathcal{H}$.*
(iii) $\operatorname{dom} f^* = \mathcal{H}$.

Then (i)⇔(ii)⇒(iii).

Proposition 14.16 *Let $f \in \Gamma_0(\mathcal{H})$. Then the following are equivalent:*

(i) *f is coercive.*
(ii) *The lower level sets $(\operatorname{lev}_{\leqslant \xi} f)_{\xi \in \mathbb{R}}$ are bounded.*
(iii) $\varliminf_{\|x\| \to +\infty} f(x)/\|x\| > 0$.
(iv) $(\exists\, (\alpha, \beta) \in \mathbb{R}_{++} \times \mathbb{R})$ $f \geqslant \alpha \|\cdot\| + \beta$.
(v) *f^* is bounded above on a neighborhood of* 0.
(vi) $0 \in \operatorname{int} \operatorname{dom} f^*$.

Proof. (i)⇔(ii): Proposition 11.12.

(ii)⇒(iii): Suppose that $\varliminf_{\|x\| \to +\infty} f(x)/\|x\| \leqslant 0$ and set $(\forall n \in \mathbb{N})$ $\alpha_n = n + 1$. Then for every $n \in \mathbb{N}$, there exists $x_n \in \mathcal{H}$ such that $\|x_n\| \geqslant \alpha_n^2$ and $f(x_n)/\|x_n\| \leqslant 1/\alpha_n$. We thus obtain a sequence $(x_n)_{n \in \mathbb{N}}$ in $\operatorname{dom} f$ such that $0 < \alpha_n/\|x_n\| \leqslant 1/\alpha_n \to 0$ and $f(x_n)/\|x_n\| \leqslant 1/\alpha_n$. Now fix $z \in \operatorname{dom} f$ and set

$$(\forall n \in \mathbb{N}) \quad y_n = \left(1 - \frac{\alpha_n}{\|x_n\|}\right) z + \frac{\alpha_n}{\|x_n\|} x_n. \tag{14.22}$$

The convexity of f implies that $\sup_{n \in \mathbb{N}} f(y_n) \leqslant |f(z)| + 1$. Therefore, $(y_n)_{n \in \mathbb{N}}$ lies in $\operatorname{lev}_{\leqslant |f(z)|+1} f$ and it is therefore bounded. On the other hand, since $\|y_n\| \geqslant \alpha_n - \|z\| \to +\infty$, we reach a contradiction.

(iii)⇒(i): Clear.

(iii)⇔(iv)⇔(v)⇔(vi): Proposition 14.14 and Theorem 8.38. ☐

Theorem 14.17 (Moreau–Rockafellar) *Let $f \in \Gamma_0(\mathcal{H})$ and let $u \in \mathcal{H}$. Then $f - \langle \cdot \mid u \rangle$ is coercive if and only if $u \in \operatorname{int} \operatorname{dom} f^*$.*

Proof. Using Proposition 14.16 and Proposition 13.23(iii), we obtain the equivalences $f - \langle \cdot \mid u \rangle$ is coercive $\Leftrightarrow$ $0 \in \operatorname{int} \operatorname{dom}(f - \langle \cdot \mid u \rangle)^*$ $\Leftrightarrow$ $0 \in \operatorname{int} \operatorname{dom}(\tau_{-u} f^*)$ $\Leftrightarrow$ $u \in \operatorname{int} \operatorname{dom} f^*$. ☐

Corollary 14.18 *Let f and g be in $\Gamma_0(\mathcal{H})$, and suppose that f is supercoercive. Then the following hold:*

(i) *$f \,\square\, g$ is coercive if and only if g is coercive.*
(ii) *$f \,\square\, g$ is supercoercive if and only if g is supercoercive.*

Proof. Proposition 14.15 asserts that f^* is bounded on every bounded subset of $\mathcal{H}$ and that $\operatorname{dom} f^* = \mathcal{H}$. Furthermore, $f \,\square\, g = f \,\boxdot\, g \in \Gamma_0(\mathcal{H})$ by Proposition 12.14(i).

(i): Using Proposition 14.16 and Proposition 13.24(i), we obtain the equivalences $f \,\square\, g$ is coercive $\Leftrightarrow$ $0 \in \operatorname{int} \operatorname{dom}(f \,\square\, g)^*$ $\Leftrightarrow$ $0 \in \operatorname{int} \operatorname{dom}(f^* + g^*)$ $\Leftrightarrow$ $0 \in \operatorname{int}(\operatorname{dom} f^* \cap \operatorname{dom} g^*)$ $\Leftrightarrow$ $0 \in \operatorname{int} \operatorname{dom} g^*$ $\Leftrightarrow$ g is coercive.

(ii): Using Proposition 14.15 and Proposition 13.24(i), we obtain the equivalences $f \,\square\, g$ is supercoercive $\Leftrightarrow$ $(f \,\square\, g)^*$ is bounded on bounded sets $\Leftrightarrow$ $f^* + g^*$ is bounded on bounded sets $\Leftrightarrow$ g^* is bounded on bounded sets $\Leftrightarrow$ g is supercoercive. ☐

14.5 The Conjugate of a Difference

Proposition 14.19 *Let $g\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper, let $h \in \Gamma_0(\mathcal{H})$, and set*

$$f\colon \mathcal{H} \to [-\infty,+\infty] : x \mapsto \begin{cases} g(x) - h(x), & \text{if } x \in \operatorname{dom} g;\\ +\infty, & \text{if } x \notin \operatorname{dom} g. \end{cases} \tag{14.23}$$

Then

$$(\forall u \in \mathcal{H}) \quad f^*(u) = \sup_{v \in \operatorname{dom} h^*} \big(g^*(u+v) - h^*(v)\big). \tag{14.24}$$

Proof. Fix $u \in \mathcal{H}$. Using Corollary 13.38, we obtain

$$\begin{aligned} f^*(u) &= \sup_{x\in\operatorname{dom} g} \big(\langle x \mid u\rangle - g(x) + h(x)\big)\\ &= \sup_{x\in\operatorname{dom} g} \big(\langle x \mid u\rangle - g(x) + h^{**}(x)\big)\\ &= \sup_{x\in\operatorname{dom} g} \Big(\langle x \mid u\rangle - g(x) + \sup_{v\in\operatorname{dom} h^*} \big(\langle v \mid x\rangle - h^*(v)\big)\Big)\\ &= \sup_{v\in\operatorname{dom} h^*} \Big(\sup_{x\in\operatorname{dom} g} \big(\langle x \mid u+v\rangle - g(x)\big) - h^*(v)\Big)\\ &= \sup_{v\in\operatorname{dom} h^*} \big(g^*(u+v) - h^*(v)\big), \end{aligned} \tag{14.25}$$

as required. ☐

Corollary 14.20 (Toland–Singer) *Let g and h be in $\Gamma_0(\mathcal{H})$. Then*

$$\inf_{x\in\operatorname{dom} g} \big(g(x) - h(x)\big) = \inf_{v\in\operatorname{dom} h^*} \big(h^*(v) - g^*(v)\big). \tag{14.26}$$

Proof. Set $u = 0$ in (14.24). ☐

Exercises

Exercise 14.1 Let $f \in \Gamma_0(\mathcal{H})$. Show that $\operatorname{pav}(f, f^*) = (1/2)\|\cdot\|^2$.

Exercise 14.2 Let f and g be in $\Gamma_0(\mathcal{H})$. Show that

$$\operatorname{pav}(f,g) = -^1\!\left(-\frac{{}^1 f + {}^1 g}{2}\right). \tag{14.27}$$

Exercise 14.3 Prove Proposition 14.9.

Exercise 14.4 Prove Proposition 14.10.

Exercise 14.5 Let $F \in \Gamma_0(\mathcal{H}\times\mathcal{H})$. Show that $\operatorname{pav}(F, F^{*\intercal})$ is autoconjugate.

Exercise 14.6 Let $\alpha \in \mathbb{R}_{++}$ and $\beta \in \mathbb{R}_{++}$. Set $q = (1/2)\|\cdot\|^2$, $f = \alpha q$, and $g = \beta q$. Show that (14.18) becomes

$$\frac{2\alpha\beta}{\alpha+\beta}\, q \leqslant \frac{\alpha+\beta+2\alpha\beta}{2+\alpha+\beta}\, q \leqslant \frac{\alpha+\beta}{2}\, q, \tag{14.28}$$

which illustrates that the coefficient of q in the middle term (which corresponds to the proximal average) is bounded below by the harmonic mean of α and β, and bounded above by their arithmetic mean.

Exercise 14.7 Let f and g be functions in $\Gamma_0(\mathcal{H})$ such that $0 \in \operatorname{sri}(\operatorname{dom} f - \operatorname{dom} g)$ and $0 \in \operatorname{sri}(\operatorname{dom} f^* - \operatorname{dom} g^*)$, and let $\alpha \in \mathbb{R}_{++}$ be such that $\operatorname{dom} f \cap \alpha \operatorname{dom} f \neq \varnothing$ and $\operatorname{dom} f^* \cap \alpha \operatorname{dom} f^* \neq \varnothing$. Set

$$f \boldsymbol{+} g = \operatorname{pav}(f+g, f \,\square\, g) \quad \text{and} \quad \alpha \star f = \operatorname{pav}\big(\alpha f, \alpha f(\cdot/\alpha)\big), \tag{14.29}$$

and show that

$$(f \boldsymbol{+} g)^* = f^* \boldsymbol{+} g^* \quad \text{and} \quad (\alpha \star f)^* = \alpha \star f^*. \tag{14.30}$$

Exercise 14.8 Let C be a closed convex subset of $\mathcal{H}$ such that $0 \in C$. Show that C is bounded if and only if $0 \in \operatorname{int} C^{\odot}$.

Exercise 14.9 Let $h\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ be proper and suppose, for every $g \in \Gamma_0(\mathcal{H})$, that f is as in (14.23) and that (14.24) holds. Let $h\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ be proper and suppose that, for every $g \in \Gamma_0(\mathcal{H})$, (14.23)–(14.24) hold. Show that $h \in \Gamma_0(\mathcal{H})$.

Exercise 14.10 Let $g\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ be proper, and set $q = (1/2)\|\cdot\|^2$. Then it follows from Proposition 13.29 that $(g-q)^* = (q-g^*)^* - q$. Prove this result using Proposition 14.19.

Chapter 15
Fenchel–Rockafellar Duality

Of central importance in convex analysis are conditions guaranteeing that the conjugate of a sum is the infimal convolution of the conjugates. The main result in this direction is a theorem due to Attouch and Brézis. In turn, it gives rise to the Fenchel–Rockafellar duality framework for convex optimization problems. The applications we discuss include von Neumann's minimax theorem as well as several results on the closure of the sum of linear subspaces.

Throughout this chapter, $\mathcal{K}$ is a real Hilbert space.

15.1 The Attouch–Brézis Theorem

Proposition 15.1 *Let f and g be functions in $\Gamma_0(\mathcal{H})$ such that* $\operatorname{dom} f \cap \operatorname{dom} g \neq \varnothing$*. Then $f^* \mathbin{\square} g^*$ is proper and convex, and it possesses a continuous affine minorant. Moreover,*

$$(f+g)^* = (f^* \mathbin{\square} g^*)^{**} = (f^* \mathbin{\square} g^*)\breve{}. \tag{15.1}$$

Proof. Proposition 13.24(i) and Corollary 13.38 yield $(f^* \mathbin{\square} g^*)^* = f^{**} + g^{**} = f + g \in \Gamma_0(\mathcal{H})$. In turn, Proposition 13.10(iii), Proposition 12.11, and Proposition 13.12(ii) imply that $f^* \mathbin{\square} g^*$ is proper and convex, and that it possesses a continuous affine minorant. Therefore, invoking Proposition 13.45, we deduce that $(f+g)^* = (f^* \mathbin{\square} g^*)^{**} = (f^* \mathbin{\square} g^*)\breve{}$. $\square$

Proposition 15.2 *Let f and g be functions in $\Gamma_0(\mathcal{H})$ such that* $0 \in \operatorname{core}(\operatorname{dom} f - \operatorname{dom} g)$*. Then $(f+g)^* = f^* \mathbin{\boxdot} g^*$.*

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_15

Proof. We work in $\mathcal{H} \oplus \mathcal{H}$. For $\eta \in \mathbb{R}$ and $\rho \in \mathbb{R}_+$, set

$$S_{\eta,\rho} = \big\{(u,v) \in \mathcal{H} \times \mathcal{H} \;\big|\; f^*(u) + g^*(v) \leqslant \eta \text{ and } \|u+v\| \leqslant \rho\big\}. \tag{15.2}$$

Now take a and b in $\mathcal{H}$. Since $\operatorname{cone}(\operatorname{dom} f - \operatorname{dom} g) = \mathcal{H}$, there exist $x \in \operatorname{dom} f$, $y \in \operatorname{dom} g$, and $\gamma \in \mathbb{R}_{++}$ such that $a - b = \gamma(x-y)$. Now set $\beta_{a,b} = \rho\|b - \gamma y\| + \gamma(f(x) + g(y) + \eta)$ and assume that $(u,v) \in S_{\eta,\rho}$. Then, by Cauchy–Schwarz and Fenchel–Young (Proposition 13.15),

$$\begin{aligned}\langle (a,b) \mid (u,v)\rangle &= \langle a \mid u\rangle + \langle b \mid v\rangle \\ &= \langle b - \gamma y \mid u+v\rangle + \gamma(\langle x \mid u\rangle + \langle y \mid v\rangle) \\ &\leqslant \|b-\gamma y\|\,\|u+v\| + \gamma\big(f(x) + f^*(u) + g(y) + g^*(v)\big) \\ &\leqslant \beta_{a,b}.\end{aligned} \tag{15.3}$$

Thus,

$$\big(\forall (a,b) \in \mathcal{H} \times \mathcal{H}\big) \qquad \sup_{(u,v) \in S_{\eta,\rho}} |\langle (a,b) \mid (u,v)\rangle| \leqslant \max\{\beta_{a,b}, \beta_{-a,-b}\} < +\infty. \tag{15.4}$$

It follows from Lemma 2.22 applied to the linear functionals $(a,b) \mapsto \langle (a,b) \mid (u,v)\rangle$ that $\sup_{(u,v)\in S_{\eta,\rho}} \|(u,v)\| < +\infty$. Hence, $S_{\eta,\rho}$ is bounded. On the other hand, $S_{\eta,\rho}$ is closed and convex. Altogether, Theorem 3.37 implies that $S_{\eta,\rho}$ is weakly compact. Since $+\colon \mathcal{H} \times \mathcal{H} \to \mathcal{H}$ is weakly continuous, Lemma 1.20 asserts that it maps $S_{\eta,\rho}$ to the weakly compact set

$$W_{\eta,\rho} = \big\{u+v \in \mathcal{H} \;\big|\; (u,v) \in \mathcal{H} \times \mathcal{H},\; f^*(u) + g^*(v) \leqslant \eta, \text{ and } \|u+v\| \leqslant \rho\big\}. \tag{15.5}$$

Hence $W_{\eta,\rho}$ is closed and, by Lemma 1.40, so is

$$W_\eta = \bigcup_{\rho \geqslant 0} W_{\eta,\rho} = \big\{u+v \in \mathcal{H} \;\big|\; (u,v) \in \mathcal{H} \times \mathcal{H},\; f^*(u) + g^*(v) \leqslant \eta\big\}. \tag{15.6}$$

Thus, for every $\nu \in \mathbb{R}$,

$$\begin{aligned}\operatorname{lev}_{\leqslant \nu}(f^* \,\square\, g^*) &= \Big\{w \in \mathcal{H} \;\Big|\; \inf_{u \in \mathcal{H}} f^*(u) + g^*(w-u) \leqslant \nu\Big\} \\ &= \bigcap_{\eta > \nu} \big\{w \in \mathcal{H} \;\big|\; (\exists\, u \in \mathcal{H})\;\; f^*(u) + g^*(w-u) \leqslant \eta\big\} \\ &= \bigcap_{\eta>\nu} W_\eta\end{aligned} \tag{15.7}$$

is closed and we deduce from Lemma 1.24 that $f^* \,\square\, g^*$ is lower semicontinuous. On the other hand, by Proposition 15.1, $f^* \,\square\, g^*$ is proper and convex. Altogether, $f^* \,\square\, g^* \in \Gamma_0(\mathcal{H})$. Therefore, Corollary 13.38 and (15.1) imply that

$$f^* \,\square\, g^* = (f^* \,\square\, g^*)^{**} = (f+g)^*. \tag{15.8}$$

It remains to show that $f^* \,\square\, g^*$ is exact. Fix $w \in \mathcal{H}$. If $w \notin \operatorname{dom}(f^* \,\square\, g^*)$, then $f^* \,\square\, g^*$ is exact at w. Now suppose that $w \in \operatorname{dom}(f^* \,\square\, g^*)$. Set $F\colon \mathcal{H} \times \mathcal{H} \to \left]-\infty, +\infty\right] : (u,v) \mapsto f^*(u) + g^*(v)$, $C = \{(u,v) \in \mathcal{H} \times \mathcal{H} \mid u + v = w\}$, and $D = C \cap \operatorname{lev}_{\leqslant \eta} F$, where $\eta \in \left](f^* \,\square\, g^*)(w), +\infty\right[$. Then $F \in \Gamma_0(\mathcal{H} \times \mathcal{H})$, C is closed and convex, and D is nonempty. Moreover, $D \subset S_{\eta,\|w\|}$ and, as shown above, $S_{\eta,\|w\|}$ is bounded. Hence, D is bounded and it follows from Theorem 11.10 that F achieves its infimum on C. We conclude that $f^* \,\square\, g^*$ is exact at w. ☐

Theorem 15.3 (Attouch–Brézis) *Let f and g be functions in $\Gamma_0(\mathcal{H})$ such that the conical hull of* $\operatorname{dom} f - \operatorname{dom} g$ *is a closed linear subspace, i.e.,*

$$0 \in \operatorname{sri}(\operatorname{dom} f - \operatorname{dom} g). \tag{15.9}$$

Then $(f+g)^* = f^* \boxdot g^* \in \Gamma_0(\mathcal{H})$.

Proof. Since $f + g \in \Gamma_0(\mathcal{H})$, Corollary 13.38 implies that $(f+g)^* \in \Gamma_0(\mathcal{H})$. Let us fix $z \in \operatorname{dom} f \cap \operatorname{dom} g$, which is nonempty by (15.9), and let us set $\varphi\colon x \mapsto f(x+z)$ and $\psi\colon y \mapsto g(y+z)$. Note that $0 \in \operatorname{dom}\varphi \cap \operatorname{dom}\psi$ and that $\operatorname{dom}\varphi - \operatorname{dom}\psi = \operatorname{dom} f - \operatorname{dom} g$. Now set $K = \operatorname{cone}(\operatorname{dom}\varphi - \operatorname{dom}\psi) = \overline{\operatorname{span}}\,(\operatorname{dom}\varphi - \operatorname{dom}\psi)$. Then

$$\operatorname{dom}\varphi \subset K \quad \text{and} \quad \operatorname{dom}\psi \subset K. \tag{15.10}$$

It follows from (15.9) that, in the Hilbert space K, we have

$$0 \in \operatorname{core}(\operatorname{dom}\varphi|_K - \operatorname{dom}\psi|_K). \tag{15.11}$$

Now set $h = \langle z \mid \cdot \rangle$ and let $u \in \mathcal{H}$. By invoking Proposition 13.23(iii), (15.10), Proposition 13.23(vi), and (15.11), and then applying Proposition 15.2 in K to the functions $\varphi|_K \in \Gamma_0(K)$ and $\psi|_K \in \Gamma_0(K)$, we obtain

$$\begin{aligned}
(f+g)^*(u) - h(u) &= (\varphi + \psi)^*(u) \\
&= \big(\varphi|_K + \psi|_K\big)^*(P_K u) \\
&= \big((\varphi|_K)^* \boxdot (\psi|_K)^*\big)(P_K u) \\
&= \min_{v \in K} \Big((\varphi|_K)^*(v) + (\psi|_K)^*(P_K u - v)\Big) \\
&= \min_{w \in \mathcal{H}} \Big((\varphi|_K)^*(P_K w) + (\psi|_K)^*\big(P_K(u - w)\big)\Big) \\
&= \min_{w \in \mathcal{H}} \Big(\varphi^*(w) + \psi^*(u - w)\Big) \\
&= \big(\varphi^* \boxdot \psi^*\big)(u) \\
&= \big((f^* - h) \boxdot (g^* - h)\big)(u) \\
&= (f^* \boxdot g^*)(u) - h(u).
\end{aligned} \tag{15.12}$$

Consequently, $(f+g)^*(u) = (f^* \,\Box\, g^*)(u)$. ☐

Remark 15.4 The following examples show that the assumptions in Theorem 15.3 are tight.

(i) Suppose that $\mathcal{H} = \mathbb{R}^2$, let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right] : x = (\xi_1,\xi_2) \mapsto -\sqrt{\xi_1\xi_2}$ if $x \in \mathbb{R}^2_+$; $+\infty$ otherwise, and let $g = \iota_{\{0\}\times\mathbb{R}}$. Then $f+g = \iota_{\{0\}\times\mathbb{R}_+}$, $f^*(\nu_1,\nu_2) = 0$ if $\nu_1 \leqslant 1/(4\nu_2) < 0$; $+\infty$ otherwise, and $g^* = \iota_{\mathbb{R}\times\{0\}}$. Thus, $(f+g)^* = \iota_{\mathbb{R}\times\mathbb{R}_-} \neq \iota_{\mathbb{R}\times\mathbb{R}_{--}} = f^* \,\Box\, g^*$. Here $\operatorname{cone}(\operatorname{dom} f - \operatorname{dom} g) = \mathbb{R}_+ \times \mathbb{R}$ is a closed cone but it is not a closed linear subspace.

(ii) Suppose that $\mathcal{H}$ is infinite-dimensional, and let U and V be closed linear subspaces of $\mathcal{H}$ such that $U + V$ is not closed (see Example 3.41) or, equivalently, such that $U^\perp + V^\perp$ is not closed (see Corollary 15.35 below). Set $f = \iota_U$ and $g = \iota_V$. Then Proposition 6.35 implies that $(f+g)^* = \iota_{\overline{U^\perp + V^\perp}}$, whereas $f^* \,\Box\, g^* = \iota_{U^\perp + V^\perp}$. In this example, $\operatorname{cone}(\operatorname{dom} f - \operatorname{dom} g)$ is a linear subspace (equivalently, $0 \in \operatorname{ri}(\operatorname{dom} f - \operatorname{dom} g)$) that is not closed. Therefore, Theorem 15.3 fails if the strong relative interior is replaced by the relative interior.

(iii) Suppose that $\mathcal{H}$ is infinite-dimensional, let f be as in Example 9.22, and let $g = \iota_{\{0\}}$. Then $f + g = g$ and $(f+g)^* \equiv 0$. Since $f^* \equiv +\infty$ by Proposition 13.12(ii), we have $f^* \,\Box\, g^* \equiv +\infty$. Hence $(f+g)^* \neq f^* \,\Box\, g^*$ even though $\operatorname{dom} f - \operatorname{dom} g = \mathcal{H}$. Therefore, assuming the lower semicontinuity of f and g is necessary.

Proposition 15.5 *Let f and g be in $\Gamma_0(\mathcal{H})$. Suppose that* $\operatorname{dom} f \cap \operatorname{dom} g \neq \varnothing$ *and that one of the following holds:*

(i) $\operatorname{cone}(\operatorname{dom} f - \operatorname{dom} g) = \overline{\operatorname{span}}\,(\operatorname{dom} f - \operatorname{dom} g)$.
(ii) $0 \in \operatorname{core}(\operatorname{dom} f - \operatorname{dom} g)$.
(iii) $0 \in \operatorname{int}(\operatorname{dom} f - \operatorname{dom} g)$.
(iv) $\operatorname{cont} f \cap \operatorname{dom} g \neq \varnothing$.
(v) $\mathcal{H}$ *is finite-dimensional and* $\operatorname{ri}(\operatorname{dom} f) \cap \operatorname{ri}(\operatorname{dom} g) \neq \varnothing$.

Then $0 \in \operatorname{sri}(\operatorname{dom} f - \operatorname{dom} g)$, *i.e.,* (15.9) *holds.*

Proof. The assertions follow from Proposition 6.19 and Corollary 8.39. ☐

Remark 15.6 The conditions in Proposition 15.5 are not equivalent. For instance, in $\mathcal{H} = \mathbb{R}^2$, take $\operatorname{dom} f = \mathbb{R} \times \{0\}$ and $\operatorname{dom} g = \{0\} \times \mathbb{R}$. Then (iv) is not satisfied but (iii) is. On the other hand, if $\operatorname{dom} f = [0,1] \times \{0\}$ and $\operatorname{dom} g = [0,1] \times \{0\}$, then (ii) is not satisfied but (15.9) and (v) are.

The following result, which extends Proposition 12.14, provides conditions under which the infimal convolution is lower semicontinuous.

Proposition 15.7 *Let f and g be in $\Gamma_0(\mathcal{H})$. Suppose that one of the following holds:*

(i) $0 \in \operatorname{sri}(\operatorname{dom} f^* - \operatorname{dom} g^*)$.
(ii) $f + g^\vee$ *is coercive and* $0 \in \operatorname{sri}(\operatorname{dom} f - \operatorname{dom} g^\vee)$.
(iii) f *is coercive and* g *is bounded below.*
(iv) $\operatorname{dom} f^* = \mathcal{H}$.
(v) f *is supercoercive.*

Then $f \,\square\, g = f \,\boxdot\, g \in \Gamma_0(\mathcal{H})$.

Proof. (i): Apply Theorem 15.3 to f^* and g^*.

(ii): Theorem 15.3 implies that $(f + g^\vee)^* = f^* \boxdot g^{*\vee} \in \Gamma_0(\mathcal{H})$. Hence, by Proposition 14.16, $0 \in \operatorname{int}\operatorname{dom}(f + g^\vee)^* = \operatorname{int}\operatorname{dom}(f^* \boxdot g^{*\vee}) = \operatorname{int}(\operatorname{dom} f^* + \operatorname{dom} g^{*\vee}) = \operatorname{int}(\operatorname{dom} f^* - \operatorname{dom} g^*) \subset \operatorname{sri}(\operatorname{dom} f^* - \operatorname{dom} g^*)$. Now use (i).

(iii): Proposition 14.16 yields $0 \in \operatorname{int}\operatorname{dom} f^*$. On the other hand, $0 \in \operatorname{dom} g^*$ since $g^*(0) = -\inf g(\mathcal{H}) < +\infty$. Hence, $0 \in \operatorname{int}(\operatorname{dom} f^* - \operatorname{dom} g^*) \subset \operatorname{sri}(\operatorname{dom} f^* - \operatorname{dom} g^*)$. Conclude using (i).

(iv): This follows from (i).

(v): Combine (iv) and Proposition 14.15. ☐

Corollary 15.8 *Let* φ *and* ψ *be functions in* $\Gamma_0(\mathcal{H} \times \mathcal{K})$. *Set*

$$F \colon \mathcal{H} \times \mathcal{K} \to [-\infty, +\infty] : (x, y) \mapsto \big(\varphi(x, \cdot) \,\square\, \psi(x, \cdot)\big)(y), \tag{15.13}$$

and assume that

$$0 \in \operatorname{sri} Q_1(\operatorname{dom}\varphi - \operatorname{dom}\psi), \tag{15.14}$$

where $Q_1 \colon \mathcal{H} \times \mathcal{K} \to \mathcal{H} \colon (x, y) \mapsto x$. *Then*

$$F^* \colon \mathcal{H} \times \mathcal{K} \to [-\infty, +\infty] : (u, v) \mapsto \big(\varphi^*(\cdot, v) \boxdot \psi^*(\cdot, v)\big)(u). \tag{15.15}$$

Proof. Set $\boldsymbol{\mathcal{H}} = \mathcal{H} \times \mathcal{K} \times \mathcal{K}$ and define $\Phi \colon \boldsymbol{\mathcal{H}} \to \left]-\infty, +\infty\right] : (x, y_1, y_2) \mapsto \varphi(x, y_1)$ and $\Psi \colon \boldsymbol{\mathcal{H}} \to \left]-\infty, +\infty\right] : (x, y_1, y_2) \mapsto \psi(x, y_2)$. Then Φ and Ψ belong to $\Gamma_0(\boldsymbol{\mathcal{H}})$, and

$$\big(\forall (u, v_1, v_2) \in \boldsymbol{\mathcal{H}}\big) \quad \begin{cases} \Phi^*(u, v_1, v_2) = \varphi^*(u, v_1) + \iota_{\{0\}}(v_2), \\ \Psi^*(u, v_1, v_2) = \psi^*(u, v_2) + \iota_{\{0\}}(v_1). \end{cases} \tag{15.16}$$

Now define $G \in \Gamma_0(\boldsymbol{\mathcal{H}})$ by

$$G \colon (x, y_1, y_2) \mapsto \varphi(x, y_1) + \psi(x, y_2) = \Phi(x, y_1, y_2) + \Psi(x, y_1, y_2) \tag{15.17}$$

and set

$$L \colon \boldsymbol{\mathcal{H}} \to \mathcal{H} \times \mathcal{K} \colon (x, y_1, y_2) \mapsto (x, y_1 + y_2). \tag{15.18}$$

Then $L \triangleright G = F$, $L^* \colon \mathcal{H} \times \mathcal{K} \to \boldsymbol{\mathcal{H}} \colon (u, v) \mapsto (u, v, v)$, and since (15.14) is equivalent to $0 \in \operatorname{sri}(\operatorname{dom}\Phi - \operatorname{dom}\Psi)$, Theorem 15.3 yields $G^* = \Phi^* \boxdot \Psi^*$. Altogether, Proposition 13.24(iv) implies that $(\forall (u, v) \in \mathcal{H} \times \mathcal{K})$ $F^*(u, v) = (L \triangleright G)^*(u, v) = G^*(L^*(u, v)) = (\Phi^* \boxdot \Psi^*)(u, v, v)$. In view of (15.16), this is precisely (15.15). ☐

15.2 Fenchel Duality

We consider the problem of minimizing the sum of two proper functions.

Proposition 15.9 *Let f and g be proper functions from $\mathcal{H}$ to $]-\infty,+\infty]$. Then*

$$(\forall x \in \mathcal{H})(\forall u \in \mathcal{H}) \quad f(x) + g(x) \geqslant -f^*(-u) - g^*(u) \tag{15.19}$$

and

$$\inf(f+g)(\mathcal{H}) \geqslant -\inf(f^{*\vee} + g^*)(\mathcal{H}). \tag{15.20}$$

Proof. Using Proposition 13.10(ii) and Proposition 13.15, we see that

$$\begin{aligned}(\forall x \in \mathcal{H})(\forall u \in \mathcal{H}) \quad f(x) + g(x) + f^*(-u) + g^*(u) &\geqslant \langle x \mid -u\rangle + \langle x \mid u\rangle \\ &= 0,\end{aligned} \tag{15.21}$$

and the result follows. □

Definition 15.10 The *primal problem* associated with the sum of two proper functions $f\colon \mathcal{H} \to]-\infty,+\infty]$ and $g\colon \mathcal{H} \to]-\infty,+\infty]$ is

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; f(x) + g(x), \tag{15.22}$$

its *dual problem* is

$$\underset{u\in\mathcal{H}}{\text{minimize}} \;\; f^*(-u) + g^*(u), \tag{15.23}$$

the *primal optimal value* is $\mu = \inf(f+g)(\mathcal{H})$, the *dual optimal value* is $\mu^* = \inf(f^{*\vee} + g^*)(\mathcal{H})$, and the *duality gap* is

$$\Delta(f,g) = \begin{cases} 0, & \text{if } \mu = -\mu^* \in \{-\infty,+\infty\};\\ \mu + \mu^*, & \text{otherwise.}\end{cases} \tag{15.24}$$

Remark 15.11 We follow here the common convention of not emphasizing the fact that the dual problem of Definition 15.10 actually depends on the ordered pair (f,g) rather than on $\{f,g\}$.

Proposition 15.12 *Let f and g be proper functions from $\mathcal{H}$ to $]-\infty,+\infty]$, and set $\mu = \inf(f+g)(\mathcal{H})$ and $\mu^* = \inf(f^{*\vee} + g^*)(\mathcal{H})$. Then the following hold:*

(i) $\mu \geqslant -\mu^*$.
(ii) $\Delta(f,g) \in [0,+\infty]$.
(iii) $\mu = -\mu^* \Leftrightarrow \Delta(f,g) = 0$.

Proof. Clear from Proposition 15.9 and Definition 15.10. □

The next proposition describes a situation in which the duality gap (15.24) is 0 and in which the dual problem possesses a solution.

Proposition 15.13 *Let f and g be functions in $\Gamma_0(\mathcal{H})$ such that*

$$0 \in \operatorname{sri}(\operatorname{dom} f - \operatorname{dom} g). \tag{15.25}$$

Then $\inf(f+g)(\mathcal{H}) = -\min(f^{*\vee} + g^*)(\mathcal{H})$.

Proof. Proposition 13.10(i) and Theorem 15.3 imply that $\inf(f+g)(\mathcal{H}) = -(f+g)^*(0) = -(f^* \mathbin{\boxdot} g^*)(0) = -\min(f^{*\vee} + g^*)(\mathcal{H})$. ☐

Corollary 15.14 *Let $f \in \Gamma_0(\mathcal{H})$ and let K be a closed convex cone in $\mathcal{H}$ such that* $0 \in \operatorname{sri}(\operatorname{dom} f - K)$. *Then* $\inf f(K) = -\min f^*(K^\oplus)$.

Proof. Set $g = \iota_K$ in Proposition 15.13, and use Example 13.3(ii) and Definition 6.9. ☐

Corollary 15.15 *Let f and g be functions in $\Gamma_0(\mathcal{H})$ such that* $0 \in \operatorname{sri}(\operatorname{dom} f - \operatorname{dom} g)$. *Suppose that $f + g \geqslant 0$ and that $g^* = g \circ L$, where $L \in \mathcal{B}(\mathcal{H})$. Then there exists $v \in \mathcal{H}$ such that $f^*(-v) + g(Lv) \leqslant 0$.*

Proof. By Proposition 15.13, there exists $v \in \mathcal{H}$ such that $0 \leqslant \inf(f+g)(\mathcal{H}) = -f^*(-v) - g^*(v)$. Hence $0 \geqslant f^*(-v) + g(Lv)$. ☐

Corollary 15.16 *Let f and g be functions in $\Gamma_0(\mathcal{H})$ such that* $0 \in \operatorname{sri}(\operatorname{dom} f - \operatorname{dom} g)$. *Suppose that $g^* = g^\vee$ and that $f + g \geqslant 0$. Then there exists $v \in \mathcal{H}$ such that $f^*(v) + g(v) \leqslant 0$, i.e., such that $(\forall x \in \mathcal{H})\ f(x) + g(x) \geqslant g(x) + \langle x \mid v\rangle + g(v) \geqslant 0$.*

Proof. Applying Corollary 15.15 with $L = -\mathrm{Id}$, we obtain the existence of $w \in \mathcal{H}$ such that $0 \geqslant f^*(-w) + g(-w) = \sup_{x\in\mathcal{H}}(\langle x \mid -w\rangle - f(x)) + g(-w)$. Now set $v = -w$. Then Proposition 13.15 yields $(\forall x \in \mathcal{H})\ f(x) + g(x) \geqslant g(x) + \langle x \mid v\rangle + g(v) = g(x) + g^*(-v) - \langle x \mid -v\rangle \geqslant 0$. ☐

Corollary 15.17 *Let $f \in \Gamma_0(\mathcal{H})$ and set $q = (1/2)\|\cdot\|^2$. Suppose that $f + q \geqslant 0$. Then there exists a vector $w \in \mathcal{H}$ such that $(\forall x \in \mathcal{H})\ f(x) + q(x) \geqslant q(x - w)$.*

Proof. By Example 13.6, $q^* = q^\vee$. Applying Corollary 15.16 with $g = q$ yields a vector v in $\mathcal{H}$ such that $(\forall x \in \mathcal{H})\ f(x) + q(x) \geqslant q(x) + \langle x \mid v\rangle + q(v) = q(x+v)$. Hence, the conclusion follows with $w = -v$. ☐

15.3 Fenchel–Rockafellar Duality

We now turn our attention to a more general variational problem involving a linear operator.

Proposition 15.18 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper, let $g\colon \mathcal{K} \to \left]-\infty,+\infty\right]$ be proper, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$. Then*

$$(\forall x \in \mathcal{H})(\forall v \in \mathcal{K}) \quad f(x) + g(Lx) \geqslant -f^*(-L^*v) - g^*(v) \tag{15.26}$$

and

$$\inf(f + g \circ L)(\mathcal{H}) \geqslant -\inf(f^{*\vee} \circ L^* + g^*)(\mathcal{K}). \tag{15.27}$$

Proof. Using Proposition 13.10(ii) and Proposition 13.15, we see that for every $x \in \mathcal{H}$ and every $v \in \mathcal{K}$,

$$f(x) + g(Lx) + f^*(-L^*v) + g^*(v) \geqslant \langle x \mid -L^*v\rangle + \langle Lx \mid v\rangle = 0, \tag{15.28}$$

and the result follows. ☐

Definition 15.19 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$, let $g\colon \mathcal{K} \to \left]-\infty,+\infty\right]$, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$. The *primal problem* associated with the composite function $f + g \circ L$ is

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x) + g(Lx), \tag{15.29}$$

its *dual problem* is

$$\underset{v \in \mathcal{K}}{\text{minimize}} \;\; f^*(-L^*v) + g^*(v), \tag{15.30}$$

the *primal optimal value* is $\mu = \inf\,(f + g \circ L)(\mathcal{H})$, the *dual optimal value* is $\mu^* = \inf(f^{*\vee} \circ L^* + g^*)(\mathcal{K})$, and the *duality gap* is

$$\varDelta(f,g,L) = \begin{cases} 0, & \text{if } \; \mu = -\mu^* \in \{-\infty,+\infty\}; \\ \mu + \mu^*, & \text{otherwise.} \end{cases} \tag{15.31}$$

Remark 15.20 As in Remark 15.11, the dual problem depends on the ordered triple (f,g,L). We follow here the common usage of not stressing this dependency.

The next proposition extends Proposition 15.12.

Proposition 15.21 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ and $g\colon \mathcal{K} \to \left]-\infty,+\infty\right]$ be proper, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$. Set $\mu = \inf(f + g \circ L)(\mathcal{H})$ and $\mu^* = \inf(f^{*\vee} \circ L^* + g^*)(\mathcal{K})$. Then the following hold:*

(i) $\mu \geqslant -\mu^*$.
(ii) $\varDelta(f,g,L) \in [0,+\infty]$.
(iii) $\mu = -\mu^* \Leftrightarrow \varDelta(f,g,L) = 0$.

Proof. Clear from Proposition 15.18 and Definition 15.19. ☐

A zero duality gap is not automatic, but it is guaranteed under additional assumptions.

Proposition 15.22 *Let $f \in \Gamma_0(\mathcal{H})$, let $g \in \Gamma_0(\mathcal{K})$, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$. Suppose that $0 \in \operatorname{core}(\operatorname{dom} g - L(\operatorname{dom} f))$. Then*

$$\inf\big(f + g\circ L\big)(\mathcal{H}) = -\min\big(f^{*\vee}\circ L^* + g^*\big)(\mathcal{K}). \tag{15.32}$$

Proof. Set $\varphi = f \oplus g$ and $V = \operatorname{gra} L$. Now take $(x,y) \in \mathcal{H}\times\mathcal{K}$. Since $0 \in \operatorname{core}(\operatorname{dom} g - L(\operatorname{dom} f))$, there exist $\gamma \in \mathbb{R}_{++}$, $a \in \operatorname{dom} f$, and $b \in \operatorname{dom} g$ such that $y - Lx = \gamma(b - La)$. Upon setting $z = a - x/\gamma$, we obtain $x = \gamma(a-z)$ and $y = \gamma(b - Lz)$. Therefore, $(x,y) = \gamma((a,b)-(z,Lz)) \in \operatorname{cone}((\operatorname{dom}\varphi) - V)$. We have thus shown that $\operatorname{cone}((\operatorname{dom}\varphi) - V) = \mathcal{H}\times\mathcal{K}$, which implies that $0 \in \operatorname{core}(V - \operatorname{dom}\varphi) \subset \operatorname{sri}(V - \operatorname{dom}\varphi)$. It then follows from Corollary 15.14 that $\inf\varphi(V) = -\min\varphi^*(V^\perp)$. However, $V^\perp = \{(u,v) \in \mathcal{H}\times\mathcal{K} \mid u = -L^*v\}$ by Fact 2.25(vii) and, by Proposition 13.30, $\varphi^* = f^* \oplus g^*$. Therefore, $\inf(f + g\circ L)(\mathcal{H}) = \inf\varphi(V) = -\min\varphi^*(V^\perp) = -\min((f^*\circ L^*)^\vee + g^*)(\mathcal{K}) = -\min(f^{*\vee}\circ L^* + g^*)(\mathcal{K})$. ☐

Theorem 15.23 *Let $f \in \Gamma_0(\mathcal{H})$, let $g \in \Gamma_0(\mathcal{K})$, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$ be such that*

$$0 \in \operatorname{sri}\big(\operatorname{dom} g - L(\operatorname{dom} f)\big). \tag{15.33}$$

Then $\inf(f + g\circ L)(\mathcal{H}) = -\min(f^{*\vee}\circ L^* + g^*)(\mathcal{K})$.

Proof. We first consider the special case $0 \in \operatorname{dom} f$ and $0 \in \operatorname{dom} g$. Set $\mathcal{A} = \overline{\operatorname{span}}\operatorname{dom} f$ and $\mathcal{B} = \operatorname{cone}(\operatorname{dom} g - L(\operatorname{dom} f))$. By (15.33), $\mathcal{B} = \overline{\operatorname{span}}(\operatorname{dom} g - L(\operatorname{dom} f))$. Hence $\operatorname{dom} f \subset \mathcal{A}$, $\operatorname{dom} g \subset \mathcal{B}$, and, since $L(\operatorname{dom} f) \subset \mathcal{B}$, we have $L(\mathcal{A}) \subset \mathcal{B}$. It follows that $\operatorname{ran} L|_{\mathcal{A}} \subset \mathcal{B}$, and, in turn, using Fact 2.25(v), that

$$\mathcal{B}^\perp \subset (\operatorname{ran} L|_{\mathcal{A}})^\perp = \ker((L|_{\mathcal{A}})^*) = \ker P_{\mathcal{A}}L^*. \tag{15.34}$$

Therefore

$$P_{\mathcal{A}}L^* = P_{\mathcal{A}}L^*(P_{\mathcal{B}} + P_{\mathcal{B}^\perp}) = P_{\mathcal{A}}L^*P_{\mathcal{B}} = P_{\mathcal{A}}L^*|_{\mathcal{B}}P_{\mathcal{B}}. \tag{15.35}$$

Next, we observe that condition (15.33) in $\mathcal{K}$ yields

$$0 \in \operatorname{core}\big(\operatorname{dom} g|_{\mathcal{B}} - (P_{\mathcal{B}}L|_{\mathcal{A}})(\operatorname{dom} f|_{\mathcal{A}})\big) \tag{15.36}$$

in $\mathcal{B}$. Thus, using the inclusions $\operatorname{dom} f \subset \mathcal{A}$ and $L(\mathcal{A}) \subset \mathcal{B}$, (15.36), Proposition 15.22, (15.35), and Proposition 13.23(vi), we obtain

$$\begin{aligned}
\inf_{x\in\mathcal{H}}\big(f(x) + g(Lx)\big) &= \inf_{x\in\mathcal{A}}\Big(f|_{\mathcal{A}}(x) + g|_{\mathcal{B}}\big((P_{\mathcal{B}}L|_{\mathcal{A}})x\big)\Big)\\
&= -\min_{v\in\mathcal{B}}\Big((f|_{\mathcal{A}})^*\big(-(P_{\mathcal{B}}L|_{\mathcal{A}})^*v\big) + (g|_{\mathcal{B}})^*(v)\Big)\\
&= -\min_{v\in\mathcal{K}}\Big((f|_{\mathcal{A}})^*\big(-(P_{\mathcal{A}}L^*|_{\mathcal{B}})(P_{\mathcal{B}}v)\big) + (g|_{\mathcal{B}})^*\big(P_{\mathcal{B}}v\big)\Big)\\
&= -\min_{v\in\mathcal{K}}\Big((f|_{\mathcal{A}})^*(-P_{\mathcal{A}}L^*v) + (g|_{\mathcal{B}})^*\big(P_{\mathcal{B}}v\big)\Big)\\
&= -\min_{v\in\mathcal{K}}\big(f^{*\vee}(L^*v) + g^*(v)\big).
\end{aligned} \tag{15.37}$$

We now consider the general case. In view of (15.33), there exist $b \in \operatorname{dom} g$ and $a \in \operatorname{dom} f$ such that $b = La$. Set $\varphi\colon x \mapsto f(x+a)$ and $\psi\colon y \mapsto g(y+b)$. Then $0 \in \operatorname{dom}\varphi$, $0 \in \operatorname{dom}\psi$, and $\operatorname{dom}\psi - L(\operatorname{dom}\varphi) = \operatorname{dom} g - L(\operatorname{dom} f)$. We therefore apply the above special case with φ and ψ to obtain

$$
\begin{aligned}
\inf_{x\in\mathcal{H}} \big(f(x) + g(Lx)\big) &= \inf_{x\in\mathcal{H}} \big(\varphi(x) + \psi(Lx)\big) \\
&= -\min_{v\in\mathcal{K}} \big(\varphi^*(-L^*v) + \psi^*(v)\big) \\
&= -\min_{v\in\mathcal{K}} \big(f^*(-L^*v) - \langle -L^*v \mid a\rangle + g^*(v) - \langle v \mid b\rangle\,\big) \\
&= -\min_{v\in\mathcal{K}} \big(f^*(-L^*v) + g^*(v)\big), \qquad (15.38)
\end{aligned}
$$

where we have used Proposition 13.23(iii) and the identity $La = b$. ☐

The following proposition provides sufficient conditions under which (15.33) is satisfied.

Proposition 15.24 *Let $f \in \Gamma_0(\mathcal{H})$, let $g \in \Gamma_0(\mathcal{K})$, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$. Suppose that $\operatorname{dom} g \cap L(\operatorname{dom} f) \neq \varnothing$ and that one of the following holds:*

(i) $\operatorname{cone}(\operatorname{dom} g - L(\operatorname{dom} f)) = \overline{\operatorname{span}}\,(\operatorname{dom} g - L(\operatorname{dom} f))$.
(ii) $\operatorname{dom} g - L(\operatorname{dom} f)$ *is a closed linear subspace.*
(iii) $\operatorname{dom} f$ *and* $\operatorname{dom} g$ *are linear subspaces and* $\operatorname{dom} g + L(\operatorname{dom} f)$ *is closed.*
(iv) $\operatorname{dom} g$ *is a cone and* $\operatorname{dom} g - \operatorname{cone} L(\operatorname{dom} f)$ *is a closed linear subspace.*
(v) $0 \in \operatorname{core}(\operatorname{dom} g - L(\operatorname{dom} f))$.
(vi) $0 \in \operatorname{int}(\operatorname{dom} g - L(\operatorname{dom} f))$.
(vii) $\operatorname{cont} g \cap L(\operatorname{dom} f) \neq \varnothing$.
(viii) $\mathcal{K}$ *is finite-dimensional and* $(\operatorname{ri}\operatorname{dom} g) \cap (\operatorname{ri} L(\operatorname{dom} f)) \neq \varnothing$.
(ix) $\mathcal{K}$ *is finite-dimensional and* $(\operatorname{ri}\operatorname{dom} g) \cap L(\operatorname{qri}\operatorname{dom} f) \neq \varnothing$.
(x) $\mathcal{H}$ *and* $\mathcal{K}$ *are finite-dimensional and* $(\operatorname{ri}\operatorname{dom} g) \cap L(\operatorname{ri}\operatorname{dom} f) \neq \varnothing$.

Then $0 \in \operatorname{sri}(\operatorname{dom} g - L(\operatorname{dom} f))$.

Proof. The assertions follow from Proposition 8.2, Proposition 3.5, (6.8), Proposition 6.19, and Corollary 8.39. ☐

We now turn to a result that is complementary to Theorem 15.23 and that relies on several external results drawn from [313]. To formulate it, we require the notions of polyhedral (convex) set and function.

A subset of $\mathcal{H}$ is *polyhedral* if it is a finite intersection of closed half-spaces, and a function $f\colon \mathcal{H} \to [-\infty,+\infty]$ is *polyhedral* if $\operatorname{epi} f$ is a polyhedral set.

Fact 15.25 *Let $f \in \Gamma_0(\mathcal{H})$, let $g \in \Gamma_0(\mathcal{K})$ be polyhedral, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$. Suppose that $\mathcal{K}$ is finite-dimensional and that one of the following holds:*

(i) $\operatorname{dom} g \cap \operatorname{ri} L(\operatorname{dom} f) \neq \varnothing$.
(ii) $\mathcal{H}$ *is finite-dimensional,* f *is polyhedral, and* $\operatorname{dom} g \cap L(\operatorname{dom} f) \neq \varnothing$.

Then $\inf(f + g\circ L)(\mathcal{H}) = -\min((f^{*\vee} \circ L^*) + g^*)(\mathcal{K})$.

Proof. We set $\mu = \inf(f + g \circ L)(\mathcal{H})$ and, in view of Proposition 15.21(i), we assume that $\mu > -\infty$. Then

$$\begin{aligned}\mu &= \inf_{(x,y)\in \operatorname{gra} L} \big(f(x) + g(y)\big) \\ &= \inf_{y\in\mathcal{K}} \Big(g(y) + \inf_{x\in L^{-1}y} f(x)\Big) \\ &= \inf_{y\in\mathcal{K}} \big(g(y) + (L \rhd f)(y)\big). \end{aligned} \tag{15.39}$$

Proposition 12.36(i) and Proposition 13.24(iv) assert that $\operatorname{dom}(L \rhd f) = L(\operatorname{dom} f)$ and that $(L \rhd f)^* = f^* \circ L^*$. If (i) holds, then let $z \in \operatorname{dom} f$ be such that $Lz \in \operatorname{dom} g \cap \operatorname{ri} L(\operatorname{dom} f) = \operatorname{dom} g \cap \operatorname{ri} \operatorname{dom}(L \rhd f)$; otherwise, (ii) holds and we let $z \in \operatorname{dom} f$ be such that $Lz \in \operatorname{dom} g \cap L(\operatorname{dom} f) = \operatorname{dom} g \cap \operatorname{dom}(L \rhd f)$. In either case, $\mu \leqslant g(Lz) + (L \rhd f)(Lz) \leqslant g(Lz) + f(z) < +\infty$ and thus

$$\mu \in \mathbb{R}. \tag{15.40}$$

If $(L \rhd f)(Lz) = -\infty$, then there exists a sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ such that $Lx_n \equiv Lz$ and $f(x_n) \to -\infty$, which implies that $\mu \leqslant f(x_n) + g(Lx_n) = f(x_n) + g(Lz) \to -\infty$, a contradiction to (15.40). Hence

$$(L \rhd f)(Lz) \in \mathbb{R}. \tag{15.41}$$

If (i) holds, then $L \rhd f$ is convex by Proposition 12.36(ii) and proper by [313, Theorem 7.2]. If (ii) holds, then $L \rhd f$ is polyhedral by [313, Corollary 19.3.1] and also proper. Therefore, [313, Theorem 31.1] yields

$$\begin{aligned}\inf_{y\in\mathcal{K}} \big(g(y) + (L \rhd f)(y)\big) &= -\min_{v\in\mathcal{K}} \big(g^*(v) + (L \rhd f)^*(-v)\big) \\ &= -\min_{v\in\mathcal{K}} \big(g^*(v) + f^{*\vee}(L^*v)\big). \end{aligned} \tag{15.42}$$

The conclusion follows by combining (15.39) with (15.42). ☐

15.4 A Conjugation Result

We start with a general conjugation formula.

Proposition 15.26 *Let $f \in \Gamma_0(\mathcal{H})$, let $g \in \Gamma_0(\mathcal{K})$, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$ be such that $\operatorname{dom} g \cap L(\operatorname{dom} f) \neq \varnothing$. Then $(f + g \circ L)^* = (f^* \,\square\, (L^* \rhd g^*))^{**}$. Moreover, $f^* \,\square\, (L^* \rhd g^*)$ is proper and convex, and it possesses a continuous affine minorant.*

Proof. It follows from Corollary 13.38 and Corollary 13.25(i) that $(f+g\circ L)^* = ((f^*)^* + ((g^*)^* \circ (L^*)^*))^* = (f^* \,\square\, (L^* \rhd g^*))^{**}$. The remaining statements follow as in the proof of Proposition 15.1. Indeed, since $f + g \circ L \in \Gamma_0(\mathcal{H})$, we have $f + g \circ L = (f + g \circ L)^{**} = (f^* \,\square\, (L^* \triangleright g^*))^{***} = (f^* \,\square\, (L^* \triangleright g^*))^*$. Hence $f^* \,\square\, (L^* \triangleright g^*)$ is proper by Proposition 13.10(iii), convex by Proposition 12.36(ii) and Proposition 12.11, and it possesses a continuous affine minorant by Proposition 13.12(ii). ☐

We now obtain an extension of Theorem 15.3.

Theorem 15.27 *Let $f \in \Gamma_0(\mathcal{H})$, let $g \in \Gamma_0(\mathcal{K})$, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$. Suppose that one of the following holds:*

(i) $0 \in \operatorname{sri}(\operatorname{dom} g - L(\operatorname{dom} f))$.
(ii) *$\mathcal{K}$ is finite-dimensional, g is polyhedral, and* $\operatorname{dom} g \cap \operatorname{ri} L(\operatorname{dom} f) \neq \varnothing$.
(iii) *$\mathcal{H}$ and $\mathcal{K}$ are finite-dimensional, f and g are polyhedral, and* $\operatorname{dom} g \cap L(\operatorname{dom} f) \neq \varnothing$.

Then $(f + g \circ L)^* = f^* \boxdot (L^* \triangleright g^*)$. In other words,

$$\big(\forall u \in \mathcal{H}\big) \quad \big(f + g \circ L\big)^*(u) = \min_{v \in \mathcal{K}} \big(f^*(u - L^*v) + g^*(v)\big). \tag{15.43}$$

Proof. Let $u \in \mathcal{H}$. Then it follows from Theorem 15.23 or from Fact 15.25 that

$$\begin{aligned}(f + g \circ L)^*(u) &= \sup_{x \in \mathcal{H}} \big(\langle x \mid u\rangle - f(x) - g(Lx)\big)\\ &= -\inf_{x \in \mathcal{H}} \big(f(x) - \langle x \mid u\rangle + g(Lx)\big)\\ &= \min_{v \in \mathcal{K}} \big(f^*(-L^*v + u) + g^*(v)\big), \end{aligned} \tag{15.44}$$

which yields the result. ☐

Corollary 15.28 *Let $g \in \Gamma_0(\mathcal{K})$ and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$. Suppose that one of the following holds:*

(i) $0 \in \operatorname{sri}(\operatorname{dom} g - \operatorname{ran} L)$.
(ii) *$\mathcal{K}$ is finite-dimensional, g is polyhedral, and* $\operatorname{dom} g \cap \operatorname{ran} L \neq \varnothing$.

Then $(g \circ L)^* = L^* \triangleright g^*$. In other words,

$$\big(\forall u \in \mathcal{H}\big) \quad (g \circ L)^*(u) = \min_{\substack{v \in \mathcal{K}\\ L^*v = u}} g^*(v) = (L^* \triangleright g^*)(u). \tag{15.45}$$

15.5 Applications

Example 15.29 Let C be a closed convex subset of $\mathcal{H}$, let D be a closed convex subset of $\mathcal{K}$, and suppose that $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$ satisfies $0 \in \operatorname{sri}(\operatorname{bar} D - L(C))$. Then

$$\inf_{x \in C} \sup_{v \in D} \langle Lx \mid v \rangle = \max_{v \in D} \inf_{x \in C} \langle Lx \mid v \rangle . \tag{15.46}$$

Proof. Set $f = \iota_C$ and $g = \sigma_D$ in Theorem 15.23. Then Example 13.3(i) and Example 13.43(i) yield $\inf_{x \in C} \sup_{v \in D} \langle Lx \mid v \rangle = \inf_{x \in \mathcal{H}} \iota_C(x) + \sigma_D(Lx) = \inf_{x \in \mathcal{H}} f(x) + g(Lx) = -\min_{v \in \mathcal{K}} f^*(-L^*v) + g^*(v) = -\min_{v \in D} \sigma_C(-L^*v) = -\min_{v \in D} \sup_{x \in C} \langle x \mid -L^*v \rangle = \max_{v \in D} \inf_{x \in C} \langle Lx \mid v \rangle$. ☐

In Euclidean spaces, the following result is known as von Neumann's minimax theorem.

Corollary 15.30 (von Neumann) *Let C be a nonempty bounded closed subset of $\mathcal{H}$, let D be a nonempty bounded closed subset of $\mathcal{K}$, and let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$. Then*

$$\min_{x \in C} \max_{v \in D} \langle Lx \mid v \rangle = \max_{v \in D} \min_{x \in C} \langle Lx \mid v \rangle . \tag{15.47}$$

Proof. Since D is bounded, we have $\operatorname{bar} D = \mathcal{K}$ and therefore $0 \in \operatorname{sri}(\operatorname{bar} D - L(C))$. Moreover, since Proposition 11.15(ii) implies that a continuous linear functional achieves its infimum and its supremum on a nonempty bounded closed convex set, we have $\sup_{v \in D} \langle Lx \mid v \rangle = \max_{v \in D} \langle Lx \mid v \rangle$ and $\inf_{x \in C} \langle Lx \mid v \rangle = \min_{x \in C} \langle Lx \mid v \rangle$. Now set $\varphi \colon x \mapsto \max_{v \in D} \langle Lx \mid v \rangle$. Then $\varphi(0) = 0$ and, therefore, Proposition 9.3 implies that $\varphi \in \Gamma_0(\mathcal{H})$. Thus, invoking Proposition 11.15(ii) again, derive from (15.46) that

$$\begin{aligned} \min_{x \in C} \max_{v \in D} \langle Lx \mid v \rangle &= \min_{x \in C} \varphi(x) \\ &= \inf_{x \in C} \sup_{v \in D} \langle Lx \mid v \rangle \\ &= \max_{v \in D} \inf_{x \in C} \langle Lx \mid v \rangle \\ &= \max_{v \in D} \min_{x \in C} \langle Lx \mid v \rangle , \end{aligned} \tag{15.48}$$

which yields (15.47). ☐

Corollary 15.31 *Let C be a nonempty closed convex cone in $\mathcal{H}$, let D be a nonempty closed convex cone in $\mathcal{K}$, and let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$. Then the following hold:*

(i) $(C \cap L^{-1}(D))^{\ominus} = \overline{C^{\ominus} + L^*(D^{\ominus})}$.

(ii) *Suppose that $D - L(C)$ is a closed linear subspace of $\mathcal{K}$. Then the set $(C \cap L^{-1}(D))^{\ominus} = C^{\ominus} + L^*(D^{\ominus})$ is a nonempty closed convex cone.*

Proof. (i): Proposition 6.35 and Proposition 6.37(i) yield $(C \cap L^{-1}(D))^{\ominus} = \overline{C^{\ominus} + (L^{-1}(D))^{\ominus}} = \overline{C^{\ominus} + \overline{L^*(D^{\ominus})}} = \overline{C^{\ominus} + L^*(D^{\ominus})}$.

(ii): Set $f = \iota_C$ and $g = \iota_D$, and note that $0 \in \operatorname{sri}(\operatorname{dom} g - L(\operatorname{dom} f))$ by assumption. Example 13.3(ii) and Theorem 15.27(i) imply that, for every $u \in \mathcal{H}$,

$$\begin{aligned}
\iota_{(C\cap L^{-1}(D))^\ominus}(u) &= \iota^*_{C\cap L^{-1}(D)}(u)\\
&= \big(\iota_C + (\iota_D \circ L)\big)^*(u)\\
&= \big(f + g\circ L\big)^*(u)\\
&= \min_{v\in\mathcal{K}} \big(f^*(u-L^*v) + g^*(v)\big)\\
&= \min_{v\in\mathcal{K}} \big(\iota_{C^\ominus}(u-L^*v) + \iota_{D^\ominus}(v)\big)\\
&= \iota_{C^\ominus + L^*(D^\ominus)}(u).
\end{aligned} \tag{15.49}$$

Therefore, $C^\ominus \cap L^*(D^\ominus)$ is the polar cone of $C \cap L^{-1}(D)$ and hence, by Proposition 6.24(ii), it is closed. □

Corollary 15.32 *Let D be a nonempty closed convex cone in $\mathcal{K}$, let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$, and suppose that $D - \operatorname{ran} L$ is a closed linear subspace. Then $(L^{-1}(D))^\ominus = L^*(D^\ominus)$.*

Corollary 15.33 *Let C be a closed linear subspace of $\mathcal{H}$, let D be a closed linear subspace of $\mathcal{K}$, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$. Then $L(C) + D$ is closed if and only if $C^\perp + L^*(D^\perp)$ is closed.*

Corollary 15.34 *Let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$. Then $\operatorname{ran} L$ is closed if and only if $\operatorname{ran} L^*$ is closed.*

Corollary 15.35 *Let U and V be closed linear subspaces of $\mathcal{H}$. Then $U+V$ is closed if and only if $U^\perp + V^\perp$ is closed.*

Corollary 15.36 *Let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$ be an operator such that $\operatorname{ran} L$ is closed, and let V be a closed linear subspace of $\mathcal{H}$. Then $L(V)$ is closed if and only if $V + \ker L$ is closed.*

Proof. Since $\operatorname{ran} L$ is closed, Corollary 15.34 yields

$$\operatorname{ran} L^* = \overline{\operatorname{ran}}\, L^*. \tag{15.50}$$

Using Corollary 15.33, (15.50), Fact 2.25(iv), and Corollary 15.35, we obtain the equivalences $L(V)$ is closed $\Leftrightarrow$ $V^\perp + \operatorname{ran} L^*$ is closed $\Leftrightarrow$ $V^\perp + \overline{\operatorname{ran}}\, L^*$ is closed $\Leftrightarrow$ $V^\perp + (\ker L)^\perp$ is closed $\Leftrightarrow$ $V + \ker L$ is closed. □

Exercises

Exercise 15.1 Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$ be such that $\operatorname{dom} f \cap \operatorname{dom} g \neq \varnothing$. Show that $(f+g)^* = f^* \,\square\, g^*$ if and only if $\operatorname{epi} f^* + \operatorname{epi} g^*$ is closed.

Exercise 15.2 Let K_1 and K_2 be nonempty closed convex cones in $\mathcal{H}$. Use Proposition 15.1 to show that $(K_1 \cap K_2)^\ominus = \overline{K_1^\ominus + K_2^\ominus}$. Compare with Proposition 6.35.

Exercise 15.3 Let $f \in \Gamma_0(\mathcal{H})$. Show that $\operatorname{dom} f + \operatorname{dom} f^* = \mathcal{H}$.

Exercise 15.4 Provide the details for Remark 15.4(i).

Exercise 15.5 Consider the setting of Definition 15.10, and let f and g be as in Remark 15.4(i). Determine μ, μ^*, and $\Delta(f,g)$, and whether the primal and dual problems possess solutions.

Exercise 15.6 Consider the setting of Definition 15.19 with $\mathcal{H} = \mathcal{K} = \mathbb{R}$, $L = \mathrm{Id}$, and $f = g = \exp$. Prove that $\mu = \mu^* = 0$, that the primal problem has no solution, and that the dual problem has a unique solution.

Exercise 15.7 Find two functions f and g in $\Gamma_0(\mathbb{R})$ for which, in the setting of Definition 15.10, $\Delta(f,g) = 0$, the primal problem has a unique solution, and the dual problem has no solution.

Exercise 15.8 Find two functions f and g in $\Gamma_0(\mathbb{R})$ for which $\Delta(f,g) = 0$ and $\inf(f+g)(\mathcal{H}) = -\infty$.

Exercise 15.9 Find two functions f and g in $\Gamma_0(\mathbb{R})$ for which $\inf(f+g)(\mathcal{H}) = +\infty$ and $\inf(f^{*\vee} + g^*)(\mathcal{H}) = -\infty$.

Exercise 15.10 Suppose that $\mathcal{H} = \mathbb{R}$. Set $f = \iota_{\mathbb{R}_-}$ and

$$g\colon x \mapsto \begin{cases} x\ln(x) - x, & \text{if } x > 0;\\ 0, & \text{if } x = 0;\\ +\infty, & \text{if } x < 0. \end{cases} \tag{15.51}$$

Show that the primal problem (15.22) has a minimizer, that $\Delta(f,g) = 0$, and that $0 \notin \operatorname{sri}(\operatorname{dom} f - \operatorname{dom} g)$.

Exercise 15.11 Suppose that $\mathcal{H} = \mathbb{R}$. Set $f = \iota_{[0,1]}$ and $g = \iota_{[-1,0]}$. Show that the primal problem (15.22) and the dual problem (15.23) both possess minimizers, that $\Delta(f,g) = 0$, and that $0 \notin \operatorname{sri}(\operatorname{dom} f - \operatorname{dom} g)$.

Exercise 15.12 Find two functions f and g in $\Gamma_0(\mathbb{R})$ such that the following hold: the primal problem (15.22) has a minimizer, the primal optimal value is 0, $\Delta(f,g) = 0$, the dual problem (15.23) does not have a minimizer, and $\operatorname{dom} f^* = \operatorname{dom} g^* = \mathbb{R}$.

Exercise 15.13 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ and $g\colon \mathcal{K} \to \left]-\infty,+\infty\right]$ be proper, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$. Show that $\Delta(f, g\circ L) \leqslant \Delta(f,g,L)$.

Exercise 15.14 Suppose that $\mathcal{H} = \mathcal{K} = \mathbb{R}$ and set $f\colon \mathbb{R} \to \left]-\infty,+\infty\right] : x \mapsto x + \iota_{\mathbb{R}_+}(x)$, $g = \iota_{\{1\}}$, and $L = \mathrm{Id}$. Compute μ and μ^* as given in Proposition 15.21.

Exercise 15.15 Suppose that $\mathcal{H} = \mathcal{K} = \mathbb{R}$ and set $f\colon \mathbb{R} \to \left]-\infty,+\infty\right] : x \mapsto -x + \iota_{\mathbb{R}_+}(x)$, $g = \iota_{\{-1\}}$, and $L\colon \mathbb{R} \to \mathbb{R} \colon x \mapsto 0$. Compute μ and μ^* as given in Proposition 15.21.

Exercise 15.16 Suppose that $\mathcal{H} = \mathcal{K} = \mathbb{R}$ and set $f\colon \mathbb{R} \to \left]-\infty, +\infty\right] : x \mapsto -x + \iota_{\mathbb{R}_+}(x)$, $g = \iota_{\{0\}}$, and $L\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto 0$. Compute μ and μ^* as given in Proposition 15.21.

Exercise 15.17 Suppose that $\mathcal{H} = \mathcal{K} = \mathbb{R}$ and set $f\colon \mathbb{R} \to \left]-\infty, +\infty\right] : x \mapsto -x + \iota_{\mathbb{R}_+}(x)$, $g = \iota_{\left]-\infty,-1\right]}$, and $L\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto x$. Compute μ and μ^* as given in Proposition 15.21.

Exercise 15.18 Suppose that $\mathcal{H} = \ell^2(\mathbb{N})$, let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0, 1\right]$ such that $\sum_{n\in\mathbb{N}} \alpha_n^2 < +\infty$, and set $L\colon \mathcal{H} \to \mathcal{H}\colon (\xi_n)_{n\in\mathbb{N}} \mapsto (\alpha_n \xi_n)_{n\in\mathbb{N}}$. Show that $L^* = L$, that $\overline{\operatorname{ran}}\, L = \mathcal{H}$, and that $\operatorname{ran} L \neq \mathcal{H}$. Conclude that the closure operation in Corollary 15.31(i) is essential.

Exercise 15.19 Consider Corollary 15.31(ii). Find an example in which $D - L(C) = \mathcal{K}$ and $C^{\ominus} + L^*(D^{\ominus})$ is not a linear subspace.

Chapter 16
Subdifferentiability of Convex Functions

The subdifferential is a fundamental tool in the analysis of nondifferentiable convex functions. In this chapter we discuss the properties of subdifferentials and the interplay between the subdifferential and the Legendre transform. Moreover, we establish the Brøndsted–Rockafellar theorem, which asserts that the graph of the subdifferential operator is dense in the domain of the separable sum of the function and its conjugate.

16.1 Basic Properties

Definition 16.1 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper. The *subdifferential* of f is the set-valued operator

$$\partial f\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x \mapsto \big\{u \in \mathcal{H} \;\big|\; (\forall y \in \mathcal{H})\;\; \langle y - x \mid u\rangle + f(x) \leqslant f(y)\big\}. \quad (16.1)$$

Let $x \in \mathcal{H}$. Then f is *subdifferentiable* at x if $\partial f(x) \neq \varnothing$; the elements of $\partial f(x)$ are the *subgradients* of f at x.

Graphically (see Figure 16.1), a vector $u \in \mathcal{H}$ is a subgradient of a proper function $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ at $x \in \operatorname{dom} f$ if the continuous affine functional

$$f_{x,u}\colon y \mapsto \langle y - x \mid u\rangle + f(x), \quad (16.2)$$

which coincides with f at x, minorizes f; in other words, u is the "slope" of a continuous affine minorant of f that coincides with f at x.

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-48311-5_31

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_16

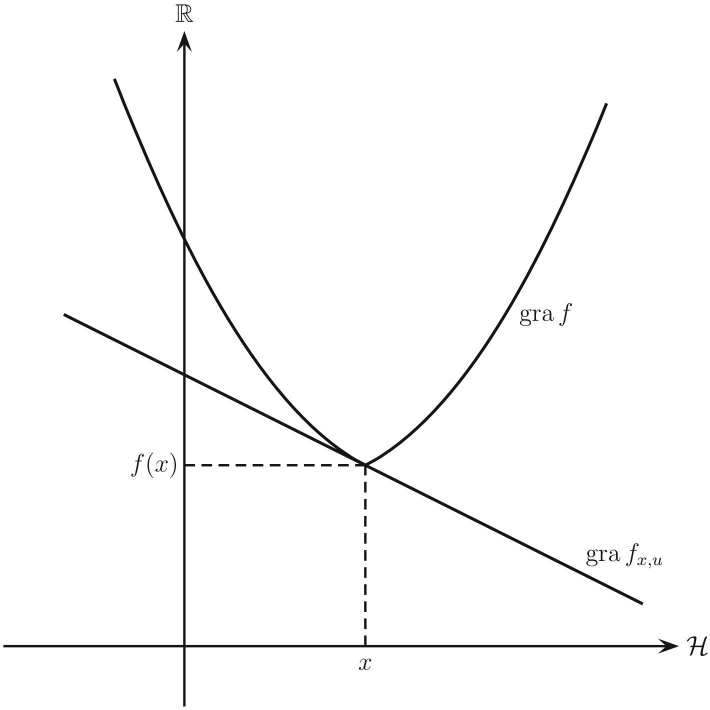

Fig. 16.1 A vector $u \in \mathcal{H}$ is a subgradient of f at x if it is the "slope" of a continuous affine minorant $f_{x,u}$ of f that coincides with f at x.

Remark 16.2 Let $f \in \Gamma_0(\mathcal{H})$. Then it follows from Theorem 9.20 that $\operatorname{dom} \partial f \neq \varnothing$. This fact will be considerably strengthened in Corollary 16.39.

Global minimizers of proper functions can be characterized by a simple but powerful principle which has its origin in the seventeenth century work of Pierre Fermat.

Theorem 16.3 (Fermat's rule) *Let* $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ *be proper. Then*

$$\operatorname{Argmin} f = \operatorname{zer} \partial f = \big\{ x \in \mathcal{H} \;\big|\; 0 \in \partial f(x) \big\}. \tag{16.3}$$

Proof. Let $x \in \mathcal{H}$. Then $x \in \operatorname{Argmin} f \Leftrightarrow (\forall y \in \mathcal{H})\ \langle y - x \mid 0 \rangle + f(x) \leqslant f(y)$ $\Leftrightarrow 0 \in \partial f(x)$. ☐

Proposition 16.4 *Let* $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ *be proper and let* $x \in \operatorname{dom} f$. *Then the following hold:*

(i) $\operatorname{dom} \partial f \subset \operatorname{dom} f$.
(ii) $\partial f(x) = \bigcap_{y \in \operatorname{dom} f} \big\{ u \in \mathcal{H} \;\big|\; \langle y - x \mid u \rangle \leqslant f(y) - f(x) \big\}$.
(iii) $\partial f(x)$ *is closed and convex.*

(iv) *Suppose that* $x \in \operatorname{dom} \partial f$. *Then* f *is lower semicontinuous and weakly lower semicontinuous at* x.

Proof. (i): Since f is proper, $f(x) = +\infty \Rightarrow \partial f(x) = \varnothing$.

(ii): Clear from (16.1).

(iii): Clear from (ii).

(iv): Take $u \in \partial f(x)$ and let $(x_a)_{a \in A}$ be a net in $\mathcal{H}$ such that $x_a \rightharpoonup x$. Then $(\forall a \in A)$ $\langle x_a - x \mid u \rangle + f(x) \leqslant f(x_a)$. Hence $f(x) \leqslant \underline{\lim} f(x_a)$. ☐

Proposition 16.5 *Let* $f \colon \mathcal{H} \to \left]-\infty, +\infty\right]$ *be proper, and let* $x \in \mathcal{H}$. *Suppose that* $x \in \operatorname{dom} \partial f$. *Then* $f^{**}(x) = f(x)$ *and* $\partial f^{**}(x) = \partial f(x)$.

Proof. Take $u \in \partial f(x)$. Then $(\forall y \in \mathcal{H})$ $\langle y - x \mid u \rangle + f(x) \leqslant f(y)$. Hence $f^{**} = \breve{f}$ by Proposition 13.45. In turn, in view of Corollary 9.10, Lemma 1.32(iv), and Proposition 16.4(iv), $f^{**}(x) = \breve{f}(x) = \bar{f}(x) = \underline{\lim}_{y \to x} f(y) = f(x)$. Thus $\langle y - x \mid u \rangle + f^{**}(x) \leqslant f^{**}(y)$, which shows that $u \in \partial f^{**}(x)$. Conversely, since $f^{**}(x) = f(x)$ and since $f^{**} \leqslant f$ by Proposition 13.16(i), we obtain $\partial f^{**}(x) \subset \partial f(x)$. ☐

Proposition 16.6 *Let* $\mathcal{K}$ *be a real Hilbert space, let* $f \colon \mathcal{H} \to \left]-\infty, +\infty\right]$ *and* $g \colon \mathcal{K} \to \left]-\infty, +\infty\right]$ *be proper, let* $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, *and let* $\lambda \in \mathbb{R}_{++}$. *Then the following hold:*

(i) $\partial(\lambda f) = \lambda \partial f$.

(ii) *Suppose that* $\operatorname{dom} g \cap L(\operatorname{dom} f) \neq \varnothing$. *Then* $\partial f + L^* \circ (\partial g) \circ L \subset \partial(f + g \circ L)$.

Proof. (i): Clear.

(ii): Take $x \in \mathcal{H}$, $u \in \partial f(x)$, and $v \in \partial g(Lx)$. Then $u + L^* v$ is a general point in $\partial f(x) + (L^* \circ (\partial g) \circ L)(x)$. It must be shown that $u + L^* v \in \partial(f + g \circ L)(x)$. It follows from (16.1) that, for every $y \in \mathcal{H}$, we have $\langle y - x \mid u \rangle + f(x) \leqslant f(y)$ and $\langle Ly - Lx \mid v \rangle + g(Lx) \leqslant g(Ly)$, hence $\langle y - x \mid L^* v \rangle + g(Lx) \leqslant g(Ly)$. Adding the first and third inequalities yields

$$(\forall y \in \mathcal{H}) \quad \langle y - x \mid u + L^* v \rangle + \big(f + g \circ L\big)(x) \leqslant \big(f + g \circ L\big)(y) \tag{16.4}$$

and, in turn, $u + L^* v \in \partial(f + g \circ L)(x)$. ☐

Proposition 16.7 *Let* $(\mathcal{H}_i)_{i \in I}$ *be a finite family of real Hilbert spaces, set* $\boldsymbol{\mathcal{H}} = \bigoplus_{i \in I} \mathcal{H}_i$, *let* $\boldsymbol{f} \colon \boldsymbol{\mathcal{H}} \to \left]-\infty, +\infty\right]$ *be proper, and let* $(x_i)_{i \in I} \in \operatorname{dom} \boldsymbol{f}$. *For every* $i \in I$, *define* $R_i \colon \mathcal{H}_i \to \boldsymbol{\mathcal{H}}$ *as follows: for every* $y \in \mathcal{H}_i$ *and every* $j \in I$, *the* j*th component of* $R_i y$ *is* y *if* $j = i$, *and* x_j *otherwise. Then*

$$\partial \boldsymbol{f}\big((x_i)_{i \in I}\big) \subset \bigtimes_{i \in I} \partial(\boldsymbol{f} \circ R_i)(x_i). \tag{16.5}$$

Proof. Set $\boldsymbol{x} = (x_i)_{i \in I}$, and take $\boldsymbol{y} = (y_i)_{i \in I} \in \boldsymbol{\mathcal{H}}$ and $\boldsymbol{u} = (u_i)_{i \in I} \in \boldsymbol{\mathcal{H}}$. Then

$$\begin{aligned}
\boldsymbol{u} \in \partial \boldsymbol{f}(\boldsymbol{x}) &\Leftrightarrow (\forall \boldsymbol{y} \in \boldsymbol{\mathcal{H}}) \quad \langle \boldsymbol{y} - \boldsymbol{x} \mid \boldsymbol{u} \rangle + \boldsymbol{f}(\boldsymbol{x}) \leqslant \boldsymbol{f}(\boldsymbol{y}) \\
&\Rightarrow (\forall i \in I)(\forall y_i \in \mathcal{H}_i) \quad \langle y_i - x_i \mid u_i \rangle + (\boldsymbol{f} \circ R_i)(x_i) \leqslant (\boldsymbol{f} \circ R_i)(y_i) \\
&\Leftrightarrow (\forall i \in I) \quad u_i \in \partial(\boldsymbol{f} \circ R_i)(x_i) \\
&\Leftrightarrow \boldsymbol{u} \in \bigtimes_{i \in I} \partial(\boldsymbol{f} \circ R_i)(x_i), \qquad (16.6)
\end{aligned}$$

which completes the proof. □

Remark 16.8 The inclusion (16.5) can be strict. Indeed, adopt the notation of Proposition 16.7 and suppose that $\mathcal{H} = \mathbb{R} \times \mathbb{R}$. Set $\boldsymbol{f} = \iota_{B(0;1)}$ and $\boldsymbol{x} = (1, 0)$. Then $\boldsymbol{f} \circ R_1 = \iota_{[-1,1]}$ and $\boldsymbol{f} \circ R_2 = \iota_{\{0\}}$. Therefore,

$$\partial \boldsymbol{f}(1, 0) = \mathbb{R}_+ \times \{0\} \neq \mathbb{R}_+ \times \mathbb{R} = \big(\partial(\boldsymbol{f} \circ R_1)(1)\big) \times \big(\partial(\boldsymbol{f} \circ R_2)(0)\big). \qquad (16.7)$$

Proposition 16.9 *Let $(\mathcal{H}_i)_{i \in I}$ be a finite family of real Hilbert spaces and, for every $i \in I$, let $f_i \colon \mathcal{H}_i \to \left]-\infty, +\infty\right]$ be proper. Then*

$$\partial \bigoplus_{i \in I} f_i = \bigtimes_{i \in I} \partial f_i. \qquad (16.8)$$

Proof. We denote by $\boldsymbol{x} = (x_i)_{i \in I}$ a generic element in $\boldsymbol{\mathcal{H}} = \bigoplus_{i \in I} \mathcal{H}_i$. Set $\boldsymbol{f} = \bigoplus_{i \in I} f_i$, and take $\boldsymbol{x}$ and $\boldsymbol{u}$ in $\boldsymbol{\mathcal{H}}$. Then Proposition 16.4(i) yields

$$\begin{aligned}
&\boldsymbol{u} \in \bigtimes_{i \in I} \partial f_i(x_i) \\
&\quad \Leftrightarrow \quad (\forall i \in I)(\forall y_i \in \mathcal{H}_i) \quad \langle y_i - x_i \mid u_i \rangle + f_i(x_i) \leqslant f_i(y_i) \qquad (16.9) \\
&\quad \Rightarrow \quad (\forall \boldsymbol{y} \in \boldsymbol{\mathcal{H}}) \quad \sum_{i \in I} \langle y_i - x_i \mid u_i \rangle + \sum_{i \in I} f_i(x_i) \leqslant \sum_{i \in I} f_i(y_i) \qquad (16.10) \\
&\quad \Leftrightarrow \quad (\forall \boldsymbol{y} \in \boldsymbol{\mathcal{H}}) \quad \langle \boldsymbol{y} - \boldsymbol{x} \mid \boldsymbol{u} \rangle + \boldsymbol{f}(\boldsymbol{x}) \leqslant \boldsymbol{f}(\boldsymbol{y}) \\
&\quad \Leftrightarrow \quad \boldsymbol{u} \in \partial \boldsymbol{f}(\boldsymbol{x}). \qquad (16.11)
\end{aligned}$$

Finally, to obtain (16.10) $\Rightarrow$ (16.9), fix $i \in I$. By forcing the coordinates of $\boldsymbol{y}$ in (16.10) to coincide with those of $\boldsymbol{x}$, except for the ith, we obtain $(\forall y_i \in \mathcal{H}_i)$ $\langle y_i - x_i \mid u_i \rangle + f_i(x_i) \leqslant f_i(y_i)$. □

The next result states that the graph of the subdifferential contains precisely those points for which the Fenchel–Young inequality (Proposition 13.15) becomes an equality.

Proposition 16.10 *Let $f \colon \mathcal{H} \to \left]-\infty, +\infty\right]$ be proper, let $x \in \mathcal{H}$, and let $u \in \mathcal{H}$. Then $u \in \partial f(x) \Leftrightarrow f(x) + f^*(u) = \langle x \mid u \rangle \Rightarrow x \in \partial f^*(u)$.*

Proof. Using (16.1) and Proposition 13.15, we get

$$\begin{aligned} u \in \partial f(x) &\Leftrightarrow (\forall y \in \operatorname{dom} f) \quad \langle y \mid u\rangle - f(y) \leqslant \langle x \mid u\rangle - f(x) \leqslant f^*(u) \\ &\Leftrightarrow f^*(u) = \sup_{y \in \operatorname{dom} f} \big(\langle y \mid u\rangle - f(y)\big) \leqslant \langle x \mid u\rangle - f(x) \leqslant f^*(u) \\ &\Leftrightarrow f(x) + f^*(u) = \langle x \mid u\rangle . \end{aligned} \tag{16.12}$$

Accordingly, using Proposition 13.16(i), we obtain $u \in \partial f(x) \Rightarrow \langle u \mid x\rangle \leqslant f^*(u) + f^{**}(x) \leqslant f^*(u) + f(x) = \langle u \mid x\rangle \Rightarrow f^*(u) + f^{**}(x) = \langle u \mid x\rangle \Rightarrow x \in \partial f^*(u)$. ☐

Corollary 16.11 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper. Then* $\operatorname{gra}\partial f \subset \operatorname{dom} f \times \operatorname{dom} f^*$.

Example 16.12 Set $f = (1/2)\|\cdot\|^2$. Then $\partial f = \mathrm{Id}$.

Proof. This follows from Proposition 16.10 and Proposition 13.19. ☐

16.2 Convex Functions

We start with a few simple examples.

Example 16.13 Let C be a nonempty convex subset of $\mathcal{H}$. Then $\partial\iota_C = N_C$.

Proof. Combine (16.1) and (6.35). ☐

Example 16.14 Let Ω be a nonempty subset of $\mathbb{R}$, set $\underline{\omega} = \inf \Omega$, set $\overline{\omega} = \sup \Omega$, and let $\xi \in \mathbb{R}$. Then

$$\partial\sigma_\Omega(\xi) = \begin{cases} \{\underline{\omega}\} \cap \mathbb{R}, & \text{if } \xi < 0; \\ \overline{\operatorname{conv}}\, \Omega, & \text{if } \xi = 0; \\ \{\overline{\omega}\} \cap \mathbb{R}, & \text{if } \xi > 0. \end{cases} \tag{16.13}$$

Example 16.15 Let $\xi \in \mathbb{R}$. Then

$$\partial|\cdot|(\xi) = \begin{cases} \{-1\}, & \text{if } \xi < 0; \\ [-1,1], & \text{if } \xi = 0; \\ \{1\}, & \text{if } \xi > 0. \end{cases} \tag{16.14}$$

Proof. Apply Example 16.14 with $\Omega = [-1,1]$ (see Example 7.9). ☐

Here is a refinement of Proposition 16.10.

Proposition 16.16 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex, let $x \in \mathcal{H}$, and let $u \in \mathcal{H}$. Then $u \in \partial f(x) \Leftrightarrow (u,-1) \in N_{\operatorname{epi} f}(x, f(x)) \Leftrightarrow f(x) + f^*(u) = \langle x \mid u\rangle \Rightarrow x \in \partial f^*(u)$.*

Proof. Note that $\operatorname{epi} f$ is nonempty and convex. Moreover, $(u,-1) \in N_{\operatorname{epi} f}(x, f(x)) \Leftrightarrow x \in \operatorname{dom} f$ and $(\forall (y,\eta) \in \operatorname{epi} f)\ \langle (y,\eta) - (x, f(x)) \mid (u,-1)\rangle \leqslant 0 \Leftrightarrow x \in \operatorname{dom} f$ and $(\forall (y,\eta) \in \operatorname{epi} f)\ \langle y - x \mid u\rangle + (\eta - f(x))(-1) \leqslant 0 \Leftrightarrow (\forall (y,\eta) \in \operatorname{epi} f)\ \langle y - x \mid u\rangle + f(x) \leqslant \eta \Leftrightarrow (\forall y \in \operatorname{dom} f)\ \langle y - x \mid u\rangle + f(x) \leqslant f(y) \Leftrightarrow u \in \partial f(x)$. In view of Proposition 16.10, the proof is complete. ☐

Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper. We have observed in Proposition 16.4(i)&(iii) that $\operatorname{dom} \partial f \subset \operatorname{dom} f$ and that, for every $x \in \operatorname{dom} f$, $\partial f(x)$ is closed and convex. Convexity of f supplies stronger statements.

Proposition 16.17 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex, and let $x \in \operatorname{dom} f$. Then the following hold:*

(i) *Suppose that $\operatorname{int}\operatorname{dom} f \neq \varnothing$ and that $x \in \operatorname{bdry}\operatorname{dom} f$. Then $\partial f(x)$ is either empty or unbounded.*
(ii) *Suppose that $x \in \operatorname{cont} f$. Then $\partial f(x)$ is nonempty and weakly compact.*
(iii) *Suppose that $x \in \operatorname{cont} f$. Then there exists $\rho \in \mathbb{R}_{++}$ such that $\partial f(B(x;\rho))$ is bounded.*
(iv) *Suppose that $\operatorname{cont} f \neq \varnothing$. Then $\operatorname{int}\operatorname{dom} f \subset \operatorname{dom} \partial f$.*

Proof. (i): By Proposition 7.5, x is a support point of $\operatorname{dom} f$. Consequently, there exists $u \in \mathcal{H} \smallsetminus \{0\}$ such that $(\forall y \in \operatorname{dom} f)\ \langle y - x \mid u\rangle \leqslant 0$. Hence $(\forall v \in \partial f(x))(\forall \lambda \in \mathbb{R}_+)\ v + \lambda u \in \partial f(x)$.

(ii)&(iii): Observe that $\operatorname{epi} f$ is nonempty and convex. Also, by Proposition 8.45, $\operatorname{int}\operatorname{epi} f \neq \varnothing$. For every $\varepsilon \in \mathbb{R}_{++}$, $(x, f(x) - \varepsilon) \notin \operatorname{epi} f$ and therefore $(x, f(x)) \in \operatorname{bdry}\operatorname{epi} f$. Using Proposition 7.5, we get $(u,\nu) \in N_{\operatorname{epi} f}(x, f(x)) \smallsetminus \{(0,0)\}$. For every $y \in \operatorname{dom} f$ and every $\eta \in \mathbb{R}_+$, we have $\langle (y, f(y) + \eta) - (x, f(x)) \mid (u,\nu)\rangle \leqslant 0$ and therefore

$$\langle y - x \mid u\rangle + (f(y) - f(x))\nu + \eta\nu \leqslant 0. \tag{16.15}$$

We first note that $\nu \leqslant 0$ since, otherwise, we get a contradiction in (16.15) by letting $\eta \to +\infty$. To show that $\nu < 0$, let us argue by contradiction. If $\nu = 0$, then (16.15) yields $\sup \langle \operatorname{dom} f - x \mid u\rangle \leqslant 0$ and, since $B(x;\varepsilon) \subset \operatorname{dom} f$ for $\varepsilon \in \mathbb{R}_{++}$ small enough, we would further deduce that $\sup \langle B(0;\varepsilon) \mid u\rangle \leqslant 0$. This would imply that $u = 0$ and, in turn, that $(u,\nu) = (0,0)$, which is impossible. Hence $\nu < 0$. Since $N_{\operatorname{epi} f}(x, f(x))$ is a cone, we also have

$$\big(u/|\nu|, -1\big) = (1/|\nu|)(u,\nu) \in N_{\operatorname{epi} f}\big(x, f(x)\big). \tag{16.16}$$

Using Proposition 16.16, we obtain $u/|\nu| \in \partial f(x)$. Hence $\partial f(x) \neq \varnothing$. By Theorem 8.38, there exist $\beta \in \mathbb{R}_{++}$ and $\rho \in \mathbb{R}_{++}$ such that f is Lipschitz continuous with constant β relative to $B(x;2\rho)$. Now take $y \in B(x;\rho)$ and $v \in \partial f(y)$. Then $(\forall z \in B(0;\rho))\ \langle z \mid v\rangle \leqslant f(y+z) - f(y) \leqslant \beta\|z\|$ and hence $\|v\| \leqslant \beta$. It follows that $\partial f(x) \subset \partial f(B(x;\rho)) \subset B(0;\beta)$. Thus, $\partial f(x)$ is bounded and, by Proposition 16.4(iii), closed and convex. It is therefore weakly compact by Theorem 3.37.

(iv): A consequence of (ii) and Corollary 8.39(i). ☐

Corollary 16.18 *Suppose that $\mathcal{H}$ is finite-dimensional, and let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex. Then the following hold:*

(i) $\varnothing \neq \operatorname{ri}\operatorname{dom} f \subset \operatorname{dom}\partial f$.
(ii) *f possesses a continuous affine minorant.*

Proof. (i): Since $\operatorname{dom} f$ is a nonempty convex set, it follows from Fact 6.14(i) that $\operatorname{ri}\operatorname{dom} f \neq \varnothing$. The inclusion follows from Corollary 8.41 and Proposition 16.17(ii), applied in the Euclidean space parallel to $\operatorname{aff}\operatorname{dom} f$.

(ii): Let $x \in \operatorname{dom}\partial f$, which is nonempty by (i), and let $u \in \partial f(x)$. Then (16.2) yields a continuous affine minorant of f. □

Corollary 16.19 *Let $h\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be convex and suppose that D is a nonempty open convex subset of $\operatorname{cont} h$. Then $(h+\iota_D)^{**}$ is the unique function $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ that satisfies*

$$f \in \Gamma_0(\mathcal{H}), \quad \operatorname{dom} f \subset \overline{D}, \quad \text{and} \quad f|_D = h|_D. \tag{16.17}$$

Moreover,

$$(\forall x \in \mathcal{H})(\forall y \in D) \quad f(x) = \begin{cases} \lim_{\alpha\downarrow 0} h\big((1-\alpha)x + \alpha y\big), & \text{if } x \in \overline{D};\\ +\infty, & \text{if } x \notin \overline{D}. \end{cases} \tag{16.18}$$

Proof. Set $g = h + \iota_D$. Then g is proper and convex, and $D \subset \operatorname{cont} g$. Using Proposition 16.17(ii) and Proposition 16.5, we see that $D \subset \operatorname{dom}\partial g$ and that

$$h|_D = g|_D = g^{**}|_D. \tag{16.19}$$

Thus $g^{**} \in \Gamma_0(\mathcal{H})$. Since $\operatorname{dom}\partial g \neq \varnothing$, Proposition 13.46(i) implies that $\operatorname{dom} g^{**} \subset \overline{\operatorname{dom} g} \subset \overline{D}$. Hence g^{**} satisfies (16.17). Now assume that $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ satisfies (16.17), fix $y \in D = \operatorname{int}\operatorname{dom} f$, and take $x \in \mathcal{H}$. If $x \notin \overline{D}$, then $f(x) = +\infty$. So assume that $x \in \overline{D}$. Proposition 3.44 implies that $\left]x,y\right] \subset D$. In view of Proposition 9.14 and (16.17), we deduce that

$$f(x) = \lim_{\alpha\downarrow 0} f\big((1-\alpha)x + \alpha y\big) = \lim_{\alpha\downarrow 0} h\big((1-\alpha)x + \alpha y\big). \tag{16.20}$$

This verifies the uniqueness of f as well as (16.18). □

Proposition 16.20 *Let $f\colon \mathcal{H} \to \mathbb{R}$ be continuous and convex. Then the following are equivalent:*

(i) *f is bounded on every bounded subset of $\mathcal{H}$.*
(ii) *f is Lipschitz continuous relative to every bounded subset of $\mathcal{H}$.*
(iii) *$\operatorname{dom}\partial f = \mathcal{H}$ and ∂f maps every bounded subset of $\mathcal{H}$ to a bounded set.*
(iv) *f^* is supercoercive.*

If $\mathcal{H}$ is finite-dimensional, then f satisfies these properties.

Proof. (i)⇒(ii): Take a bounded subset C of $\mathcal{H}$. Let $x_0 \in \mathcal{H}$ and $\rho \in \mathbb{R}_{++}$ be such that $C \subset B(x_0;\rho)$. By assumption, f is bounded on $B(x_0;2\rho)$. Proposition 8.37(ii) implies that f is Lipschitz continuous relative to $B(x_0;\rho)$, and hence relative to C.

(ii)⇒(iii): Proposition 16.17(ii) asserts that $\operatorname{dom}\partial f = \mathcal{H}$. It suffices to show that the subgradients of f are uniformly bounded on every open ball centered at 0. To this end, fix $\rho \in \mathbb{R}_{++}$, let $\lambda \in \mathbb{R}_+$ be a Lipschitz constant of f relative to $\operatorname{int} B(0;\rho)$, take $x \in \operatorname{int} B(0;\rho)$, and let $\alpha \in \mathbb{R}_{++}$ be such that $B(x;\alpha) \subset \operatorname{int} B(0;\rho)$. Now suppose that $u \in \partial f(x)$. Then

$$\big(\forall y \in B(0;1)\big) \quad \langle \alpha y \mid u\rangle \leqslant f(x+\alpha y) - f(x) \leqslant \lambda\|\alpha y\| \leqslant \lambda\alpha. \tag{16.21}$$

Taking the supremum over $y \in B(0;1)$ in (16.21), we obtain $\alpha\|u\| \leqslant \lambda\alpha$, i.e., $\|u\| \leqslant \lambda$. Thus, $\sup\|\partial f(\operatorname{int} B(0;\rho))\| \leqslant \lambda$.

(iii)⇒(i): It suffices to show that f is bounded on every closed ball centered at 0. Fix $\rho \in \mathbb{R}_{++}$. Then $\beta = \sup\|\partial f(B(0;\rho))\| < +\infty$. Now take $x \in B(0;\rho)$ and $u \in \partial f(x)$. Then $\langle 0 - x \mid u\rangle + f(x) \leqslant f(0)$ and hence $f(x) \leqslant f(0) + \langle x \mid u\rangle \leqslant f(0) + \rho\beta$. It follows that

$$\sup f\big(B(0;\rho)\big) \leqslant f(0) + \rho\beta. \tag{16.22}$$

Now take $v \in \partial f(0)$. Then $\langle x - 0 \mid v\rangle + f(0) \leqslant f(x)$ and thus $f(x) \geqslant f(0) - \|x\|\,\|v\| \geqslant f(0) - \rho\beta$. We deduce that

$$\inf f\big(B(0;\rho)\big) \geqslant f(0) - \rho\beta. \tag{16.23}$$

Altogether, (16.22) and (16.23) imply that f is bounded on $B(0;\rho)$.

(i)⇔(iv): Combine Proposition 14.15 and Corollary 13.38.

Finally, suppose that $\mathcal{H}$ is finite-dimensional. By Corollary 8.41, f satisfies (ii) and, in turn, the equivalent properties (i), (iii), and (iv). ☐

Corollary 16.21 *Suppose that $\mathcal{H}$ is finite-dimensional, and let $f \in \Gamma_0(\mathcal{H})$. Then the following are equivalent:*

(i) *f is supercoercive.*
(ii) $(\forall u \in \mathcal{H})$ *$f - \langle \cdot \mid u\rangle$ is coercive.*
(iii) $\operatorname{dom} f^* = \mathcal{H}$.

Proof. (i)⇒(iii): Proposition 14.15.

(ii)⇔(iii): Theorem 14.17.

(iii)⇒(i): By Corollary 8.40, $f^*\colon \mathcal{H} \to \mathbb{R}$ is continuous. It follows from Corollary 13.38 and Proposition 16.20 that f is supercoercive. ☐

Let us illustrate the fact that the equivalent conditions of Proposition 16.20 may fail to hold when $\mathcal{H}$ is infinite-dimensional.

Example 16.22 Suppose that $\mathcal{H} = \ell^2(I)$, where $I = \mathbb{N} \smallsetminus \{0\}$. For every $n \in I$, denote the nth standard unit vector in $\mathcal{H}$ by e_n, and set

$$f\colon \mathcal{H} \to \left]-\infty,+\infty\right] : (\xi_n)_{n\in I} \mapsto \sum_{n\in I} n\xi_n^{2n}. \tag{16.24}$$

Let $x = (\xi_n)_{n\in I} \in \mathcal{H}$. Since $\xi_n \to 0$ and $\sqrt[n]{n} \to 1$, we deduce that eventually $\xi_n^2 \leqslant 1/(\sqrt[n]{n})^3$, which is equivalent to $n\xi_n^{2n} \leqslant 1/n^2$. It follows that $f(x) \in \mathbb{R}$. Therefore, $\operatorname{dom} f = \mathcal{H}$. By Example 9.13, f is lower semicontinuous and convex. In view of Corollary 8.39(ii) and Proposition 16.17(ii), f is continuous and subdifferentiable everywhere. Note that $\sup_{n\in I} f(e_n) = \sup_{n\in I} n = +\infty$. Thus, by Proposition 16.20, f^* is not supercoercive.

Proposition 16.23 *Let $\phi\colon \mathbb{R} \to \left]-\infty,+\infty\right]$ be an even convex function that vanishes only at 0, and set $f = \phi \circ \|\cdot\|$. Then $f\colon \mathcal{H} \to \mathbb{R}$ is continuous, strictly quasiconvex, and coercive.*

Proof. Since ϕ is continuous by Corollary 8.40, so is f. It follows from Proposition 11.7(iv) that ϕ is strictly increasing on $\mathbb{R}_+$. On the other hand, it follows from Corollary 2.16 or Example 10.31(i) that $\|\cdot\|$ is strictly quasiconvex. Hence, Example 10.30(i) implies that f is strictly quasiconvex. Finally, by Corollary 16.18(i), there exists $\nu \in \partial\phi(1)$. Hence, since it follows from Proposition 11.7(i) that $\operatorname{Argmin} \phi = \{0\}$, (16.1) yields $(0-1)\nu \leqslant \phi(0) - \phi(1) < 0$ and therefore $\nu > 0$. In turn, $\phi(\eta) \geqslant (\eta - 1)\nu + \phi(1) \to +\infty$ as $\eta \to +\infty$, which implies that f is coercive. ☐

16.3 Lower Semicontinuous Convex Functions

Let us first revisit the properties of positively homogeneous convex functions.

Proposition 16.24 *Let $f \in \Gamma_0(\mathcal{H})$ be positively homogeneous. Then $f = \sigma_C$, where $C = \partial f(0)$.*

Proof. Since $f(0) = 0$, the result follows from Proposition 14.11 and Definition 16.1. ☐

A consequence of Corollary 16.19 is the following result, which shows that certain functions in $\Gamma_0(\mathcal{H})$ are uniquely determined by their behavior on the interior of their domain.

Proposition 16.25 *Let $f \in \Gamma_0(\mathcal{H})$ be such that* $\operatorname{int}\operatorname{dom} f \neq \varnothing$. *Then*

$$f = \big(f + \iota_{\operatorname{int}\operatorname{dom} f}\big)^{**}. \tag{16.25}$$

The next example illustrates the fact that in Proposition 16.25 $\operatorname{int}\operatorname{dom} f$ cannot be replaced in general by a dense convex subset of $\operatorname{dom} f$. (However, as will be seen in Corollary 16.41, $\operatorname{dom} \partial f$ is large enough to reconstruct f.)

Example 16.26 Suppose that $\mathcal{H}$ is infinite-dimensional and separable. Let $(e_n)_{n\in\mathbb{N}}$ be an orthonormal basis of $\mathcal{H}$, and let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence in $]0,1[$ such that $(\alpha_n)_{n\in\mathbb{N}} \in \ell^2(\mathbb{N})$. Define

$$f\colon \mathcal{H} \to \left]-\infty,+\infty\right] : x \mapsto \sup_{n\in\mathbb{N}} \frac{\langle x \mid e_n\rangle}{\alpha_n} \tag{16.26}$$

and set $C = \operatorname{span}\{e_n\}_{n\in\mathbb{N}}$. Then $f \in \Gamma_0(\mathcal{H})$ and $C \subset \operatorname{dom} f$. In fact, C is a convex and dense subset of $\mathcal{H}$ and $0 \leqslant f + \iota_C$. By Proposition 13.16(ii), $0 = 0^{**} \leqslant (f+\iota_C)^{**}$. On the other hand, set $z = -\sum_{n\in\mathbb{N}} \alpha_n e_n$. Then $f(z) = -1$ and $z \in (\operatorname{dom} f) \smallsetminus C$. Altogether, $(f+\iota_C)^{**} \neq f$.

Proposition 16.27 *Let* $f \in \Gamma_0(\mathcal{H})$. *Then* $\operatorname{int}\operatorname{dom} f = \operatorname{cont} f \subset \operatorname{dom}\partial f \subset \operatorname{dom} f$.

Proof. The equality follows from Corollary 8.39(ii) and the first inclusion from Proposition 16.17(ii). The second inclusion was observed in Proposition 16.4(i). ☐

Remark 16.28 Let $f \in \Gamma_0(\mathcal{H})$. Then Proposition 16.27 asserts that $\operatorname{dom}\partial f$ is sandwiched between the convex sets $\operatorname{int}\operatorname{dom} f$ and $\operatorname{dom} f$. However it may fail to be convex. For instance, suppose that $\mathcal{H} = \mathbb{R}^2$ and set $f\colon (\xi_1,\xi_2) \mapsto \max\{g(\xi_1), |\xi_2|\}$, where $g(\xi_1) = 1 - \sqrt{\xi_1}$ if $\xi_1 \geqslant 0$; $g(\xi_1) = +\infty$ if $\xi_1 < 0$. Then $\operatorname{dom}\partial f = (\mathbb{R}_+ \times \mathbb{R}) \smallsetminus (\{0\} \times \left]-1,1\right[)$.

The following theorem provides characterizations of the subdifferential of a function in $\Gamma_0(\mathcal{H})$.

Theorem 16.29 *Let* $f \in \Gamma_0(\mathcal{H})$, *let* $x \in \mathcal{H}$, *and let* $u \in \mathcal{H}$. *Then the following are equivalent:*

(i) $(x,u) \in \operatorname{gra}\partial f$.
(ii) $(u,-1) \in N_{\operatorname{epi} f}(x, f(x))$.
(iii) $f(x) + f^*(u) = \langle x \mid u\rangle$.
(iv) $(u,x) \in \operatorname{gra}\partial f^*$.

Proof. (i)⇔(ii)⇔(iii)⇒(iv): Proposition 16.16.

(iv)⇒(iii): The implication (i)⇒(iii) and Corollary 13.38 yield $(u,x) \in \operatorname{gra}\partial f^* \Rightarrow \langle u \mid x\rangle = f^*(u) + f^{**}(x) = f^*(u) + f(x)$. ☐

Corollary 16.30 *Let* $f \in \Gamma_0(\mathcal{H})$. *Then* $(\partial f)^{-1} = \partial f^*$.

Example 16.31 Let $\phi \in \Gamma_0(\mathbb{R})$ be even, set $f = \phi \circ \|\cdot\|$, and let x and u be in $\mathcal{H}$. Then the following are equivalent:

(i) $u \in \partial f(x)$.
(ii) $\|u\| \in \partial\phi(\|x\|)$ and $(\exists\, \alpha \in \mathbb{R}_+)$ $x = \alpha u$ or $u = \alpha x$.

Proof. We first derive from Lemma 1.28, Proposition 11.7(ii), and Proposition 8.21 that $f \in \Gamma_0(\mathcal{H})$. Now let $u \in \mathcal{H}$. Then it follows from Theorem 16.29, Proposition 13.15, Example 13.8, and Fact 2.11 that

$$\begin{aligned}
\text{(i)} &\Leftrightarrow f(x) + f^*(u) = \langle x \mid u \rangle \\
&\Leftrightarrow \|x\|\,\|u\| \leqslant \phi\big(\|x\|\big) + \phi^*\big(\|u\|\big) = \langle x \mid u \rangle \leqslant \|x\|\,\|u\| \\
&\Leftrightarrow \phi\big(\|x\|\big) + \phi^*\big(\|u\|\big) = \|x\|\,\|u\| = \langle x \mid u \rangle \\
&\Leftrightarrow \text{(ii)},
\end{aligned} \tag{16.27}$$

as announced. ☐

Example 16.32 Let $x \in \mathcal{H}$. Then

$$\partial\|\cdot\|(x) = \begin{cases} \big\{x/\|x\|\big\}, & \text{if } x \neq 0; \\ B(0;1), & \text{if } x = 0. \end{cases} \tag{16.28}$$

Proof. Set $\phi = |\cdot|$ in Example 16.31 and use Example 16.15. ☐

Proposition 16.33 *Let $f \in \Gamma_0(\mathcal{H})$. Then* $\operatorname{Argmin} f = \partial f^*(0)$.

Proof. Theorem 16.3 and Corollary 16.30 yield $x \in \operatorname{Argmin} f \Leftrightarrow 0 \in \partial f(x)$ $\Leftrightarrow x \in \partial f^*(0)$. ☐

Example 16.34 Let C be a nonempty closed convex subset of $\mathcal{H}$. Then $\partial\sigma_C(0) = C$.

Proof. Apply Proposition 16.33 to ι_C. ☐

Proposition 16.35 *Let $f \in \Gamma_0(\mathcal{H})$ and let $x \in \operatorname{dom} f$. Then*

$$N_{\operatorname{epi} f}\big(x, f(x)\big) = \big(N_{\operatorname{dom} f}\, x \times \{0\}\big) \cup \mathbb{R}_{++}\big(\partial f(x) \times \{-1\}\big). \tag{16.29}$$

Proof. Set $C = \operatorname{epi} f$ and let $(u,\eta) \in \mathcal{H} \times \mathbb{R}$. Using Proposition 6.47 and Proposition 9.18, we obtain the equivalences $(u,\eta) \in N_C(x, f(x)) \Leftrightarrow (x, f(x)) = P_C(x+u, f(x)+\eta) \Leftrightarrow$

$$\eta \leqslant 0 \quad \text{and} \quad (\forall y \in \operatorname{dom} f) \quad \langle y - x \mid u \rangle + \eta\big(f(y) - f(x)\big) \leqslant 0. \tag{16.30}$$

If $\eta = 0$, then (16.30) $\Leftrightarrow u \in N_{\operatorname{dom} f} x \Leftrightarrow (u,\eta) \in N_{\operatorname{dom} f} x \times \{0\}$. If $\eta < 0$, then (16.30) $\Leftrightarrow (\forall y \in \operatorname{dom} f)$ $\langle y - x \mid u/|\eta| \rangle \leqslant f(y) - f(x) \Leftrightarrow u/|\eta| \in \partial f(x)$ $\Leftrightarrow (u,\eta) = |\eta|(u/|\eta|, -1) \in \mathbb{R}_{++}(\partial f(x) \times \{-1\})$. ☐

Proposition 16.36 *Let $f \in \Gamma_0(\mathcal{H})$. Then* $\operatorname{gra} \partial f$ *is sequentially closed in* $\mathcal{H}^{\text{weak}} \times \mathcal{H}^{\text{strong}}$ *and in* $\mathcal{H}^{\text{strong}} \times \mathcal{H}^{\text{weak}}$.

Proof. Let $(x_n, u_n)_{n\in\mathbb{N}}$ be a sequence in $\operatorname{gra} \partial f$ such that $x_n \rightharpoonup x$ and $u_n \to u$. Then it follows from Theorem 16.29 that $(\forall n \in \mathbb{N})$ $f(x_n) + f^*(u_n) = \langle x_n \mid u_n \rangle$.

Hence, we derive from Proposition 13.15, Theorem 9.1, and Lemma 2.51(iii) that

$$\begin{aligned}\langle x \mid u\rangle &\leqslant f(x) + f^*(u)\\ &\leqslant \underline{\lim}\, f(x_n) + \underline{\lim}\, f^*(u_n)\\ &\leqslant \underline{\lim}\,\big(f(x_n) + f^*(u_n)\big)\\ &= \lim \langle x_n \mid u_n\rangle\\ &= \langle x \mid u\rangle\,. \end{aligned} \tag{16.31}$$

Invoking Theorem 16.29 once more, we deduce that $(x, u) \in \operatorname{gra} \partial f$, which proves the first assertion. Applying this result to f^* and then appealing to Corollary 16.30 yields the second assertion. ☐

Proposition 16.37 *Let $f \in \Gamma_0(\mathcal{H})$, and let x, u_0, and u_1 be in $\mathcal{H}$. Then the following hold:*

(i) *Suppose that $[u_0, u_1] \subset \partial f(x)$. Then f^* is affine on $[u_0, u_1]$.*
(ii) *Suppose that f^* is affine on $[u_0, u_1]$ and that $x \in \partial f^*(]u_0, u_1[)$. Then $[u_0, u_1] \subset \partial f(x)$.*

Proof. Set $(\forall \alpha \in]0, 1[)\ u_\alpha = (1 - \alpha)u_0 + \alpha u_1$.

(i): Theorem 16.29 yields $f(x) + f^*(u_0) = \langle x \mid u_0\rangle$ and $f(x) + f^*(u_1) = \langle x \mid u_1\rangle$. Now take $\alpha \in]0, 1[$. Then

$$\begin{aligned}f(x) + (1-\alpha)f^*(u_0) + \alpha f^*(u_1) &= (1-\alpha)\langle x \mid u_0\rangle + \alpha\langle x \mid u_1\rangle\\ &= \langle x \mid u_\alpha\rangle\\ &= f(x) + f^*(u_\alpha)\\ &= f(x) + f^*((1-\alpha)u_0 + \alpha u_1)\\ &\leqslant f(x) + (1-\alpha)f^*(u_0) + \alpha f^*(u_1),\end{aligned} \tag{16.32}$$

and we deduce that f^* is affine on $[u_0, u_1]$.

(ii): We have $x \in \partial f^*(u_\alpha)$ for some $\alpha \in]0, 1[$. Hence, Theorem 16.29 implies that

$$\begin{aligned}0 &= f(x) + f^*(u_\alpha) - \langle x \mid u_\alpha\rangle\\ &= (1-\alpha)\big(f(x) + f^*(u_0) - \langle x \mid u_0\rangle\big)\\ &\quad + \alpha\big(f(x) + f^*(u_1) - \langle x \mid u_1\rangle\big).\end{aligned} \tag{16.33}$$

Since the terms $f(x)+f^*(u_0)-\langle x \mid u_0\rangle$ and $f(x)+f^*(u_1)-\langle x \mid u_1\rangle$ are positive by the Fenchel–Young inequality (Proposition 13.15), they must therefore be zero. Hence, $\{u_0, u_1\} \subset \partial f(x)$ by Theorem 16.29, and we deduce from Proposition 16.4(iii) that $[u_0, u_1] \subset \partial f(x)$. ☐

Proposition 16.38 *Let $f \in \Gamma_0(\mathcal{H})$. Then* $\operatorname{gra}(f + \iota_{\operatorname{dom} \partial f})$ *is a dense subset of* $\operatorname{gra} f$*; in other words, for every $x \in \operatorname{dom} f$, there exists a sequence $(x_n)_{n\in\mathbb{N}}$ in $\operatorname{dom} \partial f$ such that $x_n \to x$ and $f(x_n) \to f(x)$.*

Proof. Take $x \in \operatorname{dom} f$ and $\varepsilon \in \mathbb{R}_{++}$, and set $(p, \pi) = P_{\operatorname{epi} f}(x, f(x) - \varepsilon)$. Proposition 9.19 implies that $\pi = f(p) > f(x) - \varepsilon$ and

$$(\forall y \in \operatorname{dom} f) \quad \langle y - p \mid x - p \rangle \leqslant \big(f(p) - f(x) + \varepsilon\big)\big(f(y) - f(p)\big). \tag{16.34}$$

In view of (16.1), we deduce that $(x - p)/(f(p) - f(x) + \varepsilon) \in \partial f(p)$, hence $p \in \operatorname{dom} \partial f$. In addition, (16.34) with $y = x$ yields $\|(x, f(x)) - (p, f(p))\|^2 = \|x - p\|^2 + |f(x) - f(p)|^2 \leqslant \varepsilon(f(x) - f(p)) < \varepsilon^2$. ☐

Corollary 16.39 *Let $f \in \Gamma_0(\mathcal{H})$. Then* $\operatorname{dom} \partial f$ *is a dense subset of* $\operatorname{dom} f$.

Corollary 16.40 *Let $f \in \Gamma_0(\mathcal{H})$. Then $f^* = (f + \iota_{\operatorname{dom} \partial f})^*$.*

Proof. This follows from Proposition 16.38, Proposition 13.16(iv), and Proposition 13.10(vi). ☐

In the light of Proposition 16.27, the following result can be viewed as a variant of Proposition 16.25.

Corollary 16.41 *Let $f \in \Gamma_0(\mathcal{H})$. Then $f = (f + \iota_{\operatorname{dom} \partial f})^{**}$.*

Proof. Combine Corollary 16.40 and Corollary 13.38. ☐

16.4 Subdifferential Calculus

In this section we establish several rules for computing subdifferentials of transformations of convex functions in terms of the subdifferentials of these functions.

Proposition 16.42 *Let $\mathcal{K}$ be a real Hilbert space, let $f \in \Gamma_0(\mathcal{H})$, let $g \in \Gamma_0(\mathcal{K})$, and let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ be such that $L(\operatorname{dom} f) \cap \operatorname{dom} g \neq \varnothing$. Suppose that $(f + g \circ L)^* = f^* \boxdot (L^* \triangleright g^*)$. Then $\partial(f + g \circ L) = \partial f + L^* \circ (\partial g) \circ L$.*

Proof. In view of Proposition 16.6(ii), it remains to establish the inclusion $\operatorname{gra} \partial(f + g \circ L) \subset \operatorname{gra} \big(\partial f + L^* \circ (\partial g) \circ L\big)$. Take $(x, u) \in \operatorname{gra} \partial(f + g \circ L)$. On the one hand, Proposition 16.10 forces

$$\big(f + g \circ L\big)(x) + \big(f + g \circ L\big)^*(u) = \langle x \mid u \rangle . \tag{16.35}$$

On the other hand, our assumption implies that there exists $v \in \mathcal{K}$ such that $(f + g \circ L)^*(u) = f^*(u - L^*v) + g^*(v)$. Altogether,

$$\big(f(x) + f^*(u - L^*v) - \langle x \mid u - L^*v \rangle\big) + \big(g(Lx) + g^*(v) - \langle x \mid L^*v \rangle\big) = 0. \tag{16.36}$$

In view of Proposition 13.15, we obtain $f(x)+f^*(u-L^*v) = \langle x \mid u - L^*v\rangle$ and $g(Lx) + g^*(v) = \langle Lx \mid v\rangle$. In turn, Proposition 16.10 yields $u - L^*v \in \partial f(x)$ and $v \in \partial g(Lx)$, hence $u \in \partial f(x) + L^*(\partial g(Lx))$. ☐

Example 16.43 Let $f \in \Gamma_0(\mathcal{H})$ and let $\gamma \in \mathbb{R}_{++}$. Then $\partial(f + (\gamma/2)\|\cdot\|^2) = \partial f + \gamma\mathrm{Id}$.

Proof. Combine Proposition 16.42, Proposition 14.1, and Example 16.12. ☐

Proposition 16.44 *Let $f \in \Gamma_0(\mathcal{H})$, and let x and p be in $\mathcal{H}$. Then*

$$p = \mathrm{Prox}_f\, x \quad \Leftrightarrow \quad x - p \in \partial f(p). \tag{16.37}$$

In other words,

$$\mathrm{Prox}_f = (\mathrm{Id} + \partial f)^{-1}. \tag{16.38}$$

Proof. We derive (16.37) from Proposition 12.26 and (16.1). Alternatively, it follows from Definition 12.23, Theorem 16.3, and Example 16.43 that $p = \mathrm{Prox}_f\, x \Leftrightarrow 0 \in \partial(f + (1/2)\|x - \cdot\|^2)(p) = \partial f(p) + p - x$. ☐

Proposition 16.45 *Let $f \in \Gamma_0(\mathcal{H})$. Then* $\mathrm{ran}(\mathrm{Id} + \partial f) = \mathcal{H}$.

Proof. We deduce from (16.38) and Definition 12.23 that $\mathrm{ran}(\mathrm{Id} + \partial f) = \mathrm{dom}\,\mathrm{Prox}_f = \mathcal{H}$. ☐

Remark 16.46 Let f and g be functions in $\Gamma_0(\mathcal{H})$ such that $\mathrm{dom}\, f \cap \mathrm{dom}\, g \neq \varnothing$. Proposition 16.42 yields $(f + g)^* = f^* \boxdot g^* \Rightarrow \partial(f + g) = \partial f + \partial g$, i.e.,

$$f^* \,\square\, g^* \text{ is exact on } \mathrm{dom}(f + g)^* \quad \Rightarrow \quad \partial(f + g) = \partial f + \partial g. \tag{16.39}$$

Almost conversely, one has (see Exercise 16.12)

$$\partial(f + g) = \partial f + \partial g \quad \Rightarrow \quad f^* \,\square\, g^* \text{ is exact on } \mathrm{dom}\,\partial(f + g)^*, \tag{16.40}$$

which raises the question whether the implication $\partial(f + g) = \partial f + \partial g \Rightarrow (f + g)^* = f^* \boxdot g^*$ holds. An example constructed in [168] shows that the answer is negative.

Theorem 16.47 *Let $\mathcal{K}$ be a real Hilbert space, let $f \in \Gamma_0(\mathcal{H})$, let $g \in \Gamma_0(\mathcal{K})$, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$. Suppose that one of the following holds:*

(i) $0 \in \mathrm{sri}(\mathrm{dom}\, g - L(\mathrm{dom}\, f))$ *(see Proposition 6.19 for special cases).*
(ii) *$\mathcal{K}$ is finite-dimensional, g is polyhedral, and* $\mathrm{dom}\, g \cap \mathrm{ri}\, L(\mathrm{dom}\, f) \neq \varnothing$.
(iii) *$\mathcal{H}$ and $\mathcal{K}$ are finite-dimensional, f and g are polyhedral, and* $\mathrm{dom}\, g \cap L(\mathrm{dom}\, f) \neq \varnothing$.

Then $\partial(f + g \circ L) = \partial f + L^* \circ (\partial g) \circ L$.

Proof. Combine Theorem 15.27 and Proposition 16.42. ☐

Corollary 16.48 *Let f and g be functions in $\Gamma_0(\mathcal{H})$ such that one of the following holds:*

(i) $0 \in \operatorname{sri}(\operatorname{dom} f - \operatorname{dom} g)$.
(ii) $\operatorname{dom} f \cap \operatorname{int} \operatorname{dom} g \neq \varnothing$.
(iii) $\operatorname{dom} g = \mathcal{H}$.
(iv) *$\mathcal{H}$ is finite-dimensional and* $\operatorname{ri} \operatorname{dom} f \cap \operatorname{ri} \operatorname{dom} g \neq \varnothing$.

Then $\partial(f+g) = \partial f + \partial g$.

Proof. (i): Clear from Theorem 16.47(i).
(ii)⇒(i): Proposition 6.19(vii).
(iii)⇒(ii): Clear.
(iv)⇒(i): Proposition 6.19(viii). ☐

The next corollary is a refinement of Proposition 16.27.

Corollary 16.49 *Let $f \in \Gamma_0(\mathcal{H})$. Then*

$$\operatorname{int} \operatorname{dom} f = \operatorname{cont} f \subset \operatorname{sri} \operatorname{dom} f \subset \operatorname{dom} \partial f \subset \operatorname{dom} f. \tag{16.41}$$

Proof. In view of Proposition 16.27, it is enough to show that $\operatorname{sri} \operatorname{dom} f \subset \operatorname{dom} \partial f$. Suppose that $y \in \operatorname{sri} \operatorname{dom} f$ and set $g = \iota_{\{y\}}$. Then $0 \in \operatorname{sri}(\operatorname{dom} f - \operatorname{dom} g)$. In addition, $f + g = \iota_{\{y\}} + f(y)$ and hence $\partial(f+g)(y) = \mathcal{H}$. On the other hand, Corollary 16.48(i) yields $\partial(f+g)(y) = \partial f(y) + \partial g(y)$. Altogether, $\partial f(y) \neq \varnothing$. ☐

Corollary 16.50 *Let m be an integer such that $m \geqslant 2$, set $I = \{1, \ldots, m\}$, and let $(f_i)_{i \in I}$ be functions in $\Gamma_0(\mathcal{H})$ such that one of the following holds:*

(i) *We have*

$$0 \in \bigcap_{i=2}^{m} \operatorname{sri}\left(\operatorname{dom} f_i - \bigcap_{j=1}^{i-1} \operatorname{dom} f_j\right). \tag{16.42}$$

(ii) *For every $i \in \{2, \ldots, m\}$, $\operatorname{dom} f_i - \bigcap_{j=1}^{i-1} \operatorname{dom} f_j$ is a closed linear subspace.*
(iii) *The sets $(\operatorname{dom} f_i)_{i \in I}$ are linear subspaces and, for every $i \in \{2, \ldots, m\}$, $\operatorname{dom} f_i + \bigcap_{j=1}^{i-1} \operatorname{dom} f_j$ is closed.*
(iv) $\operatorname{dom} f_m \cap \bigcap_{i=1}^{m-1} \operatorname{int} \operatorname{dom} f_i \neq \varnothing$.
(v) *$\mathcal{H}$ is finite-dimensional and* $\bigcap_{i \in I} \operatorname{ri} \operatorname{dom} f_i \neq \varnothing$.

Then $\partial(\sum_{i \in I} f_i) = \sum_{i \in I} \partial f_i$.

Proof. (i): We proceed by induction. For $m = 2$ and functions f_1 and f_2 in $\Gamma_0(\mathcal{H})$, (16.42) becomes $0 \in \operatorname{sri}(\operatorname{dom} f_2 - \operatorname{dom} f_1)$ and we derive from Corollary 16.48(i) that $\partial(f_2 + f_1) = \partial f_2 + \partial f_1$. Now suppose that the result is true for m functions $(f_i)_{1 \leqslant i \leqslant m}$ in $\Gamma_0(\mathcal{H})$, where $m \geqslant 2$. Let f_{m+1} be a function in $\Gamma_0(\mathcal{H})$ such that

$$0 \in \bigcap_{i=2}^{m+1} \operatorname{sri}\Bigg(\operatorname{dom} f_i - \bigcap_{j=1}^{i-1} \operatorname{dom} f_j\Bigg). \tag{16.43}$$

Then (16.42) holds and hence the induction hypothesis yields $\partial(\sum_{i=1}^m f_i) = \sum_{i=1}^m \partial f_i$. Moreover, it follows from (16.43) that

$$0 \in \operatorname{sri}\Bigg(\operatorname{dom} f_{m+1} - \bigcap_{i=1}^{m} \operatorname{dom} f_i\Bigg) = \operatorname{sri}\Bigg(\operatorname{dom} f_{m+1} - \operatorname{dom} \sum_{i=1}^{m} f_i\Bigg), \tag{16.44}$$

where $\sum_{i=1}^m f_i \in \varGamma_0(\mathcal{H})$. We therefore derive from Corollary 16.48(i) that

$$\begin{aligned}\partial\Bigg(\sum_{i=1}^{m+1} f_i\Bigg) &= \partial\Bigg(f_{m+1} + \sum_{i=1}^{m} f_i\Bigg)\\ &= \partial f_{m+1} + \partial\Bigg(\sum_{i=1}^{m} f_i\Bigg)\\ &= \partial f_{m+1} + \sum_{i=1}^{m} \partial f_i\\ &= \sum_{i=1}^{m+1} \partial f_i,\end{aligned} \tag{16.45}$$

which yields (i).

(ii): In view of Proposition 8.2 and Proposition 6.20(i), this is a special case of (i).

(iii): In view of Proposition 6.20(ii), this is a special case of (i).

(iv): In view of Proposition 8.2 and Proposition 6.20(iii), this is a special case of (i).

(v): In view of Proposition 8.2 and Proposition 6.20(iv), this is a special case of (i). ☐

The next example shows that the sum rule for subdifferentials in Corollary 16.48 fails if the domains of the functions merely intersect.

Example 16.51 Suppose that $\mathcal{H} = \mathbb{R}^2$, and set $C = B((-1,0);1)$ and $D = B((1,0);1)$. Then $C \cap D = \operatorname{dom} \iota_C \cap \operatorname{dom} \iota_D = \{(0,0)\}$ and $\partial(\iota_C + \iota_D)(0,0) = \mathbb{R}^2 \neq \mathbb{R} \times \{0\} = \partial\iota_C(0,0) + \partial\iota_D(0,0)$.

It follows from Fermat's rule (Theorem 16.3) that a function that has a minimizer is bounded below and 0 belongs to the range of its subdifferential operator. In contrast, a function without a minimizer that is unbounded below—such as a nonzero continuous linear functional—cannot have 0 in the closure of the range of its subdifferential operator. However, the following result implies that functions in $\varGamma_0(\mathcal{H})$ that are bounded below possess arbitrarily small subgradients.

Corollary 16.52 *Let $f \in \Gamma_0(\mathcal{H})$ be bounded below. Then $0 \in \overline{\operatorname{ran}}\, \partial f$.*

Proof. Fix $\varepsilon \in \mathbb{R}_{++}$. Since f is bounded below and $\|\cdot\|$ is coercive, Corollary 11.16(ii) implies that $f+\varepsilon\|\cdot\|$ has a minimizer, say z. Using Theorem 16.3, Corollary 16.48(iii), and Example 16.32, we obtain $0 \in \partial f(z) + \varepsilon B(0;1)$. $\square$

Here is another consequence of Theorem 16.47.

Corollary 16.53 *Let $g \in \Gamma_0(\mathcal{K})$ and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$. Suppose that one of the following holds:*

(i) $0 \in \operatorname{sri}(\operatorname{dom} g - \operatorname{ran} L)$.
(ii) $\mathcal{K}$ *is finite-dimensional, g is polyhedral, and* $\operatorname{dom} g \cap \operatorname{ran} L \neq \varnothing$.

Then $\partial(g \circ L) = L^* \circ (\partial g) \circ L$.

Example 16.54 Let $f \in \Gamma_0(\mathcal{H})$. Then $\partial(f^\vee) = -(\partial f)^\vee$.

Example 16.55 Let N be a strictly positive integer, set $I = \{1,\ldots,N\}$, let $(u_i)_{i\in I} \in \mathcal{H}^N$, let $(\beta_i)_{i\in I} \in \mathbb{R}^N$, and set $C = \bigcap_{i\in I} \big\{x \in \mathcal{H} \;\big|\; \langle x \mid u_i\rangle \leqslant \beta_i\big\}$. Suppose that $z \in C$ and set $J = \big\{i \in I \;\big|\; \langle z \mid u_i\rangle = \beta_i\big\}$. Then $N_C z = \sum_{j\in J} \mathbb{R}_+ u_j$.

Proof. Set $b = (\beta_i)_{i\in I}$, set $L\colon \mathcal{H} \to \mathbb{R}^N \colon x \mapsto (\langle x \mid u_i\rangle)_{i\in I}$, and set $g\colon \mathbb{R}^N \to \left]-\infty,+\infty\right] : y \mapsto \iota_{\mathbb{R}^N_-}(y-b)$. Then g is polyhedral, $\iota_C = g\circ L$, and $Lz \in \operatorname{dom} g \cap \operatorname{ran} L$. By Corollary 16.53(ii), $N_C z = \partial\iota_C(z) = \partial(g\circ L)(z) = L^*(\partial g(Lz)) = L^*(N_{\mathbb{R}^N_-}(Lz-b))$. Hence the result follows from Example 6.42(ii). $\square$

We now obtain a nonsmooth mean value theorem.

Theorem 16.56 *Let $f \in \Gamma_0(\mathcal{H})$, let x_0 and x_1 be in $\operatorname{dom} f$, and suppose that one of the following holds:*

(i) $0 \in \operatorname{sri}(\operatorname{dom} f - \operatorname{aff}\{x_0,x_1\})$.
(ii) $\operatorname{aff}\{x_0,x_1\} \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$.
(iii) $\mathcal{H}$ *is finite-dimensional and* $\operatorname{aff}\{x_0,x_1\} \cap \operatorname{ri}\operatorname{dom} f \neq \varnothing$.

Then $f(x_1) - f(x_0) \in \langle x_1 - x_0 \mid \partial f(\left]x_0,x_1\right[)\rangle$.

Proof. Set $(\forall t \in \mathbb{R})$ $x_t = (1-t)x_0 + tx_1$. Define $g\colon \mathbb{R} \to \left]-\infty,+\infty\right] : t \mapsto f(x_t)$. Then $g \in \Gamma_0(\mathbb{R})$ and $g|_{[0,1]}$ is continuous by Corollary 9.15. From Corollary 16.53(i) and Proposition 6.19(vii)&(viii), we deduce that

$$(\forall t \in \mathbb{R}) \quad \partial g(t) = \langle x_1 - x_0 \mid \partial f(x_t)\rangle\,. \tag{16.46}$$

Next, set $h\colon \mathbb{R} \to \left]-\infty,+\infty\right] : t \mapsto g(t) + (1-t)g(1) + tg(0) = f(x_t) + (1-t)f(x_1) + tf(x_0)$. On the one hand, $h \in \Gamma_0(\mathbb{R})$ and $h|_{[0,1]}$ is continuous. On the other hand, $h(0) = g(0)+g(1) = h(1)$. By convexity of h, $\max h([0,1]) = h(0)$. Altogether, the Weierstrass Theorem (Theorem 1.29) guarantees the existence of a minimizer $t \in \left]0,1\right[$ of the function $h|_{[0,1]}$. In turn, Fermat's rule (Theorem 16.3), Corollary 16.48(iv), and (16.46) yield $0 \in \partial h(t) = \partial g(t) + g(0) - g(1) = \langle x_1 - x_0 \mid \partial f(x_t)\rangle + f(x_0) - f(x_1)$. ☐

Corollary 16.57 *Let $f\colon \mathcal{H} \to \mathbb{R}$ be convex and continuous. Suppose that $\beta = \sup \|\operatorname{ran} \partial f\| < +\infty$. Then f is Lipschitz continuous with constant β.*

Proof. Combine Theorem 16.56 and Cauchy–Schwarz. ☐

The following central result can be viewed as another consequence of Corollary 16.48.

Theorem 16.58 (Brøndsted–Rockafellar) *Let $f \in \Gamma_0(\mathcal{H})$, let $(y,v) \in \operatorname{dom}(f \oplus f^*)$, and let λ and μ in $\mathbb{R}_+$ satisfy $f(y) + f^*(v) \leqslant \langle y \mid v\rangle + \lambda\mu$. Then there exists $(z,w) \in \operatorname{gra} \partial f$ such that $\|z - y\| \leqslant \lambda$ and $\|w - v\| \leqslant \mu$.*

Proof. Set $\alpha = \lambda\mu$. If $\alpha = 0$, then (y,v) has the required properties by Proposition 16.10. So assume that $\alpha > 0$, set $\beta = \lambda$, and set $h = f - \langle \cdot \mid v\rangle$. Then for every $x \in \mathcal{H}$, we have

$$h(x) = -\big(\langle x \mid v\rangle - f(x)\big) \geqslant -f^*(v) \geqslant f(y) - \langle y \mid v\rangle - \lambda\mu = h(y) - \alpha. \quad (16.47)$$

Hence h is bounded below and $\alpha \geqslant h(y) - \inf h(\mathcal{H})$. Theorem 1.46 yields a point $z \in \mathcal{H}$ such that $\|z - y\| \leqslant \beta = \lambda$ and $z \in \operatorname{Argmin}(h + (\alpha/\beta)\|\cdot - z\|)$. In turn, Theorem 16.3, Corollary 16.48(iii), and Example 16.32 imply that

$$0 \in \partial\big(h + (\alpha/\beta)\|\cdot - z\|\big)(z) = \partial f(z) - v + \mu B(0;1). \quad (16.48)$$

Thus, there exists $w \in \partial f(z)$ such that $v - w \in \mu B(0;1)$, i.e., $\|w - v\| \leqslant \mu$. ☐

Proposition 16.59 *Let $\mathcal{K}$ be a real Hilbert space, let $F \in \Gamma_0(\mathcal{H} \oplus \mathcal{K})$, and set*

$$f\colon \mathcal{H} \to [-\infty,+\infty] : x \mapsto \inf F(x,\mathcal{K}). \quad (16.49)$$

Suppose that f is proper and that $(x,y) \in \mathcal{H} \times \mathcal{K}$ satisfies $f(x) = F(x,y)$, and let $u \in \mathcal{H}$. Then $u \in \partial f(x) \Leftrightarrow (u,0) \in \partial F(x,y)$.

Proof. As seen in Proposition 13.33, $f^*(u) = F^*(u,0)$. Hence, by Proposition 16.10, $u \in \partial f(x) \Leftrightarrow f^*(u) = \langle x \mid u\rangle - f(x) \Leftrightarrow F^*(u,0) = \langle x \mid u\rangle - F(x,y)$ $\Leftrightarrow F^*(u,0) = \langle (x,y) \mid (u,0)\rangle - F(x,y) \Leftrightarrow (u,0) \in \partial F(x,y)$. ☐

Proposition 16.60 *Let $f \in \Gamma_0(\mathcal{H})$, let $\mathcal{K}$ be a real Hilbert space, let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$, suppose that $y \in \operatorname{dom}(L \triangleright f)$, and let $x \in \mathcal{H}$. Suppose that $Lx = y$. Then the following hold:*

(i) *Suppose that $(L \triangleright f)(y) = f(x)$. Then*

$$\partial(L \triangleright f)(y) = (L^*)^{-1}(\partial f(x)). \quad (16.50)$$

(ii) *Suppose that $(L^*)^{-1}(\partial f(x)) \neq \varnothing$. Then $(L \triangleright f)(y) = f(x)$.*

Proof. Let $v \in \mathcal{K}$. It follows from Proposition 13.24(iv), Proposition 13.15, and Proposition 16.10 that

$$\begin{aligned} f(x) + (L \triangleright f)^*(v) = \langle y \mid v\rangle &\Leftrightarrow f(x) + f^*(L^*v) = \langle Lx \mid v\rangle \\ &\Leftrightarrow f(x) + f^*(L^*v) = \langle x \mid L^*v\rangle \\ &\Leftrightarrow L^*v \in \partial f(x). \end{aligned} \tag{16.51}$$

(i): Proposition 16.10 and Proposition 13.24(iv) imply that

$$\begin{aligned} v \in \partial(L \triangleright f)(y) &\Leftrightarrow (L \triangleright f)(y) + (L \triangleright f)^*(v) = \langle y \mid v\rangle \\ &\Leftrightarrow f(x) + f^*(L^*v) = \langle Lx \mid v\rangle . \end{aligned} \tag{16.52}$$

To obtain (16.50), combine (16.52) with (16.51).

(ii): Suppose that $v \in (L^*)^{-1}(\partial f(x))$. Proposition 13.15 and (16.51) yield $\langle y \mid v\rangle \leqslant (L \triangleright f)(y) + (L \triangleright f)^*(v) \leqslant f(x) + (L \triangleright f)^*(v) = \langle y \mid v\rangle$. ☐

We now derive the following important rule for the subdifferential of an infimal convolution.

Proposition 16.61 *Let f and g be in $\Gamma_0(\mathcal{H})$, let $x \in \operatorname{dom}(f \,\Box\, g)$, and let $y \in \mathcal{H}$. Then the following hold:*

(i) *Suppose that $(f \,\Box\, g)(x) = f(y) + g(x - y)$. Then*

$$\partial(f \,\Box\, g)(x) = \partial f(y) \cap \partial g(x - y). \tag{16.53}$$

(ii) *Suppose that $\partial f(y) \cap \partial g(x-y) \neq \varnothing$. Then $(f \,\Box\, g)(x) = f(y) + g(x-y)$.*

Proof. Set $L\colon \mathcal{H} \oplus \mathcal{H} \to \mathcal{H}\colon (x, y) \mapsto x + y$. Then $L \in \mathcal{B}(\mathcal{H} \oplus \mathcal{H}, \mathcal{H})$ and $L^*\colon \mathcal{H} \to \mathcal{H} \oplus \mathcal{H}\colon u \mapsto (u, u)$. Proposition 12.37 implies that $L \triangleright (f \oplus g) = f \,\Box\, g$ and hence that $\operatorname{dom}(L \triangleright (f \oplus g)) = \operatorname{dom}(f \,\Box\, g)$. Thus, $L(y, x-y) = x \in \operatorname{dom}(L \triangleright (f \oplus g))$.

(i): Since $(L \triangleright (f \oplus g))(x) = (f \oplus g)(y, x - y)$, Proposition 16.60(i) and Proposition 16.9 imply that $\partial(f \,\Box\, g)(x) = (L^*)^{-1}(\partial(f \oplus g)(y, x - y)) = (L^*)^{-1}(\partial f(y) \times \partial g(x - y)) = \partial f(y) \cap \partial g(x - y)$.

(ii): The assumption that $\partial f(y) \cap \partial g(x-y) = (L^*)^{-1}(\partial f(y) \times \partial g(x-y)) = (L^*)^{-1}(\partial(f \oplus g)(y, x - y)) \neq \varnothing$ and Proposition 16.60(ii) yield $(f \,\Box\, g)(x) = (L \triangleright (f \oplus g))(x) = (f \oplus g)(y, x - y) = f(y) + g(x - y)$. ☐

Example 16.62 Let C be a nonempty closed convex subset of $\mathcal{H}$. Then

$$(\forall x \in \mathcal{H}) \;\; \partial d_C(x) = \begin{cases} \left\{\dfrac{x - P_C x}{d_C(x)}\right\}, & \text{if } \; x \notin C; \\ N_C x \cap B(0;1), & \text{if } \; x \in \operatorname{bdry} C; \\ \{0\}, & \text{if } \; x \in \operatorname{int} C. \end{cases} \tag{16.54}$$

Proof. Set $f = \iota_C$ and $g = \|\cdot\|$. Then it follows from Theorem 3.16 that $d_C(x) = f(P_C x) + g(x - P_C x)$. Therefore, Proposition 16.61(i) yields $\partial d_C(x) = \partial(f \,\Box\, g)(x) = \partial f(P_C x) \cap \partial g(x - P_C x) = N_C(P_C x) \cap (\partial\|\cdot\|)(x - P_C x)$. Since $x - P_C x \in N_C(P_C x)$ by Proposition 6.47, (16.54) follows from Example 16.32, Proposition 6.44, and Proposition 6.12(ii). ☐

Our next result concerns the subdifferential of integral functions.

Proposition 16.63 *Let $(\Omega, \mathcal{F}, \mu)$ be a complete σ-finite measure space, let $(\mathsf{H}, \langle\cdot \mid \cdot\rangle_{\mathsf{H}})$ be a separable real Hilbert space, and let $\varphi \in \Gamma_0(\mathsf{H})$. Suppose that $\mathcal{H} = L^2((\Omega, \mathcal{F}, \mu); \mathsf{H})$ and that one of the following holds:*

(i) $\mu(\Omega) < +\infty$.
(ii) $\varphi \geqslant \varphi(0) = 0$.

Set

$$f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$$
$$x \mapsto \begin{cases} \displaystyle\int_\Omega \varphi\big(x(\omega)\big)\mu(d\omega), & \text{if } \varphi \circ x \in L^1\big((\Omega, \mathcal{F}, \mu); \mathbb{R}\big); \\ +\infty, & \text{otherwise.} \end{cases} \tag{16.55}$$

Then $f \in \Gamma_0(\mathcal{H})$ and, for every $x \in \operatorname{dom} f$,

$$\partial f(x) = \big\{u \in \mathcal{H} \;\big|\; u(\omega) \in \partial\varphi\big(x(\omega)\big) \;\; \mu\text{-a.e.}\big\}. \tag{16.56}$$

Proof. Take x and u in $\mathcal{H}$. It follows from Proposition 13.15 that

$$\varphi\big(x(\omega)\big) + \varphi^*\big(u(\omega)\big) - \langle x(\omega) \mid u(\omega)\rangle_{\mathsf{H}} \geqslant 0 \quad \mu\text{-a.e.} \tag{16.57}$$

On the other hand, Proposition 13.50(i) yields $f \in \Gamma_0(\mathcal{H})$. Hence, we derive from Proposition 13.50(ii) and Theorem 16.29 that

$$\begin{aligned} u \in \partial f(x) &\Leftrightarrow f(x) + f^*(u) - \langle x \mid u\rangle = 0 \\ &\Leftrightarrow \int_\Omega \Big(\varphi\big(x(\omega)\big) + \varphi^*\big(u(\omega)\big) - \langle x(\omega) \mid u(\omega)\rangle_{\mathsf{H}}\Big)\mu(d\omega) = 0 \\ &\Leftrightarrow \varphi\big(x(\omega)\big) + \varphi^*\big(u(\omega)\big) - \langle x(\omega) \mid u(\omega)\rangle_{\mathsf{H}} = 0 \quad \mu\text{-a.e.} \\ &\Leftrightarrow u(\omega) \in \partial\varphi\big(x(\omega)\big) \quad \mu\text{-a.e.}, \end{aligned} \tag{16.58}$$

which provides (16.56). ☐

Example 16.64 Let $(\Omega, \mathcal{F}, \mu)$ be a complete σ-finite measure space and let $(\mathsf{H}, \|\cdot\|_{\mathsf{H}})$ be a separable real Hilbert space. Suppose that $\mathcal{H} = L^2((\Omega, \mathcal{F}, \mu); \mathsf{H})$. Set

$$f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$$
$$x \mapsto \begin{cases} \displaystyle\int_\Omega \|x(\omega)\|_{\mathsf{H}}\,\mu(d\omega), & \text{if } x \in L^1\big((\Omega,\mathcal{F},\mu);\mathsf{H}\big);\\ +\infty, & \text{otherwise,}\end{cases} \tag{16.59}$$

and define

$$(\forall \omega \in \Omega)\quad u(\omega) = \begin{cases} x(\omega)/\|x(\omega)\|_{\mathsf{H}}, & \text{if } x(\omega) \neq 0;\\ 0, & \text{otherwise.}\end{cases} \tag{16.60}$$

Then $u \in \partial f(x)$.

Proof. Apply Proposition 16.63 with $\varphi = \|\cdot\|_{\mathsf{H}}$ and use Example 16.32. ☐

Proposition 16.65 *Let $F \in \Gamma_0(\mathcal{H}\oplus\mathcal{H})$ be autoconjugate and let $(x,u) \in \mathcal{H}\oplus\mathcal{H}$. Then the following are equivalent:*

(i) $F(x,u) = \langle x \mid u\rangle$.
(ii) $F^*(u,x) = \langle x \mid u\rangle$.
(iii) $(u,x) \in \partial F(x,u)$.
(iv) $(x,u) = \operatorname{Prox}_F(x+u,x+u)$.

Proof. (i) $\Leftrightarrow F^{\intercal}(u,x) = \langle x \mid u\rangle \Leftrightarrow F^*(u,x) = \langle x \mid u\rangle \Leftrightarrow$ (ii). Hence, Proposition 13.36, (i), Proposition 16.10, and (16.38) yield (ii) $\Leftrightarrow F(x,u)+F^*(u,x) = 2\langle x \mid u\rangle \Leftrightarrow F(x,u)+F^*(u,x) = \langle (x,u) \mid (u,x)\rangle \Leftrightarrow (u,x) \in \partial F(x,u) \Leftrightarrow$ (iii) $\Leftrightarrow (x,u)+(u,x) \in (\mathrm{Id}+\partial F)(x,u) \Leftrightarrow$ (iv). ☐

Proposition 16.66 *Let $F \in \Gamma_0(\mathcal{H}\oplus\mathcal{H})$ and set $L\colon \mathcal{H}\oplus\mathcal{H} \to \mathcal{H}\oplus\mathcal{H}\colon (x,u) \mapsto (u,x)$. Then $\operatorname{Prox}_{F^{*\intercal}} = \mathrm{Id} - L\operatorname{Prox}_F L$.*

Proof. Since $L^* = L = L^{-1}$, it follows from Proposition 13.35 and Corollary 16.53(i) that $\mathrm{Id}+\partial F^{*\intercal} = LL + L(\partial F^*)L = L(\mathrm{Id}+\partial F^*)L$. Hence, we derive from (14.6) that $\operatorname{Prox}_{F^{*\intercal}} = (L(\mathrm{Id}+\partial F^*)L)^{-1} = L^{-1}(\mathrm{Id}+\partial F^*)^{-1}L^{-1} = L\operatorname{Prox}_{F^*} L = L(\mathrm{Id}-\operatorname{Prox}_F)L = LL - L\operatorname{Prox}_F L = \mathrm{Id} - L\operatorname{Prox}_F L$. ☐

16.5 The Subdifferential of a Composition

Definition 16.67 Let $\mathcal{K}$ be a real Hilbert space, and let $F\colon \mathcal{H} \to 2^{\mathcal{K}}$ be such that $\operatorname{gra} F$ is nonempty and convex. The *coderivative* of F at $(x,y) \in \operatorname{gra} F$ is

$$\mathsf{D}^*F(x,y)\colon \mathcal{K} \to 2^{\mathcal{H}}\colon v \mapsto \big\{u \in \mathcal{H} \;\big|\; (u,-v) \in N_{\operatorname{gra} F}(x,y)\big\}. \tag{16.61}$$

Example 16.68 Let $\mathcal{K}$ be a real Hilbert space and let C be a nonempty convex subset of $\mathcal{K}$. Set $F\colon \mathcal{H} \to 2^{\mathcal{K}}\colon x \mapsto C$ and let $(x,y) \in \operatorname{gra} F = \mathcal{H}\times C$. Then $\operatorname{gra} \mathsf{D}^*F(x,y) = -N_C y \times \{0\}$.

Proposition 16.69 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex, and set $F\colon \mathcal{H} \to 2^{\mathbb{R}}\colon x \mapsto \big\{y \in \mathbb{R} \;\big|\; f(x) \leqslant y\big\}$. Now let $x \in \operatorname{dom} f$. Then*

$$\mathsf{D}^*F(x,f(x))\colon \mathbb{R} \to \mathcal{H}\colon v \mapsto \begin{cases} v\partial f(x), & \text{if } v > 0;\\ N_{\operatorname{dom} f}\, x, & \text{if } v = 0;\\ \varnothing, & \text{if } v < 0. \end{cases} \tag{16.62}$$

Proof. Observe that $\operatorname{gra} F = \operatorname{epi} f$. Now let $u \in \mathcal{H}$ and let $v \in \mathbb{R}$. Then $u \in \mathsf{D}^*F(x,f(x))(v) \Leftrightarrow (u,-v) \in N_{\operatorname{epi} f}(x,f(x))$. If $v > 0$, then it follows from Proposition 16.16 that $u \in \mathsf{D}^*F(x,f(x))(v) \Leftrightarrow (u/v,-1) \in N_{\operatorname{epi} f}(x,f(x)) \Leftrightarrow u/v \in \partial f(x) \Leftrightarrow u \in v\partial f(x)$. If $v = 0$, then $u \in \mathsf{D}^*F(x,f(x))(0) \Leftrightarrow (u,0) \in N_{\operatorname{epi} f}(x,f(x)) \Leftrightarrow u \in N_{\operatorname{dom} f}\, x$. Finally, if $v < 0$, then $(u,-v) \notin N_{\operatorname{epi} f}(x,f(x))$ since f is convex. ☐

Proposition 16.70 *Let $\mathcal{K}$ be a real Hilbert space, let $g\colon \mathcal{H}\oplus\mathcal{K} \to \left]-\infty,+\infty\right]$ be proper and convex, let $F\colon \mathcal{H} \to 2^{\mathcal{K}}$ be such that $\operatorname{gra} F$ is convex, and set*

$$f\colon \mathcal{H} \to [-\infty,+\infty]\colon x \mapsto \inf\big\{g(x,y) \;\big|\; y \in F(x)\big\}. \tag{16.63}$$

Suppose that f is proper and let $\bar{x} \in \operatorname{dom} f$. Then

$$\bigcup_{\substack{(u,v)\in\partial g(\bar{x},\bar{y}),\\ g(\bar{x},\bar{y})=f(\bar{x}),\\ \bar{y}\in F(\bar{x})}} u + \mathsf{D}^*F(\bar{x},\bar{y})(v) \subset \partial f(\bar{x}). \tag{16.64}$$

Proof. Suppose that $(u,w,v,\bar{y}) \in \mathcal{H}\oplus\mathcal{H}\oplus\mathcal{K}\oplus\mathcal{K}$ satisfies $g(\bar{x},\bar{y}) = f(\bar{x})$, $\bar{y} \in F(\bar{x})$, $(u,v) \in \partial g(\bar{x},\bar{y})$, and $w-u \in \mathsf{D}^*F(\bar{x},\bar{y})(v)$. Then

$$(\forall (x,y) \in \operatorname{gra} F) \quad \langle x-\bar{x} \mid w-u\rangle + \langle y-\bar{y} \mid -v\rangle \leqslant 0. \tag{16.65}$$

Hence $(\forall (x,y) \in \operatorname{gra} F)$ $\langle x-\bar{x} \mid w\rangle \leqslant \langle (x,y)-(\bar{x},\bar{y}) \mid (u,v)\rangle \leqslant g(x,y) - g(\bar{x},\bar{y}) = g(x,y) - f(\bar{x})$. Thus $(\forall x \in \mathcal{H})$ $\langle x-\bar{x} \mid w\rangle \leqslant \inf_{y\in F(x)} g(x,y) - f(\bar{x}) = f(x) - f(\bar{x})$. ☐

Theorem 16.71 *Let $\mathcal{K}$ be a real Hilbert space, let $g \in \Gamma_0(\mathcal{H}\oplus\mathcal{K})$, let $F\colon \mathcal{H} \to 2^{\mathcal{K}}$ be such that $\operatorname{gra} F$ is closed and convex, and set*

$$f\colon \mathcal{H} \to [-\infty,+\infty]\colon x \mapsto \inf\big\{g(x,y) \;\big|\; y \in F(x)\big\}. \tag{16.66}$$

Suppose that f is proper, that $\bar{x} \in \operatorname{dom} f$, that $\bar{y} \in F(\bar{x})$, that $f(\bar{x}) = g(\bar{x},\bar{y})$, and that $(0,0) \in \operatorname{sri}(\operatorname{gra} F - \operatorname{dom} g)$. Then

$$\partial f(\bar{x}) = \bigcup_{(u,v)\in\partial g(\bar{x},\bar{y})} u + \mathsf{D}^*F(\bar{x},\bar{y})(v). \tag{16.67}$$

Proof. Suppose first that $w \in \partial f(\bar{x})$. Then, for every $(x,y) \in \operatorname{gra} F$, $\langle x-\bar{x} \mid w\rangle \leqslant f(x) - f(\bar{x}) = f(x) - g(\bar{x},\bar{y}) \leqslant g(x,y) - g(\bar{x},\bar{y})$. Thus, for every $(x,y) \in \mathcal{H} \oplus \mathcal{K}$, we have

$$\langle (x,y) - (\bar{x},\bar{y}) \mid (w,0)\rangle \leqslant \big(g + \iota_{\operatorname{gra} F}\big)(x,y) - \big(g + \iota_{\operatorname{gra} F}\big)(\bar{x},\bar{y}). \tag{16.68}$$

Hence, using Corollary 16.48(i),

$$(w,0) \in \partial\big(g + \iota_{\operatorname{gra} F}\big)(\bar{x},\bar{y}) = \partial g(\bar{x},\bar{y}) + N_{\operatorname{gra} F}(\bar{x},\bar{y}). \tag{16.69}$$

It follows that there exists $(u,v) \in \partial g(\bar{x},\bar{y})$ such that $(w-u,-v) \in N_{\operatorname{gra} F}(\bar{x},\bar{y})$, i.e., $w - u \in \mathsf{D}^*F(\bar{x},\bar{y})(v)$. Therefore, $w \in u + \mathsf{D}^*F(\bar{x},\bar{y})(v)$. The opposite inclusion follows from Proposition 16.70. ☐

We conclude with a formula for the subdifferential of a composition of convex functions.

Corollary 16.72 *Let $f\colon \mathcal{H} \to \mathbb{R}$ be continuous and convex, and let $\phi \in \Gamma_0(\mathbb{R})$ be increasing on* $\operatorname{ran} f$. *Suppose that* $(\operatorname{ri}(\operatorname{ran} f) + \mathbb{R}_{++}) \cap \operatorname{ri}\operatorname{dom}\phi \neq \varnothing$. *Let $\bar{x} \in \mathcal{H}$ be such that $f(\bar{x}) \in \operatorname{dom}\phi$. Then*

$$\partial(\phi \circ f)(\bar{x}) = \big\{\alpha u \;\big|\; (\alpha,u) \in \partial\phi\big(f(\bar{x})\big) \times \partial f(\bar{x})\big\}. \tag{16.70}$$

Proof. Observe that $\phi \circ f$ is convex by Proposition 8.21. Set $F\colon \mathcal{H} \to 2^{\mathbb{R}}\colon x \mapsto [f(x),+\infty[$. Since ϕ is increasing on $\operatorname{ran} f$ and lower semicontinuous, we have

$$(\forall x \in \mathcal{H}) \quad (\phi \circ f)(x) = \inf \phi\big(F(x)\big). \tag{16.71}$$

Set $g\colon \mathcal{H} \oplus \mathbb{R} \to \left]-\infty,+\infty\right] : (x,y) \mapsto \phi(y)$. On the one hand, $\operatorname{gra} F - \operatorname{dom} g = \operatorname{epi} f - (\mathcal{H} \times \operatorname{dom}\phi) = \mathcal{H} \times ((\operatorname{ran} f + \mathbb{R}_+) - \operatorname{dom}\phi)$. On the other hand, set $S = (\operatorname{ran} f + \mathbb{R}_+) - \operatorname{dom}\phi$. Then Corollary 6.15(ii) yields $0 \in \operatorname{ri} S$. Thus either $S = \{0\}$ or $0 \in \operatorname{int} S$, which implies that $\operatorname{cone}(\mathcal{H} \times S)$ is either $\mathcal{H} \oplus \{0\}$ or $\mathcal{H} \oplus \mathbb{R}$. In either case, $\operatorname{cone}(\mathcal{H} \oplus S) = \overline{\operatorname{span}}\,(\mathcal{H} \oplus S)$ and hence $(0,0) \in \operatorname{sri}(\mathcal{H} \times S)$ by (6.8). Thus,

$$(0,0) \in \operatorname{sri}(\operatorname{gra} F - \operatorname{dom} g). \tag{16.72}$$

Now set $\bar{y} = f(\bar{x})$. On the one hand, $\partial g(\bar{x},\bar{y}) = \{0\} \times \partial\phi(\bar{y}) \subset \{0\} \times \mathbb{R}_+$ because ϕ is increasing on $\operatorname{ran} f$. On the other hand, since $\operatorname{dom} f = \mathcal{H}$, Proposition 16.69 yields $\operatorname{dom} \mathsf{D}^*F(\bar{x},\bar{y}) = \mathbb{R}_+$ and $(\forall v \in \mathbb{R}_+)\ \mathsf{D}^*F(\bar{x},\bar{y}) = v\partial f(\bar{x})$. Altogether, Theorem 16.71 yields

$$\partial(\phi \circ f)(\bar{x}) = \bigcup_{(u,v) \in \partial g(\bar{x},\bar{y})} u + \mathsf{D}^*F(\bar{x},\bar{y})(v) = \bigcup_{v \in \partial\phi(\bar{y})} v\partial f(\bar{x}) \tag{16.73}$$

as claimed. ☐

Example 16.73 Let $\phi\colon \mathbb{R} \to \mathbb{R}$ be convex and even, and set $f = \phi \circ \|\cdot\|$. Then $\partial\phi(0) = [-\rho,\rho]$ for some $\rho \in \mathbb{R}_+$ and

$$(\forall x \in \mathcal{H}) \quad \partial f(x) = \begin{cases} \left\{ \dfrac{\alpha}{\|x\|} x \;\middle|\; \alpha \in \partial\phi(\|x\|) \right\}, & \text{if } x \neq 0; \\ B(0;\rho), & \text{if } x = 0. \end{cases} \tag{16.74}$$

Proof. By Corollary 8.40, ϕ is continuous. Therefore $\phi \in \Gamma_0(\mathbb{R})$, while Proposition 16.17(ii) implies that $\partial\phi(0)$ is a closed bounded interval $[-\rho, \rho]$ for some $\rho \in \mathbb{R}_+$. Furthermore, by Proposition 11.7(ii), ϕ is increasing on $\mathbb{R}_+ = \operatorname{ran} \|\cdot\|$. Altogether, the result follows from Corollary 16.72 applied to $f = \|\cdot\|$ and Example 16.32. ☐

Exercises

Exercise 16.1 Let $f : \mathbb{R} \to \left]-\infty, +\infty\right]$ be an increasing function. Suppose that $x \in \operatorname{dom} \partial f$ and that $\left]-\infty, x\right[\cap \operatorname{dom} f \neq \varnothing$. Show that $\partial f(x) \subset \mathbb{R}_+$.

Exercise 16.2 Prove Example 16.14.

Exercise 16.3 Let $f : \mathcal{H} \to \left]-\infty, +\infty\right]$ be proper, and let x and u be in $\mathcal{H}$. Show that $(x, u) \in \operatorname{gra} \partial f \Leftrightarrow (f \,\square\, \langle \cdot \mid u \rangle)(x) = f(x)$.

Exercise 16.4 Provide a function $f \in \Gamma_0(\mathbb{R})$ and a point $x \in \operatorname{dom} f$ such that $\partial f(x) = \varnothing$. Compare to Proposition 16.4(iv).

Exercise 16.5 Define $\boldsymbol{f} : \mathbb{R} \times \mathbb{R} \to \left]-\infty, +\infty\right]$ by

$$(x_1, x_2) \mapsto \begin{cases} x_1 x_2, & \text{if } x_1 > 0 \text{ and } x_2 > 0; \\ +\infty, & \text{otherwise.} \end{cases} \tag{16.75}$$

Show that (16.5) fails at every point $(x_1, x_2) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$.

Exercise 16.6 Provide f and g in $\Gamma_0(\mathbb{R})$, $x \in \operatorname{dom}(f \,\square\, g)$, and $y \in \mathbb{R}$ such that $(f \,\square\, g)(x) = f(y) + g(x - y)$ and $\partial(f \,\square\, g)(x) = \varnothing$.

Exercise 16.7 Suppose that $\mathcal{H} = \ell^2(\mathbb{N})$ and set $f : \mathcal{H} \to \left]-\infty, +\infty\right] : x = (\xi_n)_{n \in \mathbb{N}} \mapsto \sum_{n \in \mathbb{N}} n \xi_n^{2n}$. Prove the following: $f \in \Gamma_0(\mathcal{H})$, $\operatorname{dom} f = \mathcal{H}$, and f^* is not supercoercive.

Exercise 16.8 Let $f : \mathcal{H} \to \left]-\infty, +\infty\right]$ be proper. Suppose that $\operatorname{dom} f$ is convex and that $\operatorname{dom} \partial f = \operatorname{dom} f$. Show that f is convex.

Exercise 16.9 Provide an example in which $\partial f(x)$ is unbounded in Corollary 16.18.

Exercise 16.10 Prove Proposition 16.25.

Exercise 16.11 Provide details for Remark 16.28.

Exercise 16.12 Prove the implication (16.40).

Exercise 16.13 Use Corollary 16.39 to provide a different proof of Corollary 16.52.

Exercise 16.14 Is the converse of Corollary 16.52 true, i.e., if $f \in \Gamma_0(\mathcal{H})$ and $0 \in \overline{\operatorname{ran}}\,\partial f$, must f be bounded below?

Exercise 16.15 Use Theorem 16.58 to prove Corollary 16.39.

Exercise 16.16 Let C be a convex subset of $\mathcal{H}$ such that $0 \in C$ and suppose that $x \in \mathcal{H}$ satisfies $m_C(x) > 0$. Set $y = x/m_C(x)$ and assume that $y \in C$. Let $u \in \mathcal{H}$. Show that $u \in \partial m_C(x)$ if and only if $u \in N_C y$ and $\langle y \mid u\rangle = 1$.

Exercise 16.17 Let $f \in \Gamma_0(\mathcal{H})$ be sublinear and set $C = \partial f(0)$. Use Proposition 14.11 to show that $f = \sigma_C$ and that $\operatorname{epi} f^* = C \times \mathbb{R}_+$.

Exercise 16.18 Let f and g be sublinear functions in $\Gamma_0(\mathcal{H})$. Use Exercise 12.3 and Exercise 15.1 to prove the equivalence $(f+g)^* = f^* \boxdot g^* \Leftrightarrow \partial(f+g)(0) = \partial f(0) + \partial g(0)$. Compare to Remark 16.46.

Exercise 16.19 Let $\mathcal{K}$ be a real Hilbert space, let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$, let $z \in \mathcal{K}$, and set $F\colon \mathcal{H} \to \mathcal{K}\colon x \mapsto Lx+z$. Show that $(\forall (x,y) \in \operatorname{gra} F)\ \mathsf{D}^*F(x,y) = L^*$.

Exercise 16.20 Derive Proposition 16.59 from Theorem 16.71.

Chapter 17
Differentiability of Convex Functions

Fréchet differentiability, Gâteaux differentiability, directional differentiability, subdifferentiability, and continuity are notions that are closely related to each other. In this chapter, we provide fundamental results on these relationships, as well as basic results on the steepest descent direction, the Chebyshev center, and the max formula that relates the directional derivative to the support function of the subdifferential at a given point.

17.1 Directional Derivatives

Definition 17.1 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper, let $x \in \operatorname{dom} f$, and let $y \in \mathcal{H}$. The *directional derivative* of f at x in the direction y is

$$f'(x;y) = \lim_{\alpha \downarrow 0} \frac{f(x+\alpha y) - f(x)}{\alpha}, \tag{17.1}$$

provided that this limit exists in $[-\infty,+\infty]$. If $\mathcal{H} = \mathbb{R}$, then

$$f'_+(x) = f'(x;1) = \lim_{\alpha \downarrow 0} \frac{f(x+\alpha) - f(x)}{\alpha} \tag{17.2}$$

and

$$f'_-(x) = -f'(x;-1) = \lim_{\alpha \uparrow 0} \frac{f(x+\alpha) - f(x)}{\alpha} \tag{17.3}$$

are the *right* and *left derivatives* of f at x, respectively, provided they exist.

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-48311-5_31

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_17

Proposition 17.2 *Let* $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ *be proper and convex, let* $x \in \operatorname{dom} f$*, and let* $y \in \mathcal{H}$*. Then the following hold:*

(i) $f'(x;y)$ *exists in* $[-\infty,+\infty]$ *and*

$$f'(x;y) = \inf_{\alpha\in\mathbb{R}_{++}} \frac{f(x+\alpha y)-f(x)}{\alpha}. \tag{17.4}$$

(ii) $f'(x;y-x)+f(x) \leqslant f(y)$.

(iii) *Suppose that* $y \in \operatorname{dom} f$*. Then* $f'(x;y-x) \leqslant -f'(y;x-y)$.

(iv) $f'(x;\cdot)$ *is sublinear and* $f'(x;0)=0$.

(v) $f'(x;\cdot)$ *is proper, convex, and* $\operatorname{dom} f'(x;\cdot) = \operatorname{cone}(\operatorname{dom} f - x)$.

(vi) *Suppose that* $x \in \operatorname{core}\operatorname{dom} f$*. Then* $f'(x;\cdot)$ *is real-valued and sublinear.*

Proof. (i): This is a consequence of Proposition 9.27.

(ii): It follows from (i) that $f'(x;y-x) \leqslant (f(x+1(y-x))-f(x))/1 = f(y)-f(x)$.

(iii): Using (ii), we deduce that $f'(x;y-x) \leqslant f(y)-f(x) \leqslant -f'(y;x-y)$.

(iv): It is clear that $f'(x;0)=0$ and that $f'(x;\cdot)$ is positively homogeneous. Now take (y,η) and (z,ζ) in $\operatorname{epi} f'(x;\cdot)$, $\lambda \in \left]0,1\right[$, and $\varepsilon \in \mathbb{R}_{++}$. Then, for all $\alpha \in \mathbb{R}_{++}$ sufficiently small, we have $(f(x+\alpha y)-f(x))/\alpha \leqslant \eta+\varepsilon$ and $(f(x+\alpha z)-f(x))/\alpha \leqslant \zeta+\varepsilon$. For such small α, Corollary 8.12 yields

$$\begin{aligned} f\big(x+\alpha((1-\lambda)y+\lambda z)\big) - f(x) \\ &= f\big((1-\lambda)(x+\alpha y)+\lambda(x+\alpha z)\big)-f(x) \\ &\leqslant (1-\lambda)\big(f(x+\alpha y)-f(x)\big)+\lambda\big(f(x+\alpha z)-f(x)\big). \end{aligned} \tag{17.5}$$

Consequently,

$$\begin{aligned} \frac{f\big(x+\alpha((1-\lambda)y+\lambda z)\big)-f(x)}{\alpha} \\ &\leqslant (1-\lambda)\frac{f(x+\alpha y)-f(x)}{\alpha}+\lambda\frac{f(x+\alpha z)-f(x)}{\alpha} \\ &\leqslant (1-\lambda)(\eta+\varepsilon)+\lambda(\zeta+\varepsilon). \end{aligned} \tag{17.6}$$

Letting $\alpha \downarrow 0$ and then $\varepsilon \downarrow 0$, we deduce that $f'(x;(1-\lambda)y+\lambda z) \leqslant (1-\lambda)\eta+\lambda\zeta$. Therefore, $f'(x;\cdot)$ is convex.

(v): This follows from (i) and (iv).

(vi): There exists $\beta \in \mathbb{R}_{++}$ such that $[x-\beta y, x+\beta y] \subset \operatorname{dom} f$. Now, take $\alpha \in \left]0,\beta\right]$. Then, by (8.1), $f(x) \leqslant (f(x-\alpha y)+f(x+\alpha y))/2$ and thus, appealing to Proposition 9.27, we obtain

$$\begin{aligned} -\left(\frac{f(x-\beta y)-f(x)}{\beta}\right) &\leqslant -\left(\frac{f(x-\alpha y)-f(x)}{\alpha}\right) \\ &= \frac{f(x)-f(x-\alpha y)}{\alpha} \end{aligned}$$

$$\leqslant \frac{f(x+\alpha y)-f(x)}{\alpha}$$
$$\leqslant \frac{f(x+\beta y)-f(x)}{\beta}. \tag{17.7}$$

Letting $\alpha \downarrow 0$, we deduce that

$$\frac{f(x)-f(x-\beta y)}{\beta} \leqslant -f'(x;-y) \leqslant f'(x;y) \leqslant \frac{f(x+\beta y)-f(x)}{\beta}. \tag{17.8}$$

Since the leftmost and the rightmost terms in (17.8) are in $\mathbb{R}$, so are the middle terms. Therefore, $f'(x;\cdot)$ is real-valued on $\mathcal{H}$. ☐

Proposition 17.3 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex, and let $x \in \operatorname{dom} f$. Then $x \in \operatorname{Argmin} f \Leftrightarrow f'(x;\cdot) \geqslant 0$.*

Proof. Let $y \in \mathcal{H}$. We have $x \in \operatorname{Argmin} f \Rightarrow (\forall \alpha \in \mathbb{R}_{++})\ (f(x+\alpha y) - f(x))/\alpha \geqslant 0 \Rightarrow f'(x;y) \geqslant 0$. Conversely, suppose that $f'(x;\cdot) \geqslant 0$. Then, by Proposition 17.2(ii), $f(x) \leqslant f'(x;y-x) + f(x) \leqslant f(y)$ and, therefore, $x \in \operatorname{Argmin} f$. ☐

Let $x \in \operatorname{dom} f$ and suppose that $f'(x;\cdot)$ is linear and continuous on $\mathcal{H}$. Then, much like in Definition 2.54 and Remark 2.55 (which concern real-valued functions, while f maps to $\left]-\infty,+\infty\right]$), f is said to be Gâteaux differentiable at x and, by Riesz–Fréchet representation (Fact 2.24), there exists a unique vector $\nabla f(x) \in \mathcal{H}$ such that

$$(\forall y \in \mathcal{H}) \quad f'(x;y) = \langle y \mid \nabla f(x)\rangle, \tag{17.9}$$

namely the Gâteaux gradient of f at x. Alternatively, as in (2.39),

$$(\forall y \in \mathcal{H}) \quad \langle y \mid \nabla f(x)\rangle = \lim_{0 \neq \alpha \to 0} \frac{f(x+\alpha y)-f(x)}{\alpha}. \tag{17.10}$$

In a similar fashion, $\nabla^2 f(x)$, if it exists, is defined as in Remark 2.55. Furthermore, by analogy with Definition 2.56, if the convergence in (17.10) is uniform with respect to y on bounded sets, then $\nabla f(x)$ is called the Fréchet gradient of f at x; equivalently,

$$\lim_{0 \neq y \to 0} \frac{f(x+y)-f(x)-\langle y \mid \nabla f(x)\rangle}{\|y\|} = 0. \tag{17.11}$$

Lemma 2.61(ii) implies that, if f is Fréchet differentiable at $x \in \operatorname{int}\operatorname{dom} f$, then it is continuous at x.

Proposition 17.4 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be convex, let $x \in \mathcal{H}$, and suppose that f is Gâteaux differentiable at x. Then $x \in \operatorname{Argmin} f \Leftrightarrow \nabla f(x) = 0$.*

Proof. Proposition 17.3 and (17.9) yield $x \in \operatorname{Argmin} f \Leftrightarrow (\forall y \in \mathcal{H})\ \langle y \mid \nabla f(x)\rangle = f'(x;y) \geqslant 0 \Leftrightarrow \nabla f(x) = 0$. ☐

Proposition 17.5 *Let $f \in \Gamma_0(\mathcal{H})$, let $\gamma \in \mathbb{R}_{++}$, and set*

$$C = \big\{x \in \mathcal{H} \;\big|\; f(x) = {}^{\gamma}\!f(x)\big\}. \tag{17.12}$$

Then $\operatorname{Fix}\operatorname{Prox}_{\gamma f} = \operatorname{zer}\operatorname{Prox}_{f^*} = C = \operatorname{Argmin}{}^{\gamma}\!f = \operatorname{Argmin} f = \partial f^*(0)$.

Proof. Since $\operatorname{Fix}\operatorname{Prox}_f = \operatorname{Argmin} f$ by Proposition 12.29, $\operatorname{Fix}\operatorname{Prox}_{\gamma f} = \operatorname{Argmin}(\gamma f) = \operatorname{Argmin} f = \operatorname{Fix}\operatorname{Prox}_f$. On the other hand, $\operatorname{Fix}\operatorname{Prox}_f = \operatorname{zer}\operatorname{Prox}_{f^*}$ by (14.6) and $\operatorname{Argmin} f = \partial f^*(0)$ by Proposition 16.33. Finally, the identity $\operatorname{Argmin}{}^{\gamma}\!f = \operatorname{Fix}\operatorname{Prox}_{\gamma f}$ follows from Proposition 12.30 and Proposition 17.4, while the identity $\operatorname{Fix}\operatorname{Prox}_{\gamma f} = C$ follows from Remark 12.24 and Definition 12.23. □

Proposition 17.6 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be convex, let $x \in \mathcal{H}$, and suppose that f is Gâteaux differentiable at x. Then $(\forall y \in \mathcal{H})$ $\langle y - x \mid \nabla f(x)\rangle + f(x) \leqslant f(y)$.*

Proof. Combine Proposition 17.2(ii) and (17.9). □

17.2 Characterizations of Convexity

Convexity can be characterized in terms of first- and second-order differentiability properties.

Proposition 17.7 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper. Suppose that* $\operatorname{dom} f$ *is open and convex, and that f is Gâteaux differentiable on* $\operatorname{dom} f$. *Then the following are equivalent:*

(i) *f is convex.*
(ii) $(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)$ $\langle x - y \mid \nabla f(y)\rangle + f(y) \leqslant f(x)$.
(iii) $(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)$ $\langle x - y \mid \nabla f(x) - \nabla f(y)\rangle \geqslant 0$. *In other words, ∇f is* monotone.

Moreover, if f is twice Gâteaux differentiable on $\operatorname{dom} f$, *then each of the above is equivalent to*

(iv) $(\forall x \in \operatorname{dom} f)(\forall z \in \mathcal{H})$ $\big\langle z \mid \nabla^2 f(x) z\big\rangle \geqslant 0$.

Proof. Fix $x \in \operatorname{dom} f$, $y \in \operatorname{dom} f$, and $z \in \mathcal{H}$. Since $\operatorname{dom} f$ is open, there exists $\varepsilon \in \mathbb{R}_{++}$ such that $x + \varepsilon(x - y) \in \operatorname{dom} f$ and $y + \varepsilon(y - x) \in \operatorname{dom} f$. Furthermore, set $C = \left]-\varepsilon, 1+\varepsilon\right[$ and

$$\phi\colon \mathbb{R} \to \left]-\infty,+\infty\right] : \alpha \mapsto f\big(y + \alpha(x - y)\big) + \iota_C(\alpha). \tag{17.13}$$

Then ϕ is Gâteaux differentiable on C and

$$(\forall \alpha \in C) \quad \phi'(\alpha) = \big\langle x - y \mid \nabla f\big(y + \alpha(x - y)\big)\big\rangle. \tag{17.14}$$

(i)⇒(ii): Proposition 17.6.

(ii)⇒(iii): It follows from (ii) that $\langle x - y \mid \nabla f(y)\rangle + f(y) \leqslant f(x)$ and $\langle y - x \mid \nabla f(x)\rangle + f(x) \leqslant f(y)$. Adding up these two inequalities, we obtain $\langle x - y \mid \nabla f(x) - \nabla f(y)\rangle \geqslant 0$.

(iii)⇒(i): Take α and β in C such that $\alpha < \beta$, and set $y_\alpha = y + \alpha(x-y)$ and $y_\beta = y + \beta(x-y)$. Then (iii) and (17.14) imply that $\phi'(\beta) - \phi'(\alpha) = \langle y_\beta - y_\alpha \mid \nabla f(y_\beta) - \nabla f(y_\alpha)\rangle / (\beta - \alpha) \geqslant 0$. Consequently, ϕ' is increasing on C and ϕ is therefore convex by Proposition 8.14(i). In particular,

$$f\big(\alpha x + (1-\alpha)y\big) = \phi(\alpha) \leqslant \alpha\phi(1) + (1-\alpha)\phi(0) = \alpha f(x) + (1-\alpha) f(y). \tag{17.15}$$

(iii)⇒(iv): Let $z \in \mathcal{H}$. Since $\operatorname{dom} f$ is open, for $\alpha \in \mathbb{R}_{++}$ small enough, $x + \alpha z \in \operatorname{dom} f$, and (iii) yields

$$\begin{aligned}\langle z \mid \nabla f(x+\alpha z) - \nabla f(x)\rangle &= \frac{1}{\alpha}\,\langle (x+\alpha z) - x \mid \nabla f(x+\alpha z) - \nabla f(x)\rangle \\ &\geqslant 0.\end{aligned} \tag{17.16}$$

In view of (2.38), dividing by α and letting $\alpha \downarrow 0$, we obtain $\big\langle z \mid \nabla^2 f(x) z\big\rangle \geqslant 0$.

(iv)⇒(i): Note that ϕ is twice Gâteaux differentiable on C with $(\forall \alpha \in C)$ $\phi''(\alpha) = \big\langle x - y \mid \nabla^2 f\big(y + \alpha(x-y)\big)(x-y)\big\rangle \geqslant 0$. Hence, ϕ' is increasing on C and, by Proposition 8.14(i), ϕ is convex. We conclude with (17.15). ☐

Example 17.8 Let $A \in \mathcal{B}(\mathcal{H})$ and set $f\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \langle x \mid Ax\rangle$. Then f is convex if and only if $(\forall x \in \mathcal{H})$ $\langle x \mid (A + A^*)x\rangle \geqslant 0$.

Proof. Combine Example 2.57 with Proposition 17.7(iv). ☐

Proposition 17.9 *Let* $h\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ *be such that* $D = \operatorname{dom} h$ *is nonempty, open, and convex. Suppose that* h *is Fréchet differentiable on* D *and that one of the following holds:*

(i) $(\forall x \in D)(\forall y \in D)$ $\langle x - y \mid \nabla h(x) - \nabla h(y)\rangle \geqslant 0$.
(ii) h *is twice Fréchet differentiable on* D *and* $(\forall x \in D)(\forall z \in \mathcal{H})$ $\big\langle z \mid \nabla^2 h(x) z\big\rangle \geqslant 0$.

Take $y \in D$ *and set*

$$f\colon x \mapsto \begin{cases} h(x), & \text{if } x \in D;\\ \lim\limits_{\alpha\downarrow 0} h\big((1-\alpha)x + \alpha y\big), & \text{if } x \in \operatorname{bdry} D;\\ +\infty, & \text{if } x \notin \overline{D}.\end{cases} \tag{17.17}$$

Then $f \in \Gamma_0(\mathcal{H})$, $D \subset \operatorname{dom} f \subset \overline{D}$, *and* $f|_D = h|_D$.

Proof. Proposition 17.7 guarantees that h is convex. Moreover, since h is Fréchet differentiable on D, it is continuous on D by Lemma 2.61(ii). Hence, the result follows from Corollary 16.19. ☐

17.3 Characterizations of Strict Convexity

Some of the results of Section 17.2 have counterparts for strictly convex functions. However, as Example 17.13 will illustrate, subtle differences exist. The proofs of the following results are left as Exercise 17.6, Exercise 17.11, and Exercise 17.14.

Proposition 17.10 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper. Suppose that* $\operatorname{dom} f$ *is open and convex, and that f is Gâteaux differentiable on* $\operatorname{dom} f$. *Then the following are equivalent:*

(i) *f is strictly convex.*
(ii) $(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)$ $x \neq y \Rightarrow \langle x - y \mid \nabla f(y)\rangle + f(y) < f(x)$.
(iii) $(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)$ $x \neq y \Rightarrow \langle x - y \mid \nabla f(x) - \nabla f(y)\rangle > 0$.

Moreover, if f is twice Gâteaux differentiable on $\operatorname{dom} f$, *then each of the above is implied by*

(iv) $(\forall x \in \operatorname{dom} f)(\forall z \in \mathcal{H} \smallsetminus \{0\})$ $\big\langle z \mid \nabla^2 f(x) z\big\rangle > 0$.

Example 17.11 Let $A \in \mathcal{B}(\mathcal{H})$ and set $f\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \langle x \mid Ax\rangle$. Then f is strictly convex if and only if $A + A^*$ is strictly monotone.

Proof. Combine Example 2.57 with Proposition 17.10. □

Corollary 17.12, which complements Proposition 9.34, provides a convenient sufficient condition for strict convexity (see Exercise 17.13); extensions to $\mathbb{R}^N$ can be obtained as described in Remark 9.37.

Corollary 17.12 *Let $h\colon \mathbb{R} \to \left]-\infty,+\infty\right]$ be such that $D = \operatorname{dom} h$ is a nonempty open interval. Suppose that h is twice differentiable on D and such that, for every $x \in D$, $h''(x) > 0$. Take $y \in D$ and set*

$$f\colon x \mapsto \begin{cases} h(x), & \text{if } x \in D;\\[1ex] \lim\limits_{\alpha \downarrow 0} h\big((1-\alpha)x + \alpha y\big), & \text{if } x \in \operatorname{bdry} D;\\[1ex] +\infty, & \text{if } x \notin \overline{D}. \end{cases} \tag{17.18}$$

Then $f \in \Gamma_0(\mathbb{R})$, $f|_D = h|_D$, $D \subset \operatorname{dom} f \subset \overline{D}$, and f is strictly convex.

Corollary 17.12 and the following example illustrate the absence of a strictly convex counterpart to Proposition 17.9, even in the Euclidean plane.

Example 17.13 The function

$$h\colon \mathbb{R}^2 \to \left]-\infty,+\infty\right] : (\xi,\eta) \mapsto \begin{cases} \eta^2 + \eta^2/\xi, & \text{if } \xi > 0 \text{ and } \eta > 0;\\ +\infty, & \text{otherwise} \end{cases} \tag{17.19}$$

is twice Fréchet differentiable on its domain $\mathbb{R}^2_{++}$. Now let f be as in (17.17). Then

$$f\colon \mathbb{R}^2 \to \left]-\infty,+\infty\right] : (\xi,\eta) \mapsto \begin{cases} \eta^2 + \eta^2/\xi, & \text{if } \xi > 0 \text{ and } \eta \geqslant 0;\\ 0, & \text{if } \xi = \eta = 0;\\ +\infty, & \text{otherwise} \end{cases} \tag{17.20}$$

belongs to $\Gamma_0(\mathbb{R}^2)$ but f is not strictly convex.

17.4 Directional Derivatives and Subgradients

Proposition 17.14 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex, let $x \in \operatorname{dom} f$, and let $u \in \mathcal{H}$. Then the following hold:*

(i) $u \in \partial f(x) \Leftrightarrow \langle \cdot \mid u\rangle \leqslant f'(x;\cdot)$.
(ii) *Suppose that $x \in \operatorname{dom}\partial f$. Then $f'(x;\cdot)$ is proper and sublinear.*

Proof. (i): Let $\alpha \in \mathbb{R}_{++}$. Then (16.1) yields $u \in \partial f(x) \Rightarrow (\forall y \in \mathcal{H})$ $\langle y \mid u\rangle = \langle (x+\alpha y) - x \mid u\rangle/\alpha \leqslant (f(x+\alpha y) - f(x))/\alpha$. Taking the limit as $\alpha \downarrow 0$, we obtain $(\forall y \in \mathcal{H})$ $\langle y \mid u\rangle \leqslant f'(x;y)$. Conversely, it follows from Proposition 17.2(ii) and (16.1) that $(\forall y \in \mathcal{H})$ $\langle y - x \mid u\rangle \leqslant f'(x; y-x) \Rightarrow$ $(\forall y \in \mathcal{H})$ $\langle y - x \mid u\rangle \leqslant f(y) - f(x) \Rightarrow u \in \partial f(x)$.

(ii): Suppose that $u \in \partial f(x)$. Then (i) yields $f'(x;\cdot) \geqslant \langle\cdot \mid u\rangle$ and therefore $-\infty \notin f'(x;\mathcal{H})$. Hence, in view of Proposition 17.2(iv), $f'(x;\cdot)$ is proper and sublinear. ☐

Corollary 17.15 *Let C and D be convex subsets of $\mathcal{H}$. Suppose that $x \in C\cap D$ and that there exists $\varepsilon \in \mathbb{R}_{++}$ such that $C \cap B(x;\varepsilon) = D\cap B(x;\varepsilon)$. Then $N_C x = N_D x$ and $T_C x = T_D x$.*

Proof. Let $u \in \mathcal{H}$. Using Proposition 17.14(i) and Example 16.13, we obtain the equivalences $u \in N_C x \Leftrightarrow u \in \partial\iota_C(x) \Leftrightarrow \langle u \mid \cdot\rangle \leqslant \iota'_C(x;\cdot) \Leftrightarrow \langle u \mid \cdot\rangle \leqslant \iota'_D(x;\cdot) \Leftrightarrow u \in \partial\iota_D(x) \Leftrightarrow u \in N_C x$. Hence $N_C x = N_D x$ and therefore, by Proposition 6.44(i), $T_C x = N_C^{\ominus} x = N_D^{\ominus} x = T_D x$. ☐

Proposition 17.16 *Let $f\colon \mathbb{R} \to \left]-\infty,+\infty\right]$ be proper and convex, let $x \in \operatorname{dom} f$, and let $y \in \operatorname{dom} f$. Then the following hold:*

(i) $f'_-(x) \leqslant f'_+(x)$.
(ii) $\partial f(x) = [f'_-(x), f'_+(x)] \cap \mathbb{R}$.
(iii) *Suppose that $x < y$. Then $f'_+(x) \leqslant f'_-(y)$.*
(iv) *f'_- and f'_+ are increasing on* $\operatorname{dom} f$.

Proof. (i): Let $\alpha \in \mathbb{R}_{++}$. By convexity of f,

$$f(x) = f\Big(\tfrac{1}{2}(x-\alpha) + \tfrac{1}{2}(x+\alpha)\Big) \leqslant \frac{f(x-\alpha) + f(x+\alpha)}{2} \tag{17.21}$$

and therefore $-(f(x-\alpha) - f(x))/\alpha \leqslant (f(x+\alpha) - f(x))/\alpha$. Now take the limit as $\alpha \downarrow 0$.

(ii): Let $u \in \mathbb{R}$. Then, by Proposition 17.14, $u \in \partial f(x) \Leftrightarrow \pm u \leqslant f'(x;\pm 1) \Leftrightarrow [u \leqslant f'_+(x) \text{ and } -u \leqslant -f'_-(x)] \Leftrightarrow f'_-(x) \leqslant u \leqslant f'_+(x)$.

(iii): It follows from Proposition 17.2(iii)&(iv) that

$$f'_+(x) = f'(x;1) = \frac{f'(x;y-x)}{y-x} \leqslant \frac{-f'(y;x-y)}{y-x} = -f'(y;-1) = f'_-(y), \tag{17.22}$$

as claimed.

(iv): Combine (i) and (iii). ☐

Proposition 17.17 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex, and let $x \in \operatorname{dom} f$. Then $(f'(x;\cdot))^* = \iota_{\partial f(x)}$.*

Proof. Define $\varphi\colon \mathcal{H} \to [-\infty,+\infty] : y \mapsto f'(x;y)$ and let $u \in \mathcal{H}$. Then, using Proposition 17.2(i) and (13.1), we obtain

$$\begin{aligned}\varphi^*(u) &= \sup_{\alpha\in\mathbb{R}_{++}} \sup_{y\in\mathcal{H}} \left(\langle y \mid u\rangle - \frac{f(x+\alpha y)-f(x)}{\alpha} \right) \\ &= \sup_{\alpha\in\mathbb{R}_{++}} \frac{f(x) + \sup_{y\in\mathcal{H}} \big(\langle x+\alpha y \mid u\rangle - f(x+\alpha y)\big) - \langle x \mid u\rangle}{\alpha} \\ &= \sup_{\alpha\in\mathbb{R}_{++}} \frac{f(x) + f^*(u) - \langle x \mid u\rangle}{\alpha}. \end{aligned} \tag{17.23}$$

However, Proposition 16.10 asserts that $u \in \partial f(x) \Leftrightarrow f(x) + f^*(u) - \langle x \mid u\rangle = 0$. Hence, $u \in \partial f(x) \Rightarrow \varphi^*(u) = 0$ and, moreover, Fenchel–Young (Proposition 13.15) yields $u \notin \partial f(x) \Rightarrow f(x) + f^*(u) - \langle x \mid u\rangle > 0 \Rightarrow \varphi^*(u) = +\infty$. Altogether, $\varphi^* = \iota_{\partial f(x)}$. ☐

Theorem 17.18 (Max formula) *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex, and suppose that $x \in \operatorname{cont} f$. Then $f'(x;\cdot) = \max\langle \cdot \mid \partial f(x)\rangle$.*

Proof. Define $\varphi\colon \mathcal{H} \to \left]-\infty,+\infty\right] : y \mapsto f'(x;y)$ and fix $y \in \mathcal{H}$. Since $x \in \operatorname{cont} f$, Theorem 8.38 asserts that there exist $\rho \in \mathbb{R}_{++}$ and $\beta \in \mathbb{R}_{++}$ such that f is β-Lipschitz continuous relative to $B(x;\rho)$. Hence, for $\alpha \in \mathbb{R}_{++}$ small enough to ensure that $x+\alpha y \in B(x;\rho)$, we have $|f(x+\alpha y)-f(x)|/\alpha \leqslant \beta\|y\|$. Taking the limit as $\alpha \downarrow 0$, we obtain $|\varphi(y)| \leqslant \beta\|y\|$. Consequently, φ is locally bounded at every point in $\mathcal{H}$ and $\operatorname{dom}\varphi = \mathcal{H}$. Moreover, φ is convex by Proposition 17.2(v). We thus deduce from Corollary 8.39(i) that φ is continuous. Altogether, $\varphi \in \Gamma_0(\mathcal{H})$ and it follows from Theorem 13.37 and Proposition 17.17 that

$$\varphi(y) = \varphi^{**}(y) = \sup_{u\in\operatorname{dom}\varphi^*} \big(\langle u \mid y\rangle - \varphi^*(u)\big) = \sup_{u\in\partial f(x)} \langle y \mid u\rangle. \tag{17.24}$$

However, $\partial f(x)$ is nonempty and weakly compact by Proposition 16.17(ii). On the other hand, $\langle y \mid \cdot\rangle$ is weakly continuous. Hence, we derive from Theorem 1.29 that $\varphi(y) = \max_{u\in\partial f(x)} \langle y \mid u\rangle$. ☐

We now characterize Lipschitz continuity.

Corollary 17.19 *Let $f\colon \mathcal{H} \to \mathbb{R}$ be lower semicontinuous and convex, and let $\beta \in \mathbb{R}_+$. Then the following are equivalent:*

(i) *f is Lipschitz continuous with constant β.*
(ii) $\operatorname{ran} \partial f \subset B(0;\beta)$.
(iii) $\operatorname{dom} f^* \subset B(0;\beta)$.
(iv) $\operatorname{rec} f \leqslant \beta\|\cdot\|$.

Proof. Let $x \in \mathcal{H}$.

(i)⇒(ii): $(\forall y \in \mathcal{H})(\forall \alpha \in \mathbb{R}_{++})$ $(f(x+\alpha y) - f(x))/\alpha \leqslant \beta\|y\|$. Hence $(\forall y \in \mathcal{H})$ $f'(x;y) \leqslant \beta\|y\|$. Thus, using Theorem 17.18, $(\forall y \in B(0;1))$ $\max \langle y \mid \partial f(x)\rangle \leqslant \beta\|y\|$. It follows that $\sup \|\partial f(x)\| \leqslant \beta$.

(ii)⇒(i): Corollary 16.57.

(ii)⇔(iii): By Corollary 13.38, $f^* \in \Gamma_0(\mathcal{H})$. Hence, the equivalence is true because $\overline{\operatorname{dom}}\, f^* = \overline{\operatorname{dom}}\, \partial f^* = \overline{\operatorname{ran}}\, \partial f$ by Corollary 16.39 and Corollary 16.30.

(i)⇒(iv): $(\forall y \in \mathcal{H})(\forall \alpha \in \mathbb{R}_{++})$ $(f(x+\alpha y) - f(x))/\alpha \leqslant \beta\|y\|$. By Proposition 9.30(ii), $(\operatorname{rec} f)(y) \leqslant \beta\|y\|$.

(iv)⇒(i): $(\forall y \in \mathcal{H})$ $\beta\|y\| \geqslant (\operatorname{rec} f)(y) \geqslant f(x+y) - f(x)$ by Proposition 9.30(iv). □

Definition 17.20 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex, let $x \in \operatorname{dom} f$, and let $y \in \mathcal{H}$. Then y is a *descent direction* of f at x if there exists $\varepsilon \in \mathbb{R}_{++}$ such that $(\forall \alpha \in \left]0,\varepsilon\right])$ $f(x+\alpha y) < f(x)$.

Proposition 17.21 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex, let $x \in \operatorname{dom} f$, and let $y \in \mathcal{H}$. Then the following hold:*

(i) *y is a descent direction of f at x if and only if $f'(x;y) < 0$.*
(ii) *Suppose that f is differentiable at x and that $x \notin \operatorname{Argmin} f$. Then $-\nabla f(x)$ is a descent direction of f at x.*

Proof. (i): By Proposition 9.27 and Proposition 17.2(i), $(f(x+\alpha y)-f(x))/\alpha \downarrow f'(x;y)$ as $\alpha \downarrow 0$.

(ii): By (17.9) and Proposition 17.4, $f'(x;-\nabla f(x)) = -\|\nabla f(x)\|^2 < 0$. □

Proposition 17.22 (Steepest descent direction) *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex, and suppose that $x \in (\operatorname{cont} f) \smallsetminus (\operatorname{Argmin} f)$. Set $u = P_{\partial f(x)} 0$ and $z = -u/\|u\|$. Then z is the unique minimizer of $f'(x;\cdot)$ over $B(0;1)$.*

Proof. Proposition 16.17(ii), Proposition 16.4(iii), and Theorem 16.3 imply that $\partial f(x)$ is a nonempty closed convex set that does not contain 0. Hence $u \neq 0$ and Theorem 3.16 yields $\max \langle -u \mid \partial f(x) - u\rangle = 0$, i.e., $\max \langle -u \mid \partial f(x)\rangle = -\|u\|^2$. Using Theorem 17.18, we deduce that

$$f'(x;z) = \max \langle z \mid \partial f(x)\rangle = -\|u\| \tag{17.25}$$

and that

$$\big(\forall y \in B(0;1)\big) \quad f'(x;y) = \max \langle y \mid \partial f(x)\rangle \geqslant \langle y \mid u\rangle \geqslant -\|u\|. \tag{17.26}$$

Therefore, z is a minimizer of $f'(x;\cdot)$ over $B(0;1)$ and uniqueness follows from Fact 2.11. ☐

Example 17.23 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex, and suppose that $x \in \operatorname{dom} f \smallsetminus \operatorname{Argmin} f$ and that $u \in \partial f(x)$. If f is not differentiable at x, then $-u$ is not necessarily a descent direction. For instance, take $\mathcal{H} = \mathbb{R}^2$, set $f\colon (\xi_1,\xi_2) \mapsto |\xi_1| + 2|\xi_2|$, set $x = (1,0)$, and set $u = (1,\pm\delta)$, where $\delta \in [1/2, 3/2]$. Then one easily checks that $u \in \partial f(x)$ and $(\forall \alpha \in \mathbb{R}_{++})$ $f(x - \alpha u) \geqslant f(x)$.

Proposition 17.24 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex. Suppose that $x \in \operatorname{cont} f$ and that C is a nonempty subset of $\mathcal{H}$ such that $f'(x;\cdot) = \sigma_C$. Then $\partial f(x) = \overline{\operatorname{conv}}\, C$.*

Proof. In view of Theorem 17.18, we have $\sigma_C = \sigma_{\partial f(x)}$. Taking conjugates and using Example 13.43(i) and Proposition 16.4(iii), we obtain $\iota_{\overline{\operatorname{conv}}\, C} = \sigma_C^* = \sigma_{\partial f(x)}^* = \iota_{\overline{\operatorname{conv}}\, \partial f(x)} = \iota_{\partial f(x)}$. ☐

Proposition 17.25 (Chebyshev center) *Let C be a nonempty compact subset of $\mathcal{H}$, set $f\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \max_{y\in C} \|x-y\|^2$, and set*

$$\Phi_C\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x \mapsto \Big\{ r \in C \;\Big|\; \|x - r\| = \max_{y\in C} \|x-y\| \Big\}. \tag{17.27}$$

Then the following hold:

(i) *f is continuous, strongly convex, and supercoercive.*
(ii) $\operatorname{dom} \Phi_C = \mathcal{H}$, $\operatorname{gra} \Phi_C$ *is closed, and the sets* $(\Phi_C(x))_{x\in\mathcal{H}}$ *are compact.*
(iii) $(\forall x \in \mathcal{H})(\forall z \in \mathcal{H})$ $f'(x;z) = 2 \max \langle z \mid x - \Phi_C(x)\rangle$.
(iv) $(\forall x \in \mathcal{H})$ $\partial f(x) = 2(x - \overline{\operatorname{conv}}\, \Phi_C(x))$.
(v) *The function f has a unique minimizer r, called the* Chebyshev center of C, *and characterized by*

$$r \in \overline{\operatorname{conv}}\, \Phi_C(r). \tag{17.28}$$

Proof. (i): Since the functions $(\|\cdot - y\|^2)_{y\in C}$ are convex and lower semicontinuous, Proposition 9.3 implies that f is likewise. It follows from Corollary 8.39(ii) that f is continuous. Arguing similarly, we note that

$$g\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \max_{y\in C} \big(\|y\|^2 - 2\langle x \mid y\rangle \big) \tag{17.29}$$

is convex and continuous, and that $f = g + \|\cdot\|^2$ is strongly convex, hence supercoercive by Corollary 11.17.

(ii): It is clear that $\operatorname{dom} \Phi_C = \mathcal{H}$. Now let $(x_n, y_n)_{n\in\mathbb{N}}$ be a sequence in $\operatorname{gra} \Phi_C$ converging to $(x, y) \in \mathcal{H} \times \mathcal{H}$. Then $y \in C$ by closedness of C. Since f is continuous, we have $f(x) \leftarrow f(x_n) = \|x_n - y_n\|^2 \to \|x - y\|^2$. Thus, $y \in \Phi_C(x)$ and hence $\operatorname{gra} \Phi_C$ is closed. Therefore, $\Phi_C(x)$ is compact.

(iii): Let x and z be in $\mathcal{H}$, let $y \in \Phi_C(x)$, and let $t \in \mathbb{R}_{++}$. Then $f(x) = \|x-y\|^2$ and $f(x+tz) \geqslant \|x+tz-y\|^2$. Hence, $(f(x+tz)-f(x))/t \geqslant t\|x-y\|^2 +2\langle z \mid x-y\rangle$. This implies that $f'(x; z) \geqslant 2\langle z \mid x-y\rangle$ and, furthermore, that

$$f'(x; z) \geqslant 2 \max \langle z \mid x - \Phi_C(x)\rangle . \tag{17.30}$$

To establish the reverse inequality, let $(t_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}_{++}$ such that $t_n \to 0$. For every $n \in \mathbb{N}$, set $x_n = x + t_n z$, and let $r_n \in \Phi_C(x_n)$. Then $x_n \to x$. Due to the compactness of C, after passing to a subsequence and relabeling if necessary, we assume that there exists $r \in C$ such that $r_n \to r$. For every $n \in \mathbb{N}$, since $f(x_n) = \|x_n - r_n\|^2$ and $f(x) \geqslant \|x - r_n\|^2$, we have $(f(x_n) - f(x))/t_n \leqslant (\|x_n\|^2 - \|x\|^2)/t_n - 2\langle z \mid r_n\rangle$. Taking the limit as $n \to +\infty$, we deduce that

$$f'(x; z) \leqslant 2\langle z \mid x - r\rangle \leqslant 2 \max \langle z \mid x - \Phi_C(x)\rangle . \tag{17.31}$$

Combining (17.30) and (17.31), we obtain (iii).

(iv): Combine (iii) and Proposition 17.24.

(v): In view of (i) and Corollary 11.17, f has a unique minimizer over $\mathcal{H}$, say r. Finally, Theorem 16.3 and (iv) yield the characterization (17.28). ☐

Proposition 17.26 *Let $f \in \Gamma_0(\mathcal{H})$ be uniformly convex with modulus ϕ. Then the following hold:*

(i) $(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)\ f'(x; y - x) + \phi(\|y - x\|) + f(x) \leqslant f(y)$.
(ii) *f is supercoercive.*
(iii) *f has exactly one minimizer over $\mathcal{H}$.*

Proof. (i): Let $x \in \operatorname{dom} f$ and let $y \in \operatorname{dom} f$. Then (10.2) yields

$$\big(\forall \alpha \in \left]0,1\right[\big) \quad \frac{f\big(x + \alpha(y-x)\big) - f(x)}{\alpha} + (1-\alpha)\phi\big(\|y-x\|\big) \leqslant f(y) - f(x). \tag{17.32}$$

The result follows by taking the limit as $\alpha \downarrow 0$.

(ii): We assume that $\operatorname{dom} f$ is unbounded. By Remark 16.2, $\operatorname{dom} \partial f \neq \varnothing$. Now let $x \in \operatorname{dom} \partial f$, let $u \in \partial f(x)$, let $y \in \operatorname{dom} f \smallsetminus \{0\}$ be such that $\|y - x\| > 1$, and let φ be the exact modulus of convexity of f. Then it follows from (i), Proposition 10.12, and Proposition 17.14(i) that

$$-\|u\| + \|y-x\|\varphi(1) + \frac{f(x)}{\|y-x\|} \leqslant -\|u\| + \frac{\varphi\big(\|y-x\|\big)}{\|y-x\|} + \frac{f(x)}{\|y-x\|}$$

$$\leqslant \frac{\langle y - x \mid u\rangle + \varphi\big(\|y-x\|\big) + f(x)}{\|y-x\|}$$
$$\leqslant \frac{\|y\|}{\|y-x\|}\frac{f(y)}{\|y\|}. \tag{17.33}$$

Since Corollary 10.13 asserts that $\varphi(1) > 0$, it follows that $f(y)/\|y\| \to +\infty$ as $\|y\| \to +\infty$.

(iii): This follows from (ii) and Corollary 11.16. ☐

17.5 Directional Derivatives and Convexity

In this section, we provide a sufficient condition for convexity which complements Proposition 17.16.

Proposition 17.27 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper, let $x_0 \in \operatorname{dom} f$, and let $x_1 \in \mathcal{H}$ be such that $[x_0, x_1[\, \subset \operatorname{dom} f$. Suppose that $f|_{]x_0,x_1]}$ is lower semicontinuous and that, for every $x \in [x_0, x_1[$, $f'(x; x_1 - x_0)$ exists and $f'(x; x_1 - x_0) \leqslant 0$. Then $f(x_1) \leqslant f(x_0)$.*

Proof. Set $(\forall t \in \mathbb{R})\ x_t = x_0 + t(x_1 - x_0)$, set $g\colon \mathbb{R} \to \left]-\infty,+\infty\right] : t \mapsto f(x_t) - f(x_0)$, and set $I = \left]0,1\right]$. Then $g|_I$ is lower semicontinuous and

$$\big(\forall t \in [0,1[\,\big) \quad g'_+(t) = g'(t;1) = f'(x_t; x_1 - x_0) \leqslant 0. \tag{17.34}$$

Let $\varepsilon \in \mathbb{R}_{++}$. Since $g(0) = 0$, (17.34) yields $\lim_{t\downarrow 0} g(t)/t = g'_+(0) \leqslant 0$. Hence,

$$t_1 = \sup\left\{ t \in \left]0,1\right[\;\middle|\; \sup_{s \in]0,t]} \frac{g(s)}{s} \leqslant \varepsilon \right\} \in \left]0,1\right], \tag{17.35}$$

and therefore

$$\big(\forall t \in [0, t_1[\,\big) \quad g(t) \leqslant \varepsilon t. \tag{17.36}$$

Now assume that $t_1 < 1$. It follows from the lower semicontinuity of g at t_1 and (17.36) that $g(t_1) \leqslant \varepsilon t_1$. Hence, since (17.34) yields $g'_+(t_1) \leqslant 0 < \varepsilon$, there exists $t_2 \in \left]t_1, 1\right[$ such that $(\forall t \in \left]t_1, t_2\right])\ (g(t) - g(t_1))/(t - t_1) \leqslant \varepsilon$ and hence $(\forall t \in \left]t_1, t_2\right])\ g(t) \leqslant g(t_1) + \varepsilon(t - t_1) \leqslant \varepsilon t$, which contradicts the definition of t_1. Therefore, $t_1 = 1$ and (17.36) yields

$$(\forall t \in I) \quad g(t) \leqslant \varepsilon t. \tag{17.37}$$

Letting $\varepsilon \downarrow 0$, we obtain $(\forall t \in I)\ g(t) \leqslant 0$. Finally, the lower semicontinuity of $g|_I$ at 1 yields $f(x_1) - f(x_0) = g(1) \leqslant 0$. ☐

Corollary 17.28 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper, let $x_0 \in \operatorname{dom} f$, and let $x_1 \in \mathcal{H}$. Suppose that $[x_0, x_1[\subset \operatorname{dom} f$, that $f(x_0) < f(x_1)$, that $f|_{]x_0,x_1]}$ is lower semicontinuous, and that, for every $x \in [x_0, x_1[$, $f'(x; x_1 - x_0)$ exists. Then there exists $z \in [x_0, x_1[$ such that $f'(z; x_1 - x_0) > 0$.*

Proposition 17.29 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper, and suppose that $\operatorname{dom} f$ is convex, that $f|_{\operatorname{dom} f}$ is lower semicontinuous, that, for every $x \in \operatorname{dom} f$ and for every $y \in \operatorname{dom} f$, $f'(x; y-x)$ exists and $f'(x; y-x) \leqslant -f'(y; x-y)$. Then f is convex.*

Proof. Suppose that f is not convex. Then there exist a and b in $\operatorname{dom} f$ and $t_0 \in \left]0,1\right[$ such that

$$f\big(t_0 a + (1-t_0)b\big) > t_0 f(a) + (1-t_0) f(b). \tag{17.38}$$

Set

$$g\colon \mathbb{R} \to \left]-\infty,+\infty\right] : t \mapsto f\big(a + t(b-a)\big) - f(a) - t\big(f(b) - f(a)\big). \tag{17.39}$$

Then

$$(\forall t \in [0,1])(\forall s \in \mathbb{R}) \quad g'(t; s) = f'\big(a + t(b-a); s(b-a)\big) - s\big(f(b) - f(a)\big). \tag{17.40}$$

Since $g(0) = g(1) = 0$ and $g(t_0) > 0$, Corollary 17.28 asserts the existence of $t_1 \in [0, t_0[$ and $t_2 \in]t_0, 1]$ such that $g'(t_1; t_0) > 0$ and $g'(t_2; t_0 - 1) > 0$. Since $\mathbb{R}_{++} t_0 = \mathbb{R}_{++}(t_2 - t_1)$ and $\mathbb{R}_{++}(t_0 - 1) = \mathbb{R}_{++}(t_1 - t_2)$, it therefore follows from the positive homogeneity of $g'(t_1; \cdot)$ and $g'(t_2; \cdot)$ that

$$g'(t_1; t_2 - t_1) > 0 \quad \text{and} \quad g'(t_2; t_1 - t_2) > 0. \tag{17.41}$$

Now set $x = a + t_1(b-a)$ and $y = a + t_2(b-a)$. Then (17.40) yields

$$\begin{aligned} f'(x; y-x) &= f'\big(a + t_1(b-a); (t_2 - t_1)(b-a)\big) \\ &= g'(t_1; t_2 - t_1) + (t_2 - t_1)\big(f(b) - f(a)\big) \end{aligned} \tag{17.42}$$

and, similarly,

$$f'(y; x-y) = g'(t_2; t_1 - t_2) + (t_2 - t_1)\big(f(a) - f(b)\big). \tag{17.43}$$

Since, by assumption, $f'(x; y-x) \leqslant -f'(y; x-y)$, we obtain $g'(t_1; t_2 - t_1) \leqslant -g'(t_2; t_1 - t_2)$, which contradicts (17.41). ☐

Corollary 17.30 *Let $f\colon \mathbb{R} \to \left]-\infty,+\infty\right]$ be proper, and suppose that $I = \operatorname{dom} f$ is an interval in $\mathbb{R}$, that $f|_I$ is lower semicontinuous, that, for every $x \in I$ and for every $y \in I$, $f'_+(x)$ and $f'_-(x)$ exist with $f'_+(x) \leqslant f'_-(y)$ whenever $x < y$. Then f is convex.*

17.6 Gâteaux and Fréchet Differentiability

In this section, we explore the relationships between Gâteaux derivatives, Fréchet derivatives, and single-valued subdifferentials.

Proposition 17.31 *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex, and let $x \in \operatorname{dom} f$. Then the following hold:*

(i) *Suppose that f is Gâteaux differentiable at x. Then $\partial f(x) = \{\nabla f(x)\}$.*
(ii) *Suppose that $x \in \operatorname{cont} f$ and that $\partial f(x)$ consists of a single element u. Then f is Gâteaux differentiable at x and $u = \nabla f(x)$.*

Proof. (i): It follows from (16.1) and Proposition 17.6 that $\nabla f(x) \in \partial f(x)$. Now let $u \in \partial f(x)$. By Proposition 17.14(i) and (17.9), $\langle u - \nabla f(x) \mid u\rangle \leqslant f'(x; u - \nabla f(x)) = \langle u - \nabla f(x) \mid \nabla f(x)\rangle$; hence $\|u - \nabla f(x)\|^2 \leqslant 0$. Therefore $u = \nabla f(x)$. Altogether, $\partial f(x) = \{\nabla f(x)\}$.

(ii): Theorem 17.18 and (17.9) yield $(\forall y \in \mathcal{H})\ f'(x; y) = \langle y \mid u\rangle = \langle y \mid \nabla f(x)\rangle$. ☐

Proposition 17.32 *Let $f\colon \mathcal{H} \to \mathbb{R}$ be continuous and convex, and let $\phi\colon \mathbb{R} \to \mathbb{R}$ be convex. Suppose that ϕ is differentiable and increasing on* $\operatorname{ran} f$, *and let $x \in \mathcal{H}$. Then*

$$\partial(\phi \circ f)(x) = \phi'\big(f(x)\big)\partial f(x). \tag{17.44}$$

Proof. This follows from Corollary 16.72 and Proposition 17.31(i). ☐

Example 17.33 Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\phi\colon \mathbb{R} \to \mathbb{R}$ be a convex function which is differentiable on $\mathbb{R}_+$, and let $x \in \mathcal{H}$. Suppose that one of the following holds:

(i) ϕ is increasing on $\mathbb{R}_+$.
(ii) ϕ is even.

Then

$$\partial(\phi \circ d_C)(x) = \begin{cases} \left\{\dfrac{\phi'(d_C(x))}{d_C(x)}(x - P_C x)\right\}, & \text{if } x \notin C;\\ N_C x \cap B(0; \phi'(0)), & \text{if } x \in \operatorname{bdry} C;\\ \{0\}, & \text{if } x \in \operatorname{int} C. \end{cases} \tag{17.45}$$

Proof. (i): This follows from Proposition 17.32 and Example 16.62.

(ii): Combine (i) and Proposition 11.7(ii). ☐

The following result complements Proposition 16.7.

Proposition 17.34 *Consider Proposition 16.7 and suppose in addition that* $\boldsymbol{f}$ *is proper and convex, that* $\boldsymbol{x} = (x_i)_{i\in I} \in \operatorname{cont} \boldsymbol{f}$*, and that, for every* $i \in I$*,* $\boldsymbol{f} \circ R_i$ *is Gâteaux differentiable at* x_i*. Then* $\boldsymbol{f}$ *is Gâteaux differentiable at* $\boldsymbol{x}$ *and*

$$\nabla \boldsymbol{f}(\boldsymbol{x}) = \bigtimes_{i\in I} \nabla(\boldsymbol{f} \circ R_i)(x_i). \tag{17.46}$$

Proof. Since $\boldsymbol{x} \in \operatorname{cont} \boldsymbol{f}$, Proposition 16.17(ii) implies that $\partial \boldsymbol{f}(\boldsymbol{x}) \neq \varnothing$. In view of Proposition 16.7 and Proposition 17.31(i), we deduce that

$$\partial \boldsymbol{f}(\boldsymbol{x}) = \bigtimes_{i\in I} \big\{\nabla(\boldsymbol{f} \circ R_i)(x_i)\big\}. \tag{17.47}$$

Hence, the conclusion follows from Proposition 17.31(ii). $\square$

Proposition 17.35 *Let* $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ *be proper and convex, let* $x \in \mathcal{H}$*, and suppose that* f *is Gâteaux differentiable at* x*. Then* $f^*(\nabla f(x)) = \langle x \mid \nabla f(x)\rangle - f(x)$*.*

Proof. This follows from Proposition 16.10 and Proposition 17.31(i). $\square$

Proposition 17.36 *Let* $A \in \mathcal{B}(\mathcal{H})$ *be monotone and self-adjoint. Suppose that* $\operatorname{ran} A$ *is closed, set* $q_A\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto (1/2)\langle x \mid Ax\rangle$*, and define* $q_{A^\dagger}$ *likewise. Then the following hold:*

(i) q_A *is convex, continuous, Fréchet differentiable, and* $\nabla q_A = A$*.*
(ii) $q_{A^\dagger} = q_A \circ A^\dagger = \iota_{\ker A} \,\square\, q_A^*$*.*
(iii) $q_A^* = \iota_{\operatorname{ran} A} + q_{A^\dagger}$*.*
(iv) $q_A^* \circ A = q_A$*.*

Proof. (i): Clearly, $\operatorname{dom} q_A = \mathcal{H}$ and q_A is continuous. The convexity of q_A follows from Example 17.8. The Fréchet differentiability and the gradient formula were already observed in Example 2.57.

(ii)&(iii): Take $u \in \mathcal{H}$. Using Exercise 3.12 and Corollary 3.32(ii), we obtain $2q_{A^\dagger}(u) = \big\langle u \mid A^\dagger u\big\rangle = \big\langle u \mid A^\dagger A A^\dagger u\big\rangle = \big\langle A^{\dagger *} u \mid A A^\dagger u\big\rangle = \big\langle A^{*\dagger} u \mid A A^\dagger u\big\rangle = \big\langle A^\dagger u \mid A A^\dagger u\big\rangle = 2q_A(A^\dagger u)$. Hence $q_{A^\dagger} = q_A \circ A^\dagger$ is convex and continuous, and the first identity in (ii) holds.

Let us now verify (iii) since it will be utilized in the proof of the second identity in (ii). Set $V = \ker A$. Then it follows from Fact 2.25(iv) that $V^\perp = \overline{\operatorname{ran}}\, A^* = \operatorname{ran} A$. We assume first that $u \notin \operatorname{ran} A$, i.e., $P_V u \neq 0$. Then $q_A^*(u) \geqslant \sup_{n\in\mathbb{N}}(\langle nP_V u \mid u\rangle - (1/2)\langle nP_V u \mid A(nP_V u)\rangle) = \sup_{n\in\mathbb{N}} n\|P_V u\|^2 = +\infty = (\iota_{\operatorname{ran} A} + q_{A^\dagger})(u)$, as required. Next, assume that $u \in \operatorname{ran} A$, say $u = Az = \nabla q_A(z)$, where $z \in \mathcal{H}$. Since $AA^\dagger = P_{\operatorname{ran} A}$ by Proposition 3.30(ii), it follows from Proposition 17.35 that $q_A^*(u) = \langle z \mid u\rangle - (1/2)\langle z \mid Az\rangle = (1/2)\langle z \mid u\rangle = (1/2)\big\langle z \mid AA^\dagger u\big\rangle = (1/2)\big\langle Az \mid A^\dagger u\big\rangle = (1/2)\big\langle u \mid A^\dagger u\big\rangle = q_{A^\dagger}(u)$. This completes the proof of (iii).

The fact that $\operatorname{dom}(q_A^*)^* = \operatorname{dom} q_A = \mathcal{H}$, Theorem 15.3, Proposition 3.30(v), Corollary 3.32(i), (i), and Corollary 13.38 imply that $\iota_{\ker A} \,\square\, q_A^* = (\iota_{\operatorname{ran} A} + q_A)^* = (\iota_{\operatorname{ran} A^\dagger} + q_{A^{\dagger\dagger}})^* = (q_{A^\dagger}^*)^* = q_{A^\dagger}$.

(iv): In view of (iii), (ii), and Corollary 3.32(i), we have $q_A^* \circ A = \iota_{\operatorname{ran} A} \circ A + q_{A^\dagger} \circ A = q_{A^\dagger} \circ A^{\dagger\dagger} = q_{A^{\dagger\dagger}} = q_A$. ☐

The following example extends Example 16.32, which corresponds to the case $\phi = |\cdot|$.

Example 17.37 Set $f = \phi \circ \|\cdot\|$, where $\phi\colon \mathbb{R} \to \mathbb{R}$ is convex, even, and differentiable on $\mathbb{R} \smallsetminus \{0\}$. Then $\partial\phi(0) = [-\rho, \rho]$ for some $\rho \in \mathbb{R}_+$ and

$$(\forall x \in \mathcal{H}) \quad \partial f(x) = \begin{cases} \left\{ \dfrac{\phi'(\|x\|)}{\|x\|} x \right\}, & \text{if } x \neq 0; \\ B(0;\rho), & \text{if } x = 0. \end{cases} \tag{17.48}$$

Proof. This follows from Example 16.73 and Proposition 17.31(i). If ϕ is differentiable at 0, then (17.48) can also be derived by setting $C = \{0\}$ in Example 17.33(ii). ☐

As the next example shows, the implication in Proposition 17.31(i) is not reversible.

Example 17.38 Let C be the nonempty closed convex set of Example 6.11(iii) and let $u \in \mathcal{H}$. Then

$$0 \in C \subset \operatorname{cone} C = \operatorname{span} C \neq \overline{\operatorname{cone}}\, C = \overline{\operatorname{span}}\, C = \mathcal{H} \tag{17.49}$$

and $u \in \partial\iota_C(0) \Leftrightarrow \sup \langle C \mid u\rangle \leqslant 0 \Leftrightarrow \sup \langle \overline{\operatorname{cone}}\, C \mid u\rangle \leqslant 0 \Leftrightarrow \sup \langle \mathcal{H} \mid u\rangle \leqslant 0 \Leftrightarrow u = 0$. Therefore, $\partial\iota_C(0) = \{0\}$ and hence ι_C possesses a unique subgradient at 0. On the other hand, $0 \notin \operatorname{core} C$ and thus ι_C is not Gâteaux differentiable at 0.

Proposition 17.39 *Let $f \in \Gamma_0(\mathcal{H})$ and suppose that $x \in \operatorname{int\,dom} f$. Then the following are equivalent:*

(i) *f is Gâteaux differentiable at x.*
(ii) *Every selection of ∂f is strong-to-weak continuous at x.*
(iii) *There exists a selection of ∂f that is strong-to-weak continuous at x.*

Proof. (i)⇒(ii): Let G be a selection of ∂f and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\operatorname{int\,dom} f$ converging to x. In view of Fact 1.38 and Lemma 2.30, it is enough to show that $Gx_n \rightharpoonup Gx$. Since Proposition 16.17(iii) implies that $(Gx_n)_{n\in\mathbb{N}}$ is bounded, let $(Gx_{k_n})_{n\in\mathbb{N}}$ be a weakly convergent subsequence of $(Gx_n)_{n\in\mathbb{N}}$, say $Gx_{k_n} \rightharpoonup u$. Then $(x_{k_n}, Gx_{k_n})_{n\in\mathbb{N}}$ lies in $\operatorname{gra} \partial f$ and we derive from Proposition 16.36 that $(x, u) \in \operatorname{gra} \partial f$. By Proposition 17.31(i), $u = \nabla f(x)$ and, using Lemma 2.46, we deduce that $Gx_n \rightharpoonup \nabla f(x)$.

(ii)⇒(iii): This is clear.

(iii)⇒(i): Let G be a selection of ∂f that is strong-to-weak continuous at x, and fix $y \in \mathcal{H}$. Then there exists $\beta \in \mathbb{R}_{++}$ such that $[x, x+\beta y] \subset \operatorname{int\,dom} f$. Take $\alpha \in \,]0, \beta]$. Then $\alpha \langle y \mid Gx\rangle = \langle (x + \alpha y) - x \mid Gx\rangle \leqslant f(x + \alpha y) - f(x)$

and $-\alpha \langle y \mid G(x+\alpha y)\rangle = \langle x-(x+\alpha y) \mid G(x+\alpha y)\rangle \leqslant f(x)-f(x+\alpha y)$. Thus

$$0 \leqslant f(x+\alpha y)-f(x)-\alpha \langle y \mid Gx\rangle \leqslant \alpha \langle y \mid G(x+\alpha y)-Gx\rangle \qquad (17.50)$$

and hence

$$0 \leqslant \frac{f(x+\alpha y)-f(x)}{\alpha} - \langle y \mid Gx\rangle \leqslant \langle y \mid G(x+\alpha y)-Gx\rangle . \qquad (17.51)$$

This implies that

$$\lim_{\alpha\downarrow 0} \frac{f(x+\alpha y)-f(x)}{\alpha} = \langle y \mid Gx\rangle , \qquad (17.52)$$

and therefore that f is Gâteaux differentiable at x. ☐

Corollary 17.40 *Let $f \in \Gamma_0(\mathcal{H})$ be such that* $\operatorname{int}\operatorname{dom} f \neq \varnothing$. *Suppose that f is Gâteaux differentiable on* $\operatorname{int}\operatorname{dom} f$. *Then ∇f is strong-to-weak continuous on* $\operatorname{int}\operatorname{dom} f$.

Proposition 17.41 *Let $f \in \Gamma_0(\mathcal{H})$ and suppose that $x \in \operatorname{int}\operatorname{dom} f$. Then the following are equivalent:*

(i) *f is Fréchet differentiable at x.*
(ii) *Every selection of ∂f is continuous at x.*
(iii) *There exists a selection of ∂f that is continuous at x.*

Proof. (i)⇒(ii): Assume that this implication is false and set $u = \nabla f(x)$. Then there exist a sequence $(x_n,u_n)_{n\in\mathbb{N}}$ in $\operatorname{gra}\partial f$ and $\varepsilon \in \mathbb{R}_{++}$ such that $x_n \to x$ and

$$(\forall n \in \mathbb{N}) \quad \|u_n - u\| > 2\varepsilon. \qquad (17.53)$$

Note that

$$(\forall n \in \mathbb{N})(\forall y \in \mathcal{H}) \quad \langle y \mid u_n\rangle \leqslant \langle x_n - x \mid u_n\rangle + f(x+y) - f(x_n). \qquad (17.54)$$

The Fréchet differentiability of f at x ensures the existence of $\delta \in \mathbb{R}_{++}$ such that

$$\big(\forall y \in B(0;\delta)\big) \quad f(x+y)-f(x)-\langle y \mid u\rangle \leqslant \varepsilon\|y\|. \qquad (17.55)$$

In view of (17.53), there exists a sequence $(z_n)_{n\in\mathbb{N}}$ in $B(0;1)$ such that $(\forall n \in \mathbb{N})$ $\langle z_n \mid u_n - u\rangle > 2\varepsilon$. Hence, using (17.54), (17.55), the boundedness of $(u_n)_{n\in\mathbb{N}}$, which is guaranteed by Proposition 16.17(iii), and the continuity of f at x, which is guaranteed by Corollary 8.39(ii), we obtain

$$\begin{aligned} 2\varepsilon\delta &< \delta \langle z_n \mid u_n - u\rangle \\ &= \langle \delta z_n \mid u_n\rangle - \langle \delta z_n \mid u\rangle \\ &\leqslant \langle x_n - x \mid u_n\rangle + f(x+\delta z_n) - f(x_n) - \langle \delta z_n \mid u\rangle \end{aligned}$$

$$\begin{aligned}
&= \big(f(x+\delta z_n) - f(x) - \langle \delta z_n \mid u\rangle\big) + \langle x_n - x \mid u_n\rangle + f(x) - f(x_n) \\
&\leqslant \varepsilon\delta + \|x_n - x\|\|u_n\| + f(x) - f(x_n) \\
&\to \varepsilon\delta, \qquad (17.56)
\end{aligned}$$

which is absurd.

(ii)⇒(iii): This is clear.

(iii)⇒(i): Let G be a selection of ∂f such that G is continuous at x. There exists $\delta \in \mathbb{R}_{++}$ such that $B(x;\delta) \subset \operatorname{dom} G = \operatorname{dom} \partial f$. Now take $y \in B(0;\delta)$. Then $\langle y \mid Gx\rangle \leqslant f(x+y) - f(x)$ and $\langle -y \mid G(x+y)\rangle \leqslant f(x) - f(x+y)$. Thus

$$\begin{aligned}
0 &\leqslant f(x+y) - f(x) - \langle y \mid Gx\rangle \\
&\leqslant \langle y \mid G(x+y) - Gx\rangle \\
&\leqslant \|y\|\,\|G(x+y) - Gx\|. \qquad (17.57)
\end{aligned}$$

Since G is continuous at x, this implies that

$$\lim_{0 \neq \|y\| \to 0} \frac{f(x+y) - f(x) - \langle Gx \mid y\rangle}{\|y\|} = 0. \qquad (17.58)$$

Therefore, f is Fréchet differentiable at x. ☐

Item (ii) of the next result shows that the sufficient condition for Fréchet differentiability discussed in Fact 2.62 is also necessary for functions in $\Gamma_0(\mathcal{H})$.

Corollary 17.42 *Let $f \in \Gamma_0(\mathcal{H})$ be Gâteaux differentiable in a neighborhood D of a point $x \in \operatorname{dom} f$. Then the following hold:*

(i) *∇f is strong-to-weak continuous on D.*
(ii) *f is Fréchet differentiable at x if and only if ∇f is continuous at x.*

Proof. This is clear from Proposition 17.39 and Proposition 17.41. ☐

Corollary 17.43 *Let $f \in \Gamma_0(\mathcal{H})$ be Fréchet differentiable on* $\operatorname{int}\operatorname{dom} f$. *Then ∇f is continuous on* $\operatorname{int}\operatorname{dom} f$.

Corollary 17.44 *Suppose that $\mathcal{H}$ is finite-dimensional. Then Gâteaux and Fréchet differentiability are the same notions for functions in $\Gamma_0(\mathcal{H})$.*

Proof. Combine Proposition 17.39 and Proposition 17.41. ☐

The finite-dimensional setting allows for the following useful variant of Proposition 17.31.

Proposition 17.45 *Suppose that $\mathcal{H}$ is finite-dimensional, let $f \in \Gamma_0(\mathcal{H})$, let $x \in \mathcal{H}$, and let $u \in \mathcal{H}$. Then $\partial f(x) = \{u\}$ if and only if f is Fréchet differentiable at x and $u = \nabla f(x)$.*

Proof. Suppose first that $\partial f(x) = \{u\}$ and set $C = \operatorname{dom} f$. By Corollary 16.39, $\overline{C} = \overline{\operatorname{dom} \partial f}$. Hence, using Theorem 20.48 and Proposition 21.17, we deduce that $N_C x = N_{\overline{C}} x = \operatorname{rec}(\partial f(x)) = \operatorname{rec}\{u\} = \{0\}$. By Corollary 6.46 and Corollary 8.39(iii), $x \in \operatorname{int} C = \operatorname{cont} f$. It thus follows from Proposition 17.31(ii) that f is Gâteaux differentiable at x and that $\nabla f(x) = u$. Invoking Corollary 17.44, we obtain that f is Fréchet differentiable at x. The reverse implication is a consequence of Proposition 17.31(i). ☐

Example 17.46 Let $(u_n)_{n\in\mathbb{N}}$ be a sequence in $B(0;1)$ that converges weakly to 0 and let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}_{++}$ such that $\alpha_n \downarrow 0$. Set

$$f\colon \mathcal{H} \to \left]-\infty,+\infty\right] : x \mapsto \sup_{n\in\mathbb{N}} \big(\langle x \mid u_n\rangle - \alpha_n \big). \tag{17.59}$$

Then the following hold:

(i) $f \in \Gamma_0(\mathcal{H})$, $\operatorname{dom} f = \mathcal{H}$, and f is Lipschitz continuous with constant 1.
(ii) $f \geqslant 0$, $f(0) = 0$, and f is Gâteaux differentiable at 0 with $\nabla f(0) = 0$.
(iii) f is Fréchet differentiable at 0 if and only if $u_n \to 0$.

Proof. Fix $x \in \mathcal{H}$.

(i): As a supremum of lower semicontinuous and convex functions, the function f is also lower semicontinuous and convex by Proposition 9.3. Since $u_n \rightharpoonup 0$, it follows that $\langle x \mid u_n\rangle - \alpha_n \to 0$. Thus, $\operatorname{dom} f = \mathcal{H}$. Now let $y \in \mathcal{H}$. Then $(\forall n \in \mathbb{N})$ $\langle x \mid u_n\rangle - \alpha_n = \langle x - y \mid u_n\rangle + \langle y \mid u_n\rangle - \alpha_n \leqslant \|x - y\| + f(y)$. Hence $f(x) \leqslant \|x - y\| + f(y)$, which supplies the Lipschitz continuity of f with constant 1.

(ii): Since $0 = \lim(\langle x \mid u_n\rangle - \alpha_n) \leqslant \sup_{n\in\mathbb{N}}(\langle x \mid u_n\rangle - \alpha_n)$, it is clear that $f \geqslant 0$, that $f(0) = 0$, and that $0 \in \partial f(0)$. Hence $f'(0;\cdot) \geqslant 0$ by Proposition 17.3. Now fix $y \in \mathcal{H}$ and let $m \in \mathbb{N}$. Then there exists $\beta \in \mathbb{R}_{++}$ such that $(\forall n \in \{0,\ldots,m-1\})$ $(\forall \alpha \in \left]0,\beta\right])$ $\alpha \langle y \mid u_n\rangle < \alpha_n$. For every $\alpha \in \left]0,\beta\right]$, $\max_{0\leqslant n\leqslant m-1}(\langle y \mid u_n\rangle - \alpha_n/\alpha) < 0$ and hence

$$\begin{aligned} 0 &\leqslant f(\alpha y)/\alpha \\ &= \sup_{n\in\mathbb{N}} \big(\langle y \mid u_n\rangle - \alpha_n/\alpha \big) \\ &= \max\Big\{ \max_{0\leqslant n\leqslant m-1} \big(\langle y \mid u_n\rangle - \alpha_n/\alpha \big), \sup_{n\geqslant m} \big(\langle y \mid u_n\rangle - \alpha_n/\alpha \big) \Big\} \\ &= \sup_{n\geqslant m} \big(\langle y \mid u_n\rangle - \alpha_n/\alpha \big) \\ &\leqslant \sup_{n\geqslant m} \langle y \mid u_n\rangle . \end{aligned} \tag{17.60}$$

In view of Proposition 17.2(i), we deduce that

$$(\forall m \in \mathbb{N}) \quad 0 \leqslant f'(0;y) = \lim_{\alpha\downarrow 0} \frac{f(\alpha y)}{\alpha} \leqslant \sup_{n\geqslant m} \langle y \mid u_n\rangle . \tag{17.61}$$

Hence

$$0 \leqslant f'(0;y) \leqslant \inf_{m\in\mathbb{N}} \sup_{n\geqslant m} \langle y \mid u_n\rangle = \overline{\lim}\, \langle y \mid u_n\rangle = 0. \tag{17.62}$$

Therefore, $f'(0;\cdot) \equiv 0$, i.e., f is Gâteaux differentiable at 0 with $\nabla f(0) = 0$.

(iii): Since $f \geqslant 0 = f(0)$, it suffices to show that

$$u_n \to 0 \quad \Leftrightarrow \quad \lim_{0\neq\|y\|\to 0} \frac{f(y)}{\|y\|} = 0. \tag{17.63}$$

Assume first that $u_n \not\to 0$. Then there exist $\delta \in \mathbb{R}_{++}$ and a subsequence $(u_{k_n})_{n\in\mathbb{N}}$ of $(u_n)_{n\in\mathbb{N}}$ such that $\inf_{n\in\mathbb{N}} \|u_{k_n}\| \geqslant \delta$. Set

$$(\forall n\in\mathbb{N}) \quad y_n = \sqrt{\alpha_{k_n}} \frac{u_{k_n}}{\|u_{k_n}\|}. \tag{17.64}$$

Then $(\forall n\in\mathbb{N})$ $\|y_n\| = \sqrt{\alpha_{k_n}}$ and $f(y_n) \geqslant \langle y_n \mid u_{k_n}\rangle - \alpha_{k_n} = \sqrt{\alpha_{k_n}}\|u_{k_n}\| - \alpha_{k_n} \geqslant \sqrt{\alpha_{k_n}}\delta - \alpha_{k_n}$. Hence

$$y_n \to 0 \quad \text{and} \quad \underline{\lim}\, \frac{f(y_n)}{\|y_n\|} \geqslant \underline{\lim}\, \big(\delta - \sqrt{\alpha_{k_n}}\,\big) = \delta > 0. \tag{17.65}$$

Now assume that $u_n \to 0$ and fix $\varepsilon \in\,]0,1[$. There exists $m\in\mathbb{N}$ such that for every integer $n \geqslant m$ we have $\|u_n\| \leqslant \varepsilon$, and hence $(\forall y \in \mathcal{H} \smallsetminus \{0\})$ $\langle y \mid u_n\rangle - \alpha_n \leqslant \langle y \mid u_n\rangle \leqslant \|y\|\,\|u_n\| \leqslant \|y\|\varepsilon$. Thus

$$\big(\forall y \in \mathcal{H} \smallsetminus \{0\}\big)\big(\forall n\in\mathbb{N}\big) \quad n \geqslant m \;\Rightarrow\; \frac{\langle y \mid u_n\rangle - \alpha_n}{\|y\|} \leqslant \varepsilon. \tag{17.66}$$

Set $\delta = \min_{0\leqslant n\leqslant m-1} \alpha_n/(1-\varepsilon)$. Then

$$\big(\forall y \in B(0;\delta)\big)\big(\forall n \in \{0,\ldots,m-1\}\big) \quad \langle y \mid u_n\rangle - \alpha_n \leqslant \|y\| - \alpha_n \leqslant \varepsilon\|y\|. \tag{17.67}$$

Hence

$$\big(\forall y \in B(0;\delta) \smallsetminus \{0\}\big)\big(\forall n \in \{0,\ldots,m-1\}\big) \quad \frac{\langle y \mid u_n\rangle - \alpha_n}{\|y\|} \leqslant \varepsilon. \tag{17.68}$$

Altogether,

$$\big(\forall y \in B(0;\delta) \smallsetminus \{0\}\big) \quad \frac{f(y)}{\|y\|} = \sup_{n\in\mathbb{N}} \frac{\langle y \mid u_n\rangle - \alpha_n}{\|y\|} \leqslant \varepsilon, \tag{17.69}$$

which completes the proof of (17.63). ☐

17.7 Differentiability and Continuity

We first illustrate the fact that, at a point, a convex function may be Gâteaux differentiable but not continuous.

Example 17.47 Let C be as in Example 8.42 and set $f = \iota_C$. Then $0 \in (\operatorname{core} C) \smallsetminus (\operatorname{int} C)$, $\partial f(0) = \{0\}$ is a singleton, and f is Gâteaux differentiable at 0 with $\nabla f(0) = 0$. Moreover, f is lower semicontinuous at 0 by Proposition 16.4(iv). However, since $0 \notin \operatorname{int} \operatorname{dom} f = \operatorname{int} C$, f is not continuous at 0.

Proposition 17.48 *Let $f : \mathcal{H} \to \left]-\infty, +\infty\right]$ be convex and Gâteaux differentiable at $x \in \operatorname{dom} f$. Then the following hold:*

(i) *f is lower semicontinuous at x.*
(ii) *Suppose that $\mathcal{H}$ is finite-dimensional. Then $x \in \operatorname{int} \operatorname{dom} f$ and f is continuous on $\operatorname{int} \operatorname{dom} f$.*

Proof. (i): It follows from Proposition 17.31(i) that $x \in \operatorname{dom} \partial f$ and hence from Proposition 16.4(iv) that f is lower semicontinuous at x.

(ii): Note that $x \in \operatorname{core}(\operatorname{dom} f)$ by definition of Gâteaux differentiability. Proposition 6.12(iii) implies that $x \in \operatorname{int} \operatorname{dom} f$. By Corollary 8.39(iii), f is continuous on $\operatorname{int} \operatorname{dom} f$. ☐

The next example shows that Proposition 17.48 is sharp and that the implication in Proposition 17.31(ii) cannot be reversed even when $x \in \operatorname{int} \operatorname{dom} f$.

Example 17.49 Let f be the discontinuous linear functional of Example 2.27 and set $g = f^2$. Then g is convex, $\operatorname{dom} g = \mathcal{H}$, and g is Gâteaux differentiable at 0. By Proposition 17.48, g is lower semicontinuous at 0. However, g is neither continuous nor Fréchet differentiable at 0.

We conclude with two additional continuity results.

Proposition 17.50 *Let $f \in \Gamma_0(\mathcal{H})$ be Gâteaux differentiable at $x \in \operatorname{dom} f$. Then $x \in \operatorname{int} \operatorname{dom} f$ and f is continuous on $\operatorname{int} \operatorname{dom} f$.*

Proof. Since x is a point of Gâteaux differentiability, we have $x \in \operatorname{core} \operatorname{dom} f$. By Corollary 8.39(ii) and Fact 9.17, $\operatorname{cont} f = \operatorname{int} \operatorname{dom} f = \operatorname{core} \operatorname{dom} f$. ☐

Proposition 17.51 *Let $f : \mathcal{H} \to \left]-\infty, +\infty\right]$ be convex and suppose that it is Gâteaux differentiable on some open subset U of $\operatorname{dom} f$. Then f is continuous on $\operatorname{int} \operatorname{dom} f$.*

Proof. Observe that $U \subset \operatorname{int} \operatorname{dom} f$. Let $x \in U$, take $\rho \in \mathbb{R}_{++}$ such that $C = B(x; \rho) \subset U$, and set $g = f + \iota_C$. Proposition 17.48 implies that f is lower semicontinuous and real-valued on C. Consequently, $g \in \Gamma_0(\mathcal{H})$. Hence, Corollary 8.39(ii) asserts that g is continuous and real-valued on $\operatorname{int} \operatorname{dom} g = \operatorname{int} C$. Since $g|_C = f|_C$ and since $x \in \operatorname{int} C$, we deduce that $x \in \operatorname{cont} f$. Thus, the conclusion follows from Theorem 8.38. ☐

Exercises

Exercise 17.1 Let $A \in \mathcal{B}(\mathcal{H})$ be self-adjoint and monotone, and set $q_A\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto (1/2)\langle x \mid Ax\rangle$. Suppose that $\operatorname{ran} A$ is closed. Show that the following are equivalent:

(i) q_A is supercoercive.
(ii) $\operatorname{dom} q_A^* = \mathcal{H}$.
(iii) A is surjective.

Exercise 17.2 Let $\varphi \in \Gamma_0(\mathcal{H})$, let $\gamma \in \mathbb{R}_{++}$, and set $f = (\gamma/2)\|\cdot\|^2 - {}^{\gamma}\varphi(\gamma\cdot)$. Use Proposition 17.7 to show that f is convex (see Example 13.14 for an alternative proof).

Exercise 17.3 Provide a direct proof of the implication (ii)⇒(i) in Proposition 17.7.

Exercise 17.4 Let $A \in \mathcal{B}(\mathcal{H})$ be self-adjointand set $q_A\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto (1/2)\langle x \mid Ax\rangle$. Use Proposition 17.36 to show that A is monotone if and only if $A^\dagger$ is.

Exercise 17.5 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and let $\beta \in \mathbb{R}_{++}$. Suppose that $\operatorname{dom} f$ is open and convex, and that f is Gâteaux differentiable on $\operatorname{dom} f$. Show that the following are equivalent:

(i) f is β-strongly convex.
(ii) $(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)$
$$\langle x - y \mid \nabla f(y)\rangle + f(y) + (\beta/2)\|x-y\|^2 \leqslant f(x).$$
(iii) $(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)\ \langle x - y \mid \nabla f(x) - \nabla f(y)\rangle \geqslant \beta\|x-y\|^2$.

Exercise 17.6 Prove Proposition 17.10.

Exercise 17.7 Set $f\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto x^4$. Use condition (ii) of Proposition 17.10 to show that f is strictly convex. Since $f''(0) = 0$, this demonstrates that strict convexity does not imply condition (iv) of Proposition 17.10.

Exercise 17.8 (Bregman distance) Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be strictly convex, proper, and Gâteaux differentiable on $\operatorname{int}\operatorname{dom} f \neq \varnothing$. Show that the *Bregman distance*

$$D\colon \mathcal{H}\times\mathcal{H} \to \left]-\infty,+\infty\right]$$
$$(x,y) \mapsto \begin{cases} f(x) - f(y) - \langle x - y \mid \nabla f(y)\rangle, & \text{if } y \in \operatorname{int}\operatorname{dom} f;\\ +\infty, & \text{otherwise,}\end{cases} \tag{17.70}$$

satisfies $(\forall (x,y) \in \operatorname{int}\operatorname{dom} f \times \operatorname{int}\operatorname{dom} f)\ D(x,y) \geqslant 0$ and $[D(x,y) = 0 \Leftrightarrow x = y]$.

Exercise 17.9 Use Exercise 17.8 to show the following:

(i) $(\forall \xi \in \mathbb{R}_{++})(\forall \eta \in \mathbb{R}_{++})\ (\xi - \eta)/\eta - \ln(\xi) + \ln(\eta) \geqslant 0$.

(ii) $(\forall \xi \in \mathbb{R}_{++})\ \xi \geqslant \ln(1+\xi) \geqslant \xi/(\xi+1)$.

(iii) $(\forall \xi \in [0,1[)\ -\xi \geqslant \ln(1-\xi) \geqslant \xi/(\xi-1)$.

Exercise 17.10 Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[0,1/2]$ such that $\gamma_n \to 0$. Use Exercise 17.9 to show the following:

(i) $(\forall n \in \mathbb{N})\ \sum_{k=0}^{n} \gamma_k \geqslant \ln(\prod_{k=0}^{n}(1+\gamma_k)) \geqslant (2/3)\sum_{k=0}^{n}\gamma_k$.

(ii) $(\forall n \in \mathbb{N})\ -\sum_{k=0}^{n} \gamma_k \geqslant \ln(\prod_{k=0}^{n}(1-\gamma_k)) \geqslant -2\sum_{k=0}^{n}\gamma_k$.

(iii) $\sum_{n\in\mathbb{N}} \gamma_n < +\infty \Leftrightarrow \prod_{n\in\mathbb{N}}(1+\gamma_n) < +\infty$.

(iv) $\sum_{n\in\mathbb{N}} \gamma_n = +\infty \Leftrightarrow \prod_{n\in\mathbb{N}}(1-\gamma_n) = 0$.

Exercise 17.11 Use Proposition 17.9, Proposition 17.10, and Exercise 8.6 to prove Corollary 17.12.

Exercise 17.12 Explain why the function f constructed in Proposition 17.9 and Corollary 17.12 is independent of the choice of the vector y.

Exercise 17.13 Revisit Example 9.35 and Example 9.36 via Corollary 17.12.

Exercise 17.14 Provide the details for Example 17.13.

Exercise 17.15 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex, let $x \in \operatorname{dom} f$, and let $u \in \mathcal{H}$. Show that $u \in \partial f(x)$ if and only if $(\forall y \in \mathcal{H})$ $-f'(x;-y) \leqslant \langle y \mid u\rangle \leqslant f'(x;y)$.

Exercise 17.16 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex, let $x \in \operatorname{dom} f$, and suppose that $f'(x;\cdot)$ is continuous at $y \in \mathcal{H}$. Use Theorem 17.18 to show that $f'(x;y) = \max \langle y \mid \partial f(x)\rangle$.

Exercise 17.17 Let $f \in \Gamma_0(\mathcal{H})$ and let $\gamma \in \mathbb{R}_{++}$. Use Proposition 16.61(i) to prove (12.28), i.e.,

$$\nabla({}^{\gamma}\!f\,) = \gamma^{-1}\big(\mathrm{Id} - \mathrm{Prox}_{\gamma f}\,\big). \tag{17.71}$$

Exercise 17.18 Let $A \in \mathcal{B}(\mathcal{H})$ be monotone and self-adjoint, and set $q_A\colon x \mapsto (1/2)\langle x \mid Ax\rangle$. Show that q_A is convex and Fréchet differentiable, with $\nabla q_A = A$ and $q_A^* \circ A = q_A$. Moreover, prove that $\operatorname{ran} A \subset \operatorname{dom} q_A^* \subset \overline{\operatorname{ran}}\, A$. Compare with Proposition 17.36.

Exercise 17.19 Let f and g be in $\Gamma_0(\mathcal{H})$, and suppose that $f+g$ is Gâteaux differentiable on $\mathcal{H}$. Show that both f and g are Gâteaux differentiable.

Exercise 17.20 Let $\mathcal{H}$ and f be as in Example 16.22. Show that f is Gâteaux differentiable and $\nabla f\colon \mathcal{H} \to \mathcal{H}\colon (\xi_i)_{i\in I} \mapsto (2n^2\xi_n^{2n-1})_{n\in I}$.

Exercise 17.21 Suppose that $\mathcal{H}$ is finite-dimensional, and let C be a bounded closed convex subset of $\mathcal{H}$ such that $0 \in \operatorname{int} C$. Suppose further that, for every $y \in \operatorname{bdry} C$, $N_C y$ is a ray. Show that m_C is Gâteaux differentiable on $\mathcal{H} \smallsetminus \{0\}$.

Exercise 17.22 Consider Example 17.38, which deals with a nonempty closed convex set C and a point $x \in C$ at which ι_C is not Gâteaux differentiable, and yet $N_C x = \{0\}$. Demonstrate the impossibility of the existence of a nonempty closed convex subset D of $\mathcal{H}$ such that $\operatorname{ran} N_D = \{0\}$ and ι_D is nowhere Gâteaux differentiable.

Exercise 17.23 Let $f \in \Gamma_0(\mathcal{H})$ be Gâteaux differentiable at $x \in \operatorname{dom} f$. Show that $N_{\operatorname{epi} f}(x, f(x)) = \mathbb{R}_+(\nabla f(x), -1)$.

Chapter 18
Further Differentiability Results

Further results concerning derivatives and subgradients are collected in this chapter. The Ekeland–Lebourg theorem gives conditions under which the set of points of Fréchet differentiability is a dense G_δ subset of the domain of the function. Formulas for the subdifferential of a maximum and of an infimal convolution are provided, and the basic duality between differentiability and strict convexity is presented. Another highlight is the Baillon–Haddad theorem, which states that nonexpansiveness and firm nonexpansiveness are identical properties for gradients of convex functions. Finally, the subdifferential operator of the distance to a convex set is analyzed in detail.

18.1 The Ekeland–Lebourg Theorem

Proposition 18.1 *Let $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ be convex and suppose that $x \in \operatorname{cont} f$. Then f is Fréchet differentiable at x if and only if*

$$\begin{aligned}&(\forall \varepsilon \in \mathbb{R}_{++})(\exists\, \eta \in \mathbb{R}_{++})(\forall y \in \mathcal{H})\\ &\qquad\qquad \|y\| = 1 \;\;\Rightarrow\;\; f(x+\eta y) + f(x-\eta y) - 2f(x) < \eta\varepsilon.\end{aligned} \tag{18.1}$$

Proof. We assume that $\mathcal{H} \neq \{0\}$. Let $\varepsilon \in \mathbb{R}_{++}$ and let $y \in \mathcal{H}$ be such that $\|y\| = 1$. Assume that f is Fréchet differentiable at x. Then there exists $\eta \in \mathbb{R}_{++}$ such that

$$(\forall z \in \mathcal{H}) \quad \|z\| = 1 \;\;\Rightarrow\;\; f(x+\eta z) - f(x) - \langle \eta z \mid \nabla f(x)\rangle < (\varepsilon/2)\eta. \tag{18.2}$$

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-48311-5_31

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_18

Letting z be successively y and then $-y$ in (18.2) and adding the two resulting inequalities leads to (18.1). Conversely, let $\eta \in \mathbb{R}_{++}$ be such that (18.1) holds. Using Proposition 16.17(ii), we take $u \in \partial f(x)$. Then

$$\begin{aligned} 0 &\leqslant f(x+\eta y) - f(x) - \langle \eta y \mid u\rangle \\ &< \eta\varepsilon + f(x) - f(x-\eta y) + \langle -\eta y \mid u\rangle \\ &\leqslant \eta\varepsilon. \end{aligned} \tag{18.3}$$

Let $\alpha \in \left]0,\eta\right]$. Then Proposition 9.27 implies that $(f(x+\alpha y) - f(x))/\alpha \leqslant (f(x+\eta y)-f(x))/\eta$. Hence, since $u \in \partial f(x)$, (18.3) yields $(\forall z \in B(0;1))$ $0 \leqslant f(x+\eta z) - f(x) - \langle \eta z \mid u\rangle \leqslant \eta\|z\|\varepsilon$. It follows that f is Fréchet differentiable at x and that $\nabla f(x) = u$. □

Proposition 18.2 *Let* $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ *be proper and convex, let* $\varepsilon \in \mathbb{R}_{++}$, *and set*

$$S_\varepsilon = \bigcup_{\eta\in\mathbb{R}_{++}} \left\{ x \in \operatorname{cont} f \;\middle|\; \sup_{y\in\mathcal{H},\|y\|=1} \frac{f(x+\eta y) + f(x-\eta y) - 2f(x)}{\eta} < \varepsilon \right\}. \tag{18.4}$$

Then S_ε *is open.*

Proof. We assume that $\operatorname{cont} f \neq \varnothing$ and that $\mathcal{H} \neq \{0\}$. Take $x \in S_\varepsilon$. Since f is continuous at x, Theorem 8.38 guarantees the existence of $\eta_1 \in \mathbb{R}_{++}$ and $\beta \in \mathbb{R}_{++}$ such that

$$\big(\forall y \in B(x;\eta_1)\big)\big(\forall z \in B(x;\eta_1)\big) \quad |f(y) - f(z)| \leqslant \beta\|y - z\|. \tag{18.5}$$

Furthermore, using the definition of S_ε and Proposition 9.27, we deduce the existence of $\eta \in \left]0,\eta_1\right[$ such that

$$\sigma = \sup_{y\in\mathcal{H},\|y\|=1} \frac{f(x+\eta y) + f(x-\eta y) - 2f(x)}{\eta} < \varepsilon. \tag{18.6}$$

Take $\eta_2 \in \left[0, \min\{\eta_1 - \eta, \eta(\varepsilon-\sigma)/(4\beta)\}\right[$. Then for every $z \in B(x;\eta_2)$ and every $y \in \mathcal{H}$ such that $\|y\| = 1$, we have

$$\begin{aligned} \frac{f(z+\eta y) + f(z-\eta y) - 2f(z)}{\eta} &\leqslant \frac{f(x+\eta y) + f(x-\eta y) - 2f(x)}{\eta} \\ &\quad + \frac{f(z+\eta y) - f(x+\eta y)}{\eta} \\ &\quad + \frac{f(z-\eta y) - f(x-\eta y)}{\eta} \\ &\quad + 2\frac{f(x) - f(z)}{\eta} \\ &\leqslant \sigma + \frac{4\beta}{\eta}\|x - z\| \end{aligned}$$

$$\begin{aligned} &\leqslant \sigma + \frac{4\beta}{\eta}\eta_2 \\ &< \varepsilon. \end{aligned} \tag{18.7}$$

Consequently, $B(x;\eta_2) \subset S_\varepsilon$. □

Theorem 18.3 (Ekeland–Lebourg) *Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be convex, and suppose that $\operatorname{cont} f \neq \varnothing$. Then the set of points at which f is Fréchet differentiable is a dense G_δ subset of $\overline{\operatorname{dom}}\, f$.*

Proof. We assume that $\mathcal{H} \neq \{0\}$. Set $\mathcal{X} = \overline{\operatorname{dom}}\, f$. It follows from Theorem 8.38 that $\operatorname{cont} f = \operatorname{int}\operatorname{dom} f$ and from Proposition 3.45(iii) that $\overline{\operatorname{int}\operatorname{dom} f} = \mathcal{X}$. Take $y \in \operatorname{int}\operatorname{dom} f$. Then there exists $\rho \in \mathbb{R}_{++}$ such that $B(y;\rho) \subset \operatorname{int}\operatorname{dom} f$ and f is bounded on $B(y;\rho)$. Now let $\psi\colon \mathbb{R} \to \mathbb{R}$ be such that ψ is differentiable on $[0,\rho]$, $\psi|_{]0,\rho[} > 0$, and $\psi(0) = \psi(\rho) = 0$ (for instance, $\psi\colon t \mapsto \sin(t\pi/\rho)$). Set

$$\varphi\colon B(y;\rho) \to [0,+\infty]\colon x \mapsto \begin{cases} 1/\psi(\|x-y\|), & \text{if } 0 < \|x-y\| < \rho; \\ +\infty, & \text{otherwise,} \end{cases} \tag{18.8}$$

and $U = (\operatorname{int} B(y;\rho)) \smallsetminus \{y\}$. Then φ is proper and lower semicontinuous. Combining Example 2.65 and Fact 2.63, we see that φ is Fréchet differentiable on U. Fix $\varepsilon \in \mathbb{R}_{++}$. Applying Theorem 1.46(i)&(iii) in the metric space $B(y;\rho)$ to the function $\varphi - f$, we deduce the existence of $z_\varepsilon \in B(y;\rho)$ such that

$$(\varphi - f)(z_\varepsilon) \leqslant \varepsilon + \inf(\varphi - f)\big(B(y;\rho)\big) \tag{18.9}$$

and

$$\big(\forall x \in B(y;\rho)\big) \quad (\varphi - f)(z_\varepsilon) - (\varphi - f)(x) \leqslant \varepsilon\|x - z_\varepsilon\|. \tag{18.10}$$

Since $\varphi - f$ is bounded below and takes on the value $+\infty$ only on $B(y;\rho) \smallsetminus U$, (18.9) implies that $z_\varepsilon \in U$. Hence, φ is Fréchet differentiable at z_ε and there exists $\eta \in \mathbb{R}_{++}$ such that $B(z_\varepsilon;\eta) \subset U$ and

$$\big(\forall x \in B(z_\varepsilon;\eta)\big) \quad \varphi(x) - \varphi(z_\varepsilon) - \langle x - z_\varepsilon \mid \nabla\varphi(z_\varepsilon)\rangle \leqslant \varepsilon\|x - z_\varepsilon\|. \tag{18.11}$$

Adding (18.10) and (18.11), we obtain

$$\big(\forall x \in B(z_\varepsilon;\eta)\big) \quad f(x) - f(z_\varepsilon) - \langle x - z_\varepsilon \mid \nabla\varphi(z_\varepsilon)\rangle \leqslant 2\varepsilon\|x - z_\varepsilon\|. \tag{18.12}$$

Hence

$$(\forall r \in \mathcal{H}) \quad \|r\| = 1 \;\;\Rightarrow\;\; f(z_\varepsilon + \eta r) - f(z_\varepsilon) - \langle \eta r \mid \nabla\varphi(z_\varepsilon)\rangle \leqslant 2\varepsilon\eta. \tag{18.13}$$

Invoking the convexity of f, and considering (18.13) for r and $-r$, we deduce that for every $r \in \mathcal{H}$ such that $\|r\| = 1$, we have

$$0 \leqslant \frac{f(z_\varepsilon + \eta r) + f(z_\varepsilon - \eta r) - 2f(z_\varepsilon)}{\eta} \leqslant 4\varepsilon. \tag{18.14}$$

It follows that

$$\sup_{\substack{r\in\mathcal{H}\\ \|r\|=1}} \big(f(z_\varepsilon + \eta r) + f(z_\varepsilon - \eta r) - 2f(z_\varepsilon)\big) < 5\eta\varepsilon. \tag{18.15}$$

For the remainder of this proof, we adopt the notation (18.4). Since $z_\varepsilon \in S_{5\varepsilon}$, it follows from Proposition 18.2 that S_ε is dense and open in $\mathcal{X}$. Thus, by Corollary 1.45, the set $S = \bigcap_{n\in\mathbb{N}\smallsetminus\{0\}} S_{1/n}$ is a dense G_δ subset of $\mathcal{X}$. Using (18.15) and Proposition 18.1, we therefore conclude that f is Fréchet differentiable on S. ☐

Proposition 18.4 *Let C be a nonempty bounded subset of $\mathcal{H}$, and let $\varepsilon \in \mathbb{R}_{++}$. Then there exist $x \in \mathcal{H}$ and $\alpha \in \mathbb{R}_{++}$ such that*

$$\operatorname{diam}\big(\big\{u \in C \;\big|\; \sigma_C(x) - \alpha < \langle x \mid u\rangle\big\}\big) \leqslant \varepsilon. \tag{18.16}$$

Proof. We assume that $\mathcal{H} \neq \{0\}$. The support function σ_C is real-valued, convex, and continuous on $\mathcal{H}$ by Example 11.2. In view of Theorem 18.3, there exists a point $x \in \mathcal{H}$ at which σ_C is Fréchet differentiable. In turn, by Proposition 18.1, there exists $\delta \in \mathbb{R}_{++}$ such that

$$(\forall y \in \mathcal{H}) \quad \|y\| = 1 \quad\Rightarrow\quad \sigma_C(x+\delta y) + \sigma_C(x-\delta y) - 2\sigma_C(x) < \delta\varepsilon/3. \tag{18.17}$$

Now set $\alpha = \delta\varepsilon/3$ and assume that $\operatorname{diam}\big(\big\{u \in C \;\big|\; \sigma_C(x) - \alpha < \langle x \mid u\rangle\big\}\big) > \varepsilon$. Then there exist u and v in C such that $\langle x \mid u\rangle > \sigma_C(x) - \alpha$, $\langle x \mid v\rangle > \sigma_C(x) - \alpha$, and $\|u - v\| > \varepsilon$. Set $z = (u-v)/\|u-v\|$. Then $\|z\| = 1$ and $\langle z \mid u - v\rangle > \varepsilon$. Hence,

$$\begin{aligned}\sigma_C(x+\delta z) + \sigma_C(x-\delta z) &\geqslant \langle x + \delta z \mid u\rangle + \langle x - \delta z \mid v\rangle\\ &> 2\sigma_C(x) - 2\alpha + \delta\varepsilon\\ &= 2\sigma_C(x) + \delta\varepsilon/3,\end{aligned} \tag{18.18}$$

which contradicts (18.17). ☐

18.2 The Subdifferential of a Maximum

Theorem 18.5 (Dubovitskii and Milyutin) *Let $(f_i)_{i\in I}$ be a finite family of convex functions from $\mathcal{H}$ to $]-\infty, +\infty]$, and suppose that $x \in \bigcap_{i\in I} \operatorname{cont} f_i$. Set $f = \max_{i\in I} f_i$ and let $I(x) = \big\{i \in I \;\big|\; f_i(x) = f(x)\big\}$. Then*

$$\partial f(x) = \operatorname{conv} \bigcup_{i\in I(x)} \partial f_i(x). \tag{18.19}$$

Proof. By Proposition 16.17(ii), Proposition 3.39(ii), and Theorem 3.34, the set $\operatorname{conv} \bigcup_{i \in I(x)} \partial f_i(x)$ is weakly compact and closed. First, let $i \in I(x)$ and $u \in \partial f_i(x)$. Then $(\forall y \in \mathcal{H})\ \langle y - x \mid u \rangle \leqslant f_i(y) - f_i(x) \leqslant f(y) - f(x)$ and hence $u \in \partial f(x)$. This and Proposition 16.4(iii) imply that $\operatorname{conv} \bigcup_{i \in I(x)} \partial f_i(x) \subset \partial f(x)$. We now argue by contradiction and assume that this inclusion is strict, i.e., that there exists

$$u \in \partial f(x) \smallsetminus \operatorname{conv} \bigcup_{i \in I(x)} \partial f_i(x). \tag{18.20}$$

By Theorem 3.50 and Theorem 17.18, there exist $y \in \mathcal{H} \smallsetminus \{0\}$ and $\varepsilon \in \mathbb{R}_{++}$ such that

$$\langle y \mid u \rangle \geqslant \varepsilon + \max_{i \in I(x)} \sup \langle y \mid \partial f_i(x) \rangle = \varepsilon + \max_{i \in I(x)} f_i'(x; y). \tag{18.21}$$

Since y and ε can be rescaled, we assume that

$$x + y \in \bigcap_{i \in I} \operatorname{dom} f_i = \operatorname{dom} f. \tag{18.22}$$

Now let $(\alpha_n)_{n \in \mathbb{N}}$ be a sequence in $]0,1[$ such that $\alpha_n \downarrow 0$. Since I is finite, there exist a subsequence of $(\alpha_n)_{n \in \mathbb{N}}$, which we still denote by $(\alpha_n)_{n \in \mathbb{N}}$, and $j \in I$ such that

$$(\forall n \in \mathbb{N}) \quad f(x + \alpha_n y) = f_j(x + \alpha_n y). \tag{18.23}$$

Now let $n \in \mathbb{N}$. Then $f_j(x + \alpha_n y) \leqslant (1 - \alpha_n) f_j(x) + \alpha_n f_j(x + y)$ and thus, by (18.23) and (18.20),

$$\begin{aligned}
(1 - \alpha_n) f_j(x) &\geqslant f_j(x + \alpha_n y) - \alpha_n f_j(x + y) \\
&\geqslant f(x + \alpha_n y) - \alpha_n f(x + y) \\
&\geqslant f(x) + \langle \alpha_n y \mid u \rangle - \alpha_n f(x + y) \\
&\geqslant f_j(x) + \alpha_n \langle y \mid u \rangle - \alpha_n f(x + y).
\end{aligned} \tag{18.24}$$

Letting $n \to +\infty$ and using (18.22), we deduce that

$$f_j(x) = f(x). \tag{18.25}$$

In view of (18.23), (18.25), (18.20), and (18.21), we obtain

$$\begin{aligned}
f_j'(x; y) &\leftarrow \frac{f_j(x + \alpha_n y) - f_j(x)}{\alpha_n} \\
&= \frac{f(x + \alpha_n y) - f(x)}{\alpha_n} \\
&\geqslant \langle y \mid u \rangle \\
&\geqslant \varepsilon + f_j'(x; y),
\end{aligned} \tag{18.26}$$

which is the desired contradiction. ☐

18.3 Differentiability of Infimal Convolutions

Proposition 18.6 *Let f and g be in $\Gamma_0(\mathcal{H})$, and let x and y be in $\mathcal{H}$. Suppose that $x \in \operatorname{dom} \partial(f \,\Box\, g)$, that $(f \,\Box\, g)(x) = f(y) + g(x-y)$, and that f is Gâteaux differentiable at y. Then $\partial(f \,\Box\, g)(x) = \{\nabla f(y)\}$.*

Proof. Proposition 16.61(i) and Proposition 17.31(i) imply that

$$\partial(f \,\Box\, g)(x) = \partial f(y) \cap \partial g(x-y) = \{\nabla f(y)\} \cap \partial g(x-y) \subset \{\nabla f(y)\}. \tag{18.27}$$

Since $\partial(f \,\Box\, g)(x) \neq \varnothing$, we conclude that $\partial(f \,\Box\, g)(x) = \{\nabla f(y)\}$. ☐

Proposition 18.7 *Let f and g be in $\Gamma_0(\mathcal{H})$, and let x and y be in $\mathcal{H}$. Suppose that $f \,\Box\, g \in \Gamma_0(\mathcal{H})$, that $(f \,\Box\, g)(x) = f(y) + g(x-y)$, and that f is Gâteaux differentiable at y. Then $x \in \operatorname{cont}(f \,\Box\, g)$, $f \,\Box\, g$ is Gâteaux differentiable at x, and*

$$\nabla(f \,\Box\, g)(x) = \nabla f(y). \tag{18.28}$$

In addition, if f is Fréchet differentiable at y, then $f \,\Box\, g$ is Fréchet differentiable at x.

Proof. Proposition 17.50 and Proposition 12.6(ii) imply that $x = y + (x-y) \in (\operatorname{int}\operatorname{dom} f) + \operatorname{dom} g \subset \operatorname{dom} f + \operatorname{dom} g = \operatorname{dom}(f \,\Box\, g)$. However, $\operatorname{dom} g + \operatorname{int}\operatorname{dom} f = \bigcup_{z\in\operatorname{dom} g}(z + \operatorname{int}\operatorname{dom} f)$ is open as a union of open sets. Thus, $x \in \operatorname{int}\operatorname{dom}(f \,\Box\, g)$ and, since $f \,\Box\, g \in \Gamma_0(\mathcal{H})$, Proposition 16.27 implies that $x \in \operatorname{cont}(f \,\Box\, g) \subset \operatorname{dom} \partial(f \,\Box\, g)$. Next, we combine Proposition 18.6 and Proposition 17.31(ii) to deduce the Gâteaux differentiability of $f \,\Box\, g$ at x and (18.28). Now assume that f is Fréchet differentiable at y. Then, for every $z \in \mathcal{H} \smallsetminus \{0\}$,

$$\begin{aligned} 0 &\leqslant (f \,\Box\, g)(x+z) - (f \,\Box\, g)(x) - \langle z \mid \nabla f(y)\rangle \\ &\leqslant f(y+z) + g(x-y) - \big(f(y) + g(x-y)\big) - \langle z \mid \nabla f(y)\rangle \\ &= f(y+z) - f(y) - \langle z \mid \nabla f(y)\rangle \end{aligned} \tag{18.29}$$

and hence

$$\begin{aligned} 0 &\leqslant \lim_{0\neq\|z\|\to 0} \frac{(f \,\Box\, g)(x+z) - (f \,\Box\, g)(x) - \langle z \mid \nabla f(y)\rangle}{\|z\|} \\ &\leqslant \lim_{0\neq\|z\|\to 0} \frac{f(y+z) - f(y) - \langle z \mid \nabla f(y)\rangle}{\|z\|} \\ &= 0. \end{aligned} \tag{18.30}$$

Therefore, $f \,\Box\, g$ is Fréchet differentiable at x. ☐

Corollary 18.8 *Let f and g be in $\Gamma_0(\mathcal{H})$. Suppose that f is real-valued, supercoercive, and Fréchet differentiable on $\mathcal{H}$. Then $f \,\Box\, g$ is Fréchet differentiable on $\mathcal{H}$.*

Proof. By Proposition 12.14(i), $f \,\Box\, g = f \,\boxdot\, g \in \Gamma_0(\mathcal{H})$. Now let $x \in \mathcal{H}$. Since $\operatorname{dom}(f \,\Box\, g) = \mathcal{H}$ by Proposition 12.6(ii), there exists $y \in \mathcal{H}$ such that $(f \,\Box\, g)(x) = f(y) + g(x - y)$. Now apply Proposition 18.7. ☐

18.4 Differentiability and Strict Convexity

In this section, we examine the interplay between Gâteaux differentiability and strict convexity via duality.

Proposition 18.9 *Let $f \in \Gamma_0(\mathcal{H})$ be such that f^* is strictly convex on every nonempty convex subset of* $\operatorname{dom} \partial f^*$. *Then f is Gâteaux differentiable on* $\operatorname{int} \operatorname{dom} f$.

Proof. Suppose that $x \in \operatorname{int} \operatorname{dom} f$ and that $[u_1, u_2] \subset \partial f(x)$. Then it follows from Corollary 16.30 that $[u_1, u_2] \subset \operatorname{ran} \partial f = \operatorname{dom}(\partial f)^{-1} = \operatorname{dom} \partial f^*$. Hence, Proposition 16.37(i) implies that f^* is affine on $[u_1, u_2]$. Consequently, $u_1 = u_2$ and $\partial f(x)$ is a singleton. Furthermore, $x \in \operatorname{cont} f$ by Corollary 8.39(ii) and the conclusion thus follows from Proposition 17.31(ii). ☐

Proposition 18.10 *Let $f \in \Gamma_0(\mathcal{H})$ be Gâteaux differentiable on* $\operatorname{int} \operatorname{dom} f$. *Then f^* is strictly convex on every nonempty convex subset of* $\nabla f(\operatorname{int} \operatorname{dom} f)$.

Proof. Assume to the contrary that there exist two distinct points u_1 and u_2 such that f^* is affine on $[u_1, u_2] \subset \nabla f(\operatorname{int} \operatorname{dom} f)$. Choose $x \in \operatorname{int} \operatorname{dom} f$ such that $\nabla f(x) \in \left]u_1, u_2\right[$. Then Proposition 16.37(ii) implies that $[u_1, u_2] \subset \partial f(x)$, which contradicts the Gâteaux differentiability of f at x. ☐

Corollary 18.11 *Let $f \in \Gamma_0(\mathcal{H})$ be such that* $\operatorname{dom} \partial f = \operatorname{int} \operatorname{dom} f$. *Then f is Gâteaux differentiable on* $\operatorname{int} \operatorname{dom} f$ *if and only if f^* is strictly convex on every nonempty convex subset of* $\operatorname{dom} \partial f^*$, *in which case* $\operatorname{dom} \partial f^* = \nabla f(\operatorname{int} \operatorname{dom} f)$.

Proof. If f^* is strictly convex on every nonempty convex subset of $\operatorname{dom} \partial f^*$, then f is Gâteaux differentiable on $\operatorname{int} \operatorname{dom} f$ by Proposition 18.9. Let us now assume that f is Gâteaux differentiable on $\operatorname{int} \operatorname{dom} f$. Using Proposition 17.31(i) and Corollary 16.30, we see that

$$\nabla f(\operatorname{int} \operatorname{dom} f) = \partial f(\operatorname{int} \operatorname{dom} f) = \partial f(\operatorname{dom} \partial f) = \operatorname{ran} \partial f = \operatorname{dom} \partial f^*. \tag{18.31}$$

The result thus follows from Proposition 18.10. ☐

Corollary 18.12 *Let $f \in \Gamma_0(\mathcal{H})$ be such that* $\operatorname{dom} \partial f = \operatorname{int} \operatorname{dom} f$ *and* $\operatorname{dom} \partial f^* = \operatorname{int} \operatorname{dom} f^*$. *Then the following hold:*

(i) *f is Gâteaux differentiable on* $\operatorname{int} \operatorname{dom} f$ *if and only if f^* is strictly convex on* $\operatorname{int} \operatorname{dom} f^*$.

(ii) f *is strictly convex on* $\operatorname{int}\operatorname{dom} f$ *if and only if* f^* *is Gâteaux differentiable on* $\operatorname{int}\operatorname{dom} f^*$.

Proof. (i): This follows from Corollary 18.11.

(ii): Apply (i) to f^* and use Corollary 13.38. ☐

18.5 Stronger Notions of Differentiability

In this section, we investigate certain properties of the gradient of convex functions.

Theorem 18.13 *Let $f\colon \mathcal{H} \to \mathbb{R}$ be a Fréchet differentiable convex function, and let $\phi\colon \mathbb{R} \to \mathbb{R}_+$ be an even convex function that vanishes only at 0. For every $s \in \mathbb{R}$, define*

$$\psi(s) = \begin{cases} \phi(s)/|s|, & \text{if } s \neq 0;\\ 0, & \text{if } s = 0, \end{cases} \qquad \theta(s) = \int_0^1 \frac{\phi(st)}{t}\,dt,$$
$$\text{and} \quad \varrho(s) = \sup\big\{\nu \in \mathbb{R}_+ \;\big|\; 2\theta^*(\nu) \leqslant \nu s\big\}. \tag{18.32}$$

Now consider the following statements:

(i) $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ \|\nabla f(x) - \nabla f(y)\| \leqslant \psi(\|x - y\|)$.
(ii) $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ \langle x - y \mid \nabla f(x) - \nabla f(y)\rangle \leqslant \phi(\|x - y\|)$.
(iii) $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ f(y) \leqslant f(x) + \langle y - x \mid \nabla f(x)\rangle + \theta(\|x - y\|)$.
(iv) $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ f^*(\nabla f(y)) \geqslant$
$f^*(\nabla f(x)) + \langle x \mid \nabla f(y) - \nabla f(x)\rangle + \theta^*(\|\nabla f(x) - \nabla f(y)\|)$.
(v) $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ \langle x - y \mid \nabla f(x) - \nabla f(y)\rangle \geqslant 2\theta^*(\|\nabla f(x) - \nabla f(y)\|)$.
(vi) $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ \|\nabla f(x) - \nabla f(y)\| \leqslant \varrho(\|x - y\|)$.

Then (i)⇒(ii)⇒(iii)⇒(iv)⇒(v)⇒(vi).

Proof. Let x and y be in $\mathcal{H}$, and set $u = \nabla f(x)$ and $v = \nabla f(y)$.

(i)⇒(ii): Cauchy–Schwarz.

(ii)⇒(iii): Since the Gâteaux derivative of $t \mapsto f(x + t(y - x))$ is $t \mapsto \langle y - x \mid \nabla f(x + t(y - x))\rangle$, we have

$$\begin{aligned} f(y) - f(x) - \langle y - x \mid u\rangle &= \int_0^1 \langle y - x \mid \nabla f(x + t(y - x)) - \nabla f(x)\rangle\, dt\\ &= \int_0^1 \frac{1}{t}\langle t(y - x) \mid \nabla f(x + t(y - x)) - \nabla f(x)\rangle\, dt\\ &\leqslant \int_0^1 \frac{1}{t}\phi\big(\|y - x\|t\big)\, dt\\ &= \theta\big(\|y - x\|\big). \end{aligned} \tag{18.33}$$

(iii)⇒(iv): For every $z \in \mathcal{H}$, we derive from Proposition 17.35 that

$$\begin{aligned} -f(z) &\geqslant -f(x) + \langle x \mid u\rangle - \langle z \mid u\rangle - \theta\big(\|z-x\|\big) \\ &= f^*(u) - \langle z \mid u\rangle - \theta\big(\|z-x\|\big) \end{aligned} \tag{18.34}$$

and, in turn, we obtain

$$\begin{aligned} f^*(v) &\geqslant \langle z \mid v\rangle - f(z) \\ &\geqslant \langle z \mid v\rangle + f^*(u) - \langle z \mid u\rangle - \theta\big(\|z-x\|\big) \\ &= f^*(u) + \langle x \mid v-u\rangle + \langle z-x \mid v-u\rangle - \theta\big(\|z-x\|\big). \end{aligned} \tag{18.35}$$

However, since ϕ is even, so is θ and, therefore, it follows from Example 13.8 that $(\theta \circ \|\cdot\|)^* = \theta^* \circ \|\cdot\|$. Hence, we derive from (18.35) that

$$\begin{aligned} f^*(v) &\geqslant f^*(u) + \langle x \mid v-u\rangle + \sup_{z\in\mathcal{H}} \big(\langle z-x \mid v-u\rangle - \theta\big(\|z-x\|\big)\big) \\ &= f^*(u) + \langle x \mid v-u\rangle + \theta^*\big(\|v-u\|\big). \end{aligned} \tag{18.36}$$

(iv)⇒(v): We have

$$f^*(u) \geqslant f^*(v) + \langle y \mid u-v\rangle + \theta^*\big(\|u-v\|\big). \tag{18.37}$$

Adding (18.36) and (18.37) yields (v).

(v)⇒(vi): Cauchy–Schwarz. ☐

The following special case concerns convex functions with Hölder continuous gradients.

Corollary 18.14 *Let $f\colon \mathcal{H} \to \mathbb{R}$ be Fréchet differentiable and convex, let $\beta \in \mathbb{R}_{++}$, and let $p \in \,]0,1]$. Consider the following statements:*

(i) $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ \|\nabla f(x) - \nabla f(y)\| \leqslant \beta\|x-y\|^p$.
(ii) $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ \langle x-y \mid \nabla f(x) - \nabla f(y)\rangle \leqslant \beta\|x-y\|^{p+1}$.
(iii) $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ f(y) \leqslant f(x) + \langle y-x \mid \nabla f(x)\rangle + \beta(p+1)^{-1}\|x-y\|^{p+1}$.
(iv) $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ f^*(\nabla f(y)) \geqslant f^*(\nabla f(x)) + \langle x \mid \nabla f(y) - \nabla f(x)\rangle$ $+\beta^{-1/p}p(p+1)^{-1}\|\nabla f(x) - \nabla f(y)\|^{1+1/p}$.
(v) $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ \langle x-y \mid \nabla f(x) - \nabla f(y)\rangle \geqslant$ $2\beta^{-1/p}p(p+1)^{-1}\|\nabla f(x) - \nabla f(y)\|^{1+1/p}$.
(vi) $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ \|\nabla f(x) - \nabla f(y)\| \leqslant \beta((p+1)/(2p))^p\|x-y\|^p$.

Then (i)⇒(ii)⇒(iii)⇒(iv)⇒(v)⇒(vi).

Proof. This follows from Theorem 18.13 with $\phi = \beta|\cdot|^{p+1}$. Indeed, ϕ is an even convex function vanishing only at 0. Moreover, $\theta\colon t \mapsto \beta|t|^{p+1}/(p+1)$ and we deduce from Example 13.2(i) and Proposition 13.23(i) that $\theta^*\colon \nu \mapsto \beta^{-1/p}p(p+1)^{-1}|\nu|^{1+1/p}$. ☐

Next, we provide several characterizations of convex functions with Lipschitz continuous gradients.

Theorem 18.15 *Let $f\colon \mathcal{H} \to \mathbb{R}$ be continuous and convex, let $\beta \in \mathbb{R}_{++}$, and set $h = f^* - (1/\beta)q$, where $q = (1/2)\|\cdot\|^2$. Then the following are equivalent:*

(i) *f is Fréchet differentiable on $\mathcal{H}$ and ∇f is β-Lipschitz continuous.*
(ii) *f is Fréchet differentiable on $\mathcal{H}$ and*

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \langle x - y \mid \nabla f(x) - \nabla f(y)\rangle \leqslant \beta\|x - y\|^2.$$

(iii) **(Descent lemma)** *f is Fréchet differentiable on $\mathcal{H}$ and*

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad f(y) \leqslant f(x) + \langle y - x \mid \nabla f(x)\rangle + (\beta/2)\|x - y\|^2.$$

(iv) *f is Fréchet differentiable on $\mathcal{H}$ and*

$$\begin{aligned}&(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad f^*(\nabla f(y)) \geqslant \\ &\quad f^*(\nabla f(x)) + \langle x \mid \nabla f(y) - \nabla f(x)\rangle + (1/(2\beta))\|\nabla f(x) - \nabla f(y)\|^2.\end{aligned}$$

(v) *f is Fréchet differentiable on $\mathcal{H}$ and ∇f is $(1/\beta)$-cocoercive.*
(vi) *$\beta q - f$ is convex.*
(vii) *$f^* - (1/\beta)q$ is convex (i.e., f^* is $1/\beta$-strongly convex).*
(viii) *$h \in \Gamma_0(\mathcal{H})$ and $f = {}^{1/\beta}(h^*) = \beta q - {}^{\beta}h \circ \beta\mathrm{Id}$.*
(ix) *$h \in \Gamma_0(\mathcal{H})$ and $\nabla f = \mathrm{Prox}_{\beta h} \circ \beta\mathrm{Id} = \beta(\mathrm{Id} - \mathrm{Prox}_{h^*/\beta})$.*

Proof. (i)⇔(ii)⇔(iii)⇔(iv)⇔(v): Set $p = 1$ in Corollary 18.14.

(v)⇒(vi): Since $(1/\beta)\nabla f$ is firmly nonexpansive, so is $\mathrm{Id} - (1/\beta)\nabla f$ by Proposition 4.4. Hence, $\nabla(\beta q - f) = \beta\mathrm{Id} - \nabla f$ is monotone and it follows from Proposition 17.7 that $\beta q - f$ is convex.

(vi)⇔(vii): Proposition 14.2.

(vii)⇒(viii): Since $f \in \Gamma_0(\mathcal{H})$ and h is convex, Corollary 13.38 yields $h \in \Gamma_0(\mathcal{H})$ and $h^* \in \Gamma_0(\mathcal{H})$. Hence, using successively Corollary 13.38, Proposition 14.1, and Theorem 14.3(i), we obtain $f = f^{**} = (h + (1/\beta)q)^* = {}^{1/\beta}(h^*) = \beta q - {}^{\beta}h \circ \beta\mathrm{Id}$.

(viii)⇒(ix)⇒(i): Proposition 12.30. ☐

Remark 18.16 Let $f\colon \mathcal{H} \to \mathbb{R}$ be Fréchet differentiable and convex, let $\beta \in \mathbb{R}_{++}$, suppose that ∇f is β-Lipschitz continuous, and let $x \in \mathcal{H}$. Set

$$\begin{cases} a_x\colon y \mapsto f(x) + \langle y - x \mid \nabla f(x)\rangle, \\ q_x\colon y \mapsto f(x) + \langle y - x \mid \nabla f(x)\rangle + (\beta/2)\|y - x\|^2. \end{cases} \tag{18.38}$$

Then it follows from Proposition 17.7 and Theorem 18.15 that $a_x \leqslant f \leqslant q_x$. Indeed, a_x is a continuous affine minorant of f and q_x is a continuous quadratic convex majorant of f. Moreover, $a_x(x) = f(x) = q_x(x)$ and $\nabla a_x(x) = \nabla f(x) = \nabla q_x(x)$ (see Figure 18.1).

As seen in Example 4.17, nonexpansive operators are not necessarily firmly nonexpansive. Remarkably, the equivalence (i)⇔(v) in Theorem 18.15 asserts

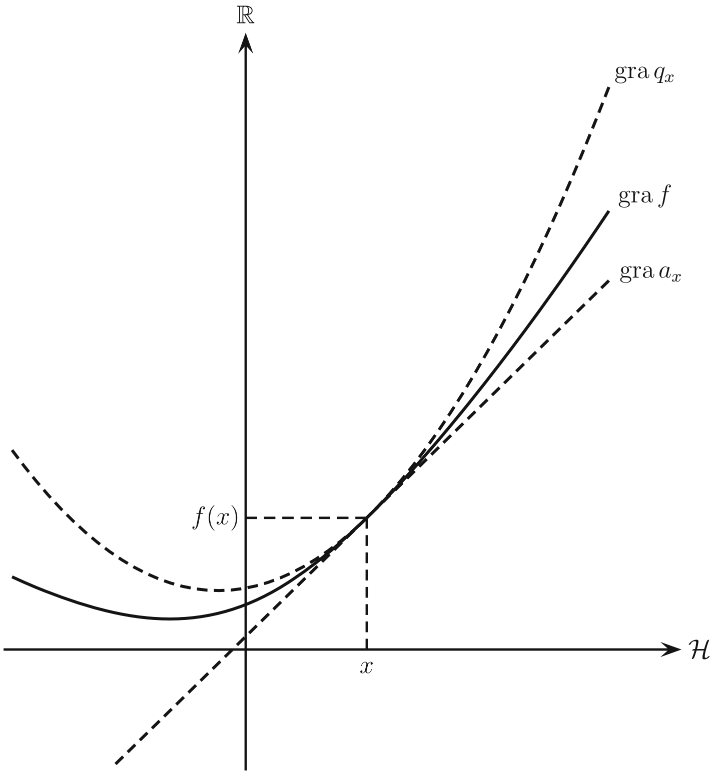

Fig. 18.1 The functions a_x and q_x of (18.38) are respectively an exact continuous affine minorant and an exact continuous quadratic majorant of f at x.

that this is true for the gradient of a convex function $f : \mathcal{H} \to \mathbb{R}$. We record this result next, which is known as the *Baillon–Haddad theorem*.

Corollary 18.17 (Baillon–Haddad) *Let $f : \mathcal{H} \to \mathbb{R}$ be a Fréchet differentiable convex function and let $\beta \in \mathbb{R}_{++}$. Then ∇f is β-Lipschitz continuous if and only if ∇f is $(1/\beta)$-cocoercive. In particular, ∇f is nonexpansive if and only if ∇f is firmly nonexpansive.*

Corollary 18.18 *Let $L \in \mathcal{B}(\mathcal{H})$ be self-adjoint and monotone, and let $x \in \mathcal{H}$. Then $\|L\| \langle Lx \mid x\rangle \geqslant \|Lx\|^2$.*

Proof. Set $f : \mathcal{H} \to \mathbb{R} : y \mapsto \langle Ly \mid y\rangle /2$. It follows from Example 2.57 that $\nabla f = L$ is Lipschitz continuous with constant $\|L\|$. Hence, the result follows from the implication (i)⇒(v) in Theorem 18.15. ☐

Corollary 18.19 (Moreau) *Let $f \in \Gamma_0(\mathcal{H})$, let $\beta \in \mathbb{R}_{++}$, and set $q = (1/2)\|\cdot\|^2$. Then f is Fréchet differentiable on $\mathcal{H}$ with a β-Lipschitz continuous gradient if and only if f is the Moreau envelope of parameter $1/\beta$ of a function in $\Gamma_0(\mathcal{H})$; more precisely,*

$$f = \left(f^* - \frac{1}{\beta}q\right)^* \,\Box\, \beta q. \tag{18.39}$$

Proof. This follows from the equivalence (i)⇔(viii) in Theorem 18.15. ☐

Corollary 18.20 (Moreau) *Let $(f_i)_{i\in I}$ be a finite family of functions in $\Gamma_0(\mathcal{H})$, let $(\alpha_i)_{i\in I}$ be a finite family of real numbers in $]0,1]$ such that $\sum_{i\in I}\alpha_i = 1$, and set $q = (1/2)\|\cdot\|^2$. Then $\sum_{i\in I}\alpha_i \operatorname{Prox}_{f_i}$ is the proximity operator of a function in $\Gamma_0(\mathcal{H})$; more precisely,*

$$\sum_{i\in I}\alpha_i \operatorname{Prox}_{f_i} = \operatorname{Prox}_h, \quad \textit{where} \quad h = \Bigl(\sum_{i\in I}\alpha_i\bigl(f_i^* \,\Box\, q\bigr)\Bigr)^* - q. \tag{18.40}$$

Proof. Set $f = \sum_{i\in I}\alpha_i(f_i^* \,\Box\, q)$. We derive from Proposition 12.15, Proposition 8.17, and Proposition 12.30 that $f\colon \mathcal{H} \to \mathbb{R}$ is convex and Fréchet differentiable on $\mathcal{H}$. Moreover, it follows from (14.7) and Proposition 12.28 that $\nabla f = \sum_{i\in I}\alpha_i \operatorname{Prox}_{f_i}$ is nonexpansive as a convex combination of nonexpansive operators. Hence, using the implication (i)⇒(ix) in Theorem 18.15 with $\beta = 1$, we deduce that $\nabla f = \operatorname{Prox}_h$, where $h = f^* - q \in \Gamma_0(\mathcal{H})$, which gives (18.40). ☐

Remark 18.21 Suppose that $I = \{1,2\}$ and that $\alpha_1 = \alpha_2 = 1/2$ in Corollary 18.20. Then, in view of Corollary 14.8(iv), the function h in (18.40) is the proximal average $\operatorname{pav}(f_1, f_2)$.

18.6 Differentiability of the Distance to a Set

Proposition 18.22 *Let C be a nonempty closed convex subset of $\mathcal{H}$ and let $x \in C$. Then the following are equivalent:*

(i) *d_C is Gâteaux differentiable at x and $\nabla d_C(x) = 0$.*
(ii) *$x \notin \operatorname{spts} C$, i.e., $(\forall u \in \mathcal{H}\smallsetminus\{0\})\ \sigma_C(u) > \langle x \mid u\rangle$.*
(iii) *$T_C x = \mathcal{H}$, i.e., $\overline{\operatorname{cone}}\,(C - x) = \mathcal{H}$.*

Proof. (i)⇒(ii): Suppose that $u \in \mathcal{H}\smallsetminus\{0\}$. Since σ_C and $\langle\cdot \mid u\rangle$ are positively homogeneous, we assume that $\|u\| = 1$. Let $y \in \mathcal{H}$. Then, since $d_C(x) = 0$, Cauchy–Schwarz yields

$$\begin{aligned}
0 &= \langle y \mid \nabla d_C(x)\rangle \\
&= \lim_{\alpha\downarrow 0}\alpha^{-1}\bigl(d_C(x+\alpha y) - d_C(x)\bigr) \\
&= \lim_{\alpha\downarrow 0}\alpha^{-1}\inf_{z\in C}\|x+\alpha y - z\| \\
&\geqslant \varlimsup_{\alpha\downarrow 0}\alpha^{-1}\inf_{z\in C}\langle x+\alpha y - z \mid u\rangle
\end{aligned}$$

$$\begin{aligned} &= \langle y \mid u\rangle + \overline{\lim_{\alpha\downarrow 0}}(-\alpha^{-1}) \sup_{z\in C} \langle z - x \mid u\rangle \\ &= \langle y \mid u\rangle - \underline{\lim}_{\alpha\downarrow 0} \alpha^{-1}\big(\sigma_C(u) - \langle x \mid u\rangle\big) \\ &= \begin{cases} \langle y \mid u\rangle, & \text{if } \sigma_C(u) = \langle x \mid u\rangle; \\ -\infty, & \text{if } \sigma_C(u) > \langle x \mid u\rangle. \end{cases} \end{aligned} \tag{18.41}$$

Thus, if $\sigma_C(u) = \langle x \mid u\rangle$, then $\langle y \mid u\rangle \leqslant 0$. Since y is an arbitrary point in $\mathcal{H}$, we obtain $u = 0$, in contradiction to our hypothesis. Hence $\sigma_C(u) > \langle x \mid u\rangle$.

(ii)⇒(iii): Assume that $\overline{\operatorname{cone}}\,(C - x) = \overline{\bigcup_{\lambda\in\mathbb{R}_{++}} \lambda(C-x)} \neq \mathcal{H}$ and take $y \in \mathcal{H} \smallsetminus \overline{\operatorname{cone}}\,(C - x)$. By Theorem 3.50, there exists $u \in \mathcal{H} \smallsetminus \{0\}$ such that $\langle y \mid u\rangle > \sup_{\gamma\in\mathbb{R}_{++}} \gamma \sup \langle C - x \mid u\rangle = \sup_{\gamma\in\mathbb{R}_{++}} \big(\gamma(\sigma_C(u) - \langle x \mid u\rangle)\big)$. We deduce that $\sigma_C(u) = \langle x \mid u\rangle$, which is impossible.

(iii)⇒(i): It follows from Example 16.62 that $\operatorname{dom}\partial d_C = \mathcal{H}$. Hence, in view of Proposition 17.31(ii), it suffices to check that $\partial d_C(x) = \{0\}$. Take $u \in \partial d_C(x)$. Then $(\forall y \in C)$ $\langle y - x \mid u\rangle \leqslant d_C(y) - d_C(x) = 0$. Hence $\sup \langle C - x \mid u\rangle \leqslant 0$, which implies that $\sup_{\lambda\in\mathbb{R}_{++}} \sup \langle \lambda(C-x) \mid u\rangle \leqslant 0$ and thus that $\sup \langle \mathcal{H} \mid u\rangle = \sup \langle \overline{\operatorname{cone}}\,(C - x) \mid u\rangle \leqslant 0$. Therefore, $u = 0$. ☐

The next result complements Corollary 12.31 and Example 16.62.

Proposition 18.23 *Let C be a nonempty closed convex subset of $\mathcal{H}$ and let $x \in \mathcal{H}$. Then exactly one of the following holds:*

(i) *Suppose that $x \in \operatorname{int} C$. Then d_C is Fréchet differentiable at x with $\nabla d_C(x) = 0$.*

(ii) *Suppose that $x \in \operatorname{bdry} C$. Then the following hold:*

  (a) *Suppose that $x \notin \operatorname{spts} C$. Then d_C is not Fréchet differentiable at x, but d_C is Gâteaux differentiable at x with $\nabla d_C(x) = 0$.*

  (b) *Suppose that $x \in \operatorname{spts} C$. Then d_C is not Gâteaux differentiable at x and $\partial d_C(x) = N_C x \cap B(0;1)$.*

(iii) *Suppose that $x \notin C$. Then d_C is Fréchet differentiable at x with $\nabla d_C(x) = (x - P_C x)/d_C(x)$.*

Proof. (i): There exists $\varepsilon \in \mathbb{R}_{++}$ such that $d_C|_{B(x;\varepsilon)} = 0$, which implies that d_C is Fréchet differentiable at x with $\nabla d_C(x) = 0$.

(ii)(a): Proposition 18.22 implies that $\nabla d_C(x) = 0$. Since Theorem 7.4 yields $\overline{\operatorname{spts}}\, C = \operatorname{bdry} C$, there exists a sequence $(x_n)_{n\in\mathbb{N}}$ of support points of C such that $x_n \to x$. Let $(u_n)_{n\in\mathbb{N}}$ be such that $(\forall n \in \mathbb{N})$ $\sigma_C(u_n) = \langle x_n \mid u_n\rangle$ and $\|u_n\| = 1$. By Example 16.62, $(\forall n \in \mathbb{N})$ $u_n \in N_C x_n \cap B(0;1) = \partial d_C(x_n)$. Altogether, $(x_n, u_n)_{n\in\mathbb{N}}$ is a sequence in $\operatorname{gra}\partial d_C$ such that $x_n \to x$ and $u_n \not\to 0 = \nabla d_C(x)$. In view of Proposition 17.41, we conclude that d_C is not Fréchet differentiable at x.

(ii)(b): Let $u \in N_C x \smallsetminus \{0\}$. Then $\{0, u/\|u\|\} \subset N_C x \cap B(0;1) = \partial d_C(x)$ by Example 16.62. Thus, d_C is not Gâteaux differentiable at x by Proposition 17.31(i).

(iii): Set $f = \|\cdot\|$ and $g = \iota_C$. Then $d_C = f \,\square\, g$ and $d_C(x) = f(x - P_C x) + g(P_C x)$. Since $\|\cdot\|$ is Fréchet differentiable at $x - P_C x$ with $\nabla\|\cdot\|(x - P_C x) = (x - P_C x)/\|x - P_C x\|$ by Example 2.65, it follows from Proposition 18.7 that d_C is Fréchet differentiable at x and that $\nabla d_C(x) = (x - P_C x)/d_C(x)$. Alternatively, combine Example 16.62, Proposition 17.31(ii), Proposition 4.16, and Fact 2.62. $\square$

Exercises

Exercise 18.1 Let $(u_a)_{a\in A}$ be a family in $\mathcal{H}$, let $(\rho_a)_{a\in A}$ be a family in $]-\infty, +\infty]$, and set $f = \sup_{a\in A}(\langle\cdot \mid u_a\rangle - \rho_a)$. Show that $\operatorname{ran} \partial f \subset \overline{\operatorname{conv}}\, \{u_a\}_{a\in A}$.

Exercise 18.2 Let $f\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto |x|$. Compute $\partial f(0)$ (see also Example 16.32) in two different ways, using Theorem 18.5 and Proposition 18.23(ii)(b).

Exercise 18.3 Provide an example that illustrates the fact that Theorem 18.5 fails if the continuity assumption is omitted.

Exercise 18.4 Let $f \in \Gamma_0(\mathcal{H})$ and let $\gamma \in \mathbb{R}_{++}$. Use Corollary 18.8 to establish the Fréchet differentiability of the Moreau envelope ${}^{\gamma}f$ (see Proposition 12.30).

Exercise 18.5 By providing a counterexample, show that the equivalence in Corollary 18.11 fails if the assumption $\operatorname{dom} \partial f = \operatorname{int} \operatorname{dom} f$ is removed.

Exercise 18.6 Let $\mathcal{K}$ be a real Hilbert space, let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, let $r \in \mathcal{K}$, and let C be a nonempty closed convex subset of $\mathcal{H}$. Set $f\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \|Lx - r\|^2 + \|x\|^2 + d_C^2(x)$. Show that f is the Moreau envelope of a function in $\Gamma_0(\mathcal{H})$.

Exercise 18.7 (Legendre function) Let $f \in \Gamma_0(\mathcal{H})$ be such that $\operatorname{dom} \partial f = \operatorname{int} \operatorname{dom} f$ and $\operatorname{dom} \partial f^* = \operatorname{int} \operatorname{dom} f^*$. Suppose that f is Gâteaux differentiable on $\operatorname{int} \operatorname{dom} f$ and that f^* is Gâteaux differentiable on $\operatorname{int} \operatorname{dom} f^*$. These assumptions together mean that f is a *Legendre function*, as is f^*. Show that both f and f^* are strictly convex on the interior of their domains, respectively. Furthermore, show that $\nabla f\colon \operatorname{int} \operatorname{dom} f \to \operatorname{int} \operatorname{dom} f^*$ is a bijection, with inverse ∇f^*.

Exercise 18.8 Provide five examples of Legendre functions (see Exercise 18.7) and five examples of functions in $\Gamma_0(\mathcal{H})$ that are not Legendre functions.

Exercise 18.9 Let $A \in \mathcal{B}(\mathcal{H})$ be monotone, self-adjoint, and surjective. Using Proposition 17.36 and its notation, deduce from Exercise 18.7 that q_A is strictly convex, that q_A^* is strictly convex, and that A is bijective.

Exercise 18.10 Use Fact 2.63 and Corollary 12.31 to give an alternative proof of Proposition 18.23(iii).

Exercise 18.11 Use Example 7.7 and Proposition 18.23 to construct a function $f \in \Gamma_0(\mathcal{H})$ and a point $x \in \mathcal{H}$ such that $\operatorname{dom} f = \mathcal{H}$, f is Gâteaux differentiable at x, but f is not Fréchet differentiable at x.

Exercise 18.12 Suppose that $\mathcal{H}$ is infinite-dimensional, that C is as in Example 6.18, and that Q is as in Exercise 8.18. Set

$$f\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto m_Q\big(\|x\|^2, d_C(x)\big). \tag{18.42}$$

Prove the following:

(i) f is continuous, convex, and $\min f(\mathcal{H}) = 0 = f(0)$.
(ii) f is Gâteaux differentiable at 0 and $\nabla f(0) = 0$.
(iii) f is Fréchet differentiable on $\mathcal{H} \smallsetminus \{0\}$.
(iv) f is not Fréchet differentiable at 0.

Chapter 19
Duality in Convex Optimization

A convex optimization problem can be paired with a dual problem involving the conjugates of the functions appearing in its (primal) formulation. In this chapter, we study the interplay between primal and dual problems in the context of Fenchel–Rockafellar duality and, more generally, for bivariate functions. The latter approach leads naturally to saddle points and Lagrangians. Special attention is given to minimization under equality constraints and under inequality constraints. We start with a discussion of instances in which all primal solutions can be recovered from an arbitrary dual solution.

Throughout this chapter, $\mathcal{K}$ is a real Hilbert space.

19.1 Primal Solutions via Dual Solutions

In this section we investigate the connections between primal and dual solutions in the context of the minimization problem discussed in Definition 15.19. We recall that given $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$, $g\colon \mathcal{K} \to \left]-\infty, +\infty\right]$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, the primal problem associated with the function $f + g \circ L$ is

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x) + g(Lx), \tag{19.1}$$

its dual problem is

$$\underset{v \in \mathcal{K}}{\text{minimize}} \;\; f^*(-L^*v) + g^*(v), \tag{19.2}$$

the primal optimal value is $\mu = \inf\,(f + g \circ L)(\mathcal{H})$, and the dual optimal value is $\mu^* = \inf(f^{*\vee} \circ L^* + g^*)(\mathcal{K})$. We saw in Proposition 15.21 that $\mu \geqslant -\mu^*$. A solution to (19.1) is called a *primal solution*, and a solution to (19.2) is

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-48311-5_31

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_19

called a *dual solution*. As noted in Remark 15.20, in principle, dual solutions depend on the ordered triple (f, g, L), and we follow the common convention adopted in Definition 15.19.

Theorem 19.1 *Let $f \in \Gamma_0(\mathcal{H})$, let $g \in \Gamma_0(\mathcal{K})$, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$ be such that $\operatorname{dom} g \cap L(\operatorname{dom} f) \neq \varnothing$. Furthermore, set $\mu = \inf(f + g \circ L)(\mathcal{H})$, set $\mu^* = \inf(f^{*\vee} \circ L^* + g^*)(\mathcal{K})$, let $x \in \mathcal{H}$, and let $v \in \mathcal{K}$. Then the following are equivalent:*

(i) *x is a primal solution, v is a dual solution, and $\mu = -\mu^*$.*
(ii) *$-L^*v \in \partial f(x)$ and $v \in \partial g(Lx)$.*
(iii) *$x \in \partial f^*(-L^*v) \cap L^{-1}(\partial g^*(v))$.*

Proof. (i)⇔(ii): Since $f + g \circ L$ is proper, we derive from Proposition 13.15 and Theorem 16.29 the equivalences

$$\begin{aligned}
\text{(i)} &\Leftrightarrow f(x) + g(Lx) = \mu = -\mu^* = -\big(f^*(-L^*v) + g^*(v)\big) \\
&\Leftrightarrow \big(f(x) + f^*(-L^*v) - \langle x \mid -L^*v\rangle\,\big) + \big(g(Lx) + g^*(v) - \langle Lx \mid v\rangle\,\big) = 0 \\
&\Leftrightarrow -L^*v \in \partial f(x) \ \text{ and } v \in \partial g(Lx).
\end{aligned} \tag{19.3}$$

(ii)⇔(iii): Corollary 16.30. ☐

Corollary 19.2 *Let $f \in \Gamma_0(\mathcal{H})$, let $g \in \Gamma_0(\mathcal{K})$, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$ be such that $\inf(f + g \circ L)(\mathcal{H}) = -\min(f^{*\vee} \circ L^* + g^*)(\mathcal{K})$ (e.g., one of the conditions listed in Theorem 15.23, Proposition 15.24, or Fact 15.25 holds), and let v be an arbitrary solution to the dual problem (19.2). Then the (possibly empty) set of primal solutions is*

$$\operatorname{Argmin}(f + g \circ L) = \partial f^*(-L^*v) \cap L^{-1}\big(\partial g^*(v)\big). \tag{19.4}$$

Example 19.3 Let $z \in \mathcal{H}$, let $h \in \Gamma_0(\mathcal{H})$, and let I be a nonempty set. For every $i \in I$, let $\mathcal{K}_i$ be a real Hilbert space, let $r_i \in \mathcal{K}_i$, let $g_i \in \Gamma_0(\mathcal{K}_i)$, and suppose that $L_i \in \mathcal{B}(\mathcal{H},\mathcal{K}_i) \smallsetminus \{0\}$. Then we define the dual of the problem

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; h(x) - \langle x \mid z\rangle + \sum_{i \in I} g_i(L_i x - r_i) \tag{19.5}$$

as

$$\underset{(v_i)_{i \in I} \in \bigoplus_{i \in I} \mathcal{K}_i}{\text{minimize}} \;\; h^*\Big(z - \sum_{i \in I} L_i^* v_i\Big) + \sum_{i \in I} \big(g_i^*(v_i) + \langle v_i \mid r_i\rangle\,\big). \tag{19.6}$$

This dual is constructed to be the Fenchel–Rockafellar dual of (19.5) formulated in terms of $f = h - \langle \cdot \mid z\rangle$, $\mathcal{K} = \bigoplus_{i \in I} \mathcal{K}_i$, $L\colon \mathcal{H} \to \mathcal{K}\colon x \mapsto (L_i x)_{i \in I}$, and $g = \bigoplus_{i \in I} g_i(\cdot - r_i)$. In this setting, $L^*\colon \mathcal{K} \to \mathcal{H}\colon (y_i)_{i \in I} \mapsto \sum_{i \in I} L_i^* y_i$ and it follows from Proposition 13.23(iii) and Proposition 13.30 that (19.5) reduces to the primal problem (19.1) and that (19.6) reduces to the dual problem (19.2).

Next, we present an instance in which the primal solution is uniquely determined by a dual solution.

Proposition 19.4 *Let $f \in \Gamma_0(\mathcal{H})$, let $g \in \Gamma_0(\mathcal{K})$, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$ be such that the following hold:*

(i) $\inf(f + g \circ L)(\mathcal{H}) = -\min(f^{*\vee} \circ L^* + g^*)(\mathcal{K})$ *(e.g., one of the conditions listed in Theorem 15.23, Proposition 15.24, or Fact 15.25 holds).*

(ii) *There exists a solution v to the dual problem (19.2) such that f^* is Gâteaux differentiable at $-L^*v$.*

Then the primal problem (19.1) either has no solution or it has a unique solution, namely

$$x = \nabla f^*(-L^*v). \tag{19.7}$$

Proof. We derive from Corollary 19.2, (ii), and Proposition 17.31(i) that $\operatorname{Argmin}(f + g \circ L) = \partial f^*(-L^*v) \cap L^{-1}(\partial g^*(v)) \subset \{\nabla f^*(-L^*v)\}$, which yields (19.7). ☐

Here is an important application of Proposition 19.4. It implies that, if the objective function of the primal problem is strongly convex, then the dual problem can be formulated in terms of a Moreau envelope (see Definition 12.20).

Proposition 19.5 *Let $\varphi \in \Gamma_0(\mathcal{H})$, let $\psi \in \Gamma_0(\mathcal{K})$, let $z \in \mathcal{H}$, let $r \in \mathcal{K}$, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$ be such that $r \in \operatorname{sri}(L(\operatorname{dom}\varphi) - \operatorname{dom}\psi)$. Consider the problem*

$$\underset{x\in\mathcal{H}}{\text{minimize}}\ \ \varphi(x) + \psi(Lx - r) + \tfrac{1}{2}\|x - z\|^2, \tag{19.8}$$

together with the problem

$$\underset{v\in\mathcal{K}}{\text{minimize}}\ \ {}^{1}(\varphi^*)(z - L^*v) + \psi^*(v) + \langle v \mid r\rangle\,, \tag{19.9}$$

which can also be written as

$$\underset{v\in\mathcal{K}}{\text{minimize}}\ \ \tfrac{1}{2}\|z - L^*v\|^2 - {}^{1}\varphi(z - L^*v) + \psi^*(v) + \langle v \mid r\rangle\,. \tag{19.10}$$

Then (19.9) has at least one solution and, if v is a solution to (19.9), then

$$x = \operatorname{Prox}_\varphi(z - L^*v) \tag{19.11}$$

is the unique solution to (19.8).

Proof. The fact that (19.8) possesses a unique solution follows from Definition 12.23, and the fact that (19.9) and (19.10) are identical from (14.6). Now set $f = \varphi + (1/2)\|\cdot - z\|^2$ and $g = \psi(\cdot - r)$. Then (19.1) reduces to (19.8). Moreover, using Proposition 14.1 and Proposition 13.23(iii), we obtain $f^*\colon u \mapsto {}^{1}(\varphi^*)(u + z) - (1/2)\|z\|^2$ and $g^*\colon v \mapsto \psi^*(v) + \langle r \mid v\rangle$, so that (19.2) reduces to (19.9). Furthermore, it follows from (14.7) that $\nabla f^*\colon u \mapsto \mathrm{Prox}_\varphi(u+z)$. Thus, the result follows from Proposition 19.4 and Theorem 15.23. ☐

Remark 19.6 The (primal) problem (19.8) involves the sum of a composite convex function and of another convex function. Such problems are not easy to solve. By contrast, the (dual) problem (19.9) involves a Moreau envelope, i.e., an everywhere defined differentiable convex function with a Lipschitz continuous gradient (see Proposition 12.30). As will be seen in Corollary 28.9, such problems are much easier to solve (see also Section 28.7).

Corollary 19.7 *Let $\varphi \in \Gamma_0(\mathcal{H})$, let $z \in \mathcal{H}$, let $r \in \mathcal{K}$, let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$ be such that $\|L\| \leqslant 1$ and $r \in \mathrm{sri}(L(\mathrm{dom}\,\varphi))$, and set $T\colon \mathcal{K} \to \mathcal{K}\colon v \mapsto r + v - L(\mathrm{Prox}_\varphi(z - L^*v))$. Then T is firmly nonexpansive, $\mathrm{Fix}\,T \neq \varnothing$, and, for every $v \in \mathrm{Fix}\,T$, the unique solution to the problem*

$$\underset{\substack{x\in\mathcal{H}\\ Lx=r}}{\text{minimize}} \;\; \varphi(x) + \frac{1}{2}\|x - z\|^2 \tag{19.12}$$

is $x = \mathrm{Prox}_\varphi(z - L^*v)$.

Proof. Set $S = \big\{v \in \mathcal{K} \;\big|\; L(\mathrm{Prox}_\varphi(z - L^*v)) = r\big\}$. The result follows from Proposition 19.5 with $\psi = \iota_{\{0\}}$. Indeed, (19.8) turns into (19.12), and the function to minimize in (19.9) is $v \mapsto {}^1(\varphi^*)(z - L^*v) + \langle v \mid r\rangle$, the gradient of which is $v \mapsto -L(\mathrm{Prox}_\varphi(z - L^*v)) + r$ by (14.7). Hence, S is the nonempty set of solutions of (19.9). Since $v \mapsto -L(\mathrm{Prox}_\varphi(z - L^*v))$ is firmly nonexpansive by Exercise 4.10, we deduce that T is firmly nonexpansive and that $\mathrm{Fix}\,T = S$. The conclusion follows from Proposition 19.5. ☐

We close this section with several applications of Proposition 19.5.

Example 19.8 Let K be a nonempty closed convex cone in $\mathcal{H}$, let $\psi \in \Gamma_0(\mathcal{K})$ be positively homogeneous, set $D = \partial\psi(0)$, let $z \in \mathcal{H}$, let $r \in \mathcal{K}$, and suppose that $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$ is such that $r \in \mathrm{sri}(L(K) - \mathrm{dom}\,\psi)$. Consider the problem

$$\underset{x\in K}{\text{minimize}} \;\; \psi(Lx - r) + \frac{1}{2}\|x - z\|^2, \tag{19.13}$$

together with the problem

$$\underset{v\in D}{\text{minimize}} \;\; \frac{1}{2}d^2_{K^\ominus}(z - L^*v) + \langle v \mid r\rangle\,. \tag{19.14}$$

Then (19.14) has at least one solution and, if v is a solution to (19.14), then the unique solution to (19.13) is $x = P_K(z - L^*v)$.

Proof. This is an application of Proposition 19.5 with $\varphi = \iota_K$. Indeed, Example 13.3(ii) yields $\varphi^* = \iota_{K^\ominus}$, and Proposition 16.24 yields $\psi^* = \sigma_D^* = \iota_D$. ☐

Example 19.9 Let K be a nonempty closed convex cone in $\mathcal{H}$, let $z \in \mathcal{H}$, let $r \in \mathcal{K}$, and let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$. Consider the problem

$$\underset{x \in K}{\text{minimize}} \;\; \|Lx - r\| + \tfrac{1}{2}\|x - z\|^2, \tag{19.15}$$

together with the problem

$$\underset{v \in \mathcal{K},\, \|v\| \leqslant 1}{\text{minimize}} \;\; \tfrac{1}{2} d^2_{K^\ominus}(z - L^* v) + \langle v \mid r \rangle\,. \tag{19.16}$$

Then (19.16) has at least one solution and, if v is a solution to (19.16), then the unique solution to (19.15) is $x = P_K(z - L^* v)$.

Proof. This is a special case of Example 19.8 with $\psi = \|\cdot\|$. Indeed, $L(K) - \operatorname{dom} \psi = L(K) - \mathcal{K} = \mathcal{K}$ and Example 16.32 yields $D = \partial\psi(0) = B(0; 1)$. ☐

Example 19.10 Let C and D be closed convex subsets of $\mathcal{H}$, let $z \in \mathcal{H}$, and let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ be such that $0 \in \operatorname{sri}(L(C) - D)$. Consider the best approximation problem

$$\underset{x \in C, Lx \in D}{\text{minimize}} \;\; \|x - z\|, \tag{19.17}$$

together with the problem

$$\underset{v \in \mathcal{K}}{\text{minimize}} \;\; \tfrac{1}{2}\|z - L^* v\|^2 - \tfrac{1}{2} d^2_C(z - L^* v) + \sigma_D(v). \tag{19.18}$$

Then (19.18) has at least one solution and, if v is a solution to (19.18), then the unique solution to (19.17) is $x = P_C(z - L^* v)$.

Proof. Apply Proposition 19.5 with $\varphi = \iota_C$, $\psi = \iota_D$, and $r = 0$. ☐

19.2 Parametric Duality

We explore an abstract duality framework defined on the product $\mathcal{H} \times \mathcal{K}$.

Definition 19.11 Let $F\colon \mathcal{H} \times \mathcal{K} \to \left]-\infty, +\infty\right]$. The associated *primal problem* is

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; F(x, 0), \tag{19.19}$$

the associated *dual problem* is

$$\underset{v \in \mathcal{K}}{\text{minimize}} \;\; F^*(0, v), \tag{19.20}$$

and the associated *value function* is

$$\vartheta\colon \mathcal{K} \to [-\infty, +\infty] : y \mapsto \inf F(\mathcal{H}, y). \tag{19.21}$$

Furthermore, $x \in \mathcal{H}$ is a *primal solution* if it solves (19.19), and $v \in \mathcal{K}$ is a *dual solution* if it solves (19.20).

Proposition 19.12 *Let* $F\colon \mathcal{H} \times \mathcal{K} \to \left]-\infty, +\infty\right]$ *and let* ϑ *be the associated value function. Then the following hold:*

(i) $\vartheta^* = F^*(0, \cdot)$.
(ii) $-\inf F^*(0, \mathcal{K}) = \vartheta^{**}(0) \leqslant \vartheta(0) = \inf F(\mathcal{H}, 0)$.

Proof. We assume that F is proper.
(i): Apply Proposition 13.33.
(ii): The first equality follows from (i) and Proposition 13.10(i), and the second equality follows from (19.21). The inequality follows from Proposition 13.16(i) applied to ϑ. ☐

The next result gives a condition under which the inequality in Proposition 19.12(ii) becomes an equality.

Proposition 19.13 *Let* $F \in \Gamma_0(\mathcal{H} \oplus \mathcal{K})$ *and suppose that the associated value function* ϑ *is lower semicontinuous at* 0 *with* $\vartheta(0) \in \mathbb{R}$. *Then*

$$\inf F(\mathcal{H}, 0) = -\inf F^*(0, \mathcal{K}) \in \mathbb{R}. \tag{19.22}$$

Proof. Since F is convex, it follows from Proposition 8.35 that ϑ is convex. Hence, Proposition 13.44 yields $\vartheta^{**}(0) = \vartheta(0)$ and the result follows. ☐

We now describe the set of dual solutions using the subdifferential of the (biconjugate of the) value function at 0.

Proposition 19.14 *Let* $F\colon \mathcal{H} \times \mathcal{K} \to \left]-\infty, +\infty\right]$, *suppose that the associated value function* ϑ *is convex and that* ϑ^* *is proper, and denote the set of dual solutions by* U. *Then the following hold:*

(i) $U = \partial\vartheta^{**}(0)$.
(ii) $\partial\vartheta(0) \neq \varnothing$ *if and only if* ϑ *is lower semicontinuous at* 0 *with* $\vartheta(0) \in \mathbb{R}$ *and* $U \neq \varnothing$, *in which case* $U = \partial\vartheta(0)$ *and* $\inf F(\mathcal{H}, 0) = -\min F^*(0, \mathcal{K}) \in \mathbb{R}$.

Proof. (i): In view of Proposition 19.12(i), Theorem 16.3, and Corollary 16.30, we have $U = \operatorname{Argmin} \vartheta^* = (\partial\vartheta^*)^{-1}(0) = \partial\vartheta^{**}(0)$.
(ii): Suppose that $\partial\vartheta(0) \neq \varnothing$. Using Proposition 16.4(i), Proposition 16.5, (i), and Proposition 19.13, we obtain $0 \in \operatorname{dom} \vartheta$, $\vartheta^{**}(0) = \vartheta(0)$, $U = \partial\vartheta^{**}(0) = \partial\vartheta(0) \neq \varnothing$, and hence $\inf F(\mathcal{H}, 0) = -\min F^*(0, \mathcal{K}) \in \mathbb{R}$. Now assume that $0 \in \operatorname{dom} \vartheta$, that ϑ is lower semicontinuous at 0, and that $U \neq \varnothing$. By Proposition 13.44 and (i), $\vartheta(0) = \vartheta^{**}(0)$ and $U = \partial\vartheta^{**}(0) \neq \varnothing$. Now take $u \in \partial\vartheta^{**}(0)$. Since $\vartheta^{**} \leqslant \vartheta$, it follows that $(\forall x \in \mathcal{H})$ $\langle x \mid u\rangle \leqslant \vartheta^{**}(x) - \vartheta^{**}(0) \leqslant \vartheta(x) - \vartheta(0)$. Therefore, $\varnothing \neq U = \partial\vartheta^{**}(0) \subset \partial\vartheta(0)$. ☐

Proposition 19.15 *Let* $F \in \Gamma_0(\mathcal{H} \oplus \mathcal{K})$ *and let* $(x, v) \in \mathcal{H} \times \mathcal{K}$. *Then the following are equivalent:*

(i) *x is a primal solution, v is a dual solution, and* (19.22) *holds.*
(ii) $F(x,0)+F^*(0,v)=0$.
(iii) $(0,v)\in\partial F(x,0)$.
(iv) $(x,0)\in\partial F^*(0,v)$.

Proof. (i)⇒(ii): $F(x,0)=\inf F(\mathcal{H},0)=-\inf F^*(0,\mathcal{K})=-F^*(0,v)\in\mathbb{R}$. Hence $F(x,0)+F^*(0,v)=0$.

(ii)⇒(i): By Proposition 19.12(ii),

$$-F^*(0,v)\leqslant-\inf F^*(0,\mathcal{K})\leqslant\inf F(\mathcal{H},0)\leqslant F(x,0). \tag{19.23}$$

(ii)⇔(iii)⇔(iv): It follows from Theorem 16.29 that (ii) ⇔ $F(x,0)+F^*(0,v)=\langle(x,0)\mid(0,v)\rangle$ ⇔ (iii) ⇔ (iv). ☐

Definition 19.16 Let $F\colon\mathcal{H}\times\mathcal{K}\to\left]-\infty,+\infty\right]$. The *Lagrangian* of F is the function

$$\mathcal{L}\colon\mathcal{H}\times\mathcal{K}\to[-\infty,+\infty]:(x,v)\mapsto\inf_{y\in\mathcal{K}}\big(F(x,y)-\langle y\mid v\rangle\big). \tag{19.24}$$

Moreover, $(x,v)\in\mathcal{H}\times\mathcal{K}$ is a *saddle point* of $\mathcal{L}$ if

$$\sup\mathcal{L}(x,\mathcal{K})=\mathcal{L}(x,v)=\inf\mathcal{L}(\mathcal{H},v). \tag{19.25}$$

Proposition 19.17 *Let $F\colon\mathcal{H}\times\mathcal{K}\to\left]-\infty,+\infty\right]$ and let $\mathcal{L}\colon\mathcal{H}\times\mathcal{K}\to[-\infty,+\infty]$ be its Lagrangian. Then the following hold:*

(i) *For every $x\in\mathcal{H}$, $\mathcal{L}(x,\cdot)$ is upper semicontinuous and concave.*
(ii) *Suppose that $F\in\Gamma_0(\mathcal{H}\oplus\mathcal{K})$. Then $(\forall x\in\mathcal{H})$ $\sup\mathcal{L}(x,\mathcal{K})=F(x,0)$.*
(iii) *Suppose that F is convex. Then, for every $v\in\mathcal{K}$, $\mathcal{L}(\cdot,v)$ is convex.*
(iv) $(\forall v\in\mathcal{K})$ $\inf\mathcal{L}(\mathcal{H},v)=-F^*(0,v)$.

Proof. (i): Let $x\in\mathcal{H}$ and set $\varphi=F(x,\cdot)$. Then, for every $v\in\mathcal{K}$,

$$\mathcal{L}(x,v)=\inf_{y\in\mathcal{K}}\big(F(x,y)-\langle y\mid v\rangle\big)=-\sup_{y\in\mathcal{K}}\big(\langle y\mid v\rangle-\varphi(y)\big)=-\varphi^*(v). \tag{19.26}$$

Since φ^* is lower semicontinuous and convex by Proposition 13.13, we deduce that $\mathcal{L}(x,\cdot)=-\varphi^*$ is upper semicontinuous and concave.

(ii): Let $x\in\mathcal{H}$ and set $\varphi=F(x,\cdot)$. Then either $\varphi\equiv+\infty$ or $\varphi\in\Gamma_0(\mathcal{K})$. It follows from Corollary 13.38 that $\varphi^{**}=\varphi$. In view of (19.26) and Proposition 13.10(i), we obtain $\sup\mathcal{L}(x,\mathcal{K})=-\inf\varphi^*(\mathcal{K})=\varphi^{**}(0)=\varphi(0)=F(x,0)$.

(iii): Let $v\in\mathcal{K}$. Since $(x,y)\mapsto F(x,y)-\langle y\mid v\rangle$ is convex, the conclusion follows from Proposition 8.35.

(iv): For every $x\in\mathcal{H}$, $\mathcal{L}(x,v)=-\sup_{y\in\mathcal{K}}(\langle x\mid 0\rangle+\langle y\mid v\rangle-F(x,y))$. Thus, $\inf\mathcal{L}(\mathcal{H},v)=-\sup_{x\in\mathcal{H}}\sup_{y\in\mathcal{K}}(\langle(x,y)\mid(0,v)\rangle-F(x,y))=-F^*(0,v)$. ☐

Corollary 19.18 *Let $F\colon \mathcal{H}\times\mathcal{K}\to\,]-\infty,+\infty]$, let $\mathcal{L}\colon \mathcal{H}\times\mathcal{K}\to[-\infty,+\infty]$ be its Lagrangian, and let $v\in\mathcal{K}$. Then v is a dual solution if and only if*

$$\inf \mathcal{L}(\mathcal{H},v)=\sup_{w\in\mathcal{K}}\,\inf\,\mathcal{L}(\mathcal{H},w). \tag{19.27}$$

Proof. By Proposition 19.17(iv),

$$\begin{aligned} v \text{ is a dual solution} &\Leftrightarrow\; -F^*(0,v)=-\inf_{w\in\mathcal{K}}F^*(0,w)\\ &\Leftrightarrow\; \inf\mathcal{L}(\mathcal{H},v)=\sup_{w\in\mathcal{K}}-F^*(0,w)\\ &\Leftrightarrow\; \inf\mathcal{L}(\mathcal{H},v)=\sup_{w\in\mathcal{K}}\inf\mathcal{L}(\mathcal{H},w), \end{aligned} \tag{19.28}$$

and the proof is complete. ☐

Corollary 19.19 *Let $F\in\Gamma_0(\mathcal{H}\oplus\mathcal{K})$, let $\mathcal{L}$ be its Lagrangian, and let $(x,v)\in\mathcal{H}\times\mathcal{K}$. Then the following are equivalent:*

(i) *x is a primal solution, v is a dual solution, and $\inf F(\mathcal{H},0)=-\inf F^*(0,\mathcal{K})\in\mathbb{R}$.*
(ii) *$F(x,0)+F^*(0,v)=0$.*
(iii) *$(0,v)\in\partial F(x,0)$.*
(iv) *$(x,0)\in\partial F^*(0,v)$.*
(v) *(x,v) is a saddle point of $\mathcal{L}$.*

Proof. (i)⇔(ii)⇔(iii)⇔(iv): Proposition 19.15.
(ii)⇒(v): By Proposition 19.17(iv)&(ii),

$$F(x,0)=-F^*(0,v)=\inf\mathcal{L}(\mathcal{H},v)\leqslant\mathcal{L}(x,v)\leqslant\sup\mathcal{L}(x,\mathcal{K})=F(x,0). \tag{19.29}$$

(v)⇒(ii): Using Proposition 19.17(iv)&(ii), we get

$$-F^*(0,v)=\inf\mathcal{L}(\mathcal{H},v)=\mathcal{L}(x,v)=\sup\mathcal{L}(x,\mathcal{K})=F(x,0). \tag{19.30}$$

Hence, since F is proper, $-F^*(0,v)=F(x,0)\in\mathbb{R}$. ☐

The following proposition illustrates how a bivariate function can be associated with a minimization problem. The second variable of this function plays the role of a perturbation. In this particular case, we recover the Fenchel–Rockafellar duality framework discussed in Section 15.3 and in Section 19.1.

Proposition 19.20 *Let $f\in\Gamma_0(\mathcal{H})$, let $g\in\Gamma_0(\mathcal{K})$, and let $L\in\mathcal{B}(\mathcal{H},\mathcal{K})$ be such that $\operatorname{dom} g\cap L(\operatorname{dom} f)\neq\varnothing$. Set*

$$F\colon\mathcal{H}\times\mathcal{K}\to\,]-\infty,+\infty]:(x,y)\mapsto f(x)+g(Lx+y). \tag{19.31}$$

Then the following hold:

(i) $F \in \Gamma_0(\mathcal{H} \oplus \mathcal{K})$.
(ii) *The primal problem is*

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x) + g(Lx). \tag{19.32}$$

(iii) *The dual problem is*

$$\underset{v \in \mathcal{K}}{\text{minimize}} \;\; f^*(-L^*v) + g^*(v). \tag{19.33}$$

(iv) *The Lagrangian is*

$$\begin{aligned} \mathcal{L}\colon \mathcal{H} \times \mathcal{K} &\to [-\infty, +\infty] \\ (x, v) &\mapsto \begin{cases} -\infty, & \textit{if } x \in \operatorname{dom} f \textit{ and } v \notin \operatorname{dom} g^*; \\ f(x) + \langle Lx \mid v\rangle - g^*(v), & \textit{if } x \in \operatorname{dom} f \textit{ and } v \in \operatorname{dom} g^*; \\ +\infty, & \textit{if } x \notin \operatorname{dom} f. \end{cases} \end{aligned} \tag{19.34}$$

(v) *Suppose that the optimal values* μ *of* (19.32) *and* μ^* *of* (19.33) *satisfy* $\mu = -\mu^* \in \mathbb{R}$, *and let* $(\bar{x}, \bar{v}) \in \mathcal{H} \times \mathcal{K}$. *Then* $(\bar{x}, \bar{v})$ *is a saddle point of* $\mathcal{L}$ *if and only if*

$$-L^*\bar{v} \in \partial f(\bar{x}) \quad \textit{and} \quad L\bar{x} \in \partial g^*(\bar{v}). \tag{19.35}$$

Proof. (i): This follows from the assumptions on f, g, and L.

(ii): This follows from (19.19) and (19.31).

(iii): Let $v \in \mathcal{K}$. Then

$$\begin{aligned} F^*(0, v) &= \sup_{(x,y) \in \mathcal{H} \times \mathcal{K}} \big(\langle y \mid v\rangle - f(x) - g(Lx + y) \big) \\ &= \sup_{x \in \mathcal{H}} \Big(\langle x \mid -L^*v\rangle - f(x) + \sup_{z \in \mathcal{K}} \big(\langle z \mid v\rangle - g(z) \big) \Big) \\ &= \big(f^* \circ L^*\big)(-v) + g^*(v). \end{aligned} \tag{19.36}$$

Hence, the result follows from (19.20).

(iv): For every $(x, v) \in \mathcal{H} \times \mathcal{K}$, we derive from (19.24) and (19.31) that

$$\begin{aligned} \mathcal{L}(x, v) &= f(x) + \inf_{y \in \mathcal{K}} \big(g(Lx + y) - \langle y \mid v\rangle \big) \\ &= \begin{cases} f(x) + \langle Lx \mid v\rangle + \inf_{z \in \mathcal{K}} \big(g(z) - \langle z \mid v\rangle \big), & \text{if } x \in \operatorname{dom} f; \\ +\infty, & \text{if } x \notin \operatorname{dom} f \end{cases} \\ &= \begin{cases} f(x) + \langle Lx \mid v\rangle - \sup_{z \in \mathcal{K}} \big(\langle z \mid v\rangle - g(z) \big), & \text{if } x \in \operatorname{dom} f; \\ +\infty, & \text{if } x \notin \operatorname{dom} f \end{cases} \end{aligned}$$

$$= \begin{cases} f(x) + \langle Lx \mid v\rangle - g^*(v), & \text{if } x \in \operatorname{dom} f;\\ +\infty, & \text{if } x \notin \operatorname{dom} f, \end{cases} \tag{19.37}$$

which yields (19.34).

(v): Since $F \in \Gamma_0(\mathcal{H} \oplus \mathcal{K})$, we derive from Corollary 19.19 that $(\bar{x}, \bar{v})$ is a saddle point of $\mathcal{L}$ if and only if $\bar{x}$ is a solution to (19.32) and $\bar{v}$ is a solution to (19.33). Hence, the conclusion follows from Theorem 19.1. ☐

19.3 Minimization under Equality Constraints

In this section, we apply the setting of Section 19.2 to convex optimization problems with affine equality constraints.

Proposition 19.21 *Let $f \in \Gamma_0(\mathcal{H})$, let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, and let $r \in L(\operatorname{dom} f)$. Set*

$$F\colon \mathcal{H} \times \mathcal{K} \to \left]-\infty, +\infty\right] : (x, y) \mapsto \begin{cases} f(x), & \text{if } Lx = r - y;\\ +\infty, & \text{if } Lx \neq r - y. \end{cases} \tag{19.38}$$

Then the following hold:

(i) $F \in \Gamma_0(\mathcal{H} \oplus \mathcal{K})$.

(ii) *The primal problem is*

$$\underset{x\in\mathcal{H}, Lx=r}{\text{minimize}} \;\; f(x). \tag{19.39}$$

(iii) *The dual problem is*

$$\underset{v\in\mathcal{K}}{\text{minimize}} \;\; f^*(-L^*v) + \langle v \mid r\rangle\,. \tag{19.40}$$

(iv) *The Lagrangian is*

$$\begin{aligned} &\mathcal{L}\colon \mathcal{H} \times \mathcal{K} \to \left]-\infty, +\infty\right]\\ &\qquad (x, v) \mapsto \begin{cases} f(x) + \langle Lx - r \mid v\rangle\,, & \text{if } x \in \operatorname{dom} f;\\ +\infty, & \text{if } x \notin \operatorname{dom} f. \end{cases} \end{aligned} \tag{19.41}$$

(v) *Suppose that the optimal values μ of (19.39) and μ^* of (19.40) satisfy $\mu = -\mu^* \in \mathbb{R}$, and let $(\bar{x}, \bar{v}) \in \mathcal{H} \times \mathcal{K}$. Then $(\bar{x}, \bar{v})$ is a saddle point of $\mathcal{L}$ if and only if $-L^*\bar{v} \in \partial f(\bar{x})$ and $L\bar{x} = r$, in which case*

$$f(\bar{x}) = \min f\big(L^{-1}(\{r\})\big) = \min \mathcal{L}(\mathcal{H}, \bar{v}). \tag{19.42}$$

Proof. All the results, except (19.42), are obtained by setting $g = \iota_{\{r\}}$ in Proposition 19.20. To prove (19.42), note that Corollary 19.19 and Proposition 19.17 yield $f(\bar{x}) = F(\bar{x}, 0) = -F^*(0, \bar{v}) = \inf \mathcal{L}(\mathcal{H}, \bar{v}) = \mathcal{L}(\bar{x}, \bar{v})$. ☐

Remark 19.22 In the setting of Proposition 19.21, let $(\bar{x}, \bar{v})$ be a saddle point of the Lagrangian (19.41). Then $\bar{v}$ is called a *Lagrange multiplier* associated with the solution $\bar{x}$ to (19.39). In view of (19.42) and (19.41), $\bar{x}$ solves the unconstrained minimization problem

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x) + \langle x \mid L^* \bar{v} \rangle \, . \tag{19.43}$$

An important application of Proposition 19.21 is in the context of problems involving finitely many scalar affine equalities.

Corollary 19.23 *Let $f \in \Gamma_0(\mathcal{H})$, let m be a strictly positive integer, set $I = \{1, \ldots, m\}$, let $(u_i)_{i \in I} \in \mathcal{H}^m$, and let $(\rho_i)_{i \in I} \in \big\{ (\langle x \mid u_i \rangle)_{i \in I} \;\big|\; x \in \operatorname{dom} f \big\}$. Set*

$$F \colon \mathcal{H} \times \mathbb{R}^m \to \left]-\infty, +\infty\right] \tag{19.44}$$

$$\big(x, (\eta_i)_{i \in I}\big) \mapsto \begin{cases} f(x), & \text{if } (\forall i \in I) \;\; \langle x \mid u_i \rangle = \rho_i - \eta_i; \\ +\infty, & \text{otherwise.} \end{cases} \tag{19.45}$$

Then the following hold:

(i) $F \in \Gamma_0(\mathcal{H} \oplus \mathbb{R}^m)$.
(ii) *The primal problem is*

$$\underset{\substack{x \in \mathcal{H} \\ \langle x \mid u_1 \rangle = \rho_1, \ldots, \langle x \mid u_m \rangle = \rho_m}}{\text{minimize}} \qquad f(x). \tag{19.46}$$

(iii) *The dual problem is*

$$\underset{(\nu_i)_{i \in I} \in \mathbb{R}^m}{\text{minimize}} \;\; f^*\Big(-\sum_{i \in I} \nu_i u_i \Big) + \sum_{i \in I} \nu_i \rho_i. \tag{19.47}$$

(iv) *The Lagrangian is*

$$\mathcal{L} \colon \mathcal{H} \times \mathbb{R}^m \to [-\infty, +\infty]$$
$$\big(x, (\nu_i)_{i \in I}\big) \mapsto \begin{cases} f(x) + \displaystyle\sum_{i \in I} \nu_i \big(\langle x \mid u_i \rangle - \rho_i \big), & \text{if } x \in \operatorname{dom} f; \\ +\infty, & \text{if } x \notin \operatorname{dom} f. \end{cases} \tag{19.48}$$

(v) *Suppose that the optimal values μ of (19.46) and μ^* of (19.47) satisfy $\mu = -\mu^* \in \mathbb{R}$, and let $(\bar{x}, (\bar{\nu}_i)_{i \in I}) \in \mathcal{H} \times \mathbb{R}^m$. Then $(\bar{x}, (\bar{\nu}_i)_{i \in I})$ is a saddle point of $\mathcal{L}$ if and only if*

$$-\sum_{i \in I} \bar{\nu}_i u_i \in \partial f(\bar{x}) \quad \text{and} \quad (\forall i \in I) \quad \langle \bar{x} \mid u_i \rangle = \rho_i, \tag{19.49}$$

in which case $\bar{x}$ is a solution to both (19.46) and the unconstrained minimization problem

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; f(x)+\sum_{i\in I}\bar{\nu}_i\,\langle x \mid u_i\rangle. \tag{19.50}$$

Proof. Apply Proposition 19.21 with $\mathcal{K}=\mathbb{R}^m$, $r=(\rho_i)_{i\in I}$, and $L\colon x\mapsto(\langle x\mid u_i\rangle)_{i\in I}$. ☐

19.4 Minimization under Inequality Constraints

To use a fairly general notion of inequality, we require the following notion of convexity with respect to a cone.

Definition 19.24 Let K be a nonempty closed convex cone in $\mathcal{K}$ and let $R\colon\mathcal{H}\to\mathcal{K}$. Then R is *convex with respect to K* if

$$(\forall x\in\mathcal{H})(\forall y\in\mathcal{H})(\forall\alpha\in\,]0,1[)\quad R\big(\alpha x+(1-\alpha)y\big)-\alpha Rx-(1-\alpha)Ry\in K. \tag{19.51}$$

Proposition 19.25 *Let $f\in\Gamma_0(\mathcal{H})$, let K be a nonempty closed convex cone in $\mathcal{K}$, and let $R\colon\mathcal{H}\to\mathcal{K}$ be continuous, convex with respect to K, and such that $K\cap R(\operatorname{dom} f)\neq\varnothing$. Set*

$$F\colon\mathcal{H}\times\mathcal{K}\to\,]-\infty,+\infty]\colon(x,y)\mapsto\begin{cases}f(x), & \text{if } Rx\in -y+K;\\ +\infty, & \text{if } Rx\notin -y+K.\end{cases} \tag{19.52}$$

Then the following hold:

(i) $F\in\Gamma_0(\mathcal{H}\oplus\mathcal{K})$.
(ii) *The primal problem is*

$$\underset{x\in\mathcal{H},\,Rx\in K}{\text{minimize}} \;\; f(x). \tag{19.53}$$

(iii) *The dual problem is*

$$\underset{v\in K^{\ominus}}{\text{minimize}} \;\; \varphi(v),\quad \text{where}\quad \varphi\colon v\mapsto\sup_{x\in\mathcal{H}}\big(-\langle Rx\mid v\rangle-f(x)\big). \tag{19.54}$$

(iv) *The Lagrangian is*

$$\mathcal{L}\colon\mathcal{H}\times\mathcal{K}\to[-\infty,+\infty]$$
$$(x,v)\mapsto\begin{cases}-\infty, & \text{if } x\in\operatorname{dom} f \text{ and } v\notin K^{\ominus};\\ f(x)+\langle Rx\mid v\rangle, & \text{if } x\in\operatorname{dom} f \text{ and } v\in K^{\ominus};\\ +\infty, & \text{if } x\notin\operatorname{dom} f.\end{cases} \tag{19.55}$$

(v) *Let* $(\bar{x},\bar{v}) \in \mathcal{H} \times \mathcal{K}$*. Then* $(\bar{x},\bar{v})$ *is a saddle point of* $\mathcal{L}$ *if and only if*

$$\begin{cases} \bar{x} \in \operatorname{dom} f, \\ R\bar{x} \in K, \\ \bar{v} \in K^{\ominus}, \\ f(\bar{x}) = \inf\limits_{x\in\mathcal{H}} \big(f(x) + \langle Rx \mid \bar{v}\rangle \big), \end{cases} \tag{19.56}$$

in which case $\langle R\bar{x} \mid \bar{v}\rangle = 0$ *and* $\bar{x}$ *is a primal solution.*

Proof. (i): The assumptions imply that F is proper and that $F_1\colon (x,y) \mapsto f(x)$ is lower semicontinuous and convex. Hence, by Corollary 9.4, it remains to check that $F_2\colon (x,y) \mapsto \iota_K(Rx+y)$ is likewise, which amounts to showing that $C = \big\{(x,y) \in \mathcal{H}\times\mathcal{K} \;\big|\; Rx+y \in K\big\}$ is closed and convex. The closedness of C follows from the continuity of R and the closedness of K. Now take (x_1,y_1) and (x_2,y_2) in C, and take $\alpha \in \left]0,1\right[$. Then $Rx_1+y_1 \in K$, $Rx_2+y_2 \in K$, and, by convexity of K, $\alpha Rx_1 + (1-\alpha)Rx_2 + (\alpha y_1 + (1-\alpha)y_2) = \alpha(Rx_1 + y_1) + (1-\alpha)(Rx_2+y_2) \in K$. On the other hand, it follows from (19.51) that $R(\alpha x_1 + (1-\alpha)x_2) - \alpha Rx_1 - (1-\alpha)Rx_2 \in K$. Adding these two inclusions yields $R(\alpha x_1 + (1-\alpha)x_2) + (\alpha y_1 + (1-\alpha)y_2) \in K + K = K$. We conclude that $\alpha(x_1,y_1) + (1-\alpha)(x_2,y_2) \in C$ and hence that C is convex.

(ii): This follows from (19.19) and (19.52).

(iii): Let $v \in \mathcal{K}$. Then

$$\begin{aligned} F^*(0,v) &= \sup_{(x,y)\in\mathcal{H}\times\mathcal{K}} \big(\langle y \mid v\rangle - f(x) - \iota_K(Rx+y) \big) \\ &= \sup_{x\in\mathcal{H}} \Bigg(\langle Rx \mid -v\rangle - f(x) + \sup_{\substack{y\in\mathcal{K}\\ Rx+y\in K}} \langle Rx+y \mid v\rangle \Bigg) \\ &= \sup_{x\in\mathcal{H}} \Big(\langle Rx \mid -v\rangle - f(x) + \sup_{z\in K} \langle z \mid v\rangle \Big) \\ &= \sup_{x\in\mathcal{H}} \big(-\langle Rx \mid v\rangle - f(x)\big) + \iota_{K^{\ominus}}(v). \end{aligned} \tag{19.57}$$

Hence, the result follows from (19.20).

(iv): Let $(x,v) \in \mathcal{H}\times\mathcal{K}$. Then (19.24) and (19.52) yield

$$\mathcal{L}(x,v) = \inf_{y\in\mathcal{K}} \big(f(x) + \iota_K(Rx+y) - \langle y \mid v\rangle \big). \tag{19.58}$$

If $x \notin \operatorname{dom} f$, then, for every $y \in \mathcal{K}$, $f(x) + \iota_K(Rx+y) - \langle y \mid v\rangle = +\infty$ and, therefore, $\mathcal{L}(x,v) = +\infty$. Now suppose that $x \in \operatorname{dom} f$. Then $f(x) \in \mathbb{R}$ and we derive from (19.58) that

$$\begin{aligned} \mathcal{L}(x,v) &= f(x) + \inf_{\substack{y\in\mathcal{K}\\ Rx+y\in K}} \langle y \mid -v\rangle \\ &= f(x) + \langle Rx \mid v\rangle - \sup_{z\in K} \langle z \mid v\rangle \\ &= f(x) + \langle Rx \mid v\rangle - \iota_{K^{\ominus}}(v), \end{aligned} \tag{19.59}$$

which yields (19.55).

(v): We derive from (19.52) and (19.57) that

$$F(\bar{x},0)+F^*(0,\bar{v})=0 \Leftrightarrow \begin{cases} \bar{x}\in \operatorname{dom} f,\\ R\bar{x}\in K,\\ \bar{v}\in K^{\ominus},\\ f(\bar{x})+\sup\limits_{x\in\mathcal{H}}\big(-\langle Rx \mid \bar{v}\rangle - f(x)\big)=0 \end{cases}$$
$$\Leftrightarrow (19.56). \tag{19.60}$$

Using the equivalence (v)⇔(ii) in Corollary 19.19, we obtain the first result. Next, we derive from (19.56) that $\langle R\bar{x} \mid \bar{v}\rangle \leqslant 0$ and $f(\bar{x}) \leqslant f(\bar{x}) + \langle R\bar{x} \mid \bar{v}\rangle$. Hence, since $f(\bar{x})\in\mathbb{R}$, $\langle R\bar{x} \mid \bar{v}\rangle = 0$. Finally, $\bar{x}$ is a primal solution by Corollary 19.19. ☐

Remark 19.26 In the setting of Proposition 19.25, suppose that $(\bar{x},\bar{v})$ is a saddle point of the Lagrangian (19.55). Then $\bar{v}$ is called a *Lagrange multiplier* associated with the solution $\bar{x}$ to (19.53). It follows from Proposition 19.25(v) that $\bar{x}$ solves the unconstrained minimization problem

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f(x)+\langle Rx \mid \bar{v}\rangle . \tag{19.61}$$

Conditions for the existence of Lagrange multipliers will be discussed in Section 27.3, in Section 27.4, and in Section 27.5. For specific examples, see Chapter 29.

Example 19.27 Let $f\in\Gamma_0(\mathcal{H})$, let $z\in\mathcal{H}$, and let K be a closed convex cone in $\mathcal{H}$ such that $z\in(\operatorname{dom} f)-K$. Set

$$F\colon \mathcal{H}\times\mathcal{H}\to\left]-\infty,+\infty\right] : (x,y)\mapsto \begin{cases} f(x), & \text{if } x\in -y+z+K;\\ +\infty, & \text{if } x\notin -y+z+K. \end{cases} \tag{19.62}$$

Then the following hold:

(i) $F\in\Gamma_0(\mathcal{H}\oplus\mathcal{H})$.
(ii) The primal problem is

$$\underset{x\in z+K}{\text{minimize}}\;\; f(x). \tag{19.63}$$

(iii) The dual problem is

$$\underset{u\in K^{\ominus}}{\text{minimize}}\;\; f^*(-u)+\langle z \mid u\rangle . \tag{19.64}$$

(iv) The Lagrangian is

$$\mathcal{L}\colon \mathcal{H}\times\mathcal{H}\to[-\infty,+\infty]$$
$$(x,u)\mapsto \begin{cases} -\infty, & \text{if } x\in\operatorname{dom} f \text{ and } u\notin K^{\ominus};\\ f(x)+\langle x-z \mid u\rangle, & \text{if } x\in\operatorname{dom} f \text{ and } u\in K^{\ominus};\\ +\infty, & \text{if } x\notin\operatorname{dom} f. \end{cases} \tag{19.65}$$

(v) Let $(\bar{x}, \bar{u}) \in \mathcal{H} \times \mathcal{H}$. Then $(\bar{x}, \bar{u})$ is a saddle point of $\mathcal{L}$ if and only if

$$\begin{cases} \bar{x} \in (z + K) \cap \operatorname{dom} f, \\ \bar{u} \in K^{\ominus}, \\ \bar{x} \in \operatorname{Argmin} \big(f + \langle \cdot \mid \bar{u} \rangle \big), \end{cases} \tag{19.66}$$

in which case $\langle \bar{x} \mid \bar{u} \rangle = \langle z \mid \bar{u} \rangle$ and $\bar{x}$ is a primal solution.

Proof. Apply Proposition 19.25 to $\mathcal{K} = \mathcal{H}$ and $R\colon x \mapsto x - z$. ☐

Example 19.28 Let $\phi \in \Gamma_0(\mathbb{R})$ be such that $0 \in \operatorname{dom} \phi$, set $\gamma = \inf \phi(\mathbb{R})$, and set

$$\begin{aligned} F\colon \mathbb{R}^2 \times \mathbb{R} &\to \left]-\infty, +\infty\right] \\ \big((\xi_1, \xi_2), y\big) &\mapsto \begin{cases} \phi(\xi_2), & \text{if } \ \xi_1 - \|(\xi_1, \xi_2)\| \geqslant y; \\ +\infty, & \text{otherwise.} \end{cases} \end{aligned} \tag{19.67}$$

Then the following hold:

(i) $F \in \Gamma_0(\mathbb{R}^2 \times \mathbb{R})$.
(ii) The primal problem is

$$\underset{\substack{(\xi_1,\xi_2)\in\mathbb{R}^2 \\ \|(\xi_1,\xi_2)\| \leqslant \xi_1}}{\operatorname{minimize}} \quad \phi(\xi_2), \tag{19.68}$$

the optimal value of (19.68) is $\mu = \phi(0)$, and the set of primal solutions is $\mathbb{R}_+ \times \{0\}$.
(iii) The dual problem is

$$\underset{v \in \mathbb{R}_+}{\operatorname{minimize}} \quad \sup_{(\xi_1,\xi_2)\in\mathbb{R}^2} \Big(v\big(\xi_1 - \|(\xi_1, \xi_2)\|\big) - \phi(\xi_2) \Big), \tag{19.69}$$

the optimal value of (19.69) is $\mu^* = -\gamma$, and the set of dual solutions is $\mathbb{R}_+$.
(iv) The Lagrangian is

$$\begin{aligned} \mathcal{L}\colon \mathbb{R}^2 \times \mathbb{R} &\to [-\infty, +\infty] \\ \big((\xi_1, \xi_2), v\big) &\mapsto \begin{cases} -\infty, & \text{if } \ \xi_2 \in \operatorname{dom} \phi \text{ and } v < 0; \\ \phi(\xi_2) + v\big(\|(\xi_1, \xi_2)\| - \xi_1\big), & \text{if } \ \xi_2 \in \operatorname{dom} \phi \text{ and } v \geqslant 0; \\ +\infty, & \text{if } \ \xi_2 \notin \operatorname{dom} \phi. \end{cases} \end{aligned} \tag{19.70}$$

(v) The Lagrangian $\mathcal{L}$ has a saddle point if and only if $\phi(0) = \gamma$, in which case the set of saddle points is $(\mathbb{R}_+ \times \{0\}) \times \mathbb{R}_+$.

(vi) The value function (see (19.21)) is

$$\vartheta\colon \mathbb{R} \to [-\infty,+\infty] : y \mapsto \begin{cases} \gamma, & \text{if } y<0;\\ \phi(0), & \text{if } y=0;\\ +\infty, & \text{if } y>0. \end{cases} \tag{19.71}$$

Proof. Suppose that $\mathcal{H}=\mathbb{R}^2$, and set $f\colon \mathbb{R}^2 \to \left]-\infty,+\infty\right] : (\xi_1,\xi_2) \mapsto \phi(\xi_2)$, $\mathcal{K}=\mathbb{R}$, $K=\mathbb{R}_-$, and $R\colon \mathbb{R}^2 \to \mathbb{R}\colon (\xi_1,\xi_2) \mapsto \|(\xi_1,\xi_2)\|-\xi_1$. Then $f \in \Gamma_0(\mathcal{H})$, $(0,0) \in \mathbb{R}\times \operatorname{dom}\phi = \operatorname{dom} f$, and hence $0 \in K \cap R(\operatorname{dom} f)$. Moreover, the bivariate function F in (19.67) is that deriving from (19.52).

(i): Proposition 19.25(i).

(ii): Proposition 19.25(ii) yields (19.68). Now let $(\xi_1,\xi_2)\in\mathbb{R}^2$. Since $0 \leqslant \|(\xi_1,\xi_2)\|-\xi_1$, it follows that $\|(\xi_1,\xi_2)\| \leqslant \xi_1 \Leftrightarrow \|(\xi_1,\xi_2)\| = \xi_1 \Leftrightarrow (\xi_1,\xi_2) \in \mathbb{R}_+ \times \{0\}$, in which case $f(\xi_1,\xi_2)=\phi(0)$.

(iii): Proposition 19.25(iii) yields (19.69). Let us determine

$$\varphi\colon \mathbb{R}_+ \to \left]-\infty,+\infty\right] : v \mapsto \sup_{(\xi_1,\xi_2)\in\mathbb{R}^2} \Big(v\big(\xi_1 - \|(\xi_1,\xi_2)\|\big) - \phi(\xi_2)\Big). \tag{19.72}$$

Let $v\in\mathbb{R}_+$ and let $(\xi_1,\xi_2)\in\mathbb{R}^2$. Then $v(\xi_1-\|(\xi_1,\xi_2)\|)-\phi(\xi_2) \leqslant -\phi(\xi_2) \leqslant -\gamma$ and thus $\varphi(v) \leqslant -\gamma$. On the other hand, since $\|(\xi_1,\xi_2)\|-\xi_1 \to 0$ as $\xi_1 \to +\infty$, we deduce that $\varphi(v)=-\gamma$.

(iv): Proposition 19.25(iv).

(v): This follows from Proposition 19.25(v) and computation of φ in (iii).

(vi): The details are left as Exercise 19.9. ☐

Remark 19.29 Consider the setting of Example 19.28. If $\gamma<\phi(0)$, then $\mathcal{L}$ has no saddle point and it follows from Corollary 19.19 that the duality gap is nonzero, even though primal and dual solutions exist. The choice $\phi=\exp^\vee$, for which $\gamma=0<1=\phi(0)$, then leads to *Duffin's duality gap*.

The application presented next is referred to as the *convex programming problem*.

Corollary 19.30 *Let $f\in\Gamma_0(\mathcal{H})$, let m and p be integers such that $m>p>0$, set $I=\{1,\ldots,p\}$, set $J=\{p+1,\ldots,m\}$, let $(g_i)_{i\in I}$ be real-valued functions in $\Gamma_0(\mathcal{H})$, suppose that $(u_j)_{j\in J}$ are vectors in $\mathcal{H}$, and let $(\rho_j)_{j\in J}\in\mathbb{R}^{m-p}$ be such that*

$$\Big\{ x\in\operatorname{dom} f \;\Big|\; \max_{i\in I} g_i(x) \leqslant 0 \textit{ and } \max_{j\in J} |\langle x \mid u_j\rangle - \rho_j| = 0 \Big\} \neq \varnothing. \tag{19.73}$$

Set

$$F\colon \mathcal{H}\times\mathbb{R}^m \to \left]-\infty,+\infty\right]$$

$$\big(x, (\eta_i)_{i\in I\cup J}\big) \mapsto \begin{cases} f(x), & \text{if } \begin{cases} (\forall i \in I) \;\; g_i(x) \leqslant -\eta_i, \\ (\forall j \in J) \;\; \langle x \mid u_j\rangle = -\eta_j + \rho_j; \end{cases} \\ +\infty, & \text{otherwise.} \end{cases} \tag{19.74}$$

Then the following hold:

(i) $F \in \Gamma_0(\mathcal{H} \oplus \mathbb{R}^m)$.

(ii) *The primal problem is*

$$\underset{\substack{x\in\mathcal{H}\\ g_1(x)\leqslant 0,\ldots,g_p(x)\leqslant 0\\ \langle x|u_{p+1}\rangle=\rho_{p+1},\ldots,\langle x|u_m\rangle=\rho_m}}{\text{minimize}} \quad f(x). \tag{19.75}$$

(iii) *The dual problem is*

$$\underset{\substack{(\nu_i)_{i\in I}\in\mathbb{R}^p_+\\ (\nu_j)_{j\in J}\in\mathbb{R}^{m-p}}}{\text{minimize}} \quad \sup_{x\in\mathcal{H}} \Big(-\sum_{i\in I} \nu_i g_i(x) + \sum_{j\in J} \nu_j \big(\rho_j - \langle x \mid u_j\rangle\big) - f(x)\Big). \tag{19.76}$$

(iv) *The Lagrangian is*

$$\begin{aligned} \mathcal{L}\colon \mathcal{H}\times\mathbb{R}^m &\to [-\infty,+\infty] \\ \big(x, (\nu_i)_{i\in I\cup J}\big) &\mapsto \begin{cases} -\infty, & \text{if } x \in \operatorname{dom} f \text{ and } (\nu_i)_{i\in I} \notin \mathbb{R}^p_+; \\ f(x) + \displaystyle\sum_{i\in I} \nu_i g_i(x) + \sum_{j\in J} \nu_j\big(\langle x \mid u_j\rangle - \rho_j\big), & \\ \qquad\qquad \text{if } x \in \operatorname{dom} f \text{ and } (\nu_i)_{i\in I} \in \mathbb{R}^p_+; & \\ +\infty, & \text{if } x \notin \operatorname{dom} f. \end{cases} \end{aligned} \tag{19.77}$$

(v) *Suppose that* $(\bar{x}, (\bar{\nu}_i)_{i\in I\cup J})$ *is a saddle point of* $\mathcal{L}$. *Then* $\bar{x}$ *is a solution to both* (19.75) *and the unconstrained minimization problem*

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; f(x) + \sum_{i\in I} \bar{\nu}_i g_i(x) + \sum_{j\in J} \bar{\nu}_j \langle x \mid u_j\rangle. \tag{19.78}$$

Moreover,

$$(\forall i \in I) \quad \begin{cases} \bar{x} \in \operatorname{dom} f, \\ g_i(\bar{x}) \leqslant 0, \\ \bar{\nu}_i \geqslant 0, \\ \bar{\nu}_i g_i(\bar{x}) = 0. \end{cases} \tag{19.79}$$

Proof. Apply Proposition 19.25 with $\mathcal{K} = \mathbb{R}^m$, $K = \mathbb{R}^p_- \times \{0\}^{m-p}$ (hence $K^\ominus = \mathbb{R}^p_+ \times \mathbb{R}^{m-p}$), and

$$R\colon x \mapsto \big(g_1(x), \ldots, g_p(x), \langle x \mid u_{p+1}\rangle - \rho_{p+1}, \ldots, \langle x \mid u_m\rangle - \rho_m\big). \tag{19.80}$$

Note that the continuity of R follows from Corollary 8.39(ii). ☐

Remark 19.31 As in Remark 19.26, the parameters $(\bar{\nu}_i)_{i\in I}$ in Corollary 19.30(v) are the Lagrange multipliers associated with the solution $\bar{x}$ to (19.75). Condition (19.79) on the multipliers $(\bar{\nu}_i)_{i\in I}$ corresponding to the inequality constraints is a *complementary slackness* condition. Note that, if for some $i \in I$, $g_i(\bar{x}) < 0$, then $\bar{\nu}_i = 0$.

Exercises

Exercise 19.1 Let $F \in \Gamma_0(\mathcal{H} \oplus \mathcal{K})$ and suppose that

$$\inf F(\mathcal{H}, 0) = -\inf F^*(0, \mathcal{K}) \in \mathbb{R}. \tag{19.81}$$

Show that the associated value function defined in (19.21) is lower semicontinuous at 0.

Exercise 19.2 Let $F \in \Gamma_0(\mathcal{H} \oplus \mathcal{K})$ with associated value function ϑ. Show that $\inf F(\mathcal{H}, 0) = -\inf F^*(0, \mathcal{K}) \in \mathbb{R}$ if and only if $\vartheta(0) \in \mathbb{R}$ and ϑ is lower semicontinuous at 0.

Exercise 19.3 Let $\mathcal{L}\colon \mathcal{H} \times \mathcal{K} \to [-\infty, +\infty]$, let $(\bar{x}, \bar{v}) \in \mathcal{H} \times \mathcal{K}$, and denote the set of saddle points of $\mathcal{L}$ by S. Prove the following:

(i) $\sup\limits_{v\in\mathcal{K}} \inf\limits_{x\in\mathcal{H}} \mathcal{L}(x, v) \leqslant \inf\limits_{x\in\mathcal{H}} \sup\limits_{v\in\mathcal{K}} \mathcal{L}(x, v)$.

(ii) $(\bar{x}, \bar{v}) \in S \Leftrightarrow \inf\limits_{x\in\mathcal{H}} \mathcal{L}(x, \bar{v}) = \sup\limits_{v\in\mathcal{K}} \mathcal{L}(\bar{x}, v)$.

(iii) Suppose that $(\bar{x}, \bar{v}) \in S$. Then

$$\sup_{v\in\mathcal{K}} \inf_{x\in\mathcal{H}} \mathcal{L}(x, v) = \mathcal{L}(\bar{x}, \bar{v}) = \inf_{x\in\mathcal{H}} \sup_{v\in\mathcal{K}} \mathcal{L}(x, v). \tag{19.82}$$

Exercise 19.4 Let $f\colon \mathcal{H} \to \mathbb{R}$, and let $K \in \{\{0\}, \mathbb{R}_+, \mathbb{R}_-, \mathbb{R}\}$. In each case, what does it mean to say that f is convex with respect to K (see Definition 19.24)?

Exercise 19.5 Derive Proposition 19.25(iii) from Proposition 19.17(iv) and Proposition 19.25(iv).

Exercise 19.6 Recover Corollary 19.23 from Corollary 19.30.

Exercise 19.7 In the setting of Proposition 19.25, prove that v is a dual solution if and only if

$$\inf_{x \in \operatorname{dom} f} \mathcal{L}(x, v) = \sup_{w \in K^{\ominus}} \inf_{x \in \operatorname{dom} f} \mathcal{L}(x, w). \tag{19.83}$$

Exercise 19.8 Consider Corollary 19.30 when $\mathcal{H} = \mathbb{R}$, $f\colon x \mapsto x$, $p = 1$, $m = 2$, $g_1\colon x \mapsto x^2$, $u_2 = 0$, and $\rho_2 = 0$. Determine F, the primal problem, the dual problem, the Lagrangian, all saddle points, and the value function.

Exercise 19.9 Check the details in Example 19.28(vi).

Exercise 19.10 In the setting of Example 19.28(vi), determine when the value function ϑ is lower semicontinuous at 0 and when $\partial\vartheta(0) \neq \varnothing$. Is it possible that $\vartheta(0) = -\infty$?

Chapter 20
Monotone Operators

The theory of monotone set-valued operators plays a central role in many areas of nonlinear analysis. A prominent example of a monotone operator is the subdifferential operator investigated in Chapter 16. Single-valued monotone operators will be seen to be closely related to the firmly nonexpansive operators studied in Chapter 4. Our investigation of monotone operators will rely heavily on the Fitzpatrick function.

The conventions introduced in Section 1.2 will be used. In particular, an operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ such that, for every $x \in \mathcal{H}$, Ax is either empty or a singleton will be identified with the corresponding (at most) single-valued operator. Conversely, if D is a nonempty subset of $\mathcal{H}$ and $T\colon D \to \mathcal{H}$, then T will be identified with the set-valued operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$, where, for every $x \in \mathcal{H}$, $Ax = \{Tx\}$ if $x \in D$, and $Ax = \varnothing$ otherwise.

20.1 Monotone Operators

Definition 20.1 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$. Then A is *monotone* if

$$\big(\forall (x,u) \in \operatorname{gra} A\big)\big(\forall (y,v) \in \operatorname{gra} A\big) \quad \langle x-y \mid u-v\rangle \geqslant 0. \tag{20.1}$$

A subset of $\mathcal{H} \times \mathcal{H}$ is *monotone* if it is the graph of a monotone operator.

Proposition 20.2 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$. Then the following are equivalent:*

(i) *A is monotone.*
(ii) *A is* accretive*, i.e.,*

$$\big(\forall (x,u) \in \operatorname{gra} A\big)\big(\forall (y,v) \in \operatorname{gra} A\big)\big(\forall \alpha \in [0,1]\,\big) \\ \|x-y+\alpha(u-v)\| \geqslant \|x-y\|. \tag{20.2}$$

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-48311-5_31

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_20

(iii) *The following holds:*

$$\big(\forall (x,u) \in \operatorname{gra} A\big)\big(\forall (y,v) \in \operatorname{gra} A\big)$$
$$\|y-u\|^2 + \|x-v\|^2 \geqslant \|x-u\|^2 + \|y-v\|^2. \quad (20.3)$$

Proof. (20.2)⇔(20.1): This follows from (20.1) and Lemma 2.13(i).
(20.3)⇔(20.1): Use Lemma 2.12(i). ☐

Example 20.3 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper. Then ∂f is monotone.

Proof. Take (x,u) and (y,v) in $\operatorname{gra}\partial f$. Then, by (16.1), $\langle x-y \mid u\rangle + f(y) \geqslant f(x)$ and $\langle y-x \mid v\rangle + f(x) \geqslant f(y)$. Adding these inequalities, we conclude that $\langle x-y \mid u-v\rangle \geqslant 0$. ☐

Example 20.4 Suppose that $\mathcal{H} = \mathbb{R}$, let D be a nonempty subset of $\mathcal{H}$, and let $A\colon D \to \mathcal{H}$ be increasing. Then A is monotone.

Example 20.5 Let D be a nonempty subset of $\mathcal{H}$ and let $T\colon D \to \mathcal{H}$ be cocoercive (in particular, firmly nonexpansive). Then T is monotone.

Proof. See (4.12). ☐

Example 20.6 Let D be a nonempty subset of $\mathcal{H}$ and let $T\colon D \to \mathcal{H}$ be α-averaged, with $\alpha \in \left]0,1/2\right]$. Then T is monotone.

Proof. Combine Remark 4.37 with Example 20.5, or use the equivalence (i)⇔(iv) in Proposition 4.35. ☐

Example 20.7 Let D be a nonempty subset of $\mathcal{H}$, let $T\colon D \to \mathcal{H}$ be nonexpansive, let $\alpha \in [-1,1]$, and set $A = \operatorname{Id} + \alpha T$. Then, for every $x \in D$ and every $y \in D$,

$$\begin{aligned}\langle x-y \mid Ax-Ay\rangle &= \|x-y\|^2 + \alpha\,\langle x-y \mid Tx-Ty\rangle \\ &\geqslant \|x-y\|\big(\|x-y\| - |\alpha|\,\|Tx-Ty\|\big) \\ &\geqslant 0. \end{aligned} \quad (20.4)$$

Consequently, A is monotone.

Example 20.8 Let D be a nonempty subset of $\mathcal{H}$, let $T\colon D \to \mathcal{H}$, and set $A = \operatorname{Id} - T$. Then the following are equivalent:

(i) T is *pseudononexpansive* (or *pseudocontractive*), i.e.,

$$(\forall x \in D)(\forall y \in D) \quad \|Tx-Ty\|^2 \leqslant$$
$$\|x-y\|^2 + \|(\operatorname{Id}-T)x - (\operatorname{Id}-T)y\|^2. \quad (20.5)$$

(ii) A is monotone.

Proof. Take x and y in D. Then $\|x-y\|^2+\|(\mathrm{Id}-T)x-(\mathrm{Id}-T)y\|^2 \geqslant \|Tx-Ty\|^2 \Leftrightarrow \|x-y\|^2+\|x-y\|^2-2\langle x-y \mid Tx-Ty\rangle+\|Tx-Ty\|^2 \geqslant \|Tx-Ty\|^2 \Leftrightarrow \|x-y\|^2-\langle x-y \mid Tx-Ty\rangle \geqslant 0 \Leftrightarrow \langle x-y \mid Ax-Ay\rangle \geqslant 0$. ☐

Example 20.9 Let H be a separable real Hilbert space, let $T \in \mathbb{R}_{++}$, and suppose that $\mathcal{H} = L^2([0,T];\mathsf{H})$ (see Example 2.8). Furthermore, let A be the time-derivative operator (see Example 2.10)

$$A\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x \mapsto \begin{cases} \{x'\}, & \text{if } x \in D; \\ \varnothing, & \text{otherwise,} \end{cases} \tag{20.6}$$

where (initial condition)

$$D = \big\{x \in W^{1,2}([0,T];\mathsf{H}) \;\big|\; x(0) = \mathsf{x}_0\big\} \quad \text{for some } \mathsf{x}_0 \in \mathsf{H} \tag{20.7}$$

or (periodicity condition)

$$D = \big\{x \in W^{1,2}([0,T];\mathsf{H}) \;\big|\; x(0) = x(T)\big\}. \tag{20.8}$$

Then A is monotone.

Proof. Take x and y in $D = \operatorname{dom} A$. Then

$$\begin{aligned} \langle x-y \mid Ax-Ay\rangle &= \int_0^T \langle x(t)-y(t) \mid x'(t)-y'(t)\rangle_{\mathsf{H}}\, dt \\ &= \frac{1}{2}\int_0^T \frac{d\|x(t)-y(t)\|_{\mathsf{H}}^2}{dt}\, dt \\ &= \frac{1}{2}\big(\|x(T)-y(T)\|_{\mathsf{H}}^2 - \|x(0)-y(0)\|_{\mathsf{H}}^2\big) \\ &\geqslant 0, \end{aligned} \tag{20.9}$$

which shows that A is monotone. ☐

Further examples can be constructed via the following monotonicity-preserving operations.

Proposition 20.10 *Let $\mathcal{K}$ be a real Hilbert space, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{K} \to 2^{\mathcal{K}}$ be monotone operators, let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$, and let $\gamma \in \mathbb{R}_+$. Then the operators A^{-1}, γA, and $A + L^*BL$ are monotone.*

Proposition 20.11 *Let $(\Omega,\mathcal{F},\mu)$ be a measure space, let $(\mathsf{H},\langle\cdot \mid \cdot\rangle_{\mathsf{H}})$ be a separable real Hilbert space, and let $\mathsf{A}\colon \mathsf{H} \to 2^{\mathsf{H}}$ be a monotone operator. Suppose that $\mathcal{H} = L^2((\Omega,\mathcal{F},\mu);\mathsf{H})$ and define $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ via*

$$\operatorname{gra} A = \big\{(x,u) \in \mathcal{H}\times\mathcal{H} \;\big|\; \big(x(\omega),u(\omega)\big) \in \operatorname{gra}\mathsf{A} \;\; \mu\text{-a.e. on } \Omega\big\}. \tag{20.10}$$

Then A is monotone.

Proof. Take (x,u) and (y,v) in $\operatorname{gra} A$. By monotonicity of A,

$$\langle x(\omega) - y(\omega) \mid u(\omega) - v(\omega)\rangle_{\mathsf{H}} \geqslant 0 \quad \mu\text{-a.e. on } \Omega. \tag{20.11}$$

In view of Example 2.6, integrating these inequalities over Ω with respect to μ, we obtain

$$\langle x - y \mid u - v\rangle = \int_\Omega \langle x(\omega) - y(\omega) \mid u(\omega) - v(\omega)\rangle_{\mathsf{H}}\, \mu(d\omega) \geqslant 0, \tag{20.12}$$

which shows that A is monotone. □

Monotone operators also arise naturally in the study of best approximation and farthest-point problems.

Example 20.12 Let C be a nonempty subset of $\mathcal{H}$ and let Π_C be the set-valued projector onto C defined in (3.17). Then Π_C is monotone.

Proof. Take (x,u) and (y,v) in $\operatorname{gra} \Pi_C$. Then $\|x-u\| = d_C(x) \leqslant \|x-v\|$ and $\|y-v\| = d_C(y) \leqslant \|y-u\|$. Hence $\|x-u\|^2 + \|y-v\|^2 \leqslant \|x-v\|^2 + \|y-u\|^2$. Now expand the squares and simplify to obtain the monotonicity of Π_C. □

Example 20.13 Let C be a nonempty bounded subset of $\mathcal{H}$ and denote by

$$\Phi_C \colon \mathcal{H} \to 2^{\mathcal{H}} \colon x \mapsto \big\{u \in C \;\big|\; \|x-u\| = \sup \|x - C\|\big\} \tag{20.13}$$

its *farthest-point operator*. Then $-\Phi_C$ is monotone.

Proof. Suppose that (x,u) and (y,v) are in $\operatorname{gra} \Phi_C$. Then $\|x-u\| \geqslant \|x-v\|$ and $\|y-v\| \geqslant \|y-u\|$. Hence $\|x-u\|^2 + \|y-v\|^2 \geqslant \|x-v\|^2 + \|y-u\|^2$. Now expand the squares and simplify to see that $-\Phi_C$ is monotone. □

Next, we provide two characterizations of monotonicity.

Proposition 20.14 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and set $F = \langle\cdot \mid \cdot\rangle$. Then the following are equivalent:*

(i) *A is monotone.*
(ii) *For all finite families $(\alpha_i)_{i\in I}$ in $]0,1[$ such that $\sum_{i\in I} \alpha_i = 1$ and $(x_i,u_i)_{i\in I}$ in* $\operatorname{gra} A$*, we have*

$$F\Big(\sum_{i\in I} \alpha_i (x_i, u_i)\Big) \leqslant \sum_{i\in I} \alpha_i F(x_i, u_i). \tag{20.14}$$

(iii) *F is convex on* $\operatorname{gra} A$.

Proof. This follows from Lemma 2.14(i). □

Example 20.15 If $A\colon \mathcal{H} \to \mathcal{H}$ is linear in Definition 20.1, we recover Definition 2.23(i).

We devote the remainder of this section to linear monotone operators.

Example 20.16 Let $A \in \mathcal{B}(\mathcal{H})$. Then the following hold:

(i) A is monotone $\Leftrightarrow$ $A + A^*$ is monotone $\Leftrightarrow$ A^* is monotone.
(ii) A^*A, AA^*, $A - A^*$, and $A^* - A$ are monotone.

Proposition 20.17 *Let $A \in \mathcal{B}(\mathcal{H})$ be monotone. Then* $\ker A = \ker A^*$ *and* $\overline{\operatorname{ran}}\, A = \overline{\operatorname{ran}}\, A^*$.

Proof. Take $x \in \ker A$ and $v \in \operatorname{ran} A$, say $v = Ay$. Then $(\forall \alpha \in \mathbb{R})$ $0 \leqslant \langle \alpha x + y \mid A(\alpha x + y)\rangle = \alpha \langle x \mid v\rangle + \langle y \mid Ay\rangle$. Hence $\langle x \mid v\rangle = 0$ and thus $\ker A \subset (\operatorname{ran} A)^{\perp} = \ker A^*$ by Fact 2.25(v). Since $A^* \in \mathcal{B}(\mathcal{H})$ is also monotone, we obtain $\ker A^* \subset \ker A^{**} = \ker A$. Altogether, $\ker A = \ker A^*$ and therefore $\overline{\operatorname{ran}}\, A = \overline{\operatorname{ran}}\, A^*$ by Fact 2.25(iv). ☐

Fact 20.18 (See [305, Chapter VII]) Let A and B be self-adjoint monotone operators in $\mathcal{B}(\mathcal{H})$ such that $AB = BA$. Then AB is monotone.

Example 20.19 Set

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \ B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \ \text{and} \ \ C = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \tag{20.15}$$

Then A, B, C, and $-C$ are continuous, linear, and monotone operators from $\mathbb{R}^2$ to $\mathbb{R}^2$. However, neither AB nor C^2 is monotone. This shows that the assumptions of self-adjointness and commutativity in Fact 20.18 are important.

20.2 Maximally Monotone Operators

Definition 20.20 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be monotone. Then A is *maximally monotone* (or *maximal monotone*) if there exists no monotone operator $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ such that $\operatorname{gra} B$ properly contains $\operatorname{gra} A$, i.e., for every $(x,u) \in \mathcal{H} \times \mathcal{H}$,

$$(x,u) \in \operatorname{gra} A \quad \Leftrightarrow \quad (\forall (y,v) \in \operatorname{gra} A) \ \ \langle x - y \mid u - v\rangle \geqslant 0. \tag{20.16}$$

Theorem 20.21 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be monotone. Then there exists a* maximally monotone extension *of A, i.e., a maximally monotone operator $\widetilde{A}\colon \mathcal{H} \to 2^{\mathcal{H}}$ such that* $\operatorname{gra} A \subset \operatorname{gra} \widetilde{A}$.

Proof. Assume first that $\operatorname{gra} A \neq \varnothing$ and set

$$\mathcal{M} = \big\{B\colon \mathcal{H} \to 2^{\mathcal{H}} \ \big|\ B \text{ is monotone and } \operatorname{gra} A \subset \operatorname{gra} B\big\}. \tag{20.17}$$

Then $\mathcal{M}$ is nonempty and partially ordered via $(\forall B_1 \in \mathcal{M})(\forall B_2 \in \mathcal{M})$ $B_1 \preccurlyeq B_2 \Leftrightarrow \operatorname{gra} B_1 \subset \operatorname{gra} B_2$. Let $\mathcal{C}$ be a chain in $\mathcal{M}$. Then the operator the graph of which is $\bigcup_{C\in\mathcal{C}} \operatorname{gra} C$ is an upper bound of $\mathcal{C}$. Therefore, Zorn's lemma (Fact 1.1) guarantees the existence of a maximal element $\widetilde{A} \in \mathcal{M}$. The operator $\widetilde{A}$ possesses the required properties. Now assume that $\operatorname{gra} A = \varnothing$. Then any maximally monotone extension $\widetilde{A}$ of the operator the graph of which is $\{(0,0)\}$ is as required. ☐

The proofs of the next two propositions are left as exercises.

Proposition 20.22 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $z \in \mathcal{H}$, let $u \in \mathcal{H}$, and let $\gamma \in \mathbb{R}_{++}$. Then A^{-1} and $x \mapsto u + \gamma A(x+z)$ are maximally monotone.*

Proposition 20.23 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{K} \to 2^{\mathcal{K}}$ be maximally monotone, and set $\boldsymbol{\mathcal{H}} = \mathcal{H} \oplus \mathcal{K}$ and $\boldsymbol{A}\colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}\colon (x,y) \mapsto Ax \times By$. Then $\boldsymbol{A}$ is maximally monotone.*

Proposition 20.24 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $U \in \mathcal{B}(\mathcal{H})$ be a self-adjoint strongly monotone operator, and let $\mathcal{K}$ be the real Hilbert space obtained by endowing $\mathcal{H}$ with the scalar product $(x,y) \mapsto \langle x \mid y\rangle_{\mathcal{K}} = \big\langle U^{-1}x \mid y\big\rangle$. Then $UA\colon \mathcal{K} \to 2^{\mathcal{K}}$ is maximally monotone.*

Proof. Set $B = UA$. For every $(x,u) \in \operatorname{gra} B$ and every $(y,v) \in \operatorname{gra} B$, $U^{-1}u \in U^{-1}Bx = Ax$ and $U^{-1}v \in U^{-1}By = Ay$, and therefore

$$\langle x - y \mid u - v\rangle_{\mathcal{K}} = \big\langle x - y \mid U^{-1}u - U^{-1}v\big\rangle \geqslant 0 \tag{20.18}$$

by monotonicity of A on $\mathcal{H}$. This shows that B is monotone on $\mathcal{K}$. Now let y and v be points in $\mathcal{H}$ such that

$$\big(\forall (x,u) \in \operatorname{gra} B\big) \quad \langle x - y \mid u - v\rangle_{\mathcal{K}} \geqslant 0. \tag{20.19}$$

Then, for every $(x,u) \in \operatorname{gra} A$, $(x, Uu) \in \operatorname{gra} B$ and we derive from (20.19) that

$$\big\langle x - y \mid u - U^{-1}v\big\rangle = \langle x - y \mid Uu - v\rangle_{\mathcal{K}} \geqslant 0. \tag{20.20}$$

Since A is maximally monotone on $\mathcal{H}$, (20.20) gives $(y, U^{-1}v) \in \operatorname{gra} A$, which implies that $(y,v) \in \operatorname{gra} B$. Hence, B is maximally monotone on $\mathcal{K}$. ☐

A fundamental example of a maximally monotone operator is the subdifferential of a function in $\Gamma_0(\mathcal{H})$.

Theorem 20.25 (Moreau) *Let $f \in \Gamma_0(\mathcal{H})$. Then ∂f is maximally monotone.*

Proof. Let $(x,u) \in \mathcal{H} \times \mathcal{H}$ and suppose that

$$\big(\forall (y,v) \in \operatorname{gra} \partial f\big) \quad \langle x - y \mid u - v\rangle \geqslant 0. \tag{20.21}$$

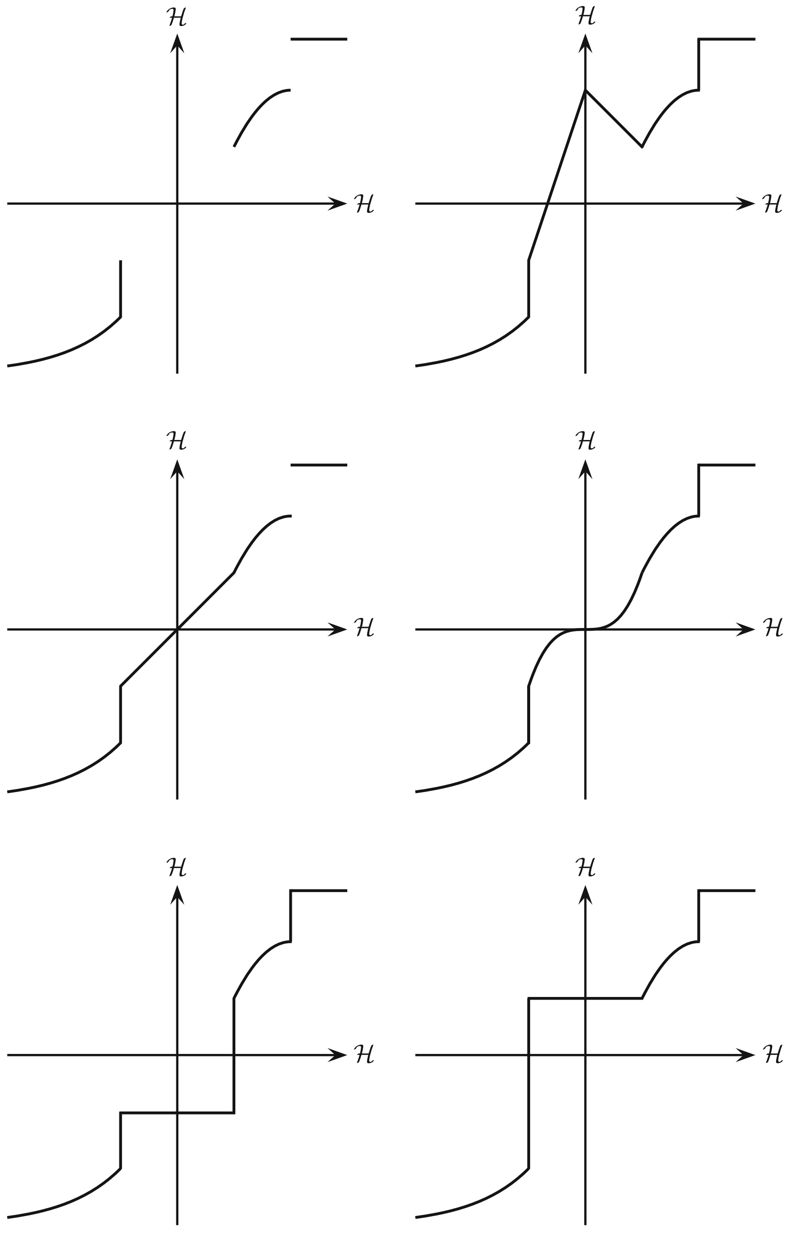

Fig. 20.1 Extensions of a monotone operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ ($\mathcal{H} = \mathbb{R}$). Top left: Graph of A. Top right: Graph of a nonmonotone extension of A. Center left: Graph of a monotone extension of A that is not maximally monotone. Center right, bottom left, and bottom right: Graphs of maximally monotone extensions of A.

In view of (20.16), we must show that $(x, u) \in \operatorname{gra} \partial f$. Set $p = \operatorname{Prox}_f(x + u)$ and $q = \operatorname{Prox}_{f^*}(x + u)$. Then it follows from (14.6) that $x + u = p + q$ and from Proposition 16.44 that $q = x + u - p \in \partial f(p)$. Hence, (20.21) yields $-\|x - p\|^2 = \langle x - p \mid u - q\rangle \geqslant 0$. Thus, $(x, u) = (p, q) \in \operatorname{gra} \partial f$. ☐

Example 20.26 Let C be a nonempty closed convex subset of $\mathcal{H}$. Then N_C is maximally monotone.

Proof. Apply Theorem 20.25 to $f = \iota_C$ and use Example 16.13. ☐

Proposition 20.27 *Let* $A\colon \mathcal{H} \to \mathcal{H}$ *be monotone and* hemicontinuous, *i.e., for every* $(x, y, z) \in \mathcal{H}^3$, $\lim_{\alpha \downarrow 0} \langle z \mid A(x + \alpha y)\rangle = \langle z \mid Ax\rangle$. *Then* A *is maximally monotone.*

Proof. Let $(x, u) \in \mathcal{H} \times \mathcal{H}$ be such that $(\forall y \in \mathcal{H})\ \langle x - y \mid u - Ay\rangle \geqslant 0$. We must show that $u = Ax$. For every $\alpha \in \left]0, 1\right]$, set $y_\alpha = x + \alpha(u - Ax)$ and observe that $\langle u - Ax \mid u - Ay_\alpha\rangle = -\langle x - y_\alpha \mid u - Ay_\alpha\rangle / \alpha \leqslant 0$. Since A is hemicontinuous, we conclude that $\|u - Ax\|^2 \leqslant 0$, i.e., that $u = Ax$. ☐

Corollary 20.28 *Let* $A\colon \mathcal{H} \to \mathcal{H}$ *be monotone and continuous. Then* A *is maximally monotone.*

Example 20.29 Let $T\colon \mathcal{H} \to \mathcal{H}$ be nonexpansive and let $\alpha \in [-1, 1]$. Then $\mathrm{Id} + \alpha T$ is maximally monotone.

Proof. Combine Example 20.7 with Corollary 20.28. ☐

Example 20.30 Let $T\colon \mathcal{H} \to \mathcal{H}$ be α-averaged, with $\alpha \in \left]0, 1/2\right]$ (in particular, firmly nonexpansive). Then T is maximally monotone.

Proof. This follows from Corollary 20.28 since T is continuous and, by Example 20.6, monotone. ☐

Example 20.31 Let $T\colon \mathcal{H} \to \mathcal{H}$ be β-cocoercive, with $\beta \in \mathbb{R}_{++}$. Then T is maximally monotone.

Proof. Since βT is firmly nonexpansive, it is maximally monotone by Example 20.30. ☐

Example 20.32 Let C be a nonempty closed convex subset of $\mathcal{H}$. Then P_C is maximally monotone.

Proof. Combine Proposition 4.16 and Example 20.30. ☐

Example 20.33 Suppose that $\mathcal{H}$ is finite-dimensional and let C be a Chebyshev subset of $\mathcal{H}$. Then P_C is maximally monotone.

Proof. This follows from Proposition 3.12, Example 20.12, and Corollary 20.28. ☐

Example 20.34 Let $A \in \mathcal{B}(\mathcal{H})$ be monotone. Then A is maximally monotone.

Proof. This follows from Corollary 20.28. ☐

Example 20.35 Let $A \in \mathcal{B}(\mathcal{H})$ be such that $A^* = -A$. Then A is maximally monotone.

Proof. We have $(\forall x \in \mathcal{H})\ \langle x \mid Ax\rangle = 0$. Hence A is monotone, and maximally so by Example 20.34. ☐

We now present some basic properties of maximally monotone operators.

Proposition 20.36 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and let $x \in \mathcal{H}$. Then Ax is closed and convex.*

Proof. We assume that $x \in \operatorname{dom} A$. Then (20.16) yields

$$Ax = \bigcap_{(y,v)\in \operatorname{gra} A} \big\{u \in \mathcal{H} \bigm| \langle x - y \mid u - v\rangle \geqslant 0\big\}, \tag{20.22}$$

which is an intersection of closed convex sets. ☐

The next two propositions address various closedness properties of the graph of a maximally monotone operator.

Proposition 20.37 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $(x_b, u_b)_{b\in B}$ be a bounded net in* $\operatorname{gra} A$*, and let $(x,u) \in \mathcal{H}\times\mathcal{H}$. Then the following hold:*

(i) *Suppose that $x_b \to x$ and $u_b \rightharpoonup u$. Then $(x,u) \in \operatorname{gra} A$.*
(ii) *Suppose that $x_b \rightharpoonup x$ and $u_b \to u$. Then $(x,u) \in \operatorname{gra} A$.*

Proof. (i): Take $(y,v) \in \operatorname{gra} A$. Then $(\forall b \in B)\ \langle x_b - y \mid u_b - v\rangle \geqslant 0$ by (20.16). In turn, Lemma 2.44 implies that $\langle x - y \mid u - v\rangle \geqslant 0$. Hence $(x,u) \in \operatorname{gra} A$ by (20.16).

(ii): Apply (i) to A^{-1}. ☐

Proposition 20.38 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone. Then the following hold:*

(i) $\operatorname{gra} A$ *is sequentially closed in $\mathcal{H}^{\text{strong}} \times \mathcal{H}^{\text{weak}}$, i.e., for every sequence $(x_n,u_n)_{n\in\mathbb{N}}$ in* $\operatorname{gra} A$ *and every $(x,u) \in \mathcal{H}\times\mathcal{H}$, if $x_n \to x$ and $u_n \rightharpoonup u$, then $(x,u) \in \operatorname{gra} A$.*
(ii) $\operatorname{gra} A$ *is sequentially closed in $\mathcal{H}^{\text{weak}} \times \mathcal{H}^{\text{strong}}$, i.e., for every sequence $(x_n,u_n)_{n\in\mathbb{N}}$ in* $\operatorname{gra} A$ *and every $(x,u) \in \mathcal{H}\times\mathcal{H}$, if $x_n \rightharpoonup x$ and $u_n \to u$, then $(x,u) \in \operatorname{gra} A$.*
(iii) $\operatorname{gra} A$ *is closed in $\mathcal{H}^{\text{strong}} \times \mathcal{H}^{\text{strong}}$.*

Proof. (i): Combine Lemma 2.46 and Proposition 20.37(i).
(ii): Apply (i) to A^{-1}.
(iii): A consequence of (i) (see Section 1.12). $\square$

As we shall see in Example 21.6 and Remark 21.7, Proposition 20.37(i) is sharp in the sense that the boundedness assumption on the net cannot be removed. Regarding Proposition 20.38(i), the next example shows that it is not possible to replace $\mathcal{H}^{\mathrm{strong}} \times \mathcal{H}^{\mathrm{weak}}$ by $\mathcal{H}^{\mathrm{weak}} \times \mathcal{H}^{\mathrm{weak}}$.

Example 20.39 The graph of a maximally monotone operator need not be sequentially closed in $\mathcal{H}^{\mathrm{weak}} \times \mathcal{H}^{\mathrm{weak}}$. Indeed, suppose that $\mathcal{H} = \ell^2(\mathbb{N})$ and set $C = B(0;1)$. Then $\mathrm{Id} - P_C$ is firmly nonexpansive by Corollary 4.18, and hence maximally monotone by Example 20.30. Consider the sequence $(x_n)_{n\in\mathbb{N}} = (e_1 + e_{2n})_{n\in\mathbb{N}}$, where $(e_n)_{n\in\mathbb{N}}$ is the sequence of standard unit vectors in $\ell^2(\mathbb{N})$. Then the sequence $(x_n, (1-1/\sqrt{2})x_n)_{n\in\mathbb{N}}$ lies in $\operatorname{gra}(\mathrm{Id} - P_C)$ and it converges weakly to $(e_1, (1-1/\sqrt{2})e_1)$. However, the weak limit $(e_1, (1-1/\sqrt{2})e_1)$ does not belong to $\operatorname{gra}(\mathrm{Id} - P_C)$.

Proposition 20.40 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and at most single-valued. Suppose that* $\operatorname{dom} A$ *is a linear subspace, that* $A|_{\operatorname{dom} A}$ *is linear, and that* $(\forall x \in \operatorname{dom} A)(\forall y \in \operatorname{dom} A)\ \langle x \mid Ay\rangle = \langle Ax \mid y\rangle$. *Set*

$$h\colon \mathcal{H} \to \left]-\infty, +\infty\right] : x \mapsto \begin{cases} \frac{1}{2}\langle x \mid Ax\rangle, & \text{if } x \in \operatorname{dom} A;\\ +\infty, & \text{otherwise,}\end{cases} \tag{20.23}$$

and

$$f\colon \mathcal{H} \to \left]-\infty, +\infty\right] : x \mapsto \sup_{y\in\operatorname{dom} A} \big(\langle x \mid Ay\rangle - h(y)\big). \tag{20.24}$$

Then the following hold:

(i) $f + \iota_{\operatorname{dom} A} = h$.
(ii) $f \in \Gamma_0(\mathcal{H})$.
(iii) $\partial f = A$.
(iv) $f = h^{**}$.

Proof. Take $x \in \operatorname{dom} A = \operatorname{dom} h$.
(i): For every $y \in \operatorname{dom} A$, $0 \leqslant \langle x - y \mid Ax - Ay\rangle = \langle x \mid Ax\rangle + \langle y \mid Ay\rangle - 2\langle x \mid Ay\rangle$, which implies that $\langle x \mid Ay\rangle - h(y) \leqslant h(x)$. Hence $f(x) \leqslant h(x)$. On the other hand, $f(x) \geqslant \langle x \mid Ax\rangle - h(x) = h(x)$. Altogether, $f + \iota_{\operatorname{dom} A} = h$.
(ii): As a supremum of continuous affine functions, $f \in \Gamma(\mathcal{H})$ by Proposition 9.3. In addition, since $f(0) = 0$, f is proper.
(iii): For every $y \in \mathcal{H}$, we have $f(x) + \langle y - x \mid Ax\rangle = \langle y \mid Ax\rangle - h(x) \leqslant f(y)$. Consequently, $Ax \in \partial f(x)$. It follows that $\operatorname{gra} A \subset \operatorname{gra} \partial f$, which implies that $A = \partial f$ since A is maximally monotone, while ∂f is monotone by Example 20.3.
(iv): Using (ii), Corollary 16.41, (iii), and (i), we see that $f = (f + \iota_{\operatorname{dom} \partial f})^{**} = (f + \iota_{\operatorname{dom} A})^{**} = h^{**}$. $\square$

Example 20.41 Suppose that $\mathcal{H}$ is infinite-dimensional and separable with an orthonormal basis $(e_n)_{n\in\mathbb{N}}$, let $(\beta_n)_{n\in\mathbb{N}}$ be a sequence in $]0,1]$ such that $\beta_n \downarrow 0$, and set $B\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \sum_{n\in\mathbb{N}} \beta_n \langle x \mid e_n\rangle e_n$. Then $B \in \mathcal{B}(\mathcal{H})$, B is maximally monotone and self-adjoint, and $\operatorname{ran} B$ is a proper linear subspace of $\mathcal{H}$ that is dense in $\mathcal{H}$. Now set $A = B^{-1}$ and

$$f\colon \mathcal{H} \to \left]-\infty,+\infty\right] : x \mapsto \sup_{y\in\operatorname{dom} A} \Big(\langle x \mid Ay\rangle - \tfrac{1}{2}\langle y \mid Ay\rangle \Big). \tag{20.25}$$

Then A is maximally monotone, $\operatorname{dom} A$ is a dense and proper linear subspace of $\mathcal{H}$, f is nowhere Gâteaux differentiable, and $\partial f = A$ is at most single-valued.

Proof. Clearly, B is linear and bounded, $\|B\| \leqslant 1$, B is self-adjoint, and B is strictly monotone and thus injective. By Example 20.34, B is maximally monotone and so is A by Proposition 20.22. Note that A is linear and single-valued on its domain, and that $\{\beta_n e_n\}_{n\in\mathbb{N}} \subset \operatorname{ran} B$. Thus, $\operatorname{dom} A = \operatorname{ran} B$ is a dense linear subspace of $\mathcal{H}$. Since $\sup_{n\in\mathbb{N}} \|Ae_n\| = \sup_{n\in\mathbb{N}} \beta_n^{-1} = +\infty$, it follows that A is not continuous and hence that $\operatorname{dom} A \neq \mathcal{H}$. Proposition 20.40(ii)&(iii) implies that $f \in \Gamma_0(\mathcal{H})$ and that $\partial f = A$ is at most single-valued. Since $\operatorname{int}\operatorname{dom} \partial f = \operatorname{int}\operatorname{dom} A = \varnothing$, f is nowhere Gâteaux differentiable by Proposition 17.50. ☐

20.3 The Partial Inverse

Definition 20.42 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a set-valued operator and let V be a closed linear subspace of $\mathcal{H}$. The *partial inverse* of A with respect to V is the operator $A_V : \mathcal{H} \to 2^{\mathcal{H}}$ defined by

$$\operatorname{gra} A_V = \big\{ (P_V x + P_{V^\perp} u, P_V u + P_{V^\perp} x) \;\big|\; (x,u) \in \operatorname{gra} A \big\}. \tag{20.26}$$

The partial inverse can be regarded as an intermediate object between an operator and its inverse.

Example 20.43 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$. Then $A_{\mathcal{H}} = A$ and $A_{\{0\}} = A^{-1}$.

Some basic properties of the partial inverse are collected below.

Proposition 20.44 (Spingarn) *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$, let V be a closed linear subspace of $\mathcal{H}$, and set*

$$L\colon \mathcal{H}\oplus\mathcal{H} \to \mathcal{H}\oplus\mathcal{H}\colon (x,u) \mapsto \big(P_V x + P_{V^\perp} u, P_V u + P_{V^\perp} x\big). \tag{20.27}$$

Then the following hold:

(i) $A_V = (A_{V^\perp})^{-1} = (A^{-1})_{V^\perp}$.
(ii) $L^{-1} = L$, $\operatorname{gra} A_V = L(\operatorname{gra} A)$, *and* $\operatorname{gra} A = L(\operatorname{gra} A_V)$.
(iii) *Let* $(x_1, u_1, x_2, u_2) \in \mathcal{H}^4$, *and set* $(y_1, v_1) = L(x_1, u_1)$ *and* $(y_2, v_2) = L(x_2, u_2)$. *Then* $\langle y_1 - y_2 \mid v_1 - v_2 \rangle = \langle x_1 - x_2 \mid u_1 - u_2 \rangle$.
(iv) A_V *is monotone if and only if* A *is.*
(v) A_V *is maximally monotone if and only if* A *is.*
(vi) *Let* $x \in \mathcal{H}$. *Then* $x \in \operatorname{zer} A_V \Leftrightarrow (P_V x, P_{V^\perp} x) \in \operatorname{gra} A$.

Proof. (i): We have

$$\begin{aligned}
\operatorname{gra} A_V &= \big\{(P_V x + P_{V^\perp} u, P_V u + P_{V^\perp} x) \;\big|\; (x,u) \in \operatorname{gra} A\big\} \\
&= \big\{(P_{V^\perp} u + P_{(V^\perp)^\perp} x, P_{V^\perp} x + P_{(V^\perp)^\perp} u) \;\big|\; (u,x) \in \operatorname{gra} A^{-1}\big\} \\
&= \operatorname{gra} (A^{-1})_{V^\perp}.
\end{aligned} \tag{20.28}$$

Likewise,

$$\begin{aligned}
\operatorname{gra} A_V &= \big\{(P_{V^\perp} u + P_{(V^\perp)^\perp} x, P_{V^\perp} x + P_{(V^\perp)^\perp} u) \;\big|\; (x,u) \in \operatorname{gra} A\big\} \\
&= \operatorname{gra} (A_{V^\perp})^{-1}.
\end{aligned} \tag{20.29}$$

(ii): Let $(x,u) \in \mathcal{H} \oplus \mathcal{H}$. Then $L(L(x,u)) = (P_V(P_V x + P_{V^\perp} u) + P_{V^\perp}(P_V u + P_{V^\perp} x), P_V(P_V u + P_{V^\perp} x) + P_{V^\perp}(P_V x + P_{V^\perp} u)) = (P_V x + P_{V^\perp} x, P_V u + P_{V^\perp} u) = (x,u)$. Thus $L^{-1} = L$. The second identity follows from (20.26) and it implies that $L(\operatorname{gra} A_V) = L^{-1}(\operatorname{gra} A_V) = L^{-1}(L(\operatorname{gra} A)) = \operatorname{gra} A$.

(iii): By orthogonality and Corollary 3.24(vi)&(v),

$$\begin{aligned}
&\langle y_1 - y_2 \mid v_1 - v_2 \rangle \\
&\quad = \langle P_V(x_1 - x_2) + P_{V^\perp}(u_1 - u_2) \mid P_V(u_1 - u_2) + P_{V^\perp}(x_1 - x_2) \rangle \\
&\quad = \langle P_V(x_1 - x_2) \mid P_V(u_1 - u_2) \rangle + \langle P_{V^\perp}(x_1 - x_2) \mid P_{V^\perp}(u_1 - u_2) \rangle \\
&\quad = \langle P_V(x_1 - x_2) \mid u_1 - u_2 \rangle + \langle P_{V^\perp}(x_1 - x_2) \mid u_1 - u_2 \rangle \\
&\quad = \langle x_1 - x_2 \mid u_1 - u_2 \rangle .
\end{aligned} \tag{20.30}$$

(iv)&(v): These follow from (ii) and (iii).

(vi): In view of (ii), $0 \in A_V x \Leftrightarrow (x,0) \in \operatorname{gra} A_V = L(\operatorname{gra} A) \Leftrightarrow (P_V x, P_{V^\perp} x) = L(x,0) = L^{-1}(x,0) \in \operatorname{gra} A$. ☐

20.4 Bivariate Functions and Maximal Monotonicity

We start with a technical fact.

Lemma 20.45 *Let* $(z,w) \in \mathcal{H} \times \mathcal{H}$, *and set*

$$\begin{aligned} G\colon \mathcal{H}\times\mathcal{H} &\to \mathbb{R} \\ (x,u) &\mapsto -\langle x \mid u\rangle + \langle z-x \mid w-u\rangle + \tfrac{1}{2}\|z-x\|^2 + \tfrac{1}{2}\|w-u\|^2 \\ &= -\langle x \mid u\rangle + \tfrac{1}{2}\|(x-z)+(u-w)\|^2 \\ &= \langle z \mid w\rangle - \langle (x,u) \mid (w,z)\rangle + \tfrac{1}{2}\|(x,u)-(z,w)\|^2 \end{aligned} \tag{20.31}$$

and $L\colon \mathcal{H}\times\mathcal{H}\to\mathcal{H}\times\mathcal{H}\colon (u,x)\mapsto(-x,-u)$. *Then* $G^* = G\circ L$. *Furthermore, let* $(x,u)\in\mathcal{H}\times\mathcal{H}$. *Then the following hold:*

(i) $G(x,u)+\langle x \mid u\rangle \geqslant 0$.
(ii) $G(x,u)+\langle x \mid u\rangle = 0 \Leftrightarrow x-z=w-u$.
(iii) $[\ G(x,u)+\langle x \mid u\rangle = 0$ *and* $\langle z-x \mid w-u\rangle \geqslant 0\ ] \Leftrightarrow (x,u)=(z,w)$.

Proof. The formula $G^*(u,x) = G(-x,-u)$ is a consequence of Proposition 13.19 (applied in $\mathcal{H}\times\mathcal{H}$) and Proposition 13.23(iii). The remaining statements follow from (20.31). ☐

Theorem 20.46 *Let* $F\colon\mathcal{H}\times\mathcal{H}\to\,]-\infty,+\infty]$ *be a convex function such that* F^* *is proper and* $F^*\geqslant\langle\cdot\mid\cdot\rangle$. *Define* $A\colon\mathcal{H}\to 2^{\mathcal{H}}$ *by*

$$\operatorname{gra} A = \big\{(x,u)\in\mathcal{H}\times\mathcal{H} \bigm| F^*(u,x) = \langle x \mid u\rangle\big\}. \tag{20.32}$$

Then the following hold:

(i) *A is monotone.*
(ii) *Suppose that* $F\geqslant\langle\cdot\mid\cdot\rangle$. *Then A is maximally monotone.*

Proof. (i): Suppose that (x,u) and (y,v) belong to $\operatorname{gra} A$. Then by convexity of F^*,

$$\begin{aligned} \langle x \mid u\rangle + \langle y \mid v\rangle &= F^*(u,x)+F^*(v,y) \\ &\geqslant 2F^*\Big(\tfrac{1}{2}(u+v),\tfrac{1}{2}(x+y)\Big) \\ &\geqslant \tfrac{1}{2}\,\langle x+y \,|\, u+v\rangle\,. \end{aligned} \tag{20.33}$$

Hence $\langle x-y \mid u-v\rangle\geqslant 0$ and A is therefore monotone.

(ii): Since F^* is proper, Proposition 13.10(iii) and Proposition 13.12(ii) imply that F is proper and that it possesses a continuous affine minorant. Using Proposition 13.46(iii) and the continuity of $\langle\cdot\mid\cdot\rangle$, we obtain $F\geqslant\langle\cdot\mid\cdot\rangle \Rightarrow (\forall(x,u)\in\mathcal{H}\times\mathcal{H})\ \underline{\lim}_{(y,v)\to(x,u)}F(y,v) \geqslant \underline{\lim}_{(y,v)\to(x,u)}\langle y \mid v\rangle$. Hence $F^{**}\geqslant\langle\cdot\mid\cdot\rangle$. Now suppose that $(z,w)\in\mathcal{H}\times\mathcal{H}$ satisfies

$$(\forall (x,u)\in\operatorname{gra} A)\quad \langle z-x \mid w-u\rangle \geqslant 0. \tag{20.34}$$

We must show that $(z,w)\in\operatorname{gra} A$. Define G and L as in Lemma 20.45, where it was observed that $G^*=G\circ L$ and that $G+\langle\cdot\mid\cdot\rangle\geqslant 0$. Since $F^{**}-\langle\cdot\mid\cdot\rangle\geqslant 0$, we see that $F^{**}+G\geqslant 0$. By Proposition 13.16(iii) and

Corollary 15.15 (applied to F^{**}, G, and L), there exists $(v,y) \in \mathcal{H}\times\mathcal{H}$ such that $0 \geqslant F^*(v,y) + (G\circ L)(-v,-y) = F^*(v,y) + G(y,v)$. The assumption that $F^* \geqslant \langle\cdot \mid \cdot\rangle$ and Lemma 20.45(i) then result in

$$0 \geqslant F^*(v,y) + G(y,v) \geqslant \langle y \mid v\rangle + G(y,v) \geqslant 0. \tag{20.35}$$

Hence

$$\langle y \mid v\rangle + G(y,v) = 0 \tag{20.36}$$

and $F^*(v,y) = \langle y \mid v\rangle$, i.e.,

$$(y,v) \in \operatorname{gra} A. \tag{20.37}$$

In view of (20.34) and (20.37), we obtain

$$\langle z-y \mid w-v\rangle \geqslant 0. \tag{20.38}$$

Lemma 20.45(iii) shows that (20.36) and (20.38) are equivalent to $(z,w) = (y,v)$. Therefore, using (20.37), we deduce that $(z,w) \in \operatorname{gra} A$. ☐

Corollary 20.47 *Let $F \in \Gamma_0(\mathcal{H}\times\mathcal{H})$ be autoconjugate and define A via*

$$\operatorname{gra} A = \big\{(x,u) \in \mathcal{H}\times\mathcal{H} \;\big|\; F(x,u) = \langle x \mid u\rangle\big\}. \tag{20.39}$$

Then A is maximally monotone.

Proof. This follows from Proposition 13.36, Proposition 16.65, and Theorem 20.46(ii). ☐

Corollary 20.47 allows us to provide an alternate proof of Theorem 20.25.

Theorem 20.48 *Let $f \in \Gamma_0(\mathcal{H})$. Then ∂f is maximally monotone.*

Proof. On the one hand, $f \oplus f^*$ is autoconjugate. On the other hand, $\big\{(x,u) \in \mathcal{H}\times\mathcal{H} \;\big|\; (f\oplus f^*)(x,u) = \langle x \mid u\rangle\big\} = \operatorname{gra}\partial f$ by Proposition 16.10. Altogether, ∂f is maximally monotone by Corollary 20.47. ☐

We now extend Proposition 20.27 and Corollary 20.28.

Proposition 20.49 (Rockafellar) *Let C be a nonempty closed convex subset of $\mathcal{H}$, and let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a monotone, at most single-valued operator such that $C \subset \operatorname{dom} A$ and A is* hemicontinuous *on C, i.e., for every $(x,y,z) \in C\times C\times\mathcal{H}$, $\lim_{\alpha\downarrow 0} \langle z \mid A((1-\alpha)x+\alpha y)\rangle = \langle z \mid Ax\rangle$. Then $A+N_C$ is maximally monotone.*

Proof. By Example 20.26, N_C is maximally monotone and hence $A+N_C$ is monotone. Now let $(z,w) \in \mathcal{H}\times\mathcal{H}$ be monotonically related to $\operatorname{gra}(A+N_C)$ and let $x \in C$. Since $N_C x$ is a cone, we have

$$(\forall u \in N_C x)(\forall \gamma \in \mathbb{R}_{++}) \qquad \gamma\langle x-z \mid u\rangle + \langle x-z \mid Ax-w\rangle \geqslant 0. \tag{20.40}$$

Letting γ tend to $+\infty$ yields $\langle x - z \mid u - 0\rangle \geqslant 0$. Since N_C is maximally monotone, we deduce that $0 \in N_C z$ and hence that $z \in C$. Next, for every $\alpha \in]0,1]$, set $x_\alpha = (1-\alpha)z + \alpha x \in C$. Then (20.40) yields

$$(\forall \alpha \in]0,1[) \quad \langle x - z \mid Ax_\alpha - w\rangle = \langle x_\alpha - z \mid Ax_\alpha - w\rangle / \alpha \geqslant 0. \tag{20.41}$$

Upon letting $\alpha \downarrow 0$, this implies that $\langle x - z \mid Az - w\rangle \geqslant 0$. Hence $w - Az \in N_C z$. Therefore, $w \in (A + N_C)z$ and we conclude that $A + N_C$ is maximally monotone. ☐

Corollary 20.50 *Let C be a nonempty closed convex subset of $\mathcal{H}$, and let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a monotone operator such that $C \subset \operatorname{dom} A$ and $A|_C$ is continuous. Then $A + N_C$ is maximally monotone.*

20.5 The Fitzpatrick Function

Definition 20.51 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be monotone. The *Fitzpatrick function* of A is

$$\begin{aligned} F_A\colon \mathcal{H} \times \mathcal{H} &\to [-\infty, +\infty] \\ (x,u) &\mapsto \sup_{(y,v)\in \operatorname{gra} A} \big(\langle y \mid u\rangle + \langle x \mid v\rangle - \langle y \mid v\rangle \big) \qquad (20.42)\\ &= \langle x \mid u\rangle - \inf_{(y,v)\in \operatorname{gra} A} \langle x - y \mid u - v\rangle . \qquad (20.43) \end{aligned}$$

Example 20.52 $F_{\mathrm{Id}}\colon \mathcal{H} \times \mathcal{H} \to]-\infty, +\infty] : (x,u) \mapsto (1/4)\|x+u\|^2$.

Example 20.53 Let $A \in \mathcal{B}(\mathcal{H})$ be such that $A^* = -A$. Then $F_A = \iota_{\operatorname{gra} A}$.

Example 20.54 Let $A \in \mathcal{B}(\mathcal{H})$ be monotone and set $q_A\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto (1/2)\langle x \mid Ax\rangle$. Then $(\forall (x,u) \in \mathcal{H} \times \mathcal{H})$ $F_A(x,u) = 2q_A^*\big(\frac{1}{2}u + \frac{1}{2}A^*x\big)$.

Proof. Take $(x,u) \in \mathcal{H} \times \mathcal{H}$. Then

$$\begin{aligned} F_A(x,u) &= \sup_{y\in\mathcal{H}} \big(\langle y \mid u\rangle + \langle x \mid Ay\rangle - \langle y \mid Ay\rangle \big) \\ &= 2 \sup_{y\in\mathcal{H}} \left(\left\langle y \,\middle|\, \tfrac{1}{2}u + \tfrac{1}{2}A^*x \right\rangle - \tfrac{1}{2}\langle y \mid Ay\rangle \right) \\ &= 2q_A^*\Big(\tfrac{1}{2}u + \tfrac{1}{2}A^*x\Big). \qquad (20.44) \end{aligned}$$

☐

Example 20.55 Let $f \in \Gamma_0(\mathcal{H})$. Then $F_{\partial f} \leqslant f \oplus f^*$ and $\operatorname{dom} f \times \operatorname{dom} f^* \subset \operatorname{dom} F_{\partial f}$.

Proof. Take $(x,u) \in \operatorname{dom} f \times \operatorname{dom} f^*$ and $(y,v) \in \operatorname{gra}\partial f$. Then $\langle y \mid u\rangle + \langle x \mid v\rangle - \langle y \mid v\rangle = (\langle y \mid u\rangle - f(y)) + (\langle x \mid v\rangle - f^*(v)) \leqslant f^*(u) + f^{**}(x)$ by Proposition 16.10 and Proposition 13.15. Hence, $F_{\partial f} \leqslant f \oplus f^*$. ☐

Proposition 20.56 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a monotone operator such that $\operatorname{gra} A \neq \varnothing$ and let $(x,u) \in \mathcal{H}\times\mathcal{H}$. Then the following hold:*

(i) *Suppose that $(x,u) \in \operatorname{gra} A$. Then $F_A(x,u) = \langle x \mid u\rangle$.*
(ii) $F_A = \big(\iota_{\operatorname{gra} A^{-1}} + \langle\cdot \mid \cdot\rangle\big)^* \in \Gamma_0(\mathcal{H}\times\mathcal{H})$.
(iii) $F_A(x,u) \leqslant \langle x \mid u\rangle$ *if and only if* $\{(x,u)\} \cup \operatorname{gra} A$ *is monotone.*
(iv) $F_A(x,u) \leqslant F_A^*(u,x)$.
(v) *Suppose that $(x,u) \in \operatorname{gra} A$. Then $F_A^*(u,x) = \langle x \mid u\rangle$.*
(vi) $F_A(x,u) = F_{A^{-1}}(u,x)$.
(vii) *Let $\gamma \in \mathbb{R}_{++}$. Then $F_{\gamma A}(x,u) = \gamma F_A(x,u/\gamma)$.*
(viii) *Suppose that $(x,u) \in \operatorname{gra} A$. Then $(x,u) = \operatorname{Prox}_{F_A}(x+u,x+u)$.*

Proof. (i): We have $\inf_{(y,v)\in\operatorname{gra} A} \langle x-y \mid u-v\rangle = 0$, and so (20.43) implies that $F_A(x,u) = \langle x \mid u\rangle$.

(ii): The identity $F_A = (\iota_{\operatorname{gra} A^{-1}} + \langle\cdot \mid \cdot\rangle)^*$ is clear from (20.42). Hence, (i) and Proposition 13.13 yield $F_A \in \Gamma_0(\mathcal{H}\times\mathcal{H})$.

(iii): Clear from (20.43).

(iv): We derive from (i) that

$$\begin{aligned}
F_A(x,u) &= \sup_{(y,v)\in\operatorname{gra} A} \big(\langle y \mid u\rangle + \langle x \mid v\rangle - \langle y \mid v\rangle\big)\\
&= \sup_{(y,v)\in\operatorname{gra} A} \big(\langle y \mid u\rangle + \langle x \mid v\rangle - F_A(y,v)\big)\\
&\leqslant \sup_{(y,v)\in\mathcal{H}\times\mathcal{H}} \big(\langle y \mid u\rangle + \langle x \mid v\rangle - F_A(y,v)\big)\\
&= \sup_{(y,v)\in\mathcal{H}\times\mathcal{H}} \big(\langle (y,v) \mid (u,x)\rangle - F_A(y,v)\big)\\
&= F_A^*(u,x).
\end{aligned} \tag{20.45}$$

(v): By (ii) and Proposition 13.16(i), $F_A^* = (\iota_{\operatorname{gra} A^{-1}} + \langle\cdot \mid \cdot\rangle)^{**} \leqslant \iota_{\operatorname{gra} A^{-1}} + \langle\cdot \mid \cdot\rangle$. This, (i), and (iv) imply that $F_A^*(u,x) \leqslant \langle x \mid u\rangle = F_A(x,u) \leqslant F_A^*(u,x)$, as required.

(vi)&(vii): Direct consequences of (20.42).

(viii): By (i) and (v), $F_A(x,u) + F_A^*(u,x) = 2\langle x \mid u\rangle = \langle (x,u) \mid (u,x)\rangle$. Hence $(u,x) \in \partial F_A(x,u)$ and thus $(x+u,x+u) \in (\operatorname{Id} + \partial F_A)(x,u)$, which yields the result. ☐

The proof of the following result is straightforward and hence omitted.

Proposition 20.57 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{K} \to 2^{\mathcal{K}}$ be monotone operators such that $\operatorname{gra} A \neq \varnothing$ and $\operatorname{gra} B \neq \varnothing$. Set $C\colon \mathcal{H}\times\mathcal{K} \to 2^{\mathcal{H}\times\mathcal{K}}\colon (x,y) \mapsto Ax \times By$, set $Q_1^{\mathcal{H}}\colon \mathcal{H}\times\mathcal{H} \to \mathcal{H}\colon (x,u) \mapsto x$, set $Q_1^{\mathcal{K}}\colon \mathcal{K}\times\mathcal{K} \to \mathcal{K}\colon (y,v) \mapsto y$, and set $Q_1^{\mathcal{H}\times\mathcal{K}}\colon \mathcal{H}\times\mathcal{K}\times\mathcal{H}\times\mathcal{K} \to \mathcal{H}\times\mathcal{K}\colon (x,y,u,v) \mapsto (x,y)$. Then the following hold:*

(i) $(\forall(x,y,u,v) \in \mathcal{H} \times \mathcal{K} \times \mathcal{H} \times \mathcal{K})\ F_C(x,y,u,v) = F_A(x,u) + F_B(y,v)$.
(ii) $Q_1^{\mathcal{H}\times\mathcal{K}} \operatorname{dom} F_C = (Q_1^{\mathcal{H}} \operatorname{dom} F_A) \times (Q_1^{\mathcal{K}} \operatorname{dom} F_B)$.

Proposition 20.58 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone. Then $F_A \geqslant \langle \cdot \mid \cdot \rangle$ and*

$$\operatorname{gra} A = \big\{(x,u) \in \mathcal{H} \times \mathcal{H} \;\big|\; F_A(x,u) = \langle x \mid u \rangle\big\}. \tag{20.46}$$

Proof. Take $(x,u) \in \mathcal{H} \times \mathcal{H}$. If $(x,u) \in \operatorname{gra} A$, then $F_A(x,u) = \langle x \mid u \rangle$ by Proposition 20.56(i). On the other hand, if $(x,u) \notin \operatorname{gra} A$, then $\{(x,u)\} \cup \operatorname{gra} A$ is not monotone, and Proposition 20.56(iii) yields $F_A(x,u) > \langle x \mid u \rangle$. ☐

Corollary 20.59 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let x and u be in $\mathcal{H}$, and let $(x_n,u_n)_{n\in\mathbb{N}}$ be a sequence in $\operatorname{gra} A$ such that $(x_n,u_n) \rightharpoonup (x,u)$. Then the following hold:*

(i) $\langle x \mid u \rangle \leqslant \varliminf \langle x_n \mid u_n \rangle$.
(ii) *Suppose that $\varliminf \langle x_n \mid u_n \rangle = \langle x \mid u \rangle$. Then $(x,u) \in \operatorname{gra} A$.*
(iii) *Suppose that $\varlimsup \langle x_n \mid u_n \rangle \leqslant \langle x \mid u \rangle$. Then $\langle x_n \mid u_n \rangle \to \langle x \mid u \rangle$ and $(x,u) \in \operatorname{gra} A$.*

Proof. (i): By Proposition 20.56(ii) and Theorem 9.1, F_A is weakly lower semicontinuous. Hence, Proposition 20.58 and the assumptions imply that

$$\langle x \mid u \rangle \leqslant F_A(x,u) \leqslant \varliminf F_A(x_n,u_n) = \varliminf \langle x_n \mid u_n \rangle. \tag{20.47}$$

(ii): In this case, (20.47) implies that $\langle x \mid u \rangle = F_A(x,u)$. By Proposition 20.58, $(x,u) \in \operatorname{gra} A$.

(iii): Using (i), we see that $\langle x \mid u \rangle \leqslant \varliminf \langle x_n \mid u_n \rangle \leqslant \varlimsup \langle x_n \mid u_n \rangle \leqslant \langle x \mid u \rangle$. Hence $\langle x_n \mid u_n \rangle \to \langle x \mid u \rangle$ and, by (ii), $(x,u) \in \operatorname{gra} A$. ☐

Proposition 20.60 *Let C and D be closed affine subspaces of $\mathcal{H}$ such that $D - D = (C-C)^{\perp}$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $(x_n,u_n)_{n\in\mathbb{N}}$ be a sequence in $\operatorname{gra} A$, and let $(x,u) \in \mathcal{H} \times \mathcal{H}$. Suppose that*

$$x_n \rightharpoonup x, \quad u_n \rightharpoonup u, \quad x_n - P_C x_n \to 0, \quad \text{and} \quad u_n - P_D u_n \to 0. \tag{20.48}$$

Then $(x,u) \in (C \times D) \cap \operatorname{gra} A$ and $\langle x_n \mid u_n \rangle \to \langle x \mid u \rangle$.

Proof. Set $V = C - C$. Since $P_C x_n \rightharpoonup x$, we derive from Corollary 3.35 that $x \in C$ and, likewise, that $u \in D$. Hence, $C = x + V$ and $D = u + V^{\perp}$. Thus, using Corollary 3.22(i),

$$P_C\colon w \mapsto P_V w + P_{V^\perp} x \quad \text{and} \quad P_D\colon w \mapsto P_{V^\perp} w + P_V u. \tag{20.49}$$

Therefore, since P_V and $P_{V^\perp}$ are weakly continuous by Proposition 4.19(i), it follows from Lemma 2.51(iii) that

$$\langle x_n \mid u_n \rangle = \langle P_V x_n + P_{V^\perp} x_n \mid P_V u_n + P_{V^\perp} u_n \rangle$$

$$\begin{aligned}
&= \langle P_V x_n \mid P_V u_n\rangle + \langle P_{V^\perp} x_n \mid P_{V^\perp} u_n\rangle \\
&= \langle P_V x_n \mid u_n - P_{V^\perp} u_n\rangle + \langle x_n - P_V x_n \mid P_{V^\perp} u_n\rangle \\
&= \langle P_V x_n \mid u_n - (P_D u_n - P_V u)\rangle + \langle x_n - (P_C x_n - P_{V^\perp} x) \mid P_{V^\perp} u_n\rangle \\
&= \langle P_V x_n \mid u_n - P_D u_n\rangle + \langle P_V x_n \mid P_V u\rangle \\
&\quad + \langle x_n - P_C x_n \mid P_{V^\perp} u_n\rangle + \langle P_{V^\perp} x \mid P_{V^\perp} u_n\rangle \\
&\to \langle P_V x \mid P_V u\rangle + \langle P_{V^\perp} x \mid P_{V^\perp} u\rangle \\
&= \langle x \mid u\rangle . \qquad (20.50)
\end{aligned}$$

Thus, the result follows from Corollary 20.59(iii). □

Proposition 20.61 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a monotone operator such that $\operatorname{gra} A \neq \varnothing$. Then the following hold:*

(i) $F_A^* = (\iota_{\operatorname{gra} A^{-1}} + \langle \cdot \mid \cdot \rangle)^{**}$.
(ii) $\operatorname{conv} \operatorname{gra} A^{-1} \subset \operatorname{dom} F_A^* \subset \overline{\operatorname{conv}} \operatorname{gra} A^{-1} \subset \overline{\operatorname{conv}} \operatorname{ran} A \times \overline{\operatorname{conv}} \operatorname{dom} A$.
(iii) $F_A^* \geqslant \langle \cdot \mid \cdot \rangle$.
(iv) *Suppose that A is maximally monotone. Then*

$$\operatorname{gra} A = \big\{ (x,u) \in \mathcal{H} \times \mathcal{H} \;\big|\; F_A^*(u,x) = \langle x \mid u\rangle \big\}. \tag{20.51}$$

Proof. (i): Proposition 20.56(ii).

(ii): Combine (i), Proposition 9.8(iv), and Proposition 13.45.

(iii)&(iv): Let B be a maximally monotone extension of A and take $(x,u) \in \mathcal{H} \times \mathcal{H}$. Since $F_A \leqslant F_B$, we derive from Proposition 13.16(ii) that $F_A^* \geqslant F_B^*$. Hence, Proposition 20.56(iv) and Proposition 20.58 imply that

$$F_A^*(u,x) \geqslant F_B^*(u,x) \geqslant F_B(x,u) \geqslant F_A(x,u) \geqslant \langle x \mid u\rangle = \langle u \mid x\rangle , \tag{20.52}$$

which yields (iii). If $(x,u) \notin \operatorname{gra} B$, then (20.52) and (20.46) yield $F_B^*(u,x) \geqslant F_B(x,u) > \langle x \mid u\rangle$. Now assume that $(x,u) \in \operatorname{gra} B$ and take $(y,v) \in \mathcal{H} \times \mathcal{H}$. Then $F_B(y,v) \geqslant \langle (y,v) \mid (u,x)\rangle - \langle x \mid u\rangle$ and hence $\langle x \mid u\rangle \geqslant \langle (y,v) \mid (u,x)\rangle - F_B(y,v)$. Taking the supremum over $(y,v) \in \mathcal{H} \times \mathcal{H}$, we obtain $\langle x \mid u\rangle \geqslant F_B^*(u,x)$. In view of (20.52), $\langle x \mid u\rangle = F_B^*(u,x)$. Thus, if A is maximally monotone, then $B = A$ and (20.51) is verified. □

Remark 20.62 Proposition 20.58 and Proposition 20.61 provide converses to Theorem 20.46.

Using the proximal average, it is possible to construct maximally monotone extensions.

Theorem 20.63 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a monotone operator such that $\operatorname{gra} A \neq \varnothing$. Set $G = \operatorname{pav}(F_A, F_A^{*\intercal})$ and define $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ via*

$$\operatorname{gra} B = \big\{ (x,u) \in \mathcal{H} \times \mathcal{H} \;\big|\; G(x,u) = \langle x \mid u\rangle \big\}. \tag{20.53}$$

Then B is a maximally monotone extension of A.

Proof. By Proposition 20.56(ii), F_A belongs to $\Gamma_0(\mathcal{H}\oplus\mathcal{H})$ and hence so does $F_A^{*\intercal}$. Using Corollary 14.8(i)&(ii), Proposition 13.35, Proposition 14.7(i), and Proposition 14.10, we obtain that $G \in \Gamma_0(\mathcal{H}\oplus\mathcal{H})$ and that $G^* = (\operatorname{pav}(F_A, F_A^{*\intercal}))^* = \operatorname{pav}(F_A^*, F_A^{*\intercal *}) = \operatorname{pav}(F_A^*, F_A^{\intercal **}) = \operatorname{pav}(F_A^*, F_A^{\intercal}) = \operatorname{pav}(F_A^{\intercal}, F_A^*) = \operatorname{pav}(F_A^{\intercal}, F_A^{*\intercal\intercal}) = (\operatorname{pav}(F_A, F_A^{*\intercal}))^{\intercal} = G^{\intercal}$. Hence G is autoconjugate and Corollary 20.47 implies that B is maximally monotone. Set $L\colon \mathcal{H}\times\mathcal{H}\to\mathcal{H}\times\mathcal{H}\colon (x,u)\mapsto(u,x)$, and let $(x,u)\in\operatorname{gra} A$. Using Corollary 14.8(iv), Proposition 20.56(viii), and Proposition 16.66, we see that

$$\begin{aligned}
\operatorname{Prox}_G(x+u,x+u) &= \tfrac{1}{2}\operatorname{Prox}_{F_A}(x+u,x+u) + \tfrac{1}{2}\operatorname{Prox}_{F_A^{*\intercal}}(x+u,x+u)\\
&= \tfrac{1}{2}(x,u) + \tfrac{1}{2}(\mathrm{Id} - L\operatorname{Prox}_{F_A} L)(x+u,x+u)\\
&= \tfrac{1}{2}(x,u) + \tfrac{1}{2}\big((x+u,x+u)-(u,x)\big)\\
&= (x,u). \qquad (20.54)
\end{aligned}$$

In view of Proposition 16.65, it follows that $G(x,u) = \langle x \mid u\rangle$ and hence that $(x,u)\in\operatorname{gra} B$. Therefore, $\operatorname{gra} A \subset \operatorname{gra} B$ and the proof is complete. ☐

Exercises

Exercise 20.1 Let $A\colon \mathcal{H}\to\mathcal{H}$ be nonexpansive and set $T = \mathrm{Id} - A$. Prove the following:

(i) T is 2-Lipschitz continuous.
(ii) Suppose that A is monotone. Then T is $\sqrt{2}$-Lipschitz continuous.
(iii) Suppose that $f\colon \mathcal{H}\to\mathbb{R}$ is convex and differentiable, and that $A = \nabla f$. Then T is 1-Lipschitz continuous.

Provide examples to illustrate that the constants are the best possible.

Exercise 20.2 Prove Proposition 20.10.

Exercise 20.3 Verify Example 20.15 and Example 20.16.

Exercise 20.4 Let $A\in\mathbb{R}^{N\times N}$ be monotone. Show that $\det A \geqslant 0$.

Exercise 20.5 Provide a proof rooted in linear algebra for Fact 20.18 when $\mathcal{H} = \mathbb{R}^N$.

Exercise 20.6 Let $A\colon \mathcal{H}\to 2^{\mathcal{H}}$ be maximally monotone and suppose that $\operatorname{gra} A$ is a convex set. Prove that $\operatorname{gra} A$ is an affine subspace.

Exercise 20.7 Prove Proposition 20.22.

Exercise 20.8 Let $A\colon \mathcal{H}\to 2^{\mathcal{H}}$ be maximally monotone and suppose that, for every $x\in\mathcal{H}$, Ax is a cone. Show that $A = N_{\overline{\operatorname{dom}} A}$.

Exercise 20.9 Deduce Proposition 20.38(i)&(ii) from Proposition 20.60.

Exercise 20.10 Suppose that $\mathcal{H}$ is infinite-dimensional. Use Exercise 18.12 to construct a maximally monotone operator $A\colon \mathcal{H} \to \mathcal{H}$ that is strong-to-weak continuous and hence hemicontinuous on $\mathcal{H}$, strong-to-strong continuous on $\mathcal{H} \smallsetminus \{0\}$, but not strong-to-strong continuous at 0.

Exercise 20.11 Provide examples of functions f and g from $\mathbb{R}$ to $\left]-\infty,+\infty\right]$ that are convex, proper, not lower semicontinuous and such that (i) ∂f is not maximally monotone and (ii) ∂g is maximally monotone.

Exercise 20.12 Verify Example 20.52.

Exercise 20.13 Verify Example 20.53.

Exercise 20.14 Consider Example 20.55. Find $f \in \Gamma_0(\mathcal{H})$ such that $\operatorname{dom} f \times \operatorname{dom} f^* = \operatorname{dom} F_{\partial f} = \overline{\operatorname{dom}}\, F_{\partial f}$.

Exercise 20.15 Consider Example 20.55 when $\mathcal{H} = \mathbb{R}$ and f is the negative Burg entropy function (defined in Example 9.36(viii)). Demonstrate that $\operatorname{dom} F_{\partial f}$ is closed and properly contains $\operatorname{dom} f \times \operatorname{dom} f^*$.

Exercise 20.16 Suppose that $\mathcal{H} = L^2([0,1])$ (see Example 2.8) and define the *Volterra integration operator* $A\colon \mathcal{H} \to \mathcal{H}$ by $(\forall x \in \mathcal{H})(\forall t \in [0,1])$ $(Ax)(t) = \int_0^t x(s)\,ds$. Show that A is continuous, linear, and monotone, and that $\operatorname{ran} A \neq \operatorname{ran} A^*$. Conclude that the closures in Proposition 20.17 are important.

Exercise 20.17 Let C be a nonempty closed convex subset of $\mathcal{H}$. Show that $F_{N_C} = \iota_C \oplus \iota_C^*$.

Exercise 20.18 Let C be a nonempty closed convex subset of $\mathcal{H}$. Show that, for every $(x,u) \in \mathcal{H} \times \mathcal{H}$,

$$F_{P_C}(x,u) = \begin{cases} \left\langle P_C\left(\tfrac{1}{2}x + \tfrac{1}{2}u\right) \;\middle|\; x+u \right\rangle - \left\| P_C\left(\tfrac{1}{2}x + \tfrac{1}{2}u\right)\right\|^2, & \text{if } u \in C;\\ +\infty, & \text{if } u \notin C. \end{cases} \tag{20.55}$$

Exercise 20.19 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, and let $\varepsilon \in \mathbb{R}_+$. The ε-enlargement of A is

$$A^\varepsilon\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x \mapsto \bigcap_{(z,w)\in \operatorname{gra} A} \big\{u \in \mathcal{H} \;\big|\; \langle x - z \mid u - w\rangle \geqslant -\varepsilon\big\}. \tag{20.56}$$

Now let (x,u) and (y,v) be in $\operatorname{gra} A^\varepsilon$. Use Proposition 20.58 to show that $\langle x-y \mid u-v\rangle \geqslant -4\varepsilon$.

Chapter 21
Finer Properties of Monotone Operators

In this chapter, we deepen our study of monotone operators. The main results are Minty's theorem, which conveniently characterizes maximal monotonicity, and the Debrunner–Flor theorem, which concerns the existence of a maximally monotone extension with a prescribed domain localization. Another highlight is the fact that the closures of the range and of the domain of a maximally monotone operator are convex, which yields the classical Bunt–Kritikos–Motzkin result on the convexity of Chebyshev sets in Euclidean spaces. Results on local boundedness, surjectivity, and single-valuedness are also presented.

21.1 Minty's Theorem

A very useful characterization of maximal monotonicity is provided by the following theorem. Recall that F_A designates the Fitzpatrick function of a monotone operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ (see Definition 20.51).

Theorem 21.1 (Minty) *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be monotone. Then A is maximally monotone if and only if* $\operatorname{ran}(\mathrm{Id} + A) = \mathcal{H}$.

Proof. Suppose first that $\operatorname{ran}(\mathrm{Id} + A) = \mathcal{H}$ and fix $(x, u) \in \mathcal{H} \times \mathcal{H}$ such that

$$\big(\forall (y, v) \in \operatorname{gra} A\big) \quad \langle x - y \mid u - v \rangle \geqslant 0. \tag{21.1}$$

Since $\operatorname{ran}(\mathrm{Id} + A) = \mathcal{H}$, there exists $(y, v) \in \mathcal{H}$ such that

$$v \in Ay \quad \text{and} \quad x + u = y + v \in (\mathrm{Id} + A)y. \tag{21.2}$$

It follows from (21.1) and (21.2) that $0 \leqslant \langle y - x \mid v - u \rangle = \langle y - x \mid x - y \rangle = -\|y-x\|^2 \leqslant 0$. Hence $y = x$ and thus $v = u$. Therefore, $(x, u) = (y, v) \in \operatorname{gra} A$,

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_21

and A is maximally monotone. Conversely, assume that A is maximally monotone. Then Proposition 20.58 implies that

$$\begin{aligned}\big(\forall(x,u)\in\mathcal{H}\times\mathcal{H}\big)\quad 2F_A(x,u)+\|(x,u)\|^2 &= 2F_A(x,u)+\|x\|^2+\|u\|^2\\ &\geqslant 2\langle x\mid u\rangle+\|x\|^2+\|u\|^2\\ &= \|x+u\|^2\\ &\geqslant 0. \end{aligned}\tag{21.3}$$

Hence, Corollary 15.17 guarantees the existence of a vector $(v,y)\in\mathcal{H}\times\mathcal{H}$ such that

$$\big(\forall(x,u)\in\mathcal{H}\times\mathcal{H}\big)\quad F_A(x,u)+\tfrac{1}{2}\|(x,u)\|^2\geqslant\tfrac{1}{2}\|(x,u)+(v,y)\|^2,\tag{21.4}$$

which yields

$$\begin{aligned}\big(\forall(x,u)\in\mathcal{H}\times\mathcal{H}\big)\quad F_A(x,u)&\geqslant\tfrac{1}{2}\|v\|^2+\langle x\mid v\rangle+\tfrac{1}{2}\|y\|^2+\langle y\mid u\rangle\\ &\geqslant -\langle y\mid v\rangle+\langle x\mid v\rangle+\langle y\mid u\rangle. \end{aligned}\tag{21.5}$$

This and Proposition 20.56(i) imply that

$$(\forall(x,u)\in\operatorname{gra}A)\quad \langle x\mid u\rangle\geqslant\tfrac{1}{2}\|v\|^2+\langle x\mid v\rangle+\tfrac{1}{2}\|y\|^2+\langle y\mid u\rangle\tag{21.6}$$
$$\geqslant -\langle y\mid v\rangle+\langle x\mid v\rangle+\langle y\mid u\rangle,\tag{21.7}$$

and hence that $\langle x-y\mid u-v\rangle\geqslant 0$. Since A is maximally monotone, we deduce that

$$v\in Ay.\tag{21.8}$$

Using (21.8) in (21.6), we obtain $2\langle y\mid v\rangle\geqslant\|v\|^2+2\langle y\mid v\rangle+\|y\|^2+2\langle y\mid v\rangle$. Hence $0\geqslant\|v\|^2+2\langle y\mid v\rangle+\|y\|^2=\|y+v\|^2$ and thus

$$-v=y.\tag{21.9}$$

Combining (21.8) and (21.9) yields $0\in(\mathrm{Id}+A)y\subset\operatorname{ran}(\mathrm{Id}+A)$. Now fix $w\in\mathcal{H}$ and define a maximally monotone operator $B\colon\mathcal{H}\to 2^{\mathcal{H}}\colon x\mapsto -w+Ax$. Then the above reasoning shows that $0\in\operatorname{ran}(\mathrm{Id}+B)$ and hence that $w\in\operatorname{ran}(\mathrm{Id}+A)$. ☐

We now provide some applications of Theorem 21.1. First, we revisit Theorem 20.25 (see also Theorem 20.48) with an alternative proof.

Theorem 21.2 *Let $f\in\Gamma_0(\mathcal{H})$. Then ∂f is maximally monotone.*

Proof. Combine Example 20.3, Proposition 16.45, and Theorem 21.1. ☐

Example 21.3 Let $\mathcal{K}$ be a real Hilbert space, let $L\in\mathcal{B}(\mathcal{H},\mathcal{K})$, and set $T=\mathrm{Id}+L^*L$ and $R=\mathrm{Id}+LL^*$. Then $T\colon\mathcal{H}\to\mathcal{H}$ and $R\colon\mathcal{K}\to\mathcal{K}$ are bijective with continuous inverses.

Proof. We focus on T since the argument for R is analogous. Set $f\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto (1/2)\|Lx\|^2$. Then it follows from Example 2.60, Theorem 21.2, and Proposition 17.31(i) that $L^*L = \nabla f$ is maximally monotone. In turn, Theorem 21.1 implies that T is surjective. Now let $x \in \mathcal{H}$. Then $Tx = 0 \Leftrightarrow \|Lx\|^2 + \|x\|^2 = \langle Tx \mid x\rangle = 0 \Leftrightarrow x = 0$. Therefore, T is injective and hence bijective. Finally, the continuity of T^{-1} follows from Fact 2.26. ☐

Proposition 21.4 *Let* H *be a separable real Hilbert space, let* $\mathsf{x}_0 \in \mathsf{H}$*, suppose that* $\mathcal{H} = L^2([0,T];\mathsf{H})$*, and let* A *be the time-derivative operator (see Example 2.10)*

$$A\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x \mapsto \begin{cases} \{x'\}, & \text{if } x \in W^{1,2}([0,T];\mathsf{H}) \text{ and } x(0) = \mathsf{x}_0;\\ \varnothing, & \text{otherwise.}\end{cases} \tag{21.10}$$

Then A *is maximally monotone.*

Proof. Monotonicity is shown in Example 20.9. To show maximality, let us fix $u \in L^2([0,T];\mathsf{H})$. In view of Theorem 21.1, we must show that there exists a solution $x \in W^{1,2}([0,T];\mathsf{H})$ to the evolution equation

$$\begin{cases} x(t) + x'(t) = u(t) & \text{a.e. on }]0,T[\\ x(0) = \mathsf{x}_0. \end{cases} \tag{21.11}$$

Let us set $v\colon [0,T] \to \mathsf{H}\colon t \mapsto e^t u(t)$. Then $v \in L^2([0,T];\mathsf{H})$ and the function $y \in W^{1,2}([0,T];\mathsf{H})$ given by

$$y(0) = \mathsf{x}_0 \quad \text{and} \quad (\forall t \in [0,T]) \quad y(t) = y(0) + \int_0^t v(s)ds \tag{21.12}$$

is therefore well defined. Now set $x\colon [0,T] \to \mathsf{H}\colon t \mapsto e^{-t}y(t)$. Then $x \in L^2([0,T];\mathsf{H})$, $x(0) = \mathsf{x}_0$, and

$$x'(t) = -e^{-t}y(t) + e^{-t}y'(t) = -x(t) + u(t) \quad \text{a.e. on }]0,T[\,. \tag{21.13}$$

Thus, $x' \in L^2([0,T];\mathsf{H})$, $x \in W^{1,2}([0,T];\mathsf{H})$, and x solves (21.11). ☐

Proposition 21.5 *Let* C *be a nonempty compact subset of* $\mathcal{H}$*. Suppose that the farthest-point operator* Φ_C *of* C *defined in (20.13) is single-valued. Then* $-\Phi_C$ *is maximally monotone and* C *is a singleton.*

Proof. It is clear that $\operatorname{dom} \Phi_C = \mathcal{H}$ since C is nonempty and compact. Set $f_C\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \|x - \Phi_C x\|$, and take x and y in $\mathcal{H}$. Then for every $z \in C$, we have $\|x - z\| \leqslant \|x - y\| + \|y - z\|$ and hence $f_C(x) = \sup \|x - C\| \leqslant \|x - y\| + \sup \|y - C\| = \|x - y\| + f_C(y)$. It follows that f_C is Lipschitz continuous with constant 1. Now let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ converging to x. Then

$$\|x_n - \Phi_C x_n\| = f_C(x_n) \to f_C(x) = \|x - \Phi_C x\|. \tag{21.14}$$

Assume that

$$\varPhi_C x_n \not\to \varPhi_C x. \tag{21.15}$$

After passing to a subsequence and relabeling if necessary, we assume that there exist $\varepsilon \in \mathbb{R}_{++}$ and $u \in C$ such that $\|\varPhi_C x_n - \varPhi_C x\| \geqslant \varepsilon$ and $\varPhi_C x_n \to u$. Taking the limit in (21.14) yields $\|x - u\| = \|x - \varPhi_C x\|$, and hence $u = \varPhi_C x$, which is impossible. Hence (21.15) is false and $\varPhi_C$ is therefore continuous. In view of Example 20.13 and Corollary 20.28, $-\varPhi_C$ is maximally monotone. By Theorem 21.1, $\operatorname{ran}(\mathrm{Id} - \varPhi_C) = \mathcal{H}$ and thus $0 \in \operatorname{ran}(\mathrm{Id} - \varPhi_C)$. We deduce the existence of $w \in \mathcal{H}$ such that $0 = \|w - \varPhi_C w\| = \sup \|w - C\|$. Hence $w \in C$ and therefore $C = \{w\}$. ☐

Example 21.6 Set $\mathbb{P} = \mathbb{N} \smallsetminus \{0\}$, suppose that $\mathcal{H} = \ell^2(\mathbb{P})$, let $(e_n)_{n\in\mathbb{P}}$ be the canonical orthonormal basis of $\mathcal{H}$, and set

$$f \colon \mathcal{H} \to \left]-\infty, +\infty\right] : x \mapsto \max\Big\{1 + \langle x \mid e_1\rangle\,,\ \sup_{2\leqslant n\in\mathbb{N}} \big\langle x \mid \sqrt{n}e_n\big\rangle\Big\}. \tag{21.16}$$

Then $f \in \varGamma_0(\mathcal{H})$ and ∂f is maximally monotone. However, $\operatorname{gra}\partial f$ is not closed in $\mathcal{H}^{\mathrm{strong}} \times \mathcal{H}^{\mathrm{weak}}$. Furthermore,

$$\operatorname{Argmin} f = \Big\{(\xi_n)_{n\in\mathbb{P}} \in \mathcal{H} \;\Big|\; \xi_1 \leqslant -1,\ \sup_{2\leqslant n\in\mathbb{N}} \xi_n \leqslant 0\Big\}. \tag{21.17}$$

Proof. It follows from Proposition 9.3 that $f \in \varGamma(\mathcal{H})$. Note that, for every integer $n \geqslant 2$, $0 = f(-e_1) < 1 = f(0) = f(e_n/\sqrt{n}) < 2 = f(e_1)$; hence $f \in \varGamma_0(\mathcal{H})$ and ∂f is maximally monotone by Theorem 21.2. Since $0 \notin \operatorname{Argmin} f$, it follows from Theorem 16.3 that

$$(0,0) \notin \operatorname{gra}\partial f. \tag{21.18}$$

Now take an integer $n \geqslant 2$ and $x \in \mathcal{H}$. Then $f(x) - f(e_n/\sqrt{n}) = f(x) - 1 \geqslant \langle x \mid \sqrt{n}e_n\rangle - \langle e_n/\sqrt{n} \mid \sqrt{n}e_n\rangle = \langle x - e_n/\sqrt{n} \mid \sqrt{n}e_n\rangle$. Furthermore, $f(x) - f(e_1) = f(x) - 2 \geqslant 1 + \langle x \mid e_1\rangle - 2 = \langle x - e_1 \mid e_1\rangle$. Consequently, $(\forall n \in \mathbb{P})$ $\sqrt{n}e_n \in \partial f(e_n/\sqrt{n})$ and thus

$$\big\{\big(e_n/\sqrt{n}, \sqrt{n}e_n\big)\big\}_{n\in\mathbb{P}} \subset \operatorname{gra}\partial f. \tag{21.19}$$

In view of Example 3.33, there exists a net $(\sqrt{n(a)}e_{n(a)})_{a\in A}$ that converges weakly to 0. Now set $C = \{0\} \cup \{e_n/\sqrt{n}\}_{n\in\mathbb{P}}$. Since $e_n/\sqrt{n} \to 0$, the set C is compact and the net $(e_{n(a)}/\sqrt{n(a)})_{a\in A}$ lies in C. Thus, by Fact 1.11, there exists a subnet $(e_{n(b)}/\sqrt{n(b)})_{b\in B}$ of $(e_{n(a)}/\sqrt{n(a)})_{a\in A}$ that converges strongly to some point in C. Since $(\sqrt{n(b)}e_{n(b)})_{b\in B}$ is a subnet of $(\sqrt{n(a)}e_{n(a)})_{a\in A}$, it is clear that

$$\sqrt{n(b)}e_{n(b)} \rightharpoonup 0. \tag{21.20}$$

We claim that

$$e_{n(b)}/\sqrt{n(b)} \to 0. \tag{21.21}$$

Assume that this is not true. Then $e_m/\sqrt{m} = \lim_{b\in B} e_{n(b)}/\sqrt{n(b)}$ for some $m \in \mathbb{P}$. Since $e_m/\sqrt{m}$ is an isolated point of C, the elements of the net $(e_{n(b)}/\sqrt{n(b)})_{b\in B}$ are eventually equal to this point. In turn, this implies that the elements of the net $(\sqrt{n(b)}e_{n(b)})_{b\in B}$ are eventually equal to $\sqrt{m}e_m$, which contradicts (21.20). We thus have verified (21.21). To sum up, (21.18)–(21.21) imply that $(e_{n(b)}/\sqrt{n(b)}, \sqrt{n(b)}e_{n(b)})_{b\in B}$ lies in $\operatorname{gra}\partial f$, that it converges to $(0,0)$ in $\mathcal{H}^{\text{strong}} \times \mathcal{H}^{\text{weak}}$, and that its limit lies outside $\operatorname{gra}\partial f$. Now fix $\varepsilon \in \mathbb{R}_{++}$ and assume that $x = (\xi_n)_{n\in\mathbb{P}} \in \ell^2(\mathbb{P})$ satisfies $f(x) \leqslant -\varepsilon$. Then for every integer $n \geqslant 2$, $\sqrt{n}\xi_n \leqslant -\varepsilon \Rightarrow \varepsilon^2/n \leqslant \xi_n^2 \Rightarrow \|x\|^2 = +\infty$, which is impossible. Hence $\inf f(\mathcal{H}) \geqslant 0$, and (21.17) is proven. ☐

Remark 21.7 The proof of Example 21.6 implies that $\operatorname{gra}\partial f$ is not closed in $\mathcal{H}^{\text{strong}} \times \mathcal{H}^{\text{weak}}$ due to the existence of an unbounded net

$$(x_b, u_b)_{b\in B} = \big(e_{n(b)}/\sqrt{n(b)}, \sqrt{n(b)}e_{n(b)}\big)_{b\in B} \tag{21.22}$$

in $\operatorname{gra}\partial f$ converging to $(0,0) \notin \operatorname{gra}\partial f$ in $\mathcal{H}^{\text{strong}} \times \mathcal{H}^{\text{weak}}$. Clearly,

$$\lim \langle x_b \mid u_b\rangle = 1 \neq \langle 0 \mid 0\rangle\,, \tag{21.23}$$

which shows that Lemma 2.44 is false without the assumption on boundedness.

21.2 The Debrunner–Flor Theorem

Theorem 21.8 (Debrunner–Flor) *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a monotone operator such that $\operatorname{gra} A \neq \varnothing$. Then*

$$(\forall w \in \mathcal{H})(\exists\, x \in \overline{\operatorname{conv}}\operatorname{dom} A) \quad 0 \leqslant \inf_{(y,v)\in\operatorname{gra} A} \langle y - x \mid v - (w - x)\rangle\,. \tag{21.24}$$

Proof. Set $C = \overline{\operatorname{conv}}\operatorname{dom} A$. In view of Proposition 20.56(iii), we must show that $(\forall w \in \mathcal{H})$ $(\exists\, x \in C)$ $F_A(x, w - x) \leqslant \langle x \mid w - x\rangle$, i.e., that

$$(\forall w \in \mathcal{H}) \quad \min_{x\in\mathcal{H}} \big(F_A(x, w - x) + \|x\|^2 - \langle x \mid w\rangle + \iota_C(x)\big) \leqslant 0. \tag{21.25}$$

Let $w \in \mathcal{H}$. We consider two cases.

(a) $w = 0$: It suffices to show that

$$\min_{x\in\mathcal{H}} \Big(F_A(x, -x) + \big(\|x\|^2 + \iota_C(x)\big)\Big) \leqslant 0. \tag{21.26}$$

Set $q = (1/2)\|\cdot\|^2$, $f\colon \mathcal{H}\times\mathcal{H} \to \left]-\infty, +\infty\right] : (y,x) \mapsto (1/2)F_A^*(2y, 2x)$, $g = (q+\iota_C)^* = q-(1/2)d_C^2$ (by Example 13.5), and $L\colon \mathcal{H}\times\mathcal{H} \to \mathcal{H}\colon (y,x) \to x-y$. We claim that

$$\inf_{(y,x)\in\mathcal{H}\times\mathcal{H}} \Big(f(y,x)+g\big(L(y,x)\big)\Big) \geqslant 0. \tag{21.27}$$

To see this, fix $(y,x) \in \mathcal{H}\times\mathcal{H}$. By Proposition 20.61(ii), $\operatorname{dom} F_A^* \subset \overline{\operatorname{conv}}\,\operatorname{ran} A \times \overline{\operatorname{conv}}\,\operatorname{dom} A$, and we thus have to consider only the case $2x \in C$. Then, since $\langle\cdot \mid \cdot\rangle \leqslant F_A^*$ by Proposition 20.61(iii), we obtain

$$\begin{aligned}
0 &= 4\langle y \mid x\rangle + \|x-y\|^2 - \|x+y\|^2 \\
&= \langle 2y \mid 2x\rangle + \|x-y\|^2 - \|(x-y)-2x\|^2 \\
&\leqslant F_A^*(2y,2x) + \|x-y\|^2 - d_C^2(x-y) \\
&= 2\big(f(y,x)+g(x-y)\big) \\
&= 2\Big(f(y,x)+g\big(L(y,x)\big)\Big).
\end{aligned} \tag{21.28}$$

Since $\operatorname{dom} g = \mathcal{H}$, Theorem 15.23 implies that

$$\min_{x\in\mathcal{H}} \big(f^*(-L^*x)+g^*(x)\big) \leqslant 0. \tag{21.29}$$

Since $f^* = (1/2)F_A$, $g^* = q + \iota_C$, and $L^*\colon \mathcal{H}\to\mathcal{H}\times\mathcal{H}\colon x\mapsto(-x,x)$, we see that (21.29) is equivalent to (21.26).

(b) $w \neq 0$: Let $B\colon \mathcal{H}\to 2^{\mathcal{H}}$ be defined via $\operatorname{gra} B = -(0,w) + \operatorname{gra} A$. The above reasoning yields a point $(x,-x) \in C\times\mathcal{H}$ such that $\{(x,-x)\}\cup\operatorname{gra} B$ is monotone. Therefore, $\{(x,w-x)\}\cup\operatorname{gra} A$ is monotone. ☐

Theorem 21.9 *Let $A\colon \mathcal{H}\to 2^{\mathcal{H}}$ be monotone. Then there exists a maximally monotone extension $\widetilde{A}$ of A such that* $\operatorname{dom}\widetilde{A} \subset \overline{\operatorname{conv}}\,\operatorname{dom} A$.

Proof. Set $C = \overline{\operatorname{conv}}\,\operatorname{dom} A$ and let $\mathcal{M}$ be the set of all monotone extensions B of A such that $\operatorname{dom} B \subset C$. Order $\mathcal{M}$ partially via $(\forall B_1\in\mathcal{M})(\forall B_2\in\mathcal{M})$ $B_1 \preceq B_2 \Leftrightarrow \operatorname{gra} B_1 \subset \operatorname{gra} B_2$. Since every chain in $\mathcal{M}$ has its union as an upper bound, Zorn's lemma (Fact 1.1) yields a maximal element $\widetilde{A}$. Now let $w\in\mathcal{H}$ and assume that $w \in \mathcal{H}\smallsetminus\operatorname{ran}(\mathrm{Id}+\widetilde{A})$. Theorem 21.8 provides $x\in C$ such that $0 \leqslant \inf_{(y,v)\in\operatorname{gra}\widetilde{A}} \langle y-x \mid v-(w-x)\rangle$. Thus, $(x,w-x)\notin\operatorname{gra}\widetilde{A}$ and hence $\{(x,w-x)\}\cup\operatorname{gra}\widetilde{A}$ is the graph of an operator in $\mathcal{M}$ that properly extends $\widetilde{A}$. This contradicts the maximality of $\widetilde{A}$. We deduce that $\operatorname{ran}(\mathrm{Id}+\widetilde{A}) = \mathcal{H}$ and, by Theorem 21.1, that $\widetilde{A}$ is maximally monotone. ☐

21.3 Domain and Range

Definition 21.10 Let $A\colon \mathcal{H}\to 2^{\mathcal{H}}$ and let $x\in\mathcal{H}$. Then A is *locally bounded* at x if there exists $\delta\in\mathbb{R}_{++}$ such that $A(B(x;\delta))$ is bounded.

Proposition 21.11 *Let $A\colon \mathcal{H}\to 2^{\mathcal{H}}$ be monotone, set $Q_1\colon \mathcal{H}\times\mathcal{H}\to\mathcal{H}\colon (x,u)\mapsto x$, and suppose that $z\in\operatorname{int} Q_1(\operatorname{dom} F_A)$. Then A is locally bounded at z.*

Proof. Define

$$f\colon \mathcal{H} \to \left]-\infty,+\infty\right] : x \mapsto \sup_{(y,v)\in \operatorname{gra} A} \frac{\langle x-y \mid v\rangle}{1+\|y\|}, \tag{21.30}$$

and observe that $f \in \Gamma(\mathcal{H})$ by Proposition 9.3. Fix $(x,u) \in \mathcal{H}\times\mathcal{H}$ and let $(y,v) \in \operatorname{gra} A$. Then $\langle x \mid v\rangle + \langle y \mid u\rangle - \langle y \mid v\rangle \leqslant F_A(x,u)$ and hence $\langle x-y \mid v\rangle \leqslant F_A(x,u) - \langle y \mid u\rangle \leqslant \max\big\{F_A(x,u), \|u\|\big\}\big(1+\|y\|\big)$. Dividing by $1+\|y\|$ and then taking the supremum over $(y,v) \in \operatorname{gra} A$ yields

$$f(x) \leqslant \max\big\{F_A(x,u), \|u\|\big\}. \tag{21.31}$$

Hence $Q_1(\operatorname{dom} F_A) \subset \operatorname{dom} f$ and thus $z \in \operatorname{int}\operatorname{dom} f$. Corollary 8.39 implies the existence of $\delta \in \mathbb{R}_{++}$ such that $\sup f(B(z;2\delta)) \leqslant 1 + f(z)$. Hence, for every $(y,v) \in \operatorname{gra} A$ and every $w \in B(0;2\delta)$, $\langle z+w-y \mid v\rangle \leqslant (1+f(z))(1+\|y\|)$ or, equivalently,

$$2\delta\|v\| + \langle z-y \mid v\rangle \leqslant \big(1+f(z)\big)\big(1+\|y\|\big). \tag{21.32}$$

Now assume that $(y,v) \in \operatorname{gra} A$ and $y \in B(z;\delta)$. Using Cauchy–Schwarz and (21.32), we deduce that

$$\begin{aligned} \delta\|v\| &= 2\delta\|v\| - \delta\|v\| \\ &\leqslant 2\delta\|v\| + \langle z-y \mid v\rangle \\ &\leqslant \big(1+f(z)\big)\big(1+\|y\|\big) \\ &\leqslant \big(1+f(z)\big)\big(1+\|y-z\|+\|z\|\big) \\ &\leqslant \big(1+f(z)\big)\big(1+\delta+\|z\|\big). \end{aligned} \tag{21.33}$$

It follows that $\sup\|A(B(z;\delta))\| \leqslant (1+f(z))(1+\delta+\|z\|)/\delta < +\infty$. ☐

Proposition 21.12 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and set $Q_1\colon \mathcal{H}\times\mathcal{H} \to \mathcal{H}\colon (x,u) \mapsto x$. Then*

$$\operatorname{int}\operatorname{dom} A \subset \operatorname{int} Q_1(\operatorname{dom} F_A) \subset \operatorname{dom} A \subset Q_1(\operatorname{dom} F_A) \subset \overline{\operatorname{dom}}\, A. \tag{21.34}$$

Consequently, $\operatorname{int}\operatorname{dom} A = \operatorname{int} Q_1(\operatorname{dom} F_A)$ *and* $\overline{\operatorname{dom} A} = \overline{Q_1(\operatorname{dom} F_A)}$. *If* $\operatorname{int}\operatorname{dom} A \neq \varnothing$*, then* $\overline{\operatorname{int}\operatorname{dom} A} = \overline{\operatorname{dom} A}$*.*

Proof. Fix $x \in \mathcal{H}$. If $x \in \operatorname{dom} A$, then there exists $u \in \mathcal{H}$ such that $(x,u) \in \operatorname{gra} A$. Hence, by Proposition 20.56(i), $F_A(x,u) = \langle x \mid u\rangle \in \mathbb{R}$, which implies that $x \in Q_1(\operatorname{dom} F_A)$. As a result, the third, and thus the first, inclusions in (21.34) are verified. Now assume that $x \in Q_1(\operatorname{dom} F_A)$, say $F_A(x,u) < +\infty$ for some $u \in \mathcal{H}$. Let $\varepsilon \in \left]0,1\right[$, set

$$\beta = \max\big\{F_A(x,u), \|u\|, \|x\|\,\|u\|\big\}, \tag{21.35}$$

and pick $\lambda \in \big]0, \varepsilon^2/(9\beta)\big[$. Define $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ via

$$\operatorname{gra} B = \big\{\big((y-x)/\lambda, v\big) \;\big|\; (y,v) \in \operatorname{gra} A\big\}. \tag{21.36}$$

Then B is maximally monotone and therefore Theorem 21.1 yields $0 \in \operatorname{ran}(\mathrm{Id} + B)$. Hence there exists $z \in \mathcal{H}$ such that $-z \in Bz$, and thus $(y,v) \in \operatorname{gra} A$ such that $(z,-z) = ((y-x)/\lambda, v)$. In turn, $\langle x \mid v\rangle + \langle y \mid u\rangle - \langle y \mid v\rangle \leqslant F_A(x,u) \leqslant \beta$ and

$$-\langle y-x \mid v\rangle + \langle y-x \mid u\rangle \leqslant \beta - \langle x \mid u\rangle \leqslant \beta + \|x\|\,\|u\| \leqslant 2\beta. \tag{21.37}$$

Consequently, $-\langle y-x \mid v\rangle = -\langle y-x \mid -z\rangle = \|y-x\|^2/\lambda$, $\langle y-x \mid u\rangle \geqslant -\|y-x\|\,\|u\| \geqslant -\beta\|y-x\|$, and we obtain

$$\|y-x\|^2 - \lambda\beta\|y-x\| - 2\lambda\beta \leqslant 0. \tag{21.38}$$

Hence $\|y-x\|$ lies between the roots of the quadratic function $\rho \mapsto \rho^2 - \lambda\beta\rho - 2\lambda\beta$ and so, in particular, $\|y-x\| \leqslant (\lambda\beta + \sqrt{(\lambda\beta)^2 + 8(\lambda\beta)})/2 \leqslant \sqrt{\lambda\beta(\lambda\beta+8)} < \sqrt{(\varepsilon^2/9)(1+8)} = \varepsilon$. Thus, the fourth inclusion in (21.34) holds, i.e.,

$$Q_1(\operatorname{dom} F_A) \subset \overline{\operatorname{dom}}\, A. \tag{21.39}$$

Let us now verify the second inclusion in (21.34). Assume that $x \in \operatorname{int} Q_1(\operatorname{dom} F_A)$. By (21.39), there exists a sequence $(x_n)_{n\in\mathbb{N}}$ in $\operatorname{dom} A$ such that $x_n \to x$. Let $(u_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that $(x_n,u_n)_{n\in\mathbb{N}}$ lies in $\operatorname{gra} A$. It follows from Proposition 21.11 that $(u_n)_{n\in\mathbb{N}}$ is bounded. Hence $(u_n)_{n\in\mathbb{N}}$ possesses a weakly convergent subsequence, say $u_{k_n} \rightharpoonup u \in \mathcal{H}$. Proposition 20.38(i) implies that $(x,u) \in \operatorname{gra} A$ and thus $x \in \operatorname{dom} A$. Finally, we assume that $\operatorname{int}\operatorname{dom} A \neq \varnothing$. Proposition 3.45(iii) applied to $Q_1(\operatorname{dom} F_A)$ implies that $\overline{\operatorname{dom}}\, A = \overline{Q_1(\operatorname{dom} F_A)} = \overline{\operatorname{int} Q_1(\operatorname{dom} F_A)} = \overline{\operatorname{int}\operatorname{dom} A}$. ☐

Example 21.13 Let C be a nonempty closed convex subset of $\mathcal{H}$ and set $Q_1 \colon \mathcal{H}\times\mathcal{H} \to \mathcal{H} \colon (x,u) \mapsto x$. Then $Q_1(\operatorname{dom} F_{N_C}) = C$.

Proof. By (21.34), $C = \operatorname{dom} N_C \subset Q_1(\operatorname{dom} F_{N_C}) \subset \overline{\operatorname{dom}}\, N_C = \overline{C} = C$. ☐

Corollary 21.14 *Let $A \colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, and set $Q_1 \colon \mathcal{H}\times\mathcal{H} \to \mathcal{H} \colon (x,u) \mapsto x$ and $Q_2 \colon \mathcal{H}\times\mathcal{H} \to \mathcal{H} \colon (x,u) \mapsto u$. Then $\overline{\operatorname{dom}}\, A = \overline{Q_1(\operatorname{dom} F_A)}$, $\operatorname{int}\operatorname{dom} A = \operatorname{int} Q_1(\operatorname{dom} F_A)$, $\overline{\operatorname{ran}}\, A = \overline{Q_2(\operatorname{dom} F_A)}$, and $\operatorname{int}\operatorname{ran} A = \operatorname{int} Q_2(\operatorname{dom} F_A)$. Consequently, the sets $\overline{\operatorname{dom}}\, A$, $\operatorname{int}\operatorname{dom} A$, $\overline{\operatorname{ran}}\, A$, and $\operatorname{int}\operatorname{ran} A$ are convex.*

Proof. The first claims follow from Proposition 21.12, applied to A and A^{-1}, and from Proposition 20.56(vi). On the one hand, since F_A is convex, its domain is convex. On the other hand, since Q_1 and Q_2 are linear, the sets $Q_1(\operatorname{dom} F_A)$ and $Q_2(\operatorname{dom} F_A)$ are convex by Proposition 3.5, and so are their closures and interiors by Proposition 3.45(i)&(ii). ☐

Remark 21.15 In Corollary 21.14, the sets $\operatorname{dom} A$ and $\operatorname{ran} A$ may fail to be convex (see Remark 16.28).

Corollary 21.16 (Bunt–Kritikos–Motzkin) *Suppose that $\mathcal{H}$ is finite-dimensional and let C be a Chebyshev subset of $\mathcal{H}$. Then C is closed and convex.*

Proof. It follows from Remark 3.11(i), Example 20.33, and Corollary 21.14, that $C = \overline{C} = \overline{\operatorname{ran}}\, P_C$ is convex. ☐

21.4 Local Boundedness and Surjectivity

Proposition 21.17 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and let $x \in \operatorname{dom} A$. Then $\operatorname{rec}(Ax) = N_{\overline{\operatorname{dom}}\, A} x$.*

Proof. Fix $u \in Ax$ and $w \in \mathcal{H}$, and assume that $w \in N_{\overline{\operatorname{dom}}\, A} x$. Then $(\forall (y,v) \in \operatorname{gra} A)$ $0 \leqslant \langle x - y \mid w \rangle \leqslant \langle x - y \mid w \rangle + \langle x - y \mid u - v \rangle = \langle x - y \mid (u + w) - v \rangle$. The maximal monotonicity of A implies that $u + w \in Ax$. Hence $w + Ax \subset Ax$, i.e., $w \in \operatorname{rec}(Ax)$. Now assume that $w \in \operatorname{rec}(Ax)$, which implies that $(\forall \rho \in \mathbb{R}_{++})$ $u + \rho w \in Ax$. Let $y \in \operatorname{dom} A$ and $v \in Ay$. Then $(\forall \rho \in \mathbb{R}_{++})$ $0 \leqslant \langle x - y \mid (u + \rho w) - v \rangle = \langle x - y \mid u - v \rangle + \rho \langle x - y \mid w \rangle$. Since ρ can be arbitrarily large, we have $0 \leqslant \langle x - y \mid w \rangle$. Therefore, $w \in N_{\overline{\operatorname{dom}}\, A} x$. ☐

The next result, combined with Theorem 21.2, provides an extension to Proposition 16.17(iii).

Theorem 21.18 (Rockafellar–Veselý) *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and let $x \in \mathcal{H}$. Then A is locally bounded at x if and only if $x \notin \operatorname{bdry} \operatorname{dom} A$.*

Proof. Let S be the set of all points at which A is locally bounded. Clearly, $\mathcal{H} \smallsetminus \overline{\operatorname{dom}}\, A \subset S$. In addition, Proposition 21.11 and Proposition 21.12 imply that $\operatorname{int} \operatorname{dom} A \subset S$. We claim that

$$S \cap \overline{\operatorname{dom}}\, A = S \cap \operatorname{dom} A. \tag{21.40}$$

Let $x \in S \cap \overline{\operatorname{dom}}\, A$. Then there exists a sequence $(x_n, u_n)_{n \in \mathbb{N}}$ in $\operatorname{gra} A$ such that $x_n \to x$ and such that $(u_n)_{n \in \mathbb{N}}$ is bounded. After passing to a subsequence if necessary, we assume that $u_n \rightharpoonup u$. Now Proposition 20.38(i) yields $(x, u) \in \operatorname{gra} A$. Hence $x \in \operatorname{dom} A$, which verifies (21.40). It remains to show that $S \cap \operatorname{bdry} \operatorname{dom} A = \varnothing$. Assume to the contrary that $x \in S \cap \operatorname{bdry} \operatorname{dom} A$ and take $\delta \in \mathbb{R}_{++}$ such that $A(B(x; 2\delta))$ is bounded. Set $C = \overline{\operatorname{dom}}\, A$, which is convex by Corollary 21.14. Theorem 7.4 guarantees the existence of a support point z of C in $(\operatorname{bdry} C) \cap B(x; \delta)$. Take $w \in N_C z \smallsetminus \{0\}$. Since $B(z; \delta) \subset B(x; 2\delta)$, $z \in S \cap \operatorname{bdry} C$. Hence, by (21.40), $z \in S \cap \operatorname{dom} A \cap \operatorname{bdry} C$. Consequently, $Az \neq \varnothing$ and, by Proposition 21.17, $w \in \operatorname{rec}(Az)$. This implies that Az is unbounded, which, since $z \in S$, is impossible. ☐

Corollary 21.19 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be monotone, and suppose that C is a compact subset of* $\operatorname{int} \operatorname{dom} A$. *Then $A(C)$ is bounded.*

Proof. We assume that A is maximally monotone. Suppose that the conclusion fails. Then $C \neq \varnothing$ and there exist a sequence $(x_n, u_n)_{n\in\mathbb{N}}$ in $\operatorname{gra} A$ and a point $x \in C$ such that $x_n \to x$ and $\|u_n\| \to +\infty$. However, in view of Theorem 21.18, this contradicts the local boundedness of A at x. $\square$

Corollary 21.20 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone. Then A is locally bounded everywhere on $\mathcal{H}$ if and only if* $\operatorname{dom} A = \mathcal{H}$.

Corollary 21.21 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and at most single-valued. Then A is strong-to-weak continuous everywhere on* $\operatorname{int}\operatorname{dom} A$.

Proof. Fix a point $x \in \operatorname{int}\operatorname{dom} A$ and a sequence $(x_n)_{n\in\mathbb{N}}$ that converges to $x \in \operatorname{int}\operatorname{dom} A$. By Theorem 21.18, $(Ax_n)_{n\in\mathbb{N}}$ is bounded. Let y be a weak sequential cluster point of $(Ax_n)_{n\in\mathbb{N}}$, say $Ax_{k_n} \rightharpoonup y$. Proposition 20.38(i) implies that $(x, y) \in \operatorname{gra} A$. Since A is at most single-valued, we deduce that $y = Ax$. It follows from Lemma 2.46 that $Ax_n \rightharpoonup Ax$. $\square$

The verification of the next result is left as Exercise 21.9.

Example 21.22 Let $\mathcal{H} = \ell^2(\mathbb{N})$ and set $A\colon \mathcal{H} \to \mathcal{H}\colon x = (\xi_n)_{n\in\mathbb{N}} \mapsto (\xi_n/2^n)_{n\in\mathbb{N}}$. Then A is maximally monotone, locally bounded everywhere on $\mathcal{H}$, and $\operatorname{dom} A = \mathcal{H}$. Now set $B = A^{-1}$. Then B is maximally monotone, nowhere locally bounded, nowhere continuous, and $\operatorname{dom} B$ is a dense proper linear subspace of $\mathcal{H}$.

Corollary 21.23 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone. Then A is surjective if and only if A^{-1} is locally bounded everywhere on $\mathcal{H}$.*

Corollary 21.24 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator such that*

$$\lim_{\|x\|\to+\infty} \inf \|Ax\| = +\infty. \tag{21.41}$$

Then A is surjective.

Proof. In view of Corollary 21.23, let us show that A^{-1} is locally bounded on $\mathcal{H}$. Assume that A^{-1} is not locally bounded at $u \in \mathcal{H}$. Then there exists a sequence $(x_n, u_n)_{n\in\mathbb{N}}$ in $\operatorname{gra} A$ such that $u_n \to u$ and $\|x_n\| \to +\infty$. Hence, $+\infty = \lim_{\|x\|\to+\infty} \inf \|Ax\| = \liminf \|Ax_n\| \leqslant \lim \|u_n\| = \|u\|$, which is impossible. $\square$

Corollary 21.25 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone with bounded domain. Then A is surjective.*

The following result complements the boundedness property of Proposition 16.20(iii).

Corollary 21.26 *Let $f \in \Gamma_0(\mathcal{H})$ and let C be a compact subset of* $\operatorname{int}\operatorname{dom} f$. *Then $\partial f(C)$ is bounded.*

Proof. We first deduce from Proposition 16.27 that $\operatorname{int}\operatorname{dom} \partial f = \operatorname{int}\operatorname{dom} f$. Hence, the claim follows from Theorem 20.25 and Corollary 21.19. $\square$

21.5 Kenderov's Theorem and Fréchet Differentiability

Theorem 21.27 (Kenderov) *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator such that* $\operatorname{int}\operatorname{dom} A \neq \varnothing$*. Then there exists a subset C of* $\operatorname{int}\operatorname{dom} A$ *that is a dense G_δ subset of* $\overline{\operatorname{dom}}\, A$ *and such that, for every point $x \in C$, Ax is a singleton and every selection of A is continuous at x.*

Proof. Let $(\varepsilon_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}_{++}$ converging to 0, fix $n \in \mathbb{N}$, and define the set C_n by requiring that $y \in C_n$ if and only if there exists an open neighborhood V of y such that $\operatorname{diam} A(V) < \varepsilon_n$. It is clear that C_n is an open subset of $\operatorname{int}\operatorname{dom} A$. To show that C_n is dense in $\operatorname{int}\operatorname{dom} A$, take $y \in \operatorname{int}\operatorname{dom} A$. Since A is locally bounded at y by Theorem 21.18, there exists an open bounded set D such that $A(D)$ is bounded. Proposition 18.4 yields $z \in \mathcal{H}$ and $\alpha \in \mathbb{R}_{++}$ such that

$$\operatorname{diam} S < \varepsilon_n, \quad \text{where} \quad S = \big\{u \in A(D) \;\big|\; \langle z \mid u\rangle > \sigma_{A(D)}(z) - \alpha\big\}. \tag{21.42}$$

Take $u_1 \in S$. Then there exists $x_1 \in D$ such that $u_1 \in Ax_1$. Now let $\gamma \in \mathbb{R}_{++}$ be small enough to satisfy $x_0 = x_1 + \gamma z \in D$, and take $u_0 \in Ax_0$. Then $u_0 \in A(D)$ and $0 \leqslant \langle x_0 - x_1 \mid u_0 - u_1\rangle = \gamma \langle z \mid u_0 - u_1\rangle$, which implies that $\langle z \mid u_0\rangle \geqslant \langle z \mid u_1\rangle$ and so $u_0 \in S$. Thus,

$$x_0 \in D \quad \text{and} \quad Ax_0 \subset S. \tag{21.43}$$

In view of Proposition 20.38(i), there exists $\rho \in \mathbb{R}_{++}$ such that

$$B(x_0;\rho) \subset D \quad \text{and} \quad A\big(B(x_0;\rho)\big) \subset \big\{u \in \mathcal{H} \;\big|\; \langle z \mid u\rangle > \sigma_{A(D)}(z) - \alpha\big\}. \tag{21.44}$$

Consequently, $A(B(x_0;\rho)) \subset S$; hence $\operatorname{diam} A(B(x_0;\rho)) < \varepsilon_n$, and thus $x_0 \in C_n$.

We have shown that for every $n \in \mathbb{N}$, C_n is a dense open subset of $\operatorname{int}\operatorname{dom} A$. Set $C = \bigcap_{n\in\mathbb{N}} C_n$. By Corollary 1.45 and Proposition 21.12, C is a dense G_δ subset of $\overline{\operatorname{dom}}\, A = \overline{\operatorname{int}\operatorname{dom} A}$. Take $x \in C$. From the definition of C, it is clear that Ax is a singleton and that every selection of A is continuous at x. ☐

Corollary 21.28 *Let $f \in \Gamma_0(\mathcal{H})$ be such that* $\operatorname{int}\operatorname{dom} f \neq \varnothing$*. Then there exists a subset C of* $\operatorname{int}\operatorname{dom} f$ *such that C is a dense G_δ subset of* $\overline{\operatorname{dom}}\, f$*, and such that f is Fréchet differentiable on C.*

Exercises

Exercise 21.1 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be monotone. Show that A is maximally monotone if and only if $\operatorname{gra} A + \operatorname{gra}(-\mathrm{Id}) = \mathcal{H} \times \mathcal{H}$.

Exercise 21.2 Let $A \in \mathbb{R}^{N\times N}$ be such that $(\forall x \in \mathbb{R}^N)$ $\langle x \mid Ax\rangle \geqslant 0$. Show that $\mathrm{Id} + A$ is surjective without appealing to Theorem 21.1.

Exercise 21.3 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $\varepsilon \in \mathbb{R}_{++}$, denote the ε-enlargement of A by A^{ε} (see Exercise 20.19), and let $(x, u) \in \operatorname{gra} A^{\varepsilon}$. Use Theorem 21.1 to show that $d_{\operatorname{gra} A}(x, u) \leqslant \sqrt{2\varepsilon}$.

Exercise 21.4 Let $f \in \Gamma_0(\mathcal{H})$ be Gâteaux differentiable on $\mathcal{H}$. Show that ∇f is strong-to-weak continuous without appealing to Corollary 17.42(i).

Exercise 21.5 Let $A\colon \mathcal{H} \to \mathcal{H}$ be monotone. Show that the following are equivalent:

(i) A is maximally monotone.
(ii) A is strong-to-weak continuous.
(iii) A is hemicontinuous.

Exercise 21.6 Let $f \in \Gamma_0(\mathcal{H})$ be Gâteaux differentiable on $\mathcal{H}$, and suppose that $\lim_{\|x\|\to+\infty} \|\nabla f(x)\| = +\infty$. Show that $\operatorname{dom} f^* = \mathcal{H}$.

Exercise 21.7 Let $A\colon \mathcal{H} \to \mathcal{H}$ be monotone and hemicontinuous. Suppose that $\mathcal{H}$ is finite-dimensional or that A is linear. Show that A is maximally monotone and continuous.

Exercise 21.8 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and at most single-valued, and suppose that $A|_{\operatorname{dom} A}$ is linear. Show that $A \in \mathcal{B}(\mathcal{H})$ if and only if $\operatorname{dom} A = \mathcal{H}$.

Exercise 21.9 Verify Example 21.22.

Exercise 21.10 Suppose that $\mathcal{H} \neq \{0\}$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a monotone operator such that $\operatorname{gra} A$ is bounded, and let $\widetilde{A}$ be a maximally monotone extension of A. Show that $\operatorname{dom} \widetilde{A}$ or $\operatorname{ran} \widetilde{A}$ is unbounded.

Exercise 21.11 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$. Show that the following are equivalent:

(i) A maps every bounded subset of $\mathcal{H}$ to a bounded set.
(ii) For every sequence $(x_n, u_n)_{n\in\mathbb{N}}$ in $\operatorname{gra} A$ such that $(x_n)_{n\in\mathbb{N}}$ is bounded, it follows that $(u_n)_{n\in\mathbb{N}}$ is bounded.
(iii) For every sequence $(u_n, x_n)_{n\in\mathbb{N}}$ in $\operatorname{gra} A^{-1}$ such that $\|u_n\| \to +\infty$, it follows that $\|x_n\| \to +\infty$.

Exercise 21.12 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone. Suppose that $\mathcal{H}$ is finite-dimensional and that $\operatorname{dom} A = \mathcal{H}$. Show that A maps every bounded subset of $\mathcal{H}$ to a bounded set.

Exercise 21.13 Provide an example of a function $f \in \Gamma_0(\mathcal{H})$ such that $\operatorname{dom} \partial f = \mathcal{H}$ and ∂f does not map every bounded subset of $\mathcal{H}$ to a bounded set. Compare to Exercise 21.12 and Corollary 21.26.

Exercise 21.14 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be monotone, and suppose that x and y are distinct points in $\mathcal{H}$. Show that $\operatorname{int}(Ax) \cap \operatorname{int}(Ay) = \varnothing$.

Exercise 21.15 Prove Corollary 21.28.

Exercise 21.16 Show that Theorem 21.27 fails if $\operatorname{int} \operatorname{dom} A = \varnothing$.

Exercise 21.17 Whenever possible, provide instances in which $f\colon \mathcal{H} \to \mathbb{R}$ is convex and continuous, and one of the following holds:

(i) f is everywhere Fréchet differentiable.
(ii) f is Fréchet differentiable except at finitely many points.
(iii) f is Fréchet differentiable except at a countably infinite set of points.
(iv) f is Fréchet differentiable except at an uncountably infinite set of points.
(v) f is Fréchet differentiable only at countably many points.

Chapter 22
Stronger Notions of Monotonicity

This chapter collects basic results on various stronger notions of monotonicity (para, strict, uniform, strong, and cyclic) and their relationships to properties of convex functions. A fundamental result is Rockafellar's characterization of maximally cyclically monotone operators as subdifferential operators and a corresponding uniqueness result for the underlying convex function.

22.1 Para, Strict, Uniform, and Strong Monotonicity

Definition 22.1 An operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ is

(i) *paramonotone* if it is monotone and

$$\begin{aligned}&(\forall (x,u) \in \operatorname{gra} A)(\forall (y,v) \in \operatorname{gra} A)\\ &\quad \langle x-y \mid u-v\rangle = 0 \quad \Rightarrow \quad (x,v) \in \operatorname{gra} A \ \text{ and } \ (y,u) \in \operatorname{gra} A; \end{aligned} \tag{22.1}$$

(ii) *strictly monotone* if

$$(\forall (x,u) \in \operatorname{gra} A)(\forall (y,v) \in \operatorname{gra} A) \quad x \neq y \quad \Rightarrow \quad \langle x-y \mid u-v\rangle > 0; \tag{22.2}$$

(iii) *uniformly monotone* with modulus $\phi\colon \mathbb{R}_+ \to [0,+\infty]$ if ϕ is increasing, vanishes only at 0, and

$$(\forall (x,u) \in \operatorname{gra} A)(\forall (y,v) \in \operatorname{gra} A) \quad \langle x-y \mid u-v\rangle \geqslant \phi\big(\|x-y\|\big); \tag{22.3}$$

(iv) *strongly monotone* with constant $\beta \in \mathbb{R}_{++}$ if $A - \beta\mathrm{Id}$ is monotone, i.e.,

$$(\forall (x,u) \in \operatorname{gra} A)(\forall (y,v) \in \operatorname{gra} A) \quad \langle x-y \mid u-v\rangle \geqslant \beta\|x-y\|^2. \tag{22.4}$$

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_22

It is clear that strong monotonicity implies uniform monotonicity, which itself implies strict monotonicity, which itself implies paramonotonicity, which itself implies monotonicity.

Proposition 22.2 *Let $\mathcal{K}$ be a real Hilbert space, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{K} \to 2^{\mathcal{K}}$ be paramonotone operators, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$. Then the following hold:*

(i) A^{-1} *is paramonotone.*
(ii) $A + L^*BL$ *is paramonotone.*

Proof. (i): Proposition 20.10 asserts that A^{-1} is monotone. Now suppose that (x_1, u_1) and (x_2, u_2) are points in $\operatorname{gra} A^{-1}$ such that $\langle x_1 - x_2 \mid u_1 - u_2\rangle = 0$. Then (u_1, x_1) and (u_2, x_2) are points in the graph of the paramonotone operator A and so are therefore (u_2, x_1) and (u_1, x_2). In turn, (x_1, u_2) and (x_2, u_1) are in $\operatorname{gra} A^{-1}$, and we conclude that A^{-1} is paramonotone.

(ii): As seen in Proposition 20.10, $A + L^*BL$ is monotone. Now suppose that (x_1, w_1) and (x_2, w_2) are points in $\operatorname{gra}(A + L^*BL)$ such that

$$\langle x_1 - x_2 \mid w_1 - w_2\rangle = 0. \tag{22.5}$$

Then there exist $u_1 \in Ax_1$, $u_2 \in Ax_2$, $v_1 \in B(Lx_1)$, and $v_2 \in B(Lx_2)$, such that $w_1 = u_1 + L^*v_1$ and $w_2 = u_2 + L^*v_2$. Thus,

$$\begin{aligned} 0 &= \langle x_1 - x_2 \mid w_1 - w_2\rangle \\ &= \langle x_1 - x_2 \mid (u_1 + L^*v_1) - (u_2 + L^*v_2)\rangle \\ &= \langle x_1 - x_2 \mid u_1 - u_2\rangle + \langle Lx_1 - Lx_2 \mid v_1 - v_2\rangle . \end{aligned} \tag{22.6}$$

Therefore, the monotonicity of A and B yields $\langle x_1 - x_2 \mid u_1 - u_2\rangle = 0$ and $\langle Lx_1 - Lx_2 \mid v_1 - v_2\rangle = 0$, respectively. In turn, the paramonotonicity of A and B implies that $(u_1, u_2) \in Ax_2 \times Ax_1$ and $(v_1, v_2) \in B(Lx_2) \times B(Lx_1)$, respectively. Altogether, $(x_2, w_1) = (x_2, u_1 + L^*v_1) \in \operatorname{gra}(A + L^*BL)$ and $(x_1, w_2) = (x_1, u_2 + L^*v_2) \in \operatorname{gra}(A + L^*BL)$, which shows that $A + L^*BL$ is paramonotone. ☐

Remark 22.3 The notions of strict, uniform, and strong monotonicity of $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ can naturally be localized to a subset C of $\operatorname{dom} A$. For instance, A is uniformly monotone on C if there exists an increasing function $\phi\colon \mathbb{R}_+ \to [0,+\infty]$ vanishing only at 0 such that

$$(\forall x \in C)(\forall y \in C)(\forall u \in Ax)(\forall v \in Ay) \;\; \langle x - y \mid u - v\rangle \geqslant \phi\big(\|x - y\|\big). \tag{22.7}$$

Example 22.4 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex. Then the following hold:

(i) ∂f is paramonotone.
(ii) Suppose that f is strictly convex. Then ∂f is strictly monotone.

(iii) Suppose that f is uniformly convex with modulus ϕ. Then ∂f is uniformly monotone with modulus 2ϕ.
(iv) Suppose that f is strongly convex with constant $\beta \in \mathbb{R}_{++}$. Then ∂f is strongly monotone with constant β.

Proof. We assume that $\operatorname{dom} \partial f$ contains at least two elements since the conclusion is clear otherwise. Take (x, u) and (y, v) in $\operatorname{gra} \partial f$, and $\alpha \in \,]0, 1[$. Then (16.1) yields

$$\alpha \langle x - y \mid v \rangle = \langle (\alpha x + (1-\alpha)y) - y \mid v \rangle \leqslant f\big(\alpha x + (1-\alpha)y\big) - f(y). \quad (22.8)$$

(i): Assume that $\langle x - y \mid u - v \rangle = 0$. It follows from Proposition 16.10 that $0 = \langle x - y \mid u - v \rangle = \langle x \mid u \rangle + \langle y \mid v \rangle - \langle x \mid v \rangle - \langle y \mid u \rangle = f(x) + f^*(u) + f(y) + f^*(v) - \langle x \mid v \rangle - \langle y \mid u \rangle = (f(x) + f^*(v) - \langle x \mid v \rangle) + (f(y) + f^*(u) - \langle y \mid u \rangle)$. Hence, by Proposition 13.15, $v \in \partial f(x)$ and $u \in \partial f(y)$.

(ii): Assume that $x \neq y$. Then (22.8) and (8.3) imply that $\langle x - y \mid v \rangle < f(x) - f(y)$. Likewise, $\langle y - x \mid u \rangle < f(y) - f(x)$. Adding these two inequalities yields (22.2).

(iii): It follows from (22.8) and (10.2) that $\langle x - y \mid v \rangle + (1-\alpha)\phi(\|x - y\|) \leqslant f(x) - f(y)$. Letting $\alpha \downarrow 0$, we obtain $\langle x - y \mid v \rangle + \phi(\|x - y\|) \leqslant f(x) - f(y)$. Likewise, $\langle y - x \mid u \rangle + \phi(\|x - y\|) \leqslant f(y) - f(x)$. Adding these two inequalities yields $\langle x - y \mid u - v \rangle \geqslant 2\phi(\|x - y\|)$.

(iv): Since f is uniformly convex with modulus $t \mapsto (1/2)\beta t^2$, we deduce from (iii) that ∂f is uniformly monotone with modulus $t \mapsto \beta t^2$, i.e., that ∂f is strongly monotone with constant β. ☐

Example 22.5 Let $f\colon \mathcal{H} \to \,]-\infty, +\infty]$ be proper and convex, let C be a nonempty subset of $\operatorname{dom} \partial f$, and suppose that f is uniformly convex on C. Then ∂f is uniformly monotone on C.

Proof. Use Remark 22.3 and proceed as in the proof of Example 22.4(iii). ☐

The next result refines Example 20.7.

Example 22.6 Let D be a nonempty subset of $\mathcal{H}$, let $T\colon D \to \mathcal{H}$, let $\alpha \in [-1, 1]$, and set $A = \operatorname{Id} + \alpha T$. Then the following hold:

(i) Suppose that T is strictly nonexpansive. Then A is strictly monotone.
(ii) Suppose that T is β-Lipschitz continuous, with $\beta \in [0, 1[$. Then A is strongly monotone with modulus $1 - |\alpha|\beta$.

Proof. (i): This follows from (20.4).

(ii): It follows from (20.4) that, for every x and y in D, $\langle x - y \mid Ax - Ay \rangle \geqslant (1 - |\alpha|\beta)\|x - y\|^2$. ☐

Example 22.7 Let D be a nonempty subset of $\mathcal{H}$, let $\beta \in \mathbb{R}_{++}$, let $T\colon D \to \mathcal{H}$, and set $A = T^{-1}$. Then T is β-cocoercive if and only if A is strongly monotone with constant β.

Proof. Take (x, u) and (y, v) in $\operatorname{gra} A$. Then $u \in T^{-1}x$ and $v \in T^{-1}y$. Hence, $x = Tu$, $y = Tv$, and $\langle u - v \mid x - y\rangle \geqslant \beta\|x - y\|^2 \Leftrightarrow \langle u - v \mid Tu - Tv\rangle \geqslant \beta\|Tu - Tv\|^2$. Thus, the conclusion follows from (4.12) and (22.4). ☐

Example 22.8 Let D be a nonempty subset of $\mathcal{H}$, let $A\colon D \to \mathcal{H}$, and suppose that one of the following holds:

(i) A is strictly monotone.
(ii) A^{-1} is strictly monotone.
(iii) A is cocoercive.

Then A is paramonotone.

Proof. (i): See Definition 22.1.
(ii): This follows from (i) and Proposition 22.2(i).
(iii): Combine Example 22.7 with (ii). ☐

The following result sharpens Example 20.7.

Example 22.9 Let D be a nonempty subset of $\mathcal{H}$, let $T\colon D \to \mathcal{H}$ be nonexpansive, let $\alpha \in [-1, 1]$, and set $A = \mathrm{Id} + \alpha T$. Then A is $\frac{1}{2}$-cocoercive and paramonotone.

Proof. Since $\alpha T = 2(\frac{1}{2}A) - \mathrm{Id}$ is nonexpansive, it follows from Proposition 4.4 that $\frac{1}{2}A$ is firmly nonexpansive, i.e., A is $\frac{1}{2}$-cocoercive. Paramonotonicity follows from Example 22.8(iii). ☐

Proposition 22.10 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be monotone. Suppose that x, y, and z belong to* $\operatorname{dom} A$*, that $x \neq y$, that $u \in Ax \cap Ay$, and that $z \in \,]x, y[$. Then the following hold:*

(i) *$Az \subset u + (x - y)^{\perp}$.*
(ii) *Suppose that A is paramonotone. Then $Az = Ax \cap Ay$.*

Proof. (i): Let $v \in Az$ and write $z = (1 - \alpha)x + \alpha y$, where $\alpha \in \,]0, 1[$. The monotonicity of A yields $0 \leqslant \langle z - x \mid v - u\rangle = \alpha \langle y - x \mid v - u\rangle$ and $0 \leqslant \langle z - y \mid v - u\rangle = (1 - \alpha) \langle x - y \mid v - u\rangle$. Hence $\langle x - y \mid v - u\rangle = 0$.
(ii): By (i), we have $\langle x - z \mid u - v\rangle = \langle y - z \mid u - v\rangle = 0$. Since A is paramonotone, we deduce that $u \in Az$ and that $v \in Ax \cap Ay$. ☐

Proposition 22.11 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and suppose that one of the following holds:*

(i) *A is uniformly monotone with a supercoercive modulus.*
(ii) *A is strongly monotone.*

Then $\lim_{\|x\|\to+\infty} \inf \|Ax\| = +\infty$ and A is surjective.

Proof. (i): Let ϕ be the modulus of uniform convexity of A and fix $(y,v) \in \operatorname{gra} A$. Then (22.3) and Cauchy–Schwarz yield

$$(\forall (x,u) \in \operatorname{gra} A) \ \ \|x-y\|\,\|u\| \geqslant \langle x-y \mid u\rangle = \langle x-y \mid u-v\rangle + \langle x-y \mid v\rangle \geqslant \phi\big(\|x-y\|\big) - \|x-y\|\,\|v\|. \tag{22.9}$$

Accordingly, since $\lim_{t\to+\infty} \phi(t)/t = +\infty$, we have $\inf_{u\in Ax}\|u\| \to +\infty$ as $\|x\| \to +\infty$. In view of Corollary 21.24, the proof is complete.

(ii): Clear from (i). ☐

Example 22.12 Let $A\colon \mathcal{H} \to \mathcal{H}$ be monotone and hemicontinuous, and let $r \in \mathcal{H}$. Suppose that one of the following holds:

(i) A is strictly monotone and $\lim_{\|x\|\to+\infty} \|Ax\| = +\infty$.
(ii) A is uniformly monotone with a supercoercive modulus.
(iii) A is strongly monotone.

Then the equation $Ax = r$ has exactly one solution.

Proof. By Proposition 20.27, A is maximally monotone.

(i): The existence of a solution follows from Corollary 21.24, and its uniqueness from (22.2).

(ii)&(iii): On the one hand, A is strictly monotone. On the other hand, $\lim_{\|x\|\to+\infty} \|Ax\| = +\infty$ by Proposition 22.11. Altogether, the conclusion follows from (i). ☐

22.2 Cyclic Monotonicity

Definition 22.13 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and let $n \in \mathbb{N}$ be such that $n \geqslant 2$. Then A is *n-cyclically monotone* if, for every $(x_1,\ldots,x_{n+1}) \in \mathcal{H}^{n+1}$ and every $(u_1,\ldots,u_n) \in \mathcal{H}^n$,

$$\left.\begin{matrix} (x_1,u_1) \in \operatorname{gra} A, \\ \vdots \\ (x_n,u_n) \in \operatorname{gra} A, \\ x_{n+1} = x_1, \end{matrix}\right\} \quad\Rightarrow\quad \sum_{i=1}^{n} \langle x_{i+1} - x_i \mid u_i\rangle \leqslant 0. \tag{22.10}$$

If A is n-cyclically monotone for every integer $n \geqslant 2$, then A is *cyclically monotone*. If A is cyclically monotone and if there exists no cyclically monotone operator $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ such that $\operatorname{gra} B$ properly contains $\operatorname{gra} A$, then A is *maximally cyclically monotone*.

It is clear that the notions of monotonicity and 2-cyclic monotonicity coincide, and that n-cyclic monotonicity implies m-cyclic monotonicity for every $m \in \{2, \ldots, n\}$. A maximally monotone operator that is cyclically monotone is also maximally cyclically monotone.

Proposition 22.14 *Let $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ be proper. Then ∂f is cyclically monotone.*

Proof. Fix an integer $n \geqslant 2$. For every $i \in \{1, \ldots, n\}$, take $(x_i, u_i) \in \operatorname{gra} \partial f$. Set $x_{n+1} = x_1$. Then (16.1) yields

$$(\forall i \in \{1, \ldots, n\}) \quad \langle x_{i+1} - x_i \mid u_i \rangle \leqslant f(x_{i+1}) - f(x_i). \tag{22.11}$$

Adding up these inequalities, we obtain $\sum_{i=1}^{n} \langle x_{i+1} - x_i \mid u_i \rangle \leqslant 0$. ☐

However, there are maximally monotone operators that are not 3-cyclically monotone.

Example 22.15 Suppose that $\mathcal{H} = \mathbb{R}^2$, and set

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \tag{22.12}$$

Then A is maximally monotone but not 3-cyclically monotone.

Proof. The maximal monotonicity of A follows from Example 20.35. Observe that $A^2 = -\mathrm{Id}$ and that $(\forall x \in \mathcal{H})$ $\|Ax\| = \|x\|$ and $\langle x \mid Ax \rangle = 0$. Now let $x_1 \in \mathcal{H} \smallsetminus \{0\}$, and set $x_2 = Ax_1$, $x_3 = Ax_2 = A^2 x_1 = -x_1$, and $x_4 = x_1$. Then

$$\begin{aligned} \sum_{i=1}^{3} \langle x_{i+1} - x_i \mid Ax_i \rangle &= \langle x_2 \mid Ax_1 \rangle + \langle x_3 \mid Ax_2 \rangle + \langle x_1 \mid Ax_3 \rangle \\ &= \|Ax_1\|^2 + \langle -x_1 \mid -x_1 \rangle + \langle x_1 \mid -Ax_1 \rangle \\ &= 2\|x_1\|^2 \\ &> 0. \end{aligned} \tag{22.13}$$

Therefore, A is not 3-cyclically monotone. ☐

Remark 22.16 The operator A of Example 22.15 is the counterclockwise rotator by $\pi/2$. More generally, let $\theta \in [0, \pi/2]$, let $n \geqslant 2$ be an integer, and let A be the counterclockwise rotator in $\mathbb{R}^2$ by θ. Then A is n-cyclically monotone if and only if $\theta \in [0, \pi/n]$; see [7] and [39].

Proposition 22.17 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be 3-cyclically monotone and maximally monotone. Then A is paramonotone.*

Proof. Let (x, u) and (y, v) be points in $\operatorname{gra} A$ such that $\langle x - y \mid u - v \rangle = 0$. Since A is 3-cyclically monotone, we have, for every $(z, w) \in \operatorname{gra} A$,

$$\begin{aligned}
0 &\geqslant \langle y - x \mid u \rangle + \langle z - y \mid v \rangle + \langle x - z \mid w \rangle \\
&= \langle y - x \mid u - v \rangle + \langle (z - y) + (y - x) \mid v \rangle + \langle x - z \mid w \rangle \\
&= \langle z - x \mid v - w \rangle . \qquad (22.14)
\end{aligned}$$

The maximal monotonicity of A implies that $(x, v) \in \operatorname{gra} A$. Likewise, we show that $(y, u) \in \operatorname{gra} A$. ☐

22.3 Rockafellar's Cyclic Monotonicity Theorem

Theorem 22.18 (Rockafellar) *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$. Then A is maximally cyclically monotone if and only if there exists $f \in \Gamma_0(\mathcal{H})$ such that $A = \partial f$.*

Proof. Suppose that $A = \partial f$ for some $f \in \Gamma_0(\mathcal{H})$. Theorem 21.2 and Proposition 22.14 imply that A is maximally monotone and cyclically monotone. Hence A is maximally cyclically monotone. Conversely, suppose that A is maximally cyclically monotone. Then $\operatorname{gra} A \neq \varnothing$. Take $(x_0, u_0) \in \operatorname{gra} A$ and set

$$\begin{aligned}
f\colon \mathcal{H} &\to [-\infty, +\infty] \\
x &\mapsto \sup_{\substack{n \in \mathbb{N} \\ n \geqslant 1}} \sup_{\substack{(x_1, u_1) \in \operatorname{gra} A \\ \vdots \\ (x_n, u_n) \in \operatorname{gra} A}} \left\{ \langle x - x_n \mid u_n \rangle + \sum_{i=0}^{n-1} \langle x_{i+1} - x_i \mid u_i \rangle \right\}. \qquad (22.15)
\end{aligned}$$

Since $\operatorname{gra} A \neq \varnothing$, we deduce that $-\infty \notin f(\mathcal{H})$. On the other hand, it follows from Proposition 9.3 that $f \in \Gamma(\mathcal{H})$. The cyclic monotonicity of A implies that $f(x_0) = 0$. Altogether, $f \in \Gamma_0(\mathcal{H})$. Now take $(x, u) \in \operatorname{gra} A$ and $\eta \in \left]-\infty, f(x)\right[$. Then there exist finitely many points $(x_1, u_1), \ldots, (x_n, u_n)$ in $\operatorname{gra} A$ such that

$$\langle x - x_n \mid u_n \rangle + \sum_{i=0}^{n-1} \langle x_{i+1} - x_i \mid u_i \rangle > \eta. \qquad (22.16)$$

Set $(x_{n+1}, u_{n+1}) = (x, u)$. Using (22.16), we deduce that, for every $y \in \mathcal{H}$,

$$\begin{aligned}
f(y) &\geqslant \langle y - x_{n+1} \mid u_{n+1} \rangle + \sum_{i=0}^{n} \langle x_{i+1} - x_i \mid u_i \rangle \\
&= \langle y - x \mid u \rangle + \langle x - x_n \mid u_n \rangle + \sum_{i=0}^{n-1} \langle x_{i+1} - x_i \mid u_i \rangle \\
&> \langle y - x \mid u \rangle + \eta. \qquad (22.17)
\end{aligned}$$

Letting $\eta \uparrow f(x)$, we deduce that $(\forall y \in \mathcal{H})\ f(y) \geqslant f(x) + \langle y - x \mid u\rangle$, i.e., $u \in \partial f(x)$. Therefore,

$$\operatorname{gra} A \subset \operatorname{gra} \partial f. \tag{22.18}$$

Since ∂f is cyclically monotone by Proposition 22.14, and since A is maximally cyclically monotone, we conclude that $A = \partial f$. ☐

Proposition 22.19 *Let f and g be functions in $\Gamma_0(\mathcal{H})$ such that $\partial f = \partial g$. Then there exists $\gamma \in \mathbb{R}$ such that $f = g + \gamma$.*

Proof. Consider first the special case in which f and g are differentiable on $\mathcal{H}$. Fix x and y in $\mathcal{H}$, and set $\varphi\colon \mathbb{R} \to \mathbb{R}\colon t \mapsto f(x + t(y - x))$ and $\psi\colon \mathbb{R} \to \mathbb{R}\colon t \mapsto g(x + t(y - x))$. We have

$$\begin{aligned}(\forall t \in \mathbb{R}) \quad \varphi'(t) &= \langle y - x \mid \nabla f(x + t(y - x))\rangle \\ &= \langle y - x \mid \nabla g(x + t(y - x))\rangle \\ &= \psi'(t).\end{aligned} \tag{22.19}$$

Hence, using Corollary 17.42, $f(y) - f(x) = \varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)dt = \int_0^1 \psi'(t)dt = \psi(1) - \psi(0) = g(y) - g(x)$. We conclude that $f - g$ is a constant. Now we turn our attention to the general case. Set $q = (1/2)\|\cdot\|^2$ and recall from Example 13.6 that $q^* = q$. Using Corollary 16.48(iii), Corollary 16.30, and Proposition 14.1, we obtain the equivalences $\partial f = \partial g \Leftrightarrow \operatorname{Id} + \partial f = \operatorname{Id} + \partial g \Leftrightarrow \partial(q + f) = \partial(q + g) \Leftrightarrow (\partial(q + f))^{-1} = (\partial(q + g))^{-1} \Leftrightarrow \partial(q + f)^* = \partial(q + g)^* \Leftrightarrow \partial(f^* \boxdot q^*) = \partial(g^* \boxdot q^*) \Leftrightarrow \partial(f^* \boxdot q) = \partial(g^* \boxdot q)$. The functions $f^* \boxdot q$ and $g^* \boxdot q$ are differentiable on $\mathcal{H}$ and their gradients coincide by Proposition 12.30 and Proposition 17.31(i). Thus, by the already verified special case, there exists $\gamma \in \mathbb{R}$ such that $f^* \boxdot q = g^* \boxdot q - \gamma$. Hence $f + q = f^{**} + q^* = (f^* \boxdot q)^* = (g^* \boxdot q - \gamma)^* = (g^* \boxdot q)^* + \gamma = g^{**} + q^* + \gamma = g + q + \gamma$, and we conclude that $f = g + \gamma$. ☐

22.4 Monotone Operators on $\mathbb{R}$

We introduce a binary relation on $\mathbb{R}^2$ via

$$\begin{aligned}&\big(\forall \boldsymbol{x}_1 = (x_1, u_1) \in \mathbb{R}^2\big)\big(\forall \boldsymbol{x}_2 = (x_2, u_2) \in \mathbb{R}^2\big)\\ &\qquad\qquad \boldsymbol{x}_1 \preccurlyeq \boldsymbol{x}_2 \quad\Leftrightarrow\quad x_1 \leqslant x_2 \text{ and } u_1 \leqslant u_2;\end{aligned} \tag{22.20}$$

and we shall write $\boldsymbol{x}_1 \prec \boldsymbol{x}_2$ if $\boldsymbol{x}_1 \preccurlyeq \boldsymbol{x}_2$ and $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$.

We leave the proof of the following results as Exercise 22.7 and Exercise 22.8. (See Section 1.3 for a discussion of order.)

Proposition 22.20 $(\mathbb{R}^2, \preccurlyeq)$ *is directed and partially ordered, but not totally ordered.*

Proposition 22.21 *Let* $A\colon \mathbb{R} \to 2^{\mathbb{R}}$ *be such that* $\operatorname{gra} A \neq \varnothing$. *Then* A *is monotone if and only if* $\operatorname{gra} A$ *is a chain in* $(\mathbb{R}^2, \preccurlyeq)$.

Theorem 22.22 *Let* $A\colon \mathbb{R} \to 2^{\mathbb{R}}$ *be such that* $\operatorname{gra} A \neq \varnothing$. *Then* A *is monotone if and only if it is cyclically monotone.*

Proof. It is clear that cyclic monotonicity implies monotonicity. We assume that A is n-cyclically monotone, and we shall show that A is $(n+1)$-cyclically monotone. To this end, let $\boldsymbol{x}_1 = (x_1, u_1), \ldots, \boldsymbol{x}_{n+1} = (x_{n+1}, u_{n+1})$ be in $\operatorname{gra} A$, and set $x_{n+2} = x_1$. It suffices to show that

$$\sum_{i=1}^{n+1} (x_{i+1} - x_i)u_i \leqslant 0. \tag{22.21}$$

The n-cyclic monotonicity of A yields $\sum_{i=1}^{n-1}(x_{i+1} - x_i)u_i + (x_1 - x_n)u_n \leqslant 0$, i.e.,

$$\sum_{i=1}^{n-1} (x_{i+1} - x_i)u_i \leqslant -(x_1 - x_n)u_n. \tag{22.22}$$

Now define $B\colon \mathbb{R} \to 2^{\mathbb{R}}$ by $\operatorname{gra} B = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n+1}\}$. Since A is monotone, so is B. By Proposition 22.21, $\operatorname{gra} B$ is a chain in $(\mathbb{R}^2, \preccurlyeq)$. Since $\operatorname{gra} B$ contains only finitely many elements, it possesses a least element, which—after cyclically relabeling if necessary—we assume to be $\boldsymbol{x}_{n+1}$. After translating $\operatorname{gra} A$ if necessary, we assume in addition that $\boldsymbol{x}_{n+1} = (x_{n+1}, u_{n+1}) = (0, 0)$, so that $\operatorname{gra} B \subset \mathbb{R}^2_+$. Using (22.22), we thus obtain

$$\begin{aligned}\sum_{i=1}^{n+1} (x_{i+1} - x_i)u_i &= \sum_{i=1}^{n-1} (x_{i+1} - x_i)u_i + (0 - x_n)u_n + (x_{n+2} - 0)0 \leqslant \\ &\quad - (x_1 - x_n)u_n - x_n u_n = -x_1 u_n \leqslant 0. \end{aligned} \tag{22.23}$$

This verifies (22.21). Therefore, by induction, A is cyclically monotone. ☐

In view of Example 22.15, the following result fails already in the Euclidean plane.

Corollary 22.23 *Let* $A\colon \mathbb{R} \to 2^{\mathbb{R}}$ *be maximally monotone. Then there exists* $f \in \Gamma_0(\mathbb{R})$ *such that* $A = \partial f$.

Proof. Combine Theorem 22.22 and Theorem 22.18. ☐

Exercises

Exercise 22.1 Provide an example of a maximally monotone operator that is not paramonotone.

Exercise 22.2 Provide an example of a maximally monotone operator that is paramonotone, but not strictly monotone.

Exercise 22.3 Provide an example of a maximally monotone operator that is strictly monotone, but not uniformly monotone.

Exercise 22.4 Provide an example of a maximally monotone operator that is uniformly monotone, but not strongly monotone.

Exercise 22.5 Suppose that $\mathcal{H}$ is finite-dimensional, and let $A \in \mathcal{B}(\mathcal{H})$. Show that A is strictly monotone if and only if it is strongly monotone.

Exercise 22.6 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and let $n \in \mathbb{N}$ be such that $n \geqslant 2$. Denote the Hilbert direct sum $\mathcal{H}^n$ by $\boldsymbol{\mathcal{H}}$ and the cyclic right-shift operator on $\boldsymbol{\mathcal{H}}$ by $\boldsymbol{R}$, so that $\boldsymbol{R} : \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}\colon (x_1, x_2, \dots, x_n) \mapsto (x_n, x_1, \dots, x_{n-1})$. Show that A is n-cyclically monotone if and only if for all $(x_1, u_1) \in \operatorname{gra} A, \dots, (x_n, u_n) \in \operatorname{gra} A$ we have

$$\|\boldsymbol{x} - \boldsymbol{u}\| \leqslant \|\boldsymbol{x} - \boldsymbol{R}\boldsymbol{u}\|, \tag{22.24}$$

where $\boldsymbol{x} = (x_1, \dots, x_n)$ and $\boldsymbol{u} = (u_1, \dots, u_n)$; equivalently,

$$\sum_{i=1}^{n} \|x_i - u_i\|^2 \leqslant \sum_{i=1}^{n} \|x_i - u_{i-1}\|^2, \tag{22.25}$$

where $u_0 = u_n$.

Exercise 22.7 Prove Proposition 22.20.

Exercise 22.8 Prove Proposition 22.21.

Exercise 22.9 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and let $n \in \mathbb{N}$ be such that $n \geqslant 2$. The *Fitzpatrick function of order* n at $(x, u) \in \mathcal{H} \times \mathcal{H}$ is defined by

$$\sup \Bigg(\langle x \mid u \rangle + \bigg(\sum_{i=1}^{n-2} \langle y_{i+1} - y_i \mid v_i \rangle \bigg) + \langle x - y_{n-1} \mid v_{n-1} \rangle + \langle y_1 - x \mid u \rangle \Bigg), \tag{22.26}$$

where the supremum is taken over $(y_1, v_1), \dots, (y_{n-1}, v_{n-1})$ in $\operatorname{gra} A$, and set $F_{A,\infty} = \sup_{n \in \{2,3,\dots\}} F_{A,n}$. Show that $F_{A,n}\colon \mathcal{H} \times \mathcal{H} \to [-\infty, +\infty]$ is lower semicontinuous and convex, and that $F_{A,n} \geqslant \langle \cdot \mid \cdot \rangle$ on $\operatorname{gra} A$. What is $F_{A,2}$?

Exercise 22.10 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and let $n \in \mathbb{N}$ be such that $n \geqslant 2$. Show that A is n-cyclically monotone if and only if $F_{A,n} = \langle \cdot \mid \cdot \rangle$ on $\operatorname{gra} A$.

Exercise 22.11 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$. Show that A is cyclically monotone if and only if $F_{A,\infty} = \langle \cdot \mid \cdot \rangle$ on $\operatorname{gra} A$.

Exercise 22.12 Let $f \in \Gamma_0(\mathcal{H})$. Show that $F_{\partial f,\infty} = f \oplus f^*$.

Exercise 22.13 Let C be a nonempty closed convex subset of $\mathcal{H}$ and let $n \in \mathbb{N}$ be such that $n \geqslant 2$. Use Exercise 20.17 and Exercise 22.12 to determine $F_{N_C,n}$ and $F_{N_C,\infty}$.

Exercise 22.14 Let $A \in \mathcal{B}(\mathcal{H})$ be such that $A^* = -A$ and let $n \in \mathbb{N}$ be such that $n \geqslant 2$. Determine $F_{A,n}$ and $F_{A,\infty}$.

Chapter 23
Resolvents of Monotone Operators

Two quite useful single-valued Lipschitz continuous operators are associated with a monotone operator, namely its resolvent and its Yosida approximation. This chapter is devoted to the investigation of these operators. It exemplifies the tight interplay between firmly nonexpansive mappings and monotone operators. Indeed, firmly nonexpansive operators with full domain can be identified with maximally monotone operators via resolvents and the Minty parameterization. When specialized to subdifferential operators, resolvents become proximity operators. Numerous calculus rules for resolvents are derived. Finally, we address the problem of finding a zero of a maximally monotone operator, via the proximal-point algorithm and via approximating curves.

23.1 Basic Identities

Definition 23.1 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and let $\gamma \in \mathbb{R}_{++}$. The *resolvent* of A is

$$J_A = (\mathrm{Id} + A)^{-1} \tag{23.1}$$

and the *Yosida approximation* of A of index γ is

$${}^{\gamma}\!A = \frac{1}{\gamma}\big(\mathrm{Id} - J_{\gamma A}\big). \tag{23.2}$$

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-48311-5_31

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_23

The following properties follow easily from the above definition and (1.7).

Proposition 23.2 *Let* $A\colon \mathcal{H} \to 2^{\mathcal{H}}$, *let* $\gamma \in \mathbb{R}_{++}$, *let* $x \in \mathcal{H}$, *and let* $p \in \mathcal{H}$. *Then the following hold:*

(i) $\operatorname{dom} J_{\gamma A} = \operatorname{dom} {}^{\gamma}A = \operatorname{ran}(\mathrm{Id} + \gamma A)$ *and* $\operatorname{ran} J_{\gamma A} = \operatorname{dom} A$.
(ii) $p \in J_{\gamma A}x \Leftrightarrow x \in p + \gamma Ap \Leftrightarrow x - p \in \gamma Ap \Leftrightarrow (p, \gamma^{-1}(x-p)) \in \operatorname{gra} A$.
(iii) $p \in {}^{\gamma}Ax \Leftrightarrow p \in A(x - \gamma p) \Leftrightarrow (x - \gamma p, p) \in \operatorname{gra} A$.

Example 23.3 Let $f \in \Gamma_0(\mathcal{H})$ and let $\gamma \in \mathbb{R}_{++}$. Then Proposition 16.44 yields

$$J_{\gamma\partial f} = \operatorname{Prox}_{\gamma f}\,. \tag{23.3}$$

In turn, it follows from Proposition 12.30 that the Yosida approximation of the subdifferential of f is the Fréchet derivative of the Moreau envelope; more precisely,

$${}^{\gamma}(\partial f) = \nabla({}^{\gamma}f\,). \tag{23.4}$$

Example 23.4 Let C be a nonempty closed convex subset of $\mathcal{H}$ and let $\gamma \in \mathbb{R}_{++}$. Setting $f = \iota_C$ in (23.3) and invoking Example 16.13, Example 12.25, (14.6), Example 13.3(i), and (23.2) yields

$$J_{N_C} = (\mathrm{Id} + N_C)^{-1} = \operatorname{Prox}_{\iota_C} = P_C = \mathrm{Id} - \operatorname{Prox}_{\sigma_C} \tag{23.5}$$

and

$${}^{\gamma}N_C = \frac{1}{\gamma}(\mathrm{Id} - P_C) = \frac{1}{\gamma}\operatorname{Prox}_{\sigma_C}\,. \tag{23.6}$$

Example 23.5 Let $\mathcal{K}$ be a real Hilbert space, let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$, let $\gamma \in \mathbb{R}_{++}$, set $\boldsymbol{\mathcal{H}} = \mathcal{H} \oplus \mathcal{K}$, and set $A\colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}\colon (x,v) \mapsto (L^*v, -Lx)$. Then, for every $(x,v) \in \boldsymbol{\mathcal{H}}$,

$$J_{\gamma A}(x,v) = \big((\mathrm{Id} + \gamma^2 L^*L)^{-1}(x - \gamma L^*v)\,,\, (\mathrm{Id} + \gamma^2 LL^*)^{-1}(v + \gamma Lx)\big). \tag{23.7}$$

Proof. It is clear that $A \in \mathcal{B}(\boldsymbol{\mathcal{H}})$ and $A^* = -A$. Therefore, A is maximally monotone by Example 20.35. Now take $(x,v) \in \boldsymbol{\mathcal{H}}$ and set $(p,q) = J_{\gamma A}(x,v)$. Then $(x,v) = (p,q) + \gamma A(p,q)$, and hence $x = p + \gamma L^*q$ and $v = q - \gamma Lp$. In turn, $Lx = Lp + \gamma LL^*q$ and $L^*v = L^*q - \gamma L^*Lp$. Thus, $x = p + \gamma L^*v + \gamma^2 L^*Lp$ and therefore $p = (\mathrm{Id} + \gamma^2 L^*L)^{-1}(x - \gamma L^*v)$. Likewise, $v = q - \gamma Lx + \gamma^2 LL^*q$ and $q = (\mathrm{Id} + \gamma^2 LL^*)^{-1}(v + \gamma Lx)$. ☐

Example 23.6 Let H be a separable real Hilbert space, let $\mathsf{x}_0 \in \mathsf{H}$, suppose that $\mathcal{H} = L^2([0,T];\mathsf{H})$, and let A be the time-derivative operator (see Example 2.10)

$$A\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x \mapsto \begin{cases} \{x'\}, & \text{if } x \in W^{1,2}([0,T];\mathsf{H}) \text{ and } x(0) = \mathsf{x}_0;\\ \varnothing, & \text{otherwise.}\end{cases} \tag{23.8}$$

Then $\operatorname{dom} J_A = \mathcal{H}$ and, for every $x \in \mathcal{H}$,

$$J_A x\colon [0,T] \to \mathsf{H}\colon t \mapsto e^{-t}\mathsf{x}_0 + \int_0^t e^{s-t}x(s)ds. \tag{23.9}$$

Proof. Let $x \in \mathcal{H}$ and set $y\colon t \mapsto e^{-t}\mathsf{x}_0 + \int_0^t e^{s-t}x(s)ds$. As shown in Proposition 21.4, A is maximally monotone; hence $\operatorname{dom} J_A = \mathcal{H}$ by Theorem 21.1, and $y \in W^{1,2}([0,T];\mathsf{H})$, $y(0) = \mathsf{x}_0$, and $x(t) = y(t) + y'(t)$ a.e. on $]0,T[$. Thus, $x = (\mathrm{Id} + A)y$ and we deduce that $y \in J_A x$. Now let $z \in J_A x$, i.e., $x = z + Az$. Then, by monotonicity of A, $0 = \langle y - z \mid x - x\rangle = \|y - z\|^2 + \langle y - z \mid Ay - Az\rangle \geqslant \|y - z\|^2$ and therefore $z = y$. ☐

Proposition 23.7 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$, let $\gamma \in \mathbb{R}_{++}$, and let $\mu \in \mathbb{R}_{++}$. Then the following hold:*

(i) $\operatorname{gra} {}^{\gamma}\!A \subset \operatorname{gra}(A \circ J_{\gamma A})$.
(ii) ${}^{\gamma}\!A = (\gamma\mathrm{Id} + A^{-1})^{-1} = (J_{\gamma^{-1}A^{-1}}) \circ \gamma^{-1}\mathrm{Id}$.
(iii) ${}^{\gamma+\mu}\!A = {}^{\gamma}({}^{\mu}\!A)$.
(iv) $J_{\gamma({}^{\mu}\!A)} = \mathrm{Id} + \gamma/(\gamma + \mu)(J_{(\gamma+\mu)A} - \mathrm{Id})$.

Proof. Let x and u be in $\mathcal{H}$.

(i): We derive from (23.2) and Proposition 23.2(ii) that $(x,u) \in \operatorname{gra} {}^{\gamma}\!A \Rightarrow (\exists\, p \in J_{\gamma A}x)\ u = \gamma^{-1}(x - p) \in Ap \Rightarrow u \in A(J_{\gamma A}x)$.

(ii): $u \in {}^{\gamma}\!Ax \Leftrightarrow \gamma u \in x - J_{\gamma A}x \Leftrightarrow x - \gamma u \in J_{\gamma A}x \Leftrightarrow x \in x - \gamma u + \gamma A(x - \gamma u) \Leftrightarrow u \in A(x - \gamma u) \Leftrightarrow x \in \gamma u + A^{-1}u \Leftrightarrow u \in (\gamma\mathrm{Id} + A^{-1})^{-1}x$. Moreover, $x \in \gamma u + A^{-1}u \Leftrightarrow \gamma^{-1}x \in u + \gamma^{-1}A^{-1}u \Leftrightarrow u \in J_{\gamma^{-1}A^{-1}}(\gamma^{-1}x)$.

(iii): Let $p \in \mathcal{H}$. Proposition 23.2 yields $p \in {}^{\gamma+\mu}\!Ax \Leftrightarrow p \in A(x - (\gamma+\mu)p) = A((x - \gamma p) - \mu p) \Leftrightarrow p \in ({}^{\mu}\!A)(x - \gamma p) \Leftrightarrow p \in {}^{\gamma}({}^{\mu}\!A)x$.

(iv): This follows from (iii), (23.1), and elementary manipulations. ☐

23.2 Monotonicity and Firm Nonexpansiveness

In this section, we focus on the close relationship between firmly nonexpansive mappings and monotone operators.

Proposition 23.8 *Let D be a nonempty subset of $\mathcal{H}$, let $T\colon D \to \mathcal{H}$, and set $A = T^{-1} - \mathrm{Id}$. Then the following hold:*

(i) $T = J_A$.
(ii) *T is firmly nonexpansive if and only if A is monotone.*
(iii) *T is firmly nonexpansive and $D = \mathcal{H}$ if and only if A is maximally monotone.*

Proof. (i): See (23.1).

(ii): Suppose that T is firmly nonexpansive, and take (x,u) and (y,v) in $\operatorname{gra} A$. Then $x + u \in T^{-1}x$, i.e., $x = T(x+u)$. Likewise, $y = T(y+v)$. Hence, Proposition 4.4(v) yields

$$\begin{aligned}&\langle x - y \mid u - v\rangle = \\ &\qquad \langle T(x+u) - T(y+v) \mid (\mathrm{Id} - T)(x+u) - (\mathrm{Id} - T)(y+v)\rangle \geqslant 0, \quad (23.10)\end{aligned}$$

which proves the monotonicity of A. Now assume that A is monotone, and let x and y be in D. Then $x - Tx \in A(Tx)$ and $y - Ty \in A(Ty)$. Hence, by monotonicity, $\langle Tx - Ty \mid (x - Tx) - (y - Ty)\rangle \geqslant 0$. By Proposition 4.4, T is firmly nonexpansive.

(iii): It follows from (ii) and Theorem 21.1 that A is maximally monotone if and only if $\mathcal{H} = \operatorname{ran}(\mathrm{Id} + A) = \operatorname{ran} T^{-1} = \operatorname{dom} T = D$. ☐

Corollary 23.9 *Let $T\colon \mathcal{H} \to \mathcal{H}$. Then T is firmly nonexpansive if and only if it is the resolvent of a maximally monotone operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$.*

Proposition 23.10 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be such that* $\operatorname{dom} A \neq \varnothing$, *set* $D = \operatorname{ran}(\mathrm{Id} + A)$, *and set* $T = J_A|_D$. *Then the following hold:*

(i) $A = T^{-1} - \mathrm{Id}$.
(ii) *A is monotone if and only if T is firmly nonexpansive.*
(iii) *A is maximally monotone if and only if T is firmly nonexpansive and $D = \mathcal{H}$.*

Proof. (i): Clear.

(ii): Assume that A is monotone, and take (x, u) and (y, v) in $\operatorname{gra} J_A$. Then $x - u \in Au$. Likewise, $y - v \in Av$. Hence, by monotonicity, $\langle u - v \mid (x - u) - (y - v)\rangle \geqslant 0$, i.e.,

$$\langle x - y \mid u - v\rangle \geqslant \|u - v\|^2. \tag{23.11}$$

In particular, for $x = y$, we obtain $u = v$. Therefore T is single-valued and we rewrite (23.11) as $\langle x - y \mid Tx - Ty\rangle \geqslant \|Tx - Ty\|^2$. In view of Proposition 4.4(iv), T is firmly nonexpansive. The reverse statement follows from Proposition 23.8(ii).

(iii): Combine (i), (ii), and Proposition 23.8(iii). ☐

Further connections between monotonicity and nonexpansiveness are listed next.

Corollary 23.11 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and let $\gamma \in \mathbb{R}_{++}$. Then the following hold:*

(i) *$J_{\gamma A}\colon \mathcal{H} \to \mathcal{H}$ and $\mathrm{Id} - J_{\gamma A}\colon \mathcal{H} \to \mathcal{H}$ are firmly nonexpansive and maximally monotone.*
(ii) *The* reflected resolvent

$$R_{\gamma A}\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto 2J_{\gamma A}x - x \tag{23.12}$$

is nonexpansive.
(iii) *${}^{\gamma}A\colon \mathcal{H} \to \mathcal{H}$ is γ-cocoercive.*
(iv) *${}^{\gamma}A$ is maximally monotone.*
(v) *${}^{\gamma}A\colon \mathcal{H} \to \mathcal{H}$ is γ^{-1}-Lipschitz continuous.*

Proof. (i): See Corollary 23.9 and Proposition 4.4 for firm nonexpansiveness, and Example 20.30 for maximal monotonicity.

(ii): Combine (i) and Proposition 4.4.

(iii): Combine (i) and (23.2).

(iv): Combine (iii) and Example 20.31.

(v): Combine (iii) and Cauchy–Schwarz. □

Proposition 23.12 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be uniformly monotone with modulus ϕ and set $D = \operatorname{ran}(\mathrm{Id} + A)$. Then*

$$(\forall x \in D)(\forall y \in D) \quad \langle x - y \mid J_A x - J_A y\rangle \geqslant \|J_A x - J_A y\|^2 + \phi\big(\|J_A x - J_A y\|\big). \quad (23.13)$$

Proof. Let $x \in D$, let $y \in D$, let $u \in \operatorname{dom} A$, and let $v \in \operatorname{dom} A$. Proposition 23.2(ii) yields $(u, v) = (J_A x, J_A y) \Leftrightarrow (x - u, y - v) \in Au \times Av \Rightarrow \langle (x - u) - (y - v) \mid u - v\rangle \geqslant \phi(\|u - v\|) \Leftrightarrow \langle x - y \mid u - v\rangle \geqslant \|u - v\|^2 + \phi(\|u - v\|)$. □

Proposition 23.13 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be monotone and let $\beta \in \mathbb{R}_{++}$. Then A is strongly monotone with constant β if and only if J_A is $(\beta + 1)$-cocoercive, in which case J_A is Lipschitz continuous with constant $1/(\beta + 1) \in \left]0, 1\right[$.*

Proof. If A is β-strongly monotone, then applying Proposition 23.12 with $\phi\colon t \mapsto \beta t^2$ shows that J_A is $(\beta + 1)$-cocoercive. Conversely, suppose that J_A is $(\beta + 1)$-cocoercive, and let x, y, u, and v be in $\mathcal{H}$. Then $(u, v) \in Ax \times Ay \Leftrightarrow ((u + x) - x, (v + y) - y) \in Ax \times Ay \Leftrightarrow (x, y) = (J_A(u + x), J_A(v + y)) \Rightarrow \langle x - y \mid (u + x) - (v + y)\rangle \geqslant (\beta + 1)\|x - y\|^2 \Leftrightarrow \langle x - y \mid u - v\rangle \geqslant \beta\|x - y\|^2$. Thus, A is β-strongly monotone. The last assertion follows from Cauchy–Schwarz. □

Proposition 23.14 *Let $\beta \in \mathbb{R}_{++}$ and let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally* β-cohypomonotone *in the sense that $A^{-1} + \beta\,\mathrm{Id}$ is maximally monotone. Let $\gamma \in \left]\beta, +\infty\right[$ and set $\lambda = 1 - \beta/\gamma$. Then $\mathrm{Id} + \lambda(J_{\gamma A} - \mathrm{Id})\colon \mathcal{H} \to \mathcal{H}$ is firmly nonexpansive.*

Proof. Set $B = {}^{\beta}A$. In view of Proposition 23.7(ii) and Proposition 20.22, B is maximally monotone. Hence, by Corollary 23.11(i), $J_{(\gamma-\beta)B}\colon \mathcal{H} \to \mathcal{H}$ is firmly nonexpansive. The result therefore follows from Proposition 23.7(iv), which provides $\mathrm{Id} + \lambda(J_{\gamma A} - \mathrm{Id}) = J_{(\gamma-\beta)B}$. □

The last two results of this section concern the problem of extending a (firmly) nonexpansive operator defined on a subset of $\mathcal{H}$ to a (firmly) nonexpansive operator defined on the whole space $\mathcal{H}$.

Theorem 23.15 *Let D be a nonempty subset of $\mathcal{H}$ and let $T\colon D \to \mathcal{H}$ be firmly nonexpansive. Then there exists a firmly nonexpansive operator $\widetilde{T}\colon \mathcal{H} \to \mathcal{H}$ such that $\widetilde{T}\big|_D = T$ and $\operatorname{ran}\widetilde{T} \subset \overline{\operatorname{conv}}\operatorname{ran} T$.*

Proof. Proposition 23.8 asserts that there exists a monotone operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ such that $\operatorname{ran}(\mathrm{Id} + A) = D$ and $T = J_A$. However, by Theorem 21.9, A admits a maximally monotone extension $\widetilde{A}\colon \mathcal{H} \to 2^{\mathcal{H}}$ such that $\operatorname{dom} \widetilde{A} \subset \overline{\operatorname{conv}}\operatorname{dom} A$. Now set $\widetilde{T} = J_{\widetilde{A}}$. Then it follows from Corollary 23.11(i) that $\widetilde{T}$ is firmly nonexpansive with $\operatorname{dom} \widetilde{T} = \mathcal{H}$. On the other hand, $\operatorname{ran} \widetilde{T} = \operatorname{dom}(\mathrm{Id} + \widetilde{A}) = \operatorname{dom} \widetilde{A} \subset \overline{\operatorname{conv}}\operatorname{dom} A = \overline{\operatorname{conv}}\operatorname{dom}(\mathrm{Id} + A) = \overline{\operatorname{conv}}\operatorname{ran} J_A = \overline{\operatorname{conv}}\operatorname{ran} T$. Finally, let $x \in D$. Then $Tx = J_A x \Rightarrow x - Tx \in A(Tx) \subset \widetilde{A}(Tx) \Rightarrow Tx = J_{\widetilde{A}}x = \widetilde{T}x$. Thus, $\widetilde{T}\big|_D = T$. ☐

Corollary 23.16 (Kirszbraun–Valentine) *Let D be a nonempty subset of $\mathcal{H}$ and let $T\colon D \to \mathcal{H}$ be a nonexpansive operator. Then there exists a nonexpansive operator $\widetilde{T}\colon \mathcal{H} \to \mathcal{H}$ such that $\widetilde{T}\big|_D = T$ and $\operatorname{ran}\widetilde{T} \subset \overline{\operatorname{conv}}\operatorname{ran} T$.*

Proof. Set $R = (\mathrm{Id}+T)/2$. Then $R\colon D \to \mathcal{H}$ is firmly nonexpansive by Proposition 4.4 and, by Theorem 23.15, it admits a firmly nonexpansive extension $\widetilde{R}\colon \mathcal{H} \to \mathcal{H}$. Hence, $\widetilde{T} = P_C \circ (2\widetilde{R} - \mathrm{Id})$, where $C = \overline{\operatorname{conv}}\operatorname{ran} T$, has the required properties. ☐

23.3 Resolvent Calculus

In this section we establish formulas for computing resolvents of transformations of maximally monotone operators.

Proposition 23.17 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $\alpha \in \mathbb{R}_+$, and let $z \in \mathcal{H}$. Then the following hold:*

(i) *Set $B = A + \alpha\mathrm{Id}$. Then $J_B = J_{(1+\alpha)^{-1}A}((1+\alpha)^{-1}\mathrm{Id})$.*
(ii) *Set $B = z + A$. Then $J_B = \tau_z J_A$.*
(iii) *Set $B = \tau_z A$. Then $J_B = z + \tau_z J_A$.*

Proof. Let x and p be in $\mathcal{H}$. Then we derive from Proposition 23.2(iii) the following equivalences:

(i): $p = J_B x \Leftrightarrow x - p \in Ap + \alpha p \Leftrightarrow x - (1+\alpha)p \in Ap \Leftrightarrow (1+\alpha)^{-1}x - p \in (1+\alpha)^{-1}Ap \Leftrightarrow p = J_{(1+\alpha)^{-1}A}((1+\alpha)^{-1}x)$.

(ii): $p = J_B x \Leftrightarrow x - p \in z + Ap \Leftrightarrow (x - z) - p \in Ap \Leftrightarrow p = J_A(x - z)$.

(iii): $p = J_B x \Leftrightarrow x - p \in A(p - z) \Leftrightarrow (x - z) - (p - z) \in A(p - z) \Leftrightarrow p - z = J_A(x - z)$. ☐

Proposition 23.18 *Let $(\mathcal{H}_i)_{i\in I}$ be a finite family of real Hilbert spaces, set $\boldsymbol{\mathcal{H}} = \bigoplus_{i\in I}\mathcal{H}_i$, and, for every $i \in I$, let $A_i\colon \mathcal{H}_i \to 2^{\mathcal{H}_i}$ be maximally monotone. Set $\boldsymbol{A}\colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}\colon (x_i)_{i\in I} \mapsto \bigtimes_{i\in I} A_i x_i$. Then $\boldsymbol{A}$ is maximally monotone and*

$$J_{\boldsymbol{A}}\colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}\colon (x_i)_{i\in I} \mapsto \big(J_{A_i}x_i\big)_{i\in I}. \tag{23.14}$$

Proof. It is clear that $\boldsymbol{A}$ is monotone. On the other hand, it follows from Theorem 21.1 that $(\forall i \in I)$ $\operatorname{ran}(\mathrm{Id} + A_i) = \mathcal{H}_i$. Hence, $\operatorname{ran}(\mathbf{Id} + \boldsymbol{A}) = \bigtimes_{i\in I} \operatorname{ran}(\mathrm{Id} + A_i) = \boldsymbol{\mathcal{H}}$, and $\boldsymbol{A}$ is therefore maximally monotone. Now let $\boldsymbol{x} = (x_i)_{i\in I} \in \boldsymbol{\mathcal{H}}$, set $\boldsymbol{p} = J_{\boldsymbol{A}}\boldsymbol{x}$, and let $(p_i)_{i\in I} \in \boldsymbol{\mathcal{H}}$. We derive from Proposition 23.2(ii) that

$$
\begin{aligned}
(\forall i \in I) \quad p_i = J_{A_i} x_i &\Leftrightarrow (\forall i \in I) \quad x_i - p_i \in A_i p_i \\
&\Leftrightarrow \boldsymbol{x} - (p_i)_{i\in I} \in (A_i p_i)_{i\in I} = \boldsymbol{A}(p_i)_{i\in I} \\
&\Leftrightarrow \boldsymbol{p} = (p_i)_{i\in I},
\end{aligned} \tag{23.15}
$$

which establishes (23.14). ☐

Given a maximally monotone operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $\gamma \in \mathbb{R}_{++}$, no simple formula is known to express $J_{\gamma A}$ in terms of J_A. However, a straightforward formula relates the graphs of these two resolvents (the proof is left as Exercise 23.17).

Proposition 23.19 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $\gamma \in \mathbb{R}_{++}$, and set $L\colon \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}\colon (x,u) \mapsto (\gamma x + (1-\gamma)u, u)$. Then $\operatorname{gra} J_{\gamma A} = L(\operatorname{gra} J_A)$.*

The next result relates the resolvent of an operator to that of its inverse.

Proposition 23.20 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and let $\gamma \in \mathbb{R}_{++}$. Then*

$$\mathrm{Id} = J_{\gamma A} + \gamma J_{\gamma^{-1}A^{-1}} \circ \gamma^{-1}\mathrm{Id}. \tag{23.16}$$

In particular,

$$J_{A^{-1}} = \mathrm{Id} - J_A. \tag{23.17}$$

Proof. Since $J_{\gamma A}$ and $J_{A^{-1}/\gamma}$ are single-valued, (23.16) follows from (23.2) and Proposition 23.7(ii). ☐

We now turn to cocoercive operators.

Proposition 23.21 *Let D be a nonempty subset of $\mathcal{H}$, let $T\colon D \to \mathcal{H}$, and let $\gamma \in \mathbb{R}_{++}$. Then the following hold:*

(i) *T is γ-cocoercive if and only if it is the Yosida approximation of index γ of a monotone operator from $\mathcal{H}$ to $2^{\mathcal{H}}$.*
(ii) *Suppose that $D = \mathcal{H}$. Then T is γ-cocoercive if and only if it is the Yosida approximation of index γ of a maximally monotone operator from $\mathcal{H}$ to $2^{\mathcal{H}}$.*

Proof. (i): T is γ-cocoercive if and only if γT is firmly nonexpansive, i.e., by Proposition 4.4, if and only if $\mathrm{Id} - \gamma T$ is firmly nonexpansive, i.e., by Proposition 23.8(i)&(ii), if and only if $\mathrm{Id} - \gamma T = J_{\gamma A}$, that is $T = {}^{\gamma}A$, for some monotone operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$.

(ii): Proceed as in (i) and invoke Proposition 23.8(iii) instead of Proposition 23.8(ii). ☐

Proposition 23.22 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $\gamma \in \mathbb{R}_{++}$, and let x, y, and v be in $\mathcal{H}$. Then*

$$(y,v) = (J_{\gamma A}x,\ {}^{\gamma}\!Ax) \quad \Leftrightarrow \quad \begin{cases} (y,v) \in \operatorname{gra} A, \\ x = y + \gamma v. \end{cases} \tag{23.18}$$

Proof. Using Proposition 23.2(ii), (23.1), and (23.2), we obtain

$$\begin{cases} y = J_{\gamma A}x, \\ v = {}^{\gamma}\!Ax \end{cases} \quad \Leftrightarrow \quad \begin{cases} (y, x-y) \in \operatorname{gra} \gamma A, \\ v = (x-y)/\gamma \end{cases} \quad \Leftrightarrow \quad \begin{cases} (y,v) \in \operatorname{gra} A, \\ x = y + \gamma v, \end{cases} \tag{23.19}$$

as required. ☐

Remark 23.23 Here are some consequences of Proposition 23.22.

(i) Set $R\colon \mathcal{H} \to \operatorname{gra} A\colon x \mapsto (J_{\gamma A}x,\ {}^{\gamma}\!Ax)$. Then R is a bijection and, more precisely, a Lipschitz homeomorphism from $\mathcal{H}$ to $\operatorname{gra} A$, viewed as a subset of $\mathcal{H} \oplus \mathcal{H}$. Indeed, for every x and y in $\mathcal{H}$, Corollary 23.11(i)&(v) yields $\|Rx - Ry\|^2 = \|J_{\gamma A}x - J_{\gamma A}y\|^2 + \|{}^{\gamma}\!Ax - {}^{\gamma}\!Ay\|^2 \leqslant \|x-y\|^2 + \gamma^{-2}\|x-y\|^2$. Hence, R is $\sqrt{1+1/\gamma^2}$-Lipschitz continuous. Conversely, if (x,u) and (y,v) are in $\operatorname{gra} A$ then, by Cauchy–Schwarz,

$$\begin{aligned} \|R^{-1}(x,u) - R^{-1}(y,v)\|^2 &= \|(x-y) + \gamma(u-v)\|^2 \\ &\leqslant (\|x-y\| + \gamma\|u-v\|)^2 \\ &\leqslant (1+\gamma^2)(\|x-y\|^2 + \|u-v\|^2) \\ &= (1+\gamma^2)\|(x,u) - (y,v)\|^2. \end{aligned} \tag{23.20}$$

Therefore, R^{-1} is $\sqrt{1+\gamma^2}$–Lipschitz continuous.

(ii) Setting $\gamma = 1$ in (i), we obtain via (23.2) and (23.17) the *Minty parameterization*

$$\mathcal{H} \to \operatorname{gra} A\colon x \mapsto (J_A x, J_{A^{-1}}x) = (J_A x, x - J_A x) \tag{23.21}$$

of $\operatorname{gra} A$.

Here is a refinement of (20.16) for points outside the graph of a maximally monotone operator.

Proposition 23.24 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, and let $(x,u) \in \mathcal{H} \times \mathcal{H} \smallsetminus \operatorname{gra} A$. Then there exists a point $(y,v) \in \operatorname{gra} A$ such that*

$$\langle x-y \mid u-v \rangle = -\|x-y\|\,\|u-v\| < 0. \tag{23.22}$$

Proof. Set $z = x+u$, $y = J_A z$, and $v = z-y$. By (23.21), $(y,v) \in \operatorname{gra} A$ and $0 = \|z-(y+v)\|^2 = \|(x-y)+(u-v)\|^2 = \|x-y\|^2 + \|u-v\|^2 + 2\langle x-y \mid u-v\rangle \geqslant$

$\|x-y\|^2+\|u-v\|^2-2\|x-y\|\,\|u-v\| = (\|x-y\|-\|u-v\|)^2 \geqslant 0$. Therefore, $\|x-y\| = \|u-v\|$ and $x-y = v-u$. ☐

We now investigate resolvents of composite operators.

Proposition 23.25 *Let $\mathcal{K}$ be a real Hilbert space, suppose that $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$ is such that LL^* is invertible, let $A\colon \mathcal{K} \to 2^{\mathcal{K}}$ be maximally monotone, and set $B = L^*AL$. Then the following hold:*

(i) $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ *is maximally monotone.*

(ii) $J_B = \mathrm{Id} - L^* \circ (LL^* + A^{-1})^{-1} \circ L$.

(iii) *Suppose that $LL^* = \mu\mathrm{Id}$ for some $\mu \in \mathbb{R}_{++}$. Then $J_B = \mathrm{Id} - L^* \circ {}^{\mu}A \circ L$.*

Proof. (i): Since $\mathcal{K} = L(L^*(\mathcal{K})) \subset L(\mathcal{H}) = \operatorname{ran} L$, we have $\operatorname{ran} L = \mathcal{K}$ and therefore $\operatorname{cone}(\operatorname{ran} L - \operatorname{dom} A) = \overline{\operatorname{span}}\,(\operatorname{ran} L - \operatorname{dom} A)$. Thus, as will be seen in Corollary 25.6, $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ is maximally monotone.

(ii): It follows from (i) and Corollary 23.11(i) that J_B is single-valued with domain $\mathcal{H}$. Now set $T = (LL^* + A^{-1})^{-1}$. Since LL^* is invertible, it is strictly monotone by Fact 2.25(vi), and so is LL^*+A^{-1}. As a result, T is single-valued on $\operatorname{dom} T = \operatorname{ran}(LL^* + A^{-1})$. On the other hand, by Example 25.17, LL^* is 3^* monotone. Hence, it follows from Corollary 25.5(i) and Corollary 25.27(ii) that $\operatorname{dom} T = \mathcal{K}$. Now let $x \in \mathcal{H}$, note that $v = T(Lx) = (LL^* + A^{-1})^{-1}(Lx)$ is well defined, and set $p = x - L^*v$. Then $Lx \in LL^*v + A^{-1}v$ and hence $Lp = L(x - L^*v) \in A^{-1}v$. In turn, $v \in A(Lp)$ and, therefore, $x - p = L^*v \in L^*(A(Lp)) = Bp$. Thus, $p = J_Bx$, as claimed.

(iii): This follows from (ii) and Proposition 23.7(ii). ☐

Corollary 23.26 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $\rho \in \mathbb{R} \smallsetminus \{0\}$, and set $B = \rho A(\rho\,\cdot)$. Then $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ is maximally monotone and*

$$J_B = \rho^{-1} J_{\rho^2 A}(\rho\,\cdot). \tag{23.23}$$

Proof. Apply Proposition 23.25(iii) to $\mathcal{K} = \mathcal{H}$ and $L = \rho\mathrm{Id}$. ☐

Corollary 23.27 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $L \in \mathcal{B}(\mathcal{H})$ be an invertible operator such that $L^{-1} = L^*$, and set $B = L^* \circ A \circ L$. Then B is maximally monotone and $J_B = L^* \circ J_A \circ L$.*

Proof. Apply Proposition 23.25(iii) to $\mathcal{K} = \mathcal{H}$ and $\mu = 1$. ☐

Corollary 23.28 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and set $B = -A^{\vee}$. Then B is maximally monotone and $J_B = -(J_A)^{\vee}$.*

Proof. Apply Corollary 23.27 to $L = -\mathrm{Id}$. ☐

Proposition 23.29 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $\mu \in \mathbb{R}_{++}$, let $\gamma \in \mathbb{R}_{++}$, and set $B = \mu^{-1}\mathrm{Id} - {}^{\mu}A = \mu^{-1}J_{\mu A}$. Then $B\colon \mathcal{H} \to \mathcal{H}$ is maximally monotone and*

$$J_{\gamma B} = \mathrm{Id} - \frac{\gamma}{\mu} J_{\frac{\mu^2}{\mu+\gamma}A} \circ \left(\frac{\mu}{\mu+\gamma}\mathrm{Id}\right). \tag{23.24}$$

Proof. It follows from Corollary 23.11(i) that $B = \mu^{-1}J_{\mu A}$ is maximally monotone and single-valued with domain $\mathcal{H}$. Hence, for every x and p in $\mathcal{H}$, Proposition 23.2(ii) yields

$$\begin{aligned} p = J_{\gamma B}x &\Leftrightarrow \frac{\mu}{\gamma}(x-p) = \mu Bp = J_{\mu A}p \\ &\Leftrightarrow p - \frac{\mu}{\gamma}(x-p) \in \mu A\Big(\frac{\mu}{\gamma}(x-p)\Big) \\ &\Leftrightarrow \frac{\mu x}{\mu+\gamma} - \frac{\mu}{\gamma}(x-p) \in \frac{\mu^2}{\mu+\gamma}A\Big(\frac{\mu}{\gamma}(x-p)\Big) \\ &\Leftrightarrow \frac{\mu}{\gamma}(x-p) = J_{\frac{\mu^2}{\mu+\gamma}A}\left(\frac{\mu x}{\mu+\gamma}\right) \\ &\Leftrightarrow p = x - \frac{\gamma}{\mu}J_{\frac{\mu^2}{\mu+\gamma}A}\left(\frac{\mu x}{\mu+\gamma}\right), \end{aligned} \tag{23.25}$$

which gives (23.24). ☐

The next proposition concerns the resolvent of the partial inverse (see Definition 20.42).

Proposition 23.30 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let V be a closed linear subspace of $\mathcal{H}$, let $x \in \mathcal{H}$, let $p \in \mathcal{H}$, and set $B = A_V$. Then the following hold:*

(i) *Let $\gamma \in \mathbb{R}_{++}$. Then*

$$p = J_{\gamma B}x \quad \Leftrightarrow \quad \frac{1}{\gamma}P_V(x-p) + P_{V^\perp}p \in A\Big(P_Vp + \frac{1}{\gamma}P_{V^\perp}(x-p)\Big). \tag{23.26}$$

(ii) $p = J_Bx \Leftrightarrow P_Vp + P_{V^\perp}(x-p) = J_Ax$.

Proof. (i): Define L as in (20.27) and recall from Proposition 20.44(v) that B is maximally monotone. It follows from Proposition 23.2(ii) and Proposition 20.44(ii) that

$$\begin{aligned} p = J_{\gamma B}x &\Leftrightarrow \big(p, \tfrac{1}{\gamma}(x-p)\big) \in \operatorname{gra} A_V \\ &\Leftrightarrow L\big(p, \tfrac{1}{\gamma}(x-p)\big) \in L(\operatorname{gra} A_V) = L^{-1}(\operatorname{gra} A_V) = \operatorname{gra} A \\ &\Leftrightarrow \big(P_Vp + \tfrac{1}{\gamma}P_{V^\perp}(x-p), \tfrac{1}{\gamma}P_V(x-p) + P_{V^\perp}p\big) \in \operatorname{gra} A. \end{aligned} \tag{23.27}$$

(ii): Set $q = P_Vp + P_{V^\perp}(x-p)$. Then using (i) with $\gamma = 1$ and Corollary 3.24(iii)&(v) yields $p = J_Bx \Leftrightarrow x - q = P_V(x-p) + P_{V^\perp}p \in Aq \Leftrightarrow q = J_Ax$. ☐

Proposition 23.31 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $\gamma \in \mathbb{R}_{++}$, let $\lambda \in \mathbb{R}_{++}$, and let $x \in \mathcal{H}$. Then the following hold:*

(i) $J_{\gamma A}x = J_{\lambda\gamma A}\big(\lambda x + (1-\lambda)J_{\gamma A}x\big)$.
(ii) *Suppose that* $\lambda \leqslant 1$. *Then* $\|J_{\lambda\gamma A}x - x\| \leqslant (2-\lambda)\|J_{\gamma A}x - x\|$.
(iii) $\|J_{\gamma A}x - J_{\lambda\gamma A}x\| \leqslant |1-\lambda|\,\|J_{\gamma A}x - x\|$.

Proof. Set $\mu = \lambda\gamma$.

(i): We have $x \in (\mathrm{Id} + \gamma A)(\mathrm{Id} + \gamma A)^{-1}x = (\mathrm{Id} + \gamma A)J_{\gamma A}x$. Hence $x - J_{\gamma A}x \in \gamma A(J_{\gamma A}x)$ and therefore $\lambda(x - J_{\gamma A}x) \in \mu A(J_{\gamma A}x)$. In turn, $\lambda x + (1-\lambda)J_{\gamma A}x \in J_{\gamma A}x + \mu A(J_{\gamma A}x) = (\mathrm{Id} + \mu A)(J_{\gamma A}x)$ and therefore $J_{\gamma A}x = J_{\mu A}\left(\lambda x + (1-\lambda)\, J_{\gamma A}x\right)$.

(ii): By (i) and Corollary 23.11(i),

$$\begin{aligned}
\|J_{\mu A}x - x\| &\leqslant \|J_{\mu A}x - J_{\gamma A}x\| + \|J_{\gamma A}x - x\| \\
&= \|J_{\mu A}x - J_{\mu A}\left(\lambda x + (1-\lambda)J_{\gamma A}x\right)\| + \|J_{\gamma A}x - x\| \\
&\leqslant \|x - \lambda x - (1-\lambda)J_{\gamma A}x\| + \|J_{\gamma A}x - x\| \\
&\leqslant (2-\lambda)\|J_{\gamma A}x - x\|. \qquad (23.28)
\end{aligned}$$

(iii): By (i) and Corollary 23.11(i), we have

$$\begin{aligned}
\|J_{\gamma A}x - J_{\mu A}x\| &= \|J_{\mu A}\big(\lambda x + (1-\lambda)J_{\gamma A}x\big) - J_{\mu A}x\| \\
&\leqslant \|\lambda x + (1-\lambda)J_{\gamma A}x - x\| \\
&= |1-\lambda|\,\|J_{\gamma A}x - x\|, \qquad (23.29)
\end{aligned}$$

which concludes the proof. ☐

There is in general no simple expression for the resolvent of the sum of two monotone operators. The following result provides a special case in which such an expression exists.

Proposition 23.32 *Let* $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ *and* $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ *be maximally monotone operators. Suppose that one of the following holds:*

(i) $(\forall y \in \operatorname{dom} B)\ By \subset B(J_A y)$.
(ii) $(\forall (x,u) \in \operatorname{gra} A)\ B(x+u) \subset Bx$.
(iii) $A + B$ *is maximally monotone and* $(\forall (x,u) \in \operatorname{gra} A)\ Bx \subset B(x+u)$.

Then $(\forall x \in \mathcal{H})\ J_{A+B}x = J_A J_B x$.

Proof. (i): Let $x \in \mathcal{H}$, and set $y = J_B x$ and $p = J_A y$. Then Proposition 23.2(ii) implies that $x - y \in By \subset Bp$ and that $y - p \in Ap$. Hence $x - p = (y-p) + (x-y) \in Ap + Bp = (A+B)p$ and therefore $p = J_{A+B}x$.

(ii): Let $y \in \mathcal{H}$. Set $x = J_A y$ and $u = y - J_A y$. By Proposition 23.2(ii), $(x,u) \in \operatorname{gra} A$. Hence, $By = B(x+u) \subset Bx = B(J_A y)$ and (i) yields the result.

(iii): Let $x \in \mathcal{H}$. Since $A+B$ is maximally monotone, $p = J_{A+B}x$ is well defined and we have $x \in p + Ap + Bp$. Hence, there exists $u \in Ap$ such that $x - (p+u) \in Bp \subset B(p+u)$. Thus $p + u = J_B x$. It follows that $J_B x - p = u \in Ap$ and therefore that $p = J_A J_B x$. ☐

Corollary 23.33 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and such that $(\forall x \in \mathcal{H})$ $x + Ax \subset \mathbb{R}_+ x$, let C be a nonempty closed convex subset of $\mathcal{H}$ such that* $\operatorname{cone}(\operatorname{dom} A - \operatorname{ran} N_C) = \overline{\operatorname{span}}\,(\operatorname{dom} A - \operatorname{ran} N_C)$*, and set $B = N_C^{-1}$. Then $J_{A+B} = J_A \circ (\mathrm{Id} - P_C) = J_A \circ \operatorname{Prox}_{\sigma_C}$.*

Proof. Since $\operatorname{dom} B = \operatorname{ran} N_C$, it follows from Corollary 25.4 below that $A + B$ is maximally monotone. Now let $(x, u) \in \operatorname{gra} A$ and let $y \in Bx$. Then $x + u = \lambda x$ for some $\lambda \in \mathbb{R}_+$, $x \in N_C y$, and therefore $\lambda x \in N_C y$. Thus $x + u \in N_C y$, i.e., $y \in B(x + u)$. Altogether, the result follows from Proposition 23.32(iii), (23.17), and (23.5). ☐

We conclude this section with a result which provides in particular an extension of the decomposition principle (23.16) involving a change of metric.

Proposition 23.34 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $U \in \mathcal{B}(\mathcal{H})$ be a self-adjoint strongly monotone operator, and let $\mathcal{K}$ be the real Hilbert space obtained by endowing $\mathcal{H}$ with the scalar product $(x, y) \mapsto \langle x \mid y\rangle_{\mathcal{K}} = \langle U^{-1}x \mid y\rangle$. Then the following hold:*

(i) *$J_{UA}\colon \mathcal{K} \to \mathcal{K}$ is firmly nonexpansive.*
(ii) *$J_{UA} = (U^{-1} + A)^{-1} \circ U^{-1}$.*
(iii) *$J_{UA} = U^{1/2} J_{U^{1/2}AU^{1/2}} U^{-1/2} = \mathrm{Id} - U J_{U^{-1}A^{-1}} U^{-1}$.*

Proof. (i): This follows from Proposition 20.24 and Corollary 23.9.

(ii): Let x and p be in $\mathcal{H}$. Then $p = J_{UA}x \Leftrightarrow x \in p + UAp \Leftrightarrow U^{-1}x \in (U^{-1} + A)p \Leftrightarrow p = (U^{-1} + A)^{-1}(U^{-1}x)$.

(iii): Let x and p be in $\mathcal{H}$. Then

$$\begin{aligned} p = J_{UA}x &\Leftrightarrow x - p \in UAp \\ &\Leftrightarrow U^{-1/2}x - U^{-1/2}p \in U^{1/2}AU^{1/2}U^{-1/2}p \\ &\Leftrightarrow U^{-1/2}p = J_{U^{1/2}AU^{1/2}}\big(U^{-1/2}x\big) \\ &\Leftrightarrow p = U^{1/2}J_{U^{1/2}AU^{1/2}}\big(U^{-1/2}x\big). \end{aligned} \tag{23.30}$$

Furthermore, by Proposition 23.25(ii),

$$J_{U^{1/2}AU^{1/2}} = \mathrm{Id} - U^{1/2}\big(U + A^{-1}\big)^{-1}U^{1/2}. \tag{23.31}$$

Hence, (23.30) yields

$$J_{UA} = \mathrm{Id} - U\big(U + A^{-1}\big)^{-1}. \tag{23.32}$$

However,

$$\begin{aligned} p = \big(U + A^{-1}\big)^{-1}x &\Leftrightarrow x \in Up + A^{-1}p \\ &\Leftrightarrow x - Up \in A^{-1}p \end{aligned}$$

$$\Leftrightarrow U^{-1}x \in \big(\mathrm{Id} + U^{-1}A^{-1}\big)p$$
$$\Leftrightarrow p = J_{U^{-1}A^{-1}}(U^{-1}x). \tag{23.33}$$

Hence, $(U + A^{-1})^{-1} = J_{U^{-1}A^{-1}}U^{-1}$ and, using (23.32), we obtain the right-hand side. ☐

23.4 Zeros of Monotone Operators

Recall that the set of zeros of a monotone operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ is denoted by

$$\operatorname{zer} A = A^{-1}0. \tag{23.34}$$

Proposition 23.35 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be strictly monotone. Then $\operatorname{zer} A$ is at most a singleton.*

Proof. Suppose that x and y are distinct points in $\operatorname{zer} A$. Then $0 \in Ax$, $0 \in Ay$, and (22.2) yields $0 = \langle x - y \mid 0 - 0\rangle > 0$, which is impossible. ☐

Proposition 23.36 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and suppose that one of the following holds:*

(i) *A^{-1} is locally bounded everywhere.*
(ii) $\lim_{\|x\|\to+\infty} \inf \|Ax\| = +\infty$.
(iii) *$\operatorname{dom} A$ is bounded.*

Then $\operatorname{zer} A \neq \varnothing$.

Proof. The conclusion follows from (i), (ii), and (iii) via Corollary 21.23, Corollary 21.24, and Corollary 21.25, respectively. ☐

Corollary 23.37 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and suppose that one of the following holds:*

(i) *A is uniformly monotone with a supercoercive modulus.*
(ii) *A is strongly monotone.*

Then $\operatorname{zer} A$ is a singleton.

Proof. (i): In view of Proposition 22.11, $\operatorname{zer} A \neq \varnothing$. Furthermore, since A is strictly monotone, we derive from Proposition 23.35 that $\operatorname{zer} A$ is a singleton.
(ii): This follows from (i) and Definition 22.1. ☐

Proposition 23.38 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be monotone and let $\gamma \in \mathbb{R}_{++}$. Then $\operatorname{Fix} J_{\gamma A} = \operatorname{zer} A = \operatorname{zer} {}^{\gamma}A$.*

Proof. Let $x \in \mathcal{H}$. Then Proposition 23.2(ii) yields $0 \in Ax \Leftrightarrow x - x \in \gamma Ax \Leftrightarrow x = J_{\gamma A}\, x \Leftrightarrow {}^{\gamma}\!Ax = 0$. ☐

Proposition 23.39 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone. Then* $\operatorname{zer} A$ *is closed and convex.*

Proof. This is a consequence of Proposition 20.36 since $\operatorname{zer} A = A^{-1}0$ and A^{-1} is maximally monotone. ☐

Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and let $\gamma \in \mathbb{R}_{++}$. Then $\operatorname{zer} A = \operatorname{zer}(\gamma A)$ and, in view of Proposition 23.38, a zero of A can be approximated iteratively by suitable resolvent iterations. Such algorithms are known as *proximal-point algorithms*.

Example 23.40 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $\gamma \in \mathbb{R}_{++}$, and let $x_0 \in \mathcal{H}$. Set

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = J_{\gamma A} x_n. \tag{23.35}$$

Then the following hold:

(i) Suppose that $\operatorname{zer} A \neq \varnothing$. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{zer} A$.

(ii) Suppose that A is strongly monotone with constant $\beta \in \mathbb{R}_{++}$. Then $(x_n)_{n\in\mathbb{N}}$ converges strongly (actually linearly) to the unique point in $\operatorname{zer} A$.

Proof. Proposition 23.38 yields $\operatorname{Fix} J_{\gamma A} = \operatorname{zer}(\gamma A) = \operatorname{zer} A$.

(i): In view of Corollary 23.11(i), the result is an application of Example 5.18 with $T = J_{\gamma A}$.

(ii): Proposition 23.13 asserts that $J_{\gamma A}$ is Lipschitz continuous with constant $1/(\beta\gamma + 1) \in \left]0,1\right[$. Therefore, the result follows by setting $T = J_{\gamma A}$ in Theorem 1.50(i)&(iii). ☐

A finer proximal-point algorithm is described in the following theorem.

Theorem 23.41 (Proximal-point algorithm) *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator such that* $\operatorname{zer} A \neq \varnothing$, *let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}_{++}$ such that $\sum_{n\in\mathbb{N}} \gamma_n^2 = +\infty$, and let $x_0 \in \mathcal{H}$. Set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = J_{\gamma_n A} x_n. \tag{23.36}$$

Then the following hold:

(i) $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* $\operatorname{zer} A$.

(ii) *Suppose that A is uniformly monotone on every bounded subset of $\mathcal{H}$ (see Remark 22.3). Then $(x_n)_{n\in\mathbb{N}}$ converges strongly to the unique point in* $\operatorname{zer} A$.

Proof. Set $(\forall n \in \mathbb{N})$ $u_n = (x_n - x_{n+1})/\gamma_n$. Then $(\forall n \in \mathbb{N})$ $u_n \in Ax_{n+1}$ and $x_{n+1} - x_{n+2} = \gamma_{n+1}u_{n+1}$. Hence, by monotonicity and Cauchy–Schwarz,

$$
\begin{aligned}
(\forall n \in \mathbb{N}) \quad 0 &\leqslant \langle x_{n+1} - x_{n+2} \mid u_n - u_{n+1} \rangle / \gamma_{n+1} \\
&= \langle u_{n+1} \mid u_n - u_{n+1} \rangle \\
&= \langle u_{n+1} \mid u_n \rangle - \|u_{n+1}\|^2 \\
&\leqslant \|u_{n+1}\|(\|u_n\| - \|u_{n+1}\|),
\end{aligned} \tag{23.37}
$$

which implies that $(\|u_n\|)_{n\in\mathbb{N}}$ converges. Now let $z \in \operatorname{zer} A$. Since Proposition 23.38 asserts that $z \in \bigcap_{n\in\mathbb{N}} \operatorname{Fix} J_{\gamma_n A}$, we deduce from (23.36), Corollary 23.11(i), and (4.1) that

$$
\begin{aligned}
(\forall n \in \mathbb{N}) \quad \|x_{n+1} - z\|^2 &= \|J_{\gamma_n A} x_n - J_{\gamma_n A} z\|^2 \\
&\leqslant \|x_n - z\|^2 - \|x_n - J_{\gamma_n A} x_n\|^2 \\
&= \|x_n - z\|^2 - \gamma_n^2 \|u_n\|^2.
\end{aligned} \tag{23.38}
$$

Thus, $(x_n)_{n\in\mathbb{N}}$ is bounded and Fejér monotone with respect to $\operatorname{zer} A$, and $\sum_{n\in\mathbb{N}} \gamma_n^2 \|u_n\|^2 < +\infty$. In turn, since $\sum_{n\in\mathbb{N}} \gamma_n^2 = +\infty$, we have $\underline{\lim}\, \|u_n\| = 0$ and therefore $u_n \to 0$ since $(\|u_n\|)_{n\in\mathbb{N}}$ converges.

(i): Let x be a weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$, say $x_{k_n+1} \rightharpoonup x$. Since $0 \leftarrow u_{k_n} \in A x_{k_n+1}$, Proposition 20.38(ii) yields $0 \in Ax$ and we derive the result from Theorem 5.5.

(ii): The assumptions imply that A is strictly monotone. Hence, by Proposition 23.35, $\operatorname{zer} A$ reduces to a singleton. Now let x be the weak limit in (i). Then $0 \in Ax$ and $(\forall n \in \mathbb{N})$ $u_n \in A x_{n+1}$. Hence, since $(x_n)_{n\in\mathbb{N}}$ lies in a bounded set, there exists an increasing function $\phi\colon \mathbb{R}_+ \to [0, +\infty]$ that vanishes only at 0 such that

$$
(\forall n \in \mathbb{N}) \quad \langle x_{n+1} - x \mid u_n \rangle \geqslant \phi\big(\|x_{n+1} - x\|\big). \tag{23.39}
$$

Since $u_n \to 0$ and $x_{n+1} \rightharpoonup x$, Lemma 2.51(iii) yields $\|x_{n+1} - x\| \to 0$. ☐

23.5 Asymptotic Behavior

In view of Proposition 20.36 and Theorem 3.16, the following notion is well defined.

Definition 23.42 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and let $x \in \operatorname{dom} A$. Then ${}^0\!Ax$ denotes the element in Ax of minimal norm.

Proposition 23.43 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $\gamma \in \mathbb{R}_{++}$, let $\mu \in \mathbb{R}_{++}$, and let $x \in \operatorname{dom} A$. Then the following hold:*

(i) $\|{}^{\gamma}\!Ax\| \leqslant \inf \|Ax\|$.
(ii) $\|{}^{\gamma+\mu}\!Ax\| \leqslant \|{}^{\gamma}\!Ax\| \leqslant \|{}^{0}\!Ax\|$.

Proof. (i): Set $p = {}^{\gamma}\!Ax$ and let $u \in Ax$. Proposition 23.2(iii) and the monotonicity of A yield $\langle u - p \mid p\rangle = \langle u - p \mid x - (x - \gamma p)\rangle / \gamma \geqslant 0$. Hence, by Cauchy–Schwarz, $\|p\| \leqslant \|u\|$.

(ii): By (i), $\|{}^{\mu}\!Ax\| \leqslant \|{}^{0}\!Ax\|$. Applying this inequality to ${}^{\gamma}\!A$, which is maximally monotone by Corollary 23.11(iv), and, using Proposition 23.7(iii), we get $\|{}^{\gamma+\mu}\!Ax\| = \|{}^{\mu}({}^{\gamma}\!A)x\| \leqslant \|{}^{0}({}^{\gamma}\!A)x\| = \|{}^{\gamma}\!Ax\|$. ☐

We now present a central perturbation result.

Theorem 23.44 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and let $x \in \mathcal{H}$. Then the inclusions*

$$\big(\forall \gamma \in \left]0,1\right[\big) \quad 0 \in Ax_\gamma + \gamma(x_\gamma - x) \tag{23.40}$$

define a unique curve $(x_\gamma)_{\gamma\in\left]0,1\right[}$. Moreover, exactly one of the following holds:

(i) $\operatorname{zer} A \neq \varnothing$ *and* $x_\gamma \to P_{\operatorname{zer} A}x$ *as* $\gamma \downarrow 0$.
(ii) $\operatorname{zer} A = \varnothing$ *and* $\|x_\gamma\| \to +\infty$ *as* $\gamma \downarrow 0$.

Proof. It follows from (23.40), Proposition 23.2(ii), and Corollary 23.9 that, for every $\gamma \in \left]0,1\right[$,

$$x_\gamma = J_{A/\gamma}\, x \tag{23.41}$$

is well defined. We now proceed in two steps.

(a) $\operatorname{zer} A \neq \varnothing \Rightarrow x_\gamma \to P_{\operatorname{zer} A}x$ as $\gamma \downarrow 0$: Set $x_0 = P_{\operatorname{zer} A}x$, which is well defined and characterized by

$$x_0 \in \operatorname{zer} A \quad \text{and} \quad (\forall z \in \operatorname{zer} A) \quad \langle z - x_0 \mid x - x_0\rangle \leqslant 0 \tag{23.42}$$

by virtue of Proposition 23.39 and Theorem 3.16. Now let $z \in \operatorname{zer} A$. Then Proposition 23.38 yields $(\forall \gamma \in \left]0,1\right[)\ z = J_{A/\gamma}z$. Therefore, it follows from (23.41) and Corollary 23.11(i) that

$$\big(\forall \gamma \in \left]0,1\right[\big) \quad \langle z - x_\gamma \mid z - x\rangle \geqslant \|z - x_\gamma\|^2. \tag{23.43}$$

Thus, by Cauchy–Schwarz, $(x_\gamma)_{\gamma\in\left]0,1\right[}$ is bounded. Now let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0,1\right[$ such that $\gamma_n \downarrow 0$ as $n \to +\infty$. Then it is enough to show that $x_{\gamma_n} \to x_0$. To this end, let y be a weak sequential cluster point of $(x_{\gamma_n})_{n\in\mathbb{N}}$, say $x_{\gamma_{k_n}} \rightharpoonup y$. Since $(x_{\gamma_{k_n}})_{n\in\mathbb{N}}$ is bounded, the sequence $(x_{\gamma_{k_n}}, \gamma_{k_n}(x - x_{\gamma_{k_n}}))_{n\in\mathbb{N}}$, which lies in $\operatorname{gra} A$ by virtue of (23.40), converges to $(y,0)$ in $\mathcal{H}^{\mathrm{weak}} \times \mathcal{H}^{\mathrm{strong}}$. Hence, it follows from Proposition 20.38(ii) that $y \in \operatorname{zer} A$. In turn, we derive from (23.43) that $0 \leftarrow \big\langle y - x \mid y - x_{\gamma_{k_n}}\big\rangle \geqslant \|y - x_{\gamma_{k_n}}\|^2$ and therefore that $x_{\gamma_{k_n}} \to y$. However, (23.43) yields $0 \geqslant \|z - x_{\gamma_{k_n}}\|^2 - \big\langle z - x_{\gamma_{k_n}} \mid z - x\big\rangle = \big\langle z - x_{\gamma_{k_n}} \mid x - x_{\gamma_{k_n}}\big\rangle \to \langle z - y \mid x - y\rangle$ and, in view of the characterization (23.42), we obtain $y = x_0$. Altogether, x_0 is the only weak sequential cluster point of the bounded sequence $(x_{\gamma_n})_{n\in\mathbb{N}}$, and it follows from Lemma 2.46 that $x_{\gamma_n} \rightharpoonup x_0$. Invoking (23.43), we obtain $\|x_{\gamma_n} - x_0\|^2 \leqslant \langle x - x_0 \mid x_{\gamma_n} - x_0\rangle \to 0$ and therefore $x_{\gamma_n} \to x_0$.

(b) $\|x_\gamma\| \not\to +\infty$ as $\gamma \downarrow 0 \Rightarrow \operatorname{zer} A \neq \varnothing$: Take a sequence $(\gamma_n)_{n\in\mathbb{N}}$ in $]0,1[$ such that $\gamma_n \downarrow 0$ and $(x_{\gamma_n})_{n\in\mathbb{N}}$ is bounded. Then (23.40) yields $0 \leftarrow \gamma_n(x - x_{\gamma_n}) \in Ax_{\gamma_n}$. Furthermore, $(x_{\gamma_n})_{n\in\mathbb{N}}$ possesses a weak sequential cluster point y, and Proposition 20.38(ii) forces $y \in \operatorname{zer} A$. ☐

We now revisit the approximating curve investigated in Proposition 4.30 and analyze it with different tools.

Corollary 23.45 *Let $T\colon \mathcal{H} \to \mathcal{H}$ be nonexpansive and let $x \in \mathcal{H}$. Then the equations*

$$\big(\forall \gamma \in \,]0,1[\,\big) \quad x_\gamma = \gamma x + (1-\gamma)Tx_\gamma \tag{23.44}$$

define a unique curve $(x_\gamma)_{\gamma\in]0,1[}$. Moreover, exactly one of the following holds:

(i) $\operatorname{Fix} T \neq \varnothing$ *and* $x_\gamma \to P_{\operatorname{Fix} T}\, x$ *as* $\gamma \downarrow 0$.
(ii) $\operatorname{Fix} T = \varnothing$ *and* $\|x_\gamma\| \to +\infty$ *as* $\gamma \downarrow 0$.

Proof. Set $A = \mathrm{Id} - T$. Then $\operatorname{zer} A = \operatorname{Fix} T$, A is maximally monotone by Example 20.29, and (23.44) becomes

$$\big(\forall \gamma \in \,]0,1[\,\big) \quad 0 = Ax_\gamma + \frac{\gamma}{1-\gamma}(x_\gamma - x). \tag{23.45}$$

Since $\lim_{\gamma\downarrow 0} \gamma/(1-\gamma) = 0$, the conclusion follows from Theorem 23.44. ☐

Corollary 23.46 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and let $x \in \mathcal{H}$. Then exactly one of the following holds:*

(i) $x \in \operatorname{dom} A$, *and* ${}^{\gamma}\!Ax \to {}^{0}\!Ax$ *and* $\|{}^{\gamma}\!Ax\| \uparrow \|{}^{0}\!Ax\|$ *as* $\gamma \downarrow 0$.
(ii) $x \notin \operatorname{dom} A$ *and* $\|{}^{\gamma}\!Ax\| \uparrow +\infty$ *as* $\gamma \downarrow 0$.

Proof. Set $B = A^{-1} - x$. Then $\operatorname{zer} B = Ax$. Moreover, B is maximally monotone and, by Theorem 23.44, the inclusions

$$\big(\forall \gamma \in \,]0,1[\,\big) \quad 0 \in Bx_\gamma + \gamma(x_\gamma - 0) \tag{23.46}$$

define a unique curve $(x_\gamma)_{\gamma\in]0,1[}$. However, it follows from Proposition 23.7(ii) that $(\forall \gamma \in \,]0,1[)\ 0 \in Bx_\gamma + \gamma x_\gamma \Leftrightarrow 0 \in A^{-1}x_\gamma - x + \gamma x_\gamma \Leftrightarrow x \in (\gamma \mathrm{Id} + A^{-1})x_\gamma \Leftrightarrow x_\gamma = {}^{\gamma}\!Ax$.

(i): Suppose that $x \in \operatorname{dom} A$. Then $\operatorname{zer} B \neq \varnothing$, and it follows from Theorem 23.44(i) that ${}^{\gamma}\!Ax = x_\gamma \to P_{\operatorname{zer} B} 0 = {}^{0}\!Ax$ as $\gamma \downarrow 0$. In turn, we derive from Proposition 23.43(ii) that $\|{}^{\gamma}\!Ax\| \uparrow \|{}^{0}\!Ax\|$ as $\gamma \downarrow 0$.

(ii): Suppose that $x \notin \operatorname{dom} A$. Then $\operatorname{zer} B = \varnothing$, and the assertion therefore follows from Theorem 23.44(ii). ☐

Remark 23.47 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and let $x \in \operatorname{dom} A$. Then (23.2) and Corollary 23.46(i) yield

$$\frac{J_{\gamma A}x - (x - \gamma\, {}^{0}\!Ax)}{\gamma} \to 0 \ \text{ as } \ \gamma \downarrow 0. \tag{23.47}$$

If A is at most single-valued, then $A|_{\operatorname{dom} A} = {}^0A|_{\operatorname{dom} A}$ and (23.47) can be expressed as

$$J_{\gamma A}x = x - \gamma Ax + o(\gamma) \ \text{ as } \ \gamma \downarrow 0. \tag{23.48}$$

Theorem 23.48 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and let $x \in \mathcal{H}$. Set*

$$(\forall \gamma \in \mathbb{R}_{++}) \quad x_\gamma = J_{\gamma A}x. \tag{23.49}$$

Then $x_\gamma \to P_{\overline{\operatorname{dom}} A}x$ as $\gamma \downarrow 0$, and exactly one of the following holds:

(i) $\operatorname{zer} A \neq \varnothing$ *and* $x_\gamma \to P_{\operatorname{zer} A}x$ *as* $\gamma \uparrow +\infty$.
(ii) $\operatorname{zer} A = \varnothing$ *and* $\|x_\gamma\| \to +\infty$ *as* $\gamma \uparrow +\infty$.

Proof. Set $D = \overline{\operatorname{dom}} A$ and $Z = \operatorname{zer} A$. It follows from Corollary 21.14 that D is nonempty, closed, and convex. Hence, by Theorem 3.16, P_D is well defined. Now let $\gamma \in \left]0,1\right[$ and let $(y,v) \in \operatorname{gra} A$. It follows from Proposition 23.2(ii) and the monotonicity of A that $\langle x_\gamma - y \mid x - x_\gamma - \gamma v\rangle \geqslant 0$, which yields

$$\|x_\gamma - y\|^2 \leqslant \langle x_\gamma - y \mid x - y\rangle + \gamma\|x_\gamma - y\|\,\|v\|. \tag{23.50}$$

As a result, $\|x_\gamma - y\| \leqslant \|x - y\| + \gamma\|v\|$, and $(x_\gamma)_{\gamma\in]0,1[}$ is therefore bounded. Now let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0,1\right[$ such that $\gamma_n \downarrow 0$ as $n \to +\infty$. Let z be a weak sequential cluster point of $(x_{\gamma_n})_{n\in\mathbb{N}}$, say $x_{\gamma_{k_n}} \rightharpoonup z$. Note that Corollary 3.35 implies that $z \in D$ since $(x_{\gamma_{k_n}})_{n\in\mathbb{N}}$ lies in $\operatorname{dom} A \subset D$. On the other hand, since $(x_{\gamma_{k_n}})_{n\in\mathbb{N}}$ is bounded, Lemma 2.42 and (23.50) yield $\|z - y\|^2 \leqslant \underline{\lim}\,\|x_{\gamma_{k_n}} - y\|^2 \leqslant \langle z - y \mid x - y\rangle$ and therefore $\langle x - z \mid y - z\rangle \leqslant 0$. Since y is an arbitrary point in $\operatorname{dom} A$, this inequality holds for every $y \in D$ and, in view of Theorem 3.16, it follows that $z = P_D x$ is the only weak sequential cluster point of the bounded sequence $(x_{\gamma_n})_{n\in\mathbb{N}}$. Hence, by Lemma 2.46, $x_{\gamma_n} \rightharpoonup P_D x$. It follows from (23.50) that

$$\overline{\lim}\,\|x_{\gamma_n} - y\|^2 \leqslant \langle P_D x - y \mid x - y\rangle\,. \tag{23.51}$$

Now set $f\colon \mathcal{H} \to \mathbb{R}\colon z \mapsto \overline{\lim}\,\|x_{\gamma_n} - z\|^2$. Then $f \geqslant 0$ and it follows from Example 8.19 that f is continuous. Now let $(y_n)_{n\in\mathbb{N}}$ be a sequence in $\operatorname{dom} A$ that converges strongly to $P_D x$. Then, using (23.51), we obtain $0 \leqslant f(P_D x) = \lim f(y_n) \leqslant \lim \langle P_D x - y_n \mid x - y_n\rangle = 0$. Hence $0 = f(P_D x) = \overline{\lim}\,\|x_{\gamma_n} - P_D x\|^2$ and therefore $x_{\gamma_n} \to P_D x$. Thus, $x_\gamma \to P_D x$ as $\gamma \downarrow 0$.

(i)&(ii): Apply Theorem 23.44 since (23.49) is equivalent to $(\forall \gamma \in \mathbb{R}_{++})$ $0 \in Ax_\gamma + \gamma^{-1}(x_\gamma - x)$. ☐

Exercises

Exercise 23.1 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and at most single-valued. Define the *Cayley transform* $C_A\colon \mathcal{H} \to \mathcal{H}$ of A by $C_A = (\mathrm{Id} - A)(\mathrm{Id} + A)^{-1}$. Show that $C_A = 2J_A - \mathrm{Id}$ and determine C_{C_A}. What happens when A is linear with $\operatorname{dom} A = \mathcal{H}$?

Exercise 23.2 Suppose that $A \in \mathcal{B}(\mathcal{H})$ satisfies $A^* = -A = A^{-1}$. Show that $J_A = \frac{1}{2}(\mathrm{Id} - A)$ and that $R_A = -A$.

Exercise 23.3 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $\gamma \in \mathbb{R}_{++}$, and let $\mu \in \mathbb{R}_{++}$. Show that

$$(\forall x \in \mathcal{H}) \quad \|J_{\gamma A}x - J_{\mu A}x\| \leqslant |\gamma - \mu| \, \| {}^{\max\{\gamma,\mu\}}Ax\|. \tag{23.52}$$

Exercise 23.4 Let $A\colon \mathcal{H} \to \mathcal{H}$ and $B\colon \mathcal{H} \to \mathcal{H}$ be maximally monotone operators such that $J_{A+B} = J_A \circ J_B$. Show that $(\forall x \in \mathcal{H})\ Bx = B(x + Ax)$, and compare with Proposition 23.32(iii).

Exercise 23.5 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and let $\gamma \in \mathbb{R}_{++}$. Show that

$$\mathrm{Id} = \gamma \mathrm{Id} \circ {}^{\gamma}A + {}^{\gamma^{-1}}(A^{-1}) \circ \gamma^{-1}\mathrm{Id} \tag{23.53}$$

and deduce that $\mathrm{Id} = {}^{1}A + {}^{1}(A^{-1})$.

Exercise 23.6 Let D be a nonempty subset of $\mathcal{H}$ and set $T = \mathrm{Id}|_D$. Determine $\widetilde{T}$, where $\widetilde{T}$ is as in the conclusion of Theorem 23.15.

Exercise 23.7 Let D be a nonempty subset of $\mathcal{H}$ and set $T = \mathrm{Id}|_D$. Determine $\operatorname{Fix}\widetilde{T}$ and $\operatorname{ran}\widetilde{T}$, where $\widetilde{T}$ is as in the conclusion of Corollary 23.16. Give examples in which (i) $\widetilde{T}$ is unique, and (ii) $\widetilde{T}$ is not unique.

Exercise 23.8 Let $S\colon \mathcal{H} \to \mathcal{H}$ be nonexpansive. Show that the set

$$\big\{(1/2)(x + Sx, x - Sx) \;\big|\; x \in \mathcal{H}\big\} \tag{23.54}$$

is the graph of a maximally monotone operator on $\mathcal{H}$.

Exercise 23.9 Show that conditions (i) and (ii) in Proposition 23.32 are equivalent.

Exercise 23.10 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone. Show that $\operatorname{dom}(A) + \operatorname{ran}(A) = \mathcal{H}$.

Exercise 23.11 Let $f \in \Gamma_0(\mathcal{H})$. Show that $\operatorname{dom}(\partial f) + \operatorname{dom}(\partial f^*) = \mathcal{H}$. Compare to Exercise 15.3.

Exercise 23.12 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone. Prove the following:

(i) J_A is injective if and only if A is at most single-valued.
(ii) J_A is surjective if and only if $\operatorname{dom} A = \mathcal{H}$.

Exercise 23.13 Let $(A_i)_{i\in I}$ be a finite family of monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$, let $(\omega_i)_{i\in I}$ be real numbers in $]0, 1]$ such that $\sum_{i\in I} \omega_i = 1$, and define the corresponding *resolvent average* by

$$A = \big(\textstyle\sum_{i \in I} \omega_i J_{A_i} \big)^{-1} - \mathrm{Id}. \tag{23.55}$$

Show that A is monotone, and that A is maximally monotone if and only if the operators $(A_i)_{\in I}$ are likewise.

Exercise 23.14 Let f and g be in $\Gamma_0(\mathcal{H})$ and denote the proximal average of f and g by h. Show that ∂h is the resolvent average of ∂f and ∂g, with weights $(1/2, 1/2)$.

Exercise 23.15 Let $(A_i)_{\in I}$ be a finite family of maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$, let $(\omega_i)_{i\in I}$ be real numbers in $]0,1]$ such that $\sum_{i\in I} \omega_i = 1$, and denote the corresponding resolvent average by $A = (\sum_{i\in I} \omega_i J_{A_i})^{-1} - \mathrm{Id}$. Show A^{-1} is the resolvent average of the family $(A_i^{-1})_{i\in I}$.

Exercise 23.16 Let $(A_i)_{\in I}$ be a finite family of maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$, let $(\omega_i)_{i\in I}$ be real numbers in $]0,1]$ such that $\sum_{i\in I} \omega_i = 1$, and consider the corresponding resolvent average $A = \big(\sum_{i\in I} \omega_i J_{A_i}\big)^{-1} - \mathrm{Id}$. Show that if some A_i is at most single-valued, then A is likewise.

Exercise 23.17 Prove Proposition 23.19.

Exercise 23.18 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $\gamma \in \mathbb{R}_{++}$, and set $L\colon \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}\colon (x, u) \mapsto (x + \gamma u, x)$. Show that $\mathrm{gra}\, J_{\gamma A} = L(\mathrm{gra}\, A)$.

Exercise 23.19 Show that the point (y, v) in Proposition 23.24 need not be unique.

Exercise 23.20 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator and let $x \in \mathcal{H}$. Use Corollary 23.37(ii) to show that the inclusions (23.40) define a unique curve $(x_\gamma)_{\gamma\in]0,1[}$.

Exercise 23.21 Without using Theorem 23.44, show directly that the equations (23.44) define a unique curve $(x_\gamma)_{\gamma\in]0,1[}$.

Exercise 23.22 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator such that $\mathrm{zer}\, A \neq \varnothing$, and let $x \in \mathcal{H}$. Show that ${}^{\gamma}\!Ax \to 0$ as $\gamma \uparrow +\infty$.

Exercise 23.23 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $x \in \mathcal{H}$, let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that $x_n \to x$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}_{++}$ such that $\gamma_n \downarrow 0$. Show that $J_{\gamma_n A} x_n \to P_{\overline{\mathrm{dom}}\, A} x$.

Exercise 23.24 Let $A \in \mathcal{B}(\mathcal{H})$ be a monotone operator such that $\mathrm{ran}\, A$ is closed. Show that

$${}^{0}\big(A^{-1}\big) = A^{\dagger}|_{\mathrm{ran}\, A}, \tag{23.56}$$

where A^{-1} is the set-valued inverse of A.

Chapter 24
Proximity Operators

Recall from Definition 12.23 and Proposition 12.28 that the proximity operator of $f \in \Gamma_0(\mathcal{H})$ is the firmly nonexpansive operator defined by

$$\operatorname{Prox}_f \colon \mathcal{H} \to \mathcal{H} \colon x \mapsto \underset{y \in \mathcal{H}}{\operatorname{argmin}} \left(f(y) + \tfrac{1}{2}\|x - y\|^2 \right). \tag{24.1}$$

In this chapter, we pursue our investigation of the properties of proximity operators and provide concrete examples of such operators.

24.1 Characterizations

Let $f \in \Gamma_0(\mathcal{H})$. We first recall from Proposition 16.44 that the proximity operator of f is characterized by

$$(\forall x \in \mathcal{H})(\forall p \in \mathcal{H}) \quad p = \operatorname{Prox}_f x \quad \Leftrightarrow \quad x - p \in \partial f(p), \tag{24.2}$$

which yields

$$\operatorname{ran}(\operatorname{Prox}_f) = \operatorname{ran}(\operatorname{Id} + \partial f)^{-1} = \operatorname{dom}(\operatorname{Id} + \partial f) = \operatorname{dom} \partial f \subset \operatorname{dom} f. \tag{24.3}$$

Furthermore, (14.6) implies that the duality relation

$$(\forall x \in \mathcal{H})(\forall p \in \mathcal{H}) \quad p = \operatorname{Prox}_f x \quad \Leftrightarrow \quad x - p = \operatorname{Prox}_{f^*} x \tag{24.4}$$

holds. Let us sharpen (24.2) in the case of a differentiable function.

Proposition 24.1 *Let $f \in \Gamma_0(\mathcal{H})$, and let x and p be in $\mathcal{H}$. Suppose that f is Gâteaux differentiable at p. Then $p = \operatorname{Prox}_f x \Leftrightarrow \nabla f(p) + p = x$.*

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-48311-5_31

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_24

Proof. This follows from (24.2) and Proposition 17.31(i). $\square$

Example 24.2 Let $L \in \mathcal{B}(\mathcal{H})$ be self-adjoint and monotone, let $u \in \mathcal{H}$, let $\alpha \in \mathbb{R}$, and set $f\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto (1/2)\langle Lx \mid x\rangle + \langle x \mid u\rangle + \alpha$. Then $f \in \Gamma_0(\mathcal{H})$ and $(\forall x \in \mathcal{H})$ $\operatorname{Prox}_f x = (\mathrm{Id} + L)^{-1}(x - u)$.

Proof. Combine Example 2.57 with Proposition 24.1. $\square$

Example 24.3 Let I be a nonempty finite set. For every $i \in I$, let $\mathcal{K}_i$ be a real Hilbert space, let $r_i \in \mathcal{K}_i$, let $L_i \in \mathcal{B}(\mathcal{H}, \mathcal{K}_i)$, and let $\lambda_i \in \mathbb{R}_{++}$. Set $f\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto (1/2)\sum_{i\in I} \lambda_i \|L_i x - r_i\|^2$. Then $f \in \Gamma_0(\mathcal{H})$ and

$$(\forall x \in \mathcal{H}) \quad \operatorname{Prox}_f x = \Big(\mathrm{Id} + \sum_{i\in I} \lambda_i L_i^* L_i\Big)^{-1}\Big(x + \sum_{i\in I} \lambda_i L_i^* r_i\Big). \tag{24.5}$$

Proof. Set $L = \sum_{i\in I} \lambda_i L_i^* L_i$, $u = -\sum_{i\in I} \lambda_i L_i^* r_i$, and $\alpha = \sum_{i\in I} \lambda_i \|r_i\|^2/2$. Then $f\colon x \mapsto \sum_{i\in I} \lambda_i \langle L_i x - r_i \mid L_i x - r_i\rangle/2 = \langle Lx \mid x\rangle/2 + \langle x \mid u\rangle + \alpha$. Hence, (24.5) follows from Example 24.2. $\square$

The next result characterizes proximity operators as gradients.

Proposition 24.4 *Let $T\colon \mathcal{H} \to \mathcal{H}$, let $f \in \Gamma_0(\mathcal{H})$, and set $q = (1/2)\|\cdot\|^2$. Then the following are equivalent:*

(i) $T = \operatorname{Prox}_f$.
(ii) $T = \nabla(f + q)^*$.
(iii) $T = \nabla(f^* \,\square\, q) = \nabla({}^1(f^*))$.

Proof. It follows from Proposition 14.1, Proposition 12.30, and (24.4) that $\nabla(f + q)^* = \nabla({}^1(f^*)) = \mathrm{Id} - \operatorname{Prox}_{f^*} = \operatorname{Prox}_f$. $\square$

Corollary 24.5 *Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$ be such that $f = g - (1/2)\|\cdot\|^2$. Then* $\operatorname{Prox}_f = \nabla g^*$.

Corollary 24.6 (Moreau) *Let $L \in \mathcal{B}(\mathcal{H})$. Then L is a proximity operator if and only if L is self-adjoint and monotone, and $\|L\| \leqslant 1$.*

Proof. If L is a proximity operator, it is a gradient by Proposition 24.4 and hence Proposition 2.58 implies that $L = L^*$. Furthermore, L is firmly nonexpansive, hence nonexpansive and monotone, i.e., $\|L\| \leqslant 1$ and $L \succcurlyeq 0$. Conversely, suppose that $L^* = L \succcurlyeq 0$, and set $f\colon x \mapsto (1/2)\langle Lx \mid x\rangle$ and $h = f^* - q$. Then it follows from Example 2.57 and Theorem 18.15 (with $\beta = 1$) that $L = \nabla f = \operatorname{Prox}_h$. $\square$

Corollary 24.7 *Let f and g be functions in $\Gamma_0(\mathcal{H})$ such that* $\operatorname{Prox}_f = \operatorname{Prox}_g$. *Then there exists $\gamma \in \mathbb{R}$ such that $f = g + \gamma$.*

Proof. Set $q = (1/2)\|\cdot\|^2$. It follows from Proposition 24.4 that $\nabla(f + q)^* = \nabla(g + q)^*$. Therefore there exists $\gamma \in \mathbb{R}$ such that $(f + q)^* = (g + q)^* - \gamma$. Upon conjugating and invoking Corollary 13.38, we obtain $f + q = g + q + \gamma$ and hence $f = g + \gamma$. $\square$

24.2 Basic Properties

Let $f \in \Gamma_0(\mathcal{H})$. Then ∂f is maximally monotone (Theorem 20.25) and, as seen in Example 23.3,

$$(\forall \gamma \in \mathbb{R}_{++}) \quad J_{\gamma \partial f} = \operatorname{Prox}_{\gamma f}. \tag{24.6}$$

In particular, if f is Gâteaux differentiable at $x \in \mathcal{H}$, then we derive from (23.48) that

$$\operatorname{Prox}_{\gamma f} x = x - \gamma \nabla f(x) + o(\gamma) \ \text{ as } \ \gamma \downarrow 0. \tag{24.7}$$

In general, (24.6) can be used to derive properties of proximity operators from those of resolvents. For instance, in view of Example 23.3 and Corollary 16.30, setting $A = \partial f$ in (23.16) yields Theorem 14.3(ii), namely

$$(\forall x \in \mathcal{H})(\forall \gamma \in \mathbb{R}_{++}) \quad x = \operatorname{Prox}_{\gamma f} x + \gamma \operatorname{Prox}_{\gamma^{-1} f^*} \big(\gamma^{-1} x\big). \tag{24.8}$$

Here are further illustrations.

Proposition 24.8 *Let $f \in \Gamma_0(\mathcal{H})$, let $x \in \mathcal{H}$, let $z \in \mathcal{H}$, and let $\gamma \in \mathbb{R}_{++}$. Then the following hold:*

(i) *Set $g = f + (\alpha/2)\|\cdot - z\|^2 + \langle \cdot \mid u\rangle + \beta$, where $u \in \mathcal{H}$, $\alpha \in \mathbb{R}_+$, and $\beta \in \mathbb{R}$. Then $\operatorname{Prox}_{\gamma g} x = \operatorname{Prox}_{\gamma(\gamma\alpha+1)^{-1} f}\big((\gamma\alpha+1)^{-1}(x + \gamma(\alpha z - u))\big)$.*

(ii) *Set $g = \tau_z f$. Then $\operatorname{Prox}_{\gamma g} x = z + \operatorname{Prox}_{\gamma f}(x - z)$.*

(iii) *Set $g = \tau_z f - \langle \cdot \mid z\rangle$. Then $\operatorname{Prox}_g x = z + \operatorname{Prox}_f x$.*

(iv) *Set $g = f \circ L$, where $L \in \mathcal{B}(\mathcal{H})$ is an invertible operator such that $L^{-1} = L^*$. Then $\operatorname{Prox}_{\gamma g} x = L^*(\operatorname{Prox}_{\gamma f}(Lx))$.*

(v) *Set $g = f(\mu \cdot - z)$, where $\mu \in \mathbb{R} \smallsetminus \{0\}$. Then*

$$\operatorname{Prox}_{\gamma g} x = \frac{1}{\mu}\big(z + \operatorname{Prox}_{\gamma \mu^2 f}(\mu x - z)\big).$$

(vi) *Set $g = f^{\vee}$. Then $\operatorname{Prox}_g x = -\operatorname{Prox}_f(-x)$.*

(vii) *Set $g = {}^{\mu}f$, where $\mu \in \mathbb{R}_{++}$. Then*

$$\operatorname{Prox}_{\gamma g} x = x + \frac{\gamma}{\gamma + \mu}\big(\operatorname{Prox}_{(\gamma+\mu) f} x - x\big).$$

(viii) *Set $g = (2\mu)^{-1}\|\cdot\|^2 - {}^{\mu}f$, where $\mu \in \mathbb{R}_{++}$. Then*

$$\operatorname{Prox}_{\gamma g} x = x - \frac{\gamma}{\mu} \operatorname{Prox}_{\frac{\mu^2}{\mu+\gamma} f}\left(\frac{\mu}{\mu + \gamma} x\right).$$

(ix) *Set $g = f^*$. Then $\operatorname{Prox}_{\gamma g} x = x - \gamma \operatorname{Prox}_{\gamma^{-1} f}(\gamma^{-1} x)$.*

Proof. These properties are specializations to $A = \partial f$ and $B = \partial g$ of some of the results of Section 23.3.

(i): Proposition 23.17(i)&(ii) (note that Corollary 16.48(iii) and Proposition 17.31(i) imply that $\partial g = \partial f + \alpha(\operatorname{Id} - z) + u$).

(ii): Proposition 23.17(iii).

(iii): Combine (i) (with $\alpha = \beta = 0$) and (ii).

(iv): Corollary 23.27 (note that, since $\operatorname{ran} L = \mathcal{H}$, Corollary 16.53(i) yields $\partial g = L^* \circ (\partial f) \circ L$).

(v): Corollary 23.26 and (ii).

(vi): Corollary 23.28.

(vii): Proposition 23.7(iv) and (23.4).

(viii): Example 13.14 asserts that $\varphi = (\mu/2)\|\cdot\|^2 - {}^{\mu}f(\mu\cdot) \in \Gamma_0(\mathcal{H})$. Therefore $g = \varphi(\mu^{-1}\cdot) \in \Gamma_0(\mathcal{H})$. Furthermore, (23.4) yields $\nabla g = \mu^{-1}\mathrm{Id} - {}^{\mu}(\partial f)$. Hence, the result follows from Proposition 23.29.

(ix): See (24.8). ☐

Example 24.9 Consider the generalized Huber function of Example 13.7, namely

$$f\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \begin{cases} \rho\|x\| - \dfrac{\rho^2}{2}, & \text{if } \|x\| > \rho;\\[2mm] \dfrac{\|x\|^2}{2}, & \text{if } \|x\| \leqslant \rho, \end{cases} \tag{24.9}$$

and let $\gamma \in \mathbb{R}_{++}$. Then

$$(\forall x \in \mathcal{H}) \quad \operatorname{Prox}_{\gamma f} x = \begin{cases} \left(1 - \dfrac{\gamma\rho}{\|x\|}\right)x, & \text{if } \|x\| > (\gamma+1)\rho;\\[2mm] \dfrac{1}{\gamma+1}x, & \text{if } \|x\| \leqslant (\gamma+1)\rho. \end{cases} \tag{24.10}$$

Proof. Combine Example 14.5 and Proposition 24.8(vii). ☐

Proposition 24.10 *Let $f \in \Gamma_0(\mathcal{H})$ be even. Then* Prox_f *is odd and* $\operatorname{Prox}_f 0 = 0$.

Proof. The first claim follows from Proposition 24.8(vi) and it implies the second. ☐

Proposition 24.11 *Let $(\mathcal{H}_i)_{i\in I}$ be a finite family of real Hilbert spaces, set $\boldsymbol{\mathcal{H}} = \bigoplus_{i\in I} \mathcal{H}_i$, and, for every $i \in I$, let $f_i \in \Gamma_0(\mathcal{H}_i)$ and $x_i \in \mathcal{H}_i$. Set $\boldsymbol{x} = (x_i)_{i\in I}$ and $\boldsymbol{f} = \bigoplus_{i\in I} f_i$. Then $\operatorname{Prox}_{\boldsymbol{f}} \boldsymbol{x} = (\operatorname{Prox}_{f_i} x_i)_{i\in I}$.*

Proof. It is clear that $\boldsymbol{f} \in \Gamma_0(\boldsymbol{\mathcal{H}})$. The result therefore follows from Proposition 23.18, with $(\forall i \in I)\ A_i = \partial f_i$. ☐

Proposition 24.12 *Let $(\phi_i)_{i\in I}$ be a family of functions in $\Gamma_0(\mathbb{R})$ such that $(\forall i \in I)\ \phi_i \geqslant \phi_i(0) = 0$. Suppose that $\mathcal{H} = \ell^2(I)$ and set $f = \bigoplus_{i\in I} \phi_i$. Then the following hold:*

(i) $f \in \Gamma_0(\mathcal{H})$.

(ii) $(\forall (\xi_i)_{i\in I} \in \mathcal{H})\ \operatorname{Prox}_f (\xi_i)_{i\in I} = (\operatorname{Prox}_{\phi_i} \xi_i)_{i\in I}$.

(iii) Prox_f *is weakly sequentially continuous.*

Proof. (i): For every $i \in I$, define $\varphi_i \colon \mathcal{H} \to \left]-\infty,+\infty\right] : (\xi_j)_{j\in I} \mapsto \phi_i(\xi_i)$. Then the family $(\varphi_i)_{i\in I}$ lies in $\Gamma_0(\mathcal{H})$ and, since $f(0) = 0$, it follows from Corollary 9.4(ii) that $f = \sum_{i\in I} \varphi_i \in \Gamma_0(\mathcal{H})$.

(ii): Let $x = (\xi_i)_{i\in I} \in \mathcal{H}$, and set $p = (\pi_i)_{i\in I} = (\operatorname{Prox}_{\phi_i} \xi_i)_{i\in I}$. For every $i \in I$, since $0 \in \operatorname{Argmin} \phi_i$, it follows from Theorem 16.3 that $0 - 0 \in \partial\phi_i(0)$ and therefore from (16.37) that $\operatorname{Prox}_{\phi_i} 0 = 0$. Hence, we derive from Proposition 12.28 that

$$(\forall i \in I) \quad |\pi_i|^2 = |\operatorname{Prox}_{\phi_i} \xi_i - \operatorname{Prox}_{\phi_i} 0|^2 \leqslant |\xi_i - 0|^2 = |\xi_i|^2. \tag{24.11}$$

This shows that $p \in \mathcal{H}$. Now let $y = (\eta_i)_{i\in I} \in \mathcal{H}$. It follows from Proposition 12.26 that

$$(\forall i \in I) \quad (\eta_i - \pi_i)(\xi_i - \pi_i) + \phi_i(\pi_i) \leqslant \phi_i(\eta_i). \tag{24.12}$$

Summing over I, we obtain $\langle y - p \mid x - p\rangle + f(p) \leqslant f(y)$. In view of (12.25), we conclude that $p = \operatorname{Prox}_f x$.

(iii): Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ and let $x \in \mathcal{H}$ be such that $x_n \rightharpoonup x$. Then Proposition 2.50 asserts that $(x_n)_{n\in\mathbb{N}}$ is bounded. On the other hand, arguing as in (24.11), since $0 \in \operatorname{Argmin} f$, $\operatorname{Prox}_f 0 = 0$ and, by nonexpansiveness of Prox_f, we obtain $(\forall n \in \mathbb{N})\ \|\operatorname{Prox}_f x_n\| \leqslant \|x_n\|$. Hence,

$$\big(\operatorname{Prox}_f x_n\big)_{n\in\mathbb{N}} \text{ is bounded.} \tag{24.13}$$

Now set $x = (\xi_i)_{i\in I}$ and $(\forall n \in \mathbb{N})\ x_n = (\xi_{i,n})_{i\in I}$. Denoting the standard unit vectors in $\mathcal{H}$ by $(e_i)_{i\in I}$, we obtain $(\forall i \in I)\ \xi_{i,n} = \langle x_n \mid e_i\rangle \to \langle x \mid e_i\rangle = \xi_i$. Since, by Proposition 12.28, the operators $(\operatorname{Prox}_{\phi_i})_{i\in I}$ are continuous, (ii) yields $(\forall i \in I)\ \langle \operatorname{Prox}_f x_n \mid e_i\rangle = \operatorname{Prox}_{\phi_i} \xi_{i,n} \to \operatorname{Prox}_{\phi_i} \xi_i = \langle \operatorname{Prox}_f x \mid e_i\rangle$. It then follows from (24.13) and Proposition 2.50 that $\operatorname{Prox}_f x_n \rightharpoonup \operatorname{Prox}_f x$. ☐

Proposition 24.13 *Let $(\Omega, \mathcal{F}, \mu)$ be a complete σ-finite measure space, let $(\mathsf{H}, \langle\cdot \mid \cdot\rangle_{\mathsf{H}})$ be a separable real Hilbert space, and let $\varphi \in \Gamma_0(\mathsf{H})$. Suppose that $\mathcal{H} = L^2((\Omega, \mathcal{F}, \mu); \mathsf{H})$ and that one of the following holds:*

(i) $\mu(\Omega) < +\infty$.
(ii) $\varphi \geqslant \varphi(0) = 0$.

Set

$$\begin{aligned} f\colon\ & \mathcal{H} \to \left]-\infty,+\infty\right] \\ & x \mapsto \begin{cases} \displaystyle\int_\Omega \varphi\big(x(\omega)\big)\mu(d\omega), & \text{if } \varphi \circ x \in L^1\big((\Omega, \mathcal{F}, \mu); \mathbb{R}\big); \\ +\infty, & \text{otherwise.} \end{cases} \end{aligned} \tag{24.14}$$

Let $x \in \mathcal{H}$ and define, for μ-almost every $\omega \in \Omega$, $p(\omega) = \operatorname{Prox}_\varphi x(\omega)$. Then $p = \operatorname{Prox}_f x$.

Proof. By Proposition 9.40, $f \in \Gamma_0(\mathcal{H})$. Now take x and p in $\mathcal{H}$. Then it follows from (24.2) and Proposition 16.63 that $p(\omega) = \operatorname{Prox}_\varphi x(\omega)$ μ-a.e. $\Leftrightarrow$ $x(\omega) - p(\omega) \in \partial\varphi(p(\omega))$ μ-a.e. $\Leftrightarrow$ $x - p \in \partial f(p)$. ☐

Proposition 24.14 *Let $\mathcal{K}$ be a real Hilbert space, let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$ be such that $LL^* = \mu\mathrm{Id}$ for some $\mu \in \mathbb{R}_{++}$, let $f \in \Gamma_0(\mathcal{K})$, and let $x \in \mathcal{H}$. Then $f \circ L \in \Gamma_0(\mathcal{H})$ and $\operatorname{Prox}_{f\circ L} x = x + \mu^{-1}L^*(\operatorname{Prox}_{\mu f}(Lx) - Lx)$.*

Proof. Since $\operatorname{ran} L = \mathcal{K}$ by Fact 2.25(vi) and Fact 2.26, Corollary 16.53(i) yields $\partial(f \circ L) = L^* \circ (\partial f) \circ L$. Hence, the result follows by setting $A = \partial f$ in Proposition 23.25(iii) and using (23.2). ☐

Corollary 24.15 *Suppose that $u \in \mathcal{H}\smallsetminus\{0\}$, let $\phi \in \Gamma_0(\mathbb{R})$, set $g = \phi(\langle \cdot \mid u\rangle)$, and let $x \in \mathcal{H}$. Then*

$$\operatorname{Prox}_g x = x + \frac{\operatorname{Prox}_{\|u\|^2\phi} \langle x \mid u\rangle - \langle x \mid u\rangle}{\|u\|^2} u. \tag{24.15}$$

Proof. Apply Proposition 24.14 to $\mathcal{K} = \mathbb{R}$, $f = \phi$, and $L = \langle \cdot \mid u\rangle$. ☐

Proposition 24.16 *Suppose that $\mathcal{H}$ is infinite-dimensional and separable, let $(e_k)_{k\in\mathbb{N}}$ be an orthonormal basis of $\mathcal{H}$, let $(\phi_k)_{k\in\mathbb{N}}$ be a family of functions in $\Gamma_0(\mathbb{R})$ such that $(\forall k \in \mathbb{N})$ $\phi_k \geqslant \phi_k(0) = 0$, and set*

$$f\colon \mathcal{H} \to \left]-\infty,+\infty\right] : x \mapsto \sum_{k\in\mathbb{N}} \phi_k\big(\langle x \mid e_k\rangle\big). \tag{24.16}$$

Then $f \in \Gamma_0(\mathcal{H})$ and

$$(\forall x \in \mathcal{H}) \quad \operatorname{Prox}_f x = \sum_{k\in\mathbb{N}} \big(\operatorname{Prox}_{\phi_k} \langle x \mid e_k\rangle\big) e_k. \tag{24.17}$$

Proof. Set $L\colon \mathcal{H} \to \ell^2(\mathbb{N})\colon x \mapsto (\langle x \mid e_k\rangle)_{k\in\mathbb{N}}$. Then L is linear, bounded, and invertible with $L^{-1} = L^*\colon \ell^2(\mathbb{N}) \to \mathcal{H}\colon (\xi_k)_{k\in\mathbb{N}} \mapsto \sum_{k\in\mathbb{N}} \xi_k e_k$. Now set $\varphi\colon \ell^2(\mathbb{N}) \to \left]-\infty,+\infty\right] : (\xi_k)_{k\in\mathbb{N}} \mapsto \sum_{k\in\mathbb{N}} \phi_k(\xi_k)$. Then $f = \varphi \circ L$ and we derive from Proposition 24.14 that $\operatorname{Prox}_f = L^* \circ \operatorname{Prox}_\varphi \circ L$. In turn, using Proposition 24.12(i)&(ii), we obtain the announced results. ☐

Proposition 24.17 *Suppose that $\mathcal{H} = \mathbb{R}^N$ and set $I = \{1,\ldots,N\}$. Let $f \in \Gamma_0(\mathcal{H})$ be such that $\operatorname{dom}\partial f \subset \mathbb{R}^N_{++}$, set $u = (\mu_i)_{i\in I} \in \mathbb{R}^N_+$, set $g\colon \mathcal{H} \to \mathbb{R}\colon (\xi_i)_{i\in I} \mapsto \sum_{i\in I} \mu_i|\xi_i|$, and let $x \in \mathcal{H}$. Then $\operatorname{Prox}_{f+g} x = \operatorname{Prox}_f(x - u)$.*

Proof. Set $p = \operatorname{Prox}_{f+g} x$. Then (24.3) and Corollary 16.48(iii) imply that $p \in \operatorname{dom}\partial(f + g) = \operatorname{dom}(\partial f + \partial g) = \operatorname{dom}\partial f \subset \mathbb{R}^N_{++}$ and, in turn, that $\partial g(p) = \{u\}$. It therefore follows from (24.2) that $x - p \in \partial(f + g)(p) = \partial f(p) + u$, i.e., $p = \operatorname{Prox}_f(x - u)$. ☐

Next, we describe a situation in which the proximity operator of a sum of functions reduces to a composition of their proximity operators.

Proposition 24.18 *Let f and g be in $\Gamma_0(\mathcal{H})$. Suppose that one of the following holds:*

(i) $(\forall y \in \operatorname{dom} \partial g)\ \partial g(y) \subset \partial g(\operatorname{Prox}_f y)$.
(ii) $(\forall (x,u) \in \operatorname{gra} \partial f)\ \partial g(x+u) \subset \partial g(x)$.
(iii) $\operatorname{dom} g$ *is open,* g *is Gâteaux differentiable on* $\operatorname{dom} g$, $\operatorname{dom} \partial f \subset \operatorname{dom} g$, *and* $(\forall y \in \operatorname{dom} g)\ \nabla g(y) = \nabla g(\operatorname{Prox}_f y)$.
(iv) $0 \in \operatorname{sri}(\operatorname{dom} f - \operatorname{dom} g)$ *and* $(\forall (x,u) \in \operatorname{gra} \partial f)\ \partial g(x) \subset \partial g(x+u)$.

Then $\operatorname{Prox}_{f+g} = \operatorname{Prox}_f \circ \operatorname{Prox}_g$.

Proof. Set $A = \partial f$ and $B = \partial g$. Then A and B are maximally monotone by Theorem 20.25 and we can apply Proposition 23.32 to them.

(i)&(ii): These follow respectively from (i) and (ii) in Proposition 23.32.

(iii): In view of Proposition 17.31(i), this is a special case of (i).

(iv): Since $f + g \in \Gamma_0(\mathcal{H})$, it follows from Corollary 16.48 and Theorem 20.25 that $A + B = \partial(f+g)$ is maximally monotone. The claim is therefore a consequence of Proposition 23.32(iii). ☐

Proposition 24.19 *Let $f \in \Gamma_0(\mathcal{H})$ be positively homogeneous, let $\gamma \in \mathbb{R}_{++}$, and let $x \in \mathcal{H}$. Then* $\operatorname{Prox}_{\gamma f} x = x - P_{\gamma \partial f(0)}\, x$.

Proof. By Proposition 16.24, $f = \sigma_{\partial f(0)}$. Hence, Theorem 14.3(ii), Proposition 16.4(iii), and Example 13.43(i) yield $\operatorname{Prox}_{\gamma f} x = x - \gamma \operatorname{Prox}_{f^*/\gamma}(x/\gamma) = x - \gamma P_{\partial f(0)}(x/\gamma) = x - P_{\gamma \partial f(0)}\, x$. ☐

As a special case of Proposition 24.19, we recover the soft thresholder of Example 4.17(i).

Example 24.20 Let $\gamma \in \mathbb{R}_{++}$ and let $x \in \mathcal{H}$. Then

$$\operatorname{Prox}_{\gamma\|\cdot\|} x = \left(1 - \frac{\gamma}{\max\{\|x\|, \gamma\}}\right) x. \tag{24.18}$$

If $\mathcal{H} = \mathbb{R}$, then (24.18) reduces to the soft thresholder on $[-\gamma, \gamma]$, namely

$$\begin{aligned}\operatorname{Prox}_{\gamma|\cdot|} x &= \operatorname{soft}_{[-\gamma,\gamma]} x = \operatorname{sign}(x) \max\{|x| - \gamma, 0\} \\ &= \begin{cases} x + \gamma, & \text{if } x < -\gamma; \\ 0, & \text{if } -\gamma \leqslant x \leqslant \gamma; \\ x - \gamma, & \text{if } x > \gamma. \end{cases}\end{aligned} \tag{24.19}$$

Proof. Apply Proposition 24.19 to $f = \|\cdot\|$, and use Example 16.32 and Example 3.18. ☐

Example 24.21 Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a complete probability space, suppose that $\mathcal{H} = L^2(\Omega, \mathcal{F}, \mathsf{P})$ (see Example 2.9), and set $f \colon \mathcal{H} \to \mathbb{R} \colon X \mapsto \mathsf{E}|X|$. Let $X \in \mathcal{H}$, let $\gamma \in \mathbb{R}_{++}$, and define, for P-almost every $\omega \in \Omega$, $P(\omega) = \operatorname{soft}_{[-\gamma,\gamma]} X(\omega) = \operatorname{sign}(X(\omega)) \max\{|X(\omega)| - \gamma, 0\}$. Then $P = \operatorname{Prox}_{\gamma f} X$.

Proof. Apply Proposition 24.13 with $\mu = \mathsf{P}$, $\mathsf{H} = \mathbb{R}$, and $\varphi = \gamma|\cdot|$, and use (24.19). ☐

Example 24.22 Suppose that $\mathcal{H} = \mathbb{R}^N$, set $I = \{1,\ldots,N\}$, let $\gamma \in \mathbb{R}_{++}$, and let $x = (\xi_i)_{i\in I} \in \mathcal{H}$. Then $\operatorname{Prox}_{\gamma\|\cdot\|_1} x = (\operatorname{sign}(\xi_i)\max\{|\xi_i| - \gamma, 0\})_{i\in I}$.

Proof. Combine Proposition 24.12(ii) and (24.19). ☐

Example 24.23 Suppose that $\mathcal{H} = \mathbb{R}^N$, set $I = \{1,\ldots,N\}$, let $\gamma \in \mathbb{R}_{++}$, set $C = \big\{x \in \mathcal{H} \;\big|\; \|x\|_1 \leqslant 1\big\}$, and let $x \in \mathcal{H}$. Then $\operatorname{Prox}_{\gamma\|\cdot\|_\infty} x = x - \gamma P_C(\gamma^{-1}x)$, where P_C is provided by Example 29.28.

Proof. In view of Proposition 24.8(ix) and Example 13.32, $\operatorname{Prox}_{\gamma\|\cdot\|_\infty} x = x - \gamma \operatorname{Prox}_{\gamma^{-1}\|\cdot\|_\infty^*}(\gamma^{-1}x) = x - \gamma \operatorname{Prox}_{\iota_C}(\gamma^{-1}x) = x - \gamma P_C(\gamma^{-1}x)$. ☐

The next result examines proximity operators when the scalar product is modified.

Proposition 24.24 *Let $f \in \Gamma_0(\mathcal{H})$, let $U \in \mathcal{B}(\mathcal{H})$ be a self-adjoint strongly monotone operator, and let*

$$\operatorname{Prox}_f^U \colon \mathcal{H} \to \mathcal{H} \colon x \mapsto \underset{y\in\mathcal{H}}{\operatorname{argmin}} \;\Big(f(y) + \tfrac{1}{2}\langle U(x-y) \mid x-y\rangle \Big) \tag{24.20}$$

be the proximity operator of f relative to the metric induced by U. Then the following hold:

(i) $\operatorname{Prox}_f^U = J_{U^{-1}\partial f}$.
(ii) $\operatorname{Prox}_f^U = U^{-1/2}\operatorname{Prox}_{f\circ U^{-1/2}} U^{1/2} = \operatorname{Id} - U^{-1}\operatorname{Prox}_{f^*}^{U^{-1}} U$.

Proof. (i): Let x and p be in $\mathcal{H}$. Then it follows from Theorem 16.3 and Corollary 16.48(iii) that $p = \operatorname{Prox}_f^U x \Leftrightarrow 0 \in \partial f(p) + U(p-x) \Leftrightarrow x \in p + U^{-1}\partial f(p) \Leftrightarrow p = J_{U^{-1}\partial f}x$.

(ii): Apply Proposition 23.34(iii) to $A = \partial f$, using (i) and the facts that, by Corollary 16.53(i), $\partial(f\circ U^{-1/2}) = (U^{-1/2})^*\circ(\partial f)\circ U^{-1/2} = U^{-1/2}\circ(\partial f)\circ U^{-1/2}$ and that, by Corollary 16.30, $J_{UA^{-1}} = J_{U\partial f^*} = \operatorname{Prox}_{f^*}^{U^{-1}}$. ☐

Example 24.25 Suppose that $\mathcal{H} = \mathbb{R}^N$, where $N \in \{2,3,\ldots\}$. Set $I = \{1,\ldots,N\}$, set $f \colon \mathcal{H} \to \mathbb{R} \colon (\xi_i)_{i\in I} \mapsto \max\{\xi_i\}_{i\in I}$, let $x \in \mathcal{H}$, set $(\xi_i)_{i\in I} = x_\downarrow$ (see Fact 2.18), and let $P \in \mathbb{R}^{N\times N}$ be a permutation matrix such that $Px = x_\downarrow$. Set

$$(\forall i \in I)\quad \eta_i = \frac{-1 + \sum_{k=1}^{i}\xi_k}{i}, \tag{24.21}$$

consider the (ordered) list of inequalities

$$\eta_1 > \xi_2,\ \eta_2 > \xi_3,\ \ldots,\ \eta_{N-1} > \xi_N,\ \eta_N \leqslant \xi_N, \tag{24.22}$$

and let n be the smallest integer in I for which the nth inequality is satisfied. Then

$$\operatorname{Prox}_f x_{\downarrow} = [\eta_n, \dots, \eta_n, \xi_{n+1}, \dots, \xi_N]^{\mathsf{T}} \tag{24.23}$$

and

$$\operatorname{Prox}_f x = P^{\mathsf{T}} \operatorname{Prox}_f x_{\downarrow}. \tag{24.24}$$

Proof. Denote the vector on the right-hand side of (24.23) by $p = (\pi_i)_{i\in I}$, and the standard unit vectors of $\mathcal{H}$ by $(e_i)_{i\in I}$. Note that n is well defined because, if the second-to-last inequality in (24.22) fails, then the last one must hold. To establish (24.23), we consider N cases.

- $n = 1$: Observe that $p = x_{\downarrow} - e_1$ and that $\pi_1 = \eta_1 = \xi_1 - 1 > \pi_2 = \xi_2 \geqslant \pi_3 = \xi_3 \geqslant \cdots \geqslant \pi_N = \xi_N$. Hence $\nabla f(p) = e_1$ and $p + \nabla f(p) = x_{\downarrow}$. Therefore, Proposition 24.1 yields $p = \operatorname{Prox}_f x_{\downarrow}$.
- $n = 2$: We have $\eta_1 \leqslant \xi_2$ and $\eta_2 > \xi_3$. The latter inequality implies that $\pi_1 = \eta_2 = \pi_2 > \pi_3 = \xi_3 \geqslant \cdots \geqslant \pi_N = \xi_N$. It follows from Theorem 18.5 that $\partial f(p) = [e_1, e_2]$. The inequality $\eta_1 \leqslant \xi_2$ yields $\delta_2 = \xi_2 - \pi_2 \geqslant 0$ and thus $\delta_1 = \xi_1 - \pi_1 = \xi_1 - \pi_2 \geqslant \xi_2 - \pi_2 \geqslant 0$. Furthermore, $\delta_1 + \delta_2 = 1$. It follows that $\delta_1 e_1 + \delta_2 e_2 \in [e_1, e_2]$ and thus $x_{\downarrow} = p + \delta_1 e_1 + \delta_2 e_2 \in (\mathrm{Id} + \partial f)(p)$. Therefore, $p = \operatorname{Prox}_f x_{\downarrow}$.
- $3 \leqslant n \leqslant N$: These remaining $N - 2$ cases are dealt with analogously.

Finally, (24.24) follows from Proposition 24.8(iv). ☐

The following result provides information about the gap between function values and optimal values.

Proposition 24.26 (Güler) *Let $f \in \Gamma_0(\mathcal{H})$, let $x \in \mathcal{H}$, and suppose that $z \in \operatorname{Argmin} f$ and that $\operatorname{Prox}_f x \notin \operatorname{Argmin} f$. Then*

$$\frac{f(\operatorname{Prox}_f x) - f(z)}{f(x) - f(z)} \leqslant 1 - \frac{2}{3}\frac{f(\operatorname{Prox}_f x) - f(z)}{\|x - z\|^2} \tag{24.25}$$

or, equivalently,

$$\frac{2}{3\|x - z\|^2} \leqslant \frac{1}{f(\operatorname{Prox}_f x) - f(z)} - \frac{1}{f(x) - f(z)}. \tag{24.26}$$

Proof. Set $p = \operatorname{Prox}_f x$ and observe that Proposition 12.29 implies that $x \notin \operatorname{Argmin} f$. Hence $x \neq z$ and $f(z) < f(x)$. It follows from Proposition 12.26 and Cauchy–Schwarz that

$$\begin{aligned} f(z) &\geqslant f(p) + \langle x - p \mid z - p\rangle \\ &= f(p) + \langle x - p \mid z - x\rangle + \langle x - p \mid x - p\rangle \\ &\geqslant f(p) - \|x - p\|\|z - x\| + \|x - p\|^2. \end{aligned} \tag{24.27}$$

Hence

$$\frac{f(p) - f(z)}{\|x - z\|} \leqslant \|x - p\|. \tag{24.28}$$

On the other hand, Proposition 12.27 yields

$$\|x - p\|^2 + f(p) \leqslant f(x). \tag{24.29}$$

Altogether,

$$\big(f(x) - f(z)\big) - \big(f(p) - f(z)\big) = f(x) - f(p) \geqslant \left(\frac{f(p) - f(z)}{\|x - z\|}\right)^2. \tag{24.30}$$

Hence

$$\frac{1}{f(x) - f(z)} \leqslant \frac{1}{f(p) - f(z)} \left(1 + \frac{f(p) - f(z)}{\|x - z\|^2}\right)^{-1}. \tag{24.31}$$

Using (12.23), we obtain $f(p) \leqslant f(p) + (1/2)\|x - p\|^2 \leqslant f(z) + (1/2)\|x - z\|^2$. Thus

$$\frac{f(p) - f(z)}{\|x - z\|^2} \leqslant \frac{1}{2}. \tag{24.32}$$

Set $(\forall t \in \mathbb{R})$ $g(t) = 1/(1 + t)$ if $t > -1$, and $g(t) = +\infty$ otherwise. Then $g(0) = 1$ and $g(1/2) = 2/3$. Since g is convex, we deduce that

$$\big(\forall t \in [0, 1/2]\big) \quad (1 + t)^{-1} \leqslant 1 - \frac{2}{3}t. \tag{24.33}$$

Combining (24.31)–(24.33), we obtain

$$\frac{f(p) - f(z)}{f(x) - f(z)} \leqslant \left(1 + \frac{f(p) - f(z)}{\|x - z\|^2}\right)^{-1} \leqslant 1 - \frac{2}{3}\frac{f(p) - f(z)}{\|x - z\|^2}, \tag{24.34}$$

which completes the proof. ☐

24.3 Distance-Based Functions

We start with a general result.

Proposition 24.27 *Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\phi\colon \mathbb{R} \to \mathbb{R}$ be an even convex function that is differentiable on $\mathbb{R} \smallsetminus \{0\}$, set $f = \phi \circ d_C$, and let $x \in \mathcal{H}$. Then*

$$\operatorname{Prox}_f x = \begin{cases} x + \dfrac{\operatorname{Prox}_{\phi^*} d_C(x)}{d_C(x)}(P_C x - x), & \text{if } d_C(x) > \max \partial\phi(0); \\ P_C x, & \text{if } d_C(x) \leqslant \max \partial\phi(0). \end{cases} \tag{24.35}$$

Proof. It follows from Corollary 8.40 and Proposition 11.7(ii) that ϕ is continuous and increasing on $\mathbb{R}_+$. On the other hand, it follows from Corollary 12.12 that d_C is convex. Hence, Proposition 8.21 asserts that $\phi \circ d_C$ is a real-valued continuous convex function and therefore $f \in \Gamma_0(\mathcal{H})$. Now set $p = \operatorname{Prox}_f x$. Since ϕ is a real-valued even convex function, we derive from Example 17.37 that $\partial\phi(0) = [-\beta, \beta]$ for some $\beta \in \mathbb{R}_+$. We establish two intermediate results.

(a) $p \in C \Rightarrow d_C(x) \leqslant \beta$: Suppose that $p \in C$ and let $y \in C$. Then $f(y) = \phi(d_C(y)) = \phi(0)$ and, in particular, $f(p) = \phi(0)$. Hence, it follows from (24.2) and (16.1) that

$$\langle y - p \mid x - p\rangle + \phi(0) = \langle y - p \mid x - p\rangle + f(p) \leqslant f(y) = \phi(0). \tag{24.36}$$

Consequently, $\langle y - p \mid x - p\rangle \leqslant 0$ and, in view of (3.10), we get

$$p = P_C x. \tag{24.37}$$

Now let $u \in \partial f(p)$. Since $p \in C$, $d_C(p) = 0$ and, by (7.4), $\sigma_C(u) \geqslant \langle p \mid u\rangle$. Hence, Proposition 13.15 and Example 13.26 yield

$$-0\,\|u\| = 0 \leqslant \sigma_C(u) - \langle p \mid u\rangle = \sigma_C(u) - f(p) - f^*(u) = -\phi(0) - \phi^*(\|u\|). \tag{24.38}$$

We therefore deduce from Proposition 13.15 that $\|u\| \in \partial\phi(0)$. Thus, $u \in \partial f(p) \Rightarrow \|u\| \leqslant \beta$. Since (24.2) asserts that $x - p \in \partial f(p)$, we obtain $\|x - p\| \leqslant \beta$ and hence, since $p \in C$, $d_C(x) \leqslant \|x - p\| \leqslant \beta$.

(b) $p \notin C \Rightarrow d_C(x) > \beta$: Suppose that $p \notin C$. Since C is closed, $d_C(p) > 0$ and ϕ is therefore differentiable at $d_C(p)$. It follows from (24.2), Fact 2.63, and Proposition 18.23(iii) that

$$x - p = f'(p) = \frac{\phi'(d_C(p))}{d_C(p)}(p - P_C p). \tag{24.39}$$

Since $\phi' \geqslant 0$ on $\mathbb{R}_{++}$, upon taking the norm, we obtain

$$\|p - x\| = \phi'\big(d_C(p)\big) \tag{24.40}$$

and therefore

$$p - x = \frac{\|p - x\|}{d_C(p)}(P_C p - p). \tag{24.41}$$

In turn, appealing to Proposition 4.16 and (3.10), we obtain

$$\begin{aligned}\|P_C p - P_C x\|^2 &\leqslant \langle p - x \mid P_C p - P_C x\rangle \\ &= \frac{\|p - x\|}{d_C(p)}\,\langle P_C p - p \mid P_C p - P_C x\rangle \\ &\leqslant 0,\end{aligned} \tag{24.42}$$

from which we deduce that

$$P_C p = P_C x. \tag{24.43}$$

Hence, (24.41) becomes

$$p - x = \frac{\|p - x\|}{\|p - P_C x\|}(P_C x - p), \tag{24.44}$$

which can be rewritten as

$$p - x = \frac{\|p - x\|}{\|p - x\| + \|p - P_C x\|}(P_C x - x). \tag{24.45}$$

Taking the norm yields

$$\|p - x\| = \frac{\|p - x\|}{\|p - x\| + \|p - P_C x\|} d_C(x), \tag{24.46}$$

and it follows from (24.43) that

$$d_C(x) = \|p - x\| + \|p - P_C x\| = \|p - x\| + d_C(p). \tag{24.47}$$

Therefore, in the light of (24.40), we obtain

$$d_C(x) - d_C(p) = \|p - x\| = \phi'(d_C(p)) \tag{24.48}$$

and we derive from Proposition 24.1 that

$$0 < d_C(p) = \operatorname{Prox}_\phi d_C(x) \tag{24.49}$$

and hence from (24.4) that

$$d_C(x) - d_C(p) = d_C(x) - \operatorname{Prox}_\phi d_C(x) = \operatorname{Prox}_{\phi^*} d_C(x). \tag{24.50}$$

In turn, (24.48) results in

$$\|p - x\| = d_C(x) - d_C(p) = \operatorname{Prox}_{\phi^*} d_C(x). \tag{24.51}$$

To sum up, coming back to (24.45) and invoking (24.47) and (24.51), we obtain

$$\begin{aligned} p &= x + \frac{\|p - x\|}{\|p - x\| + \|p - P_C x\|}(P_C x - x) \\ &= x + \frac{\operatorname{Prox}_{\phi^*} d_C(x)}{d_C(x)}(P_C x - x). \end{aligned} \tag{24.52}$$

Furthermore, since $d_C(p) \neq 0$ by (24.49), we derive from (24.2) that $d_C(x) \notin \partial\phi(0)$ and hence that $d_C(x) > \beta$.

Upon combining (a) and (b), we obtain

$$p \in C \quad \Leftrightarrow \quad d_C(x) \leqslant \beta. \tag{24.53}$$

Altogether, (24.35) follows from (24.37), (24.52), and (24.53). □

Example 24.28 Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\gamma \in \mathbb{R}_{++}$, and let $x \in \mathcal{H}$. Then

$$\operatorname{Prox}_{\gamma d_C} x = \begin{cases} x + \dfrac{\gamma}{d_C(x)}(P_C x - x), & \text{if } d_C(x) > \gamma; \\ P_C x, & \text{if } d_C(x) \leqslant \gamma. \end{cases} \tag{24.54}$$

Proof. Apply Proposition 24.27 with $\phi = \gamma|\cdot|$ and use Example 16.15. □

Example 24.29 Let $\varepsilon \in \mathbb{R}_{++}$, set $f = \max\{\|\cdot\| - \varepsilon, 0\}$, let $\gamma \in \mathbb{R}_{++}$, and let $x \in \mathcal{H}$. Then

$$\operatorname{Prox}_{\gamma f} x = \begin{cases} (1 - \gamma/\|x\|)x, & \text{if } \|x\| > \gamma + \varepsilon; \\ (\varepsilon/\|x\|)x, & \text{if } \varepsilon < \|x\| \leqslant \gamma + \varepsilon; \\ x, & \text{if } \|x\| \leqslant \varepsilon. \end{cases} \tag{24.55}$$

If $\mathcal{H} = \mathbb{R}$, f is *Vapnik's ε-insensitive loss* and (24.55) becomes

$$\operatorname{Prox}_{\gamma f} x = \begin{cases} x - \gamma \operatorname{sign}(x), & \text{if } |x| > \gamma + \varepsilon; \\ \varepsilon \operatorname{sign}(x), & \text{if } \varepsilon < |x| \leqslant \gamma + \varepsilon; \\ x, & \text{if } |x| \leqslant \varepsilon. \end{cases} \tag{24.56}$$

Proof. Set $C = B(0;\varepsilon)$. As seen in (3.13), $f = d_C$. Hence, the result is obtained by combining Example 24.28 and Example 3.18. □

Proposition 24.30 *Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\phi \in \Gamma_0(\mathbb{R})$ be an even function such that ϕ^* is differentiable on $\mathbb{R} \smallsetminus \{0\}$, set $f = \sigma_C + \phi \circ \|\cdot\|$, and let $x \in \mathcal{H}$. Then*

$$\operatorname{Prox}_f x = \begin{cases} \dfrac{\operatorname{Prox}_\phi d_C(x)}{d_C(x)}(x - P_C x), & \text{if } d_C(x) > \max \operatorname{Argmin} \phi; \\ x - P_C x, & \text{if } d_C(x) \leqslant \max \operatorname{Argmin} \phi. \end{cases} \tag{24.57}$$

Proof. We derive from Proposition 13.21 that ϕ^* is a convex even function which is differentiable on $\mathbb{R} \smallsetminus \{0\}$, from Corollary 13.38 that $\phi^{**} = \phi$, and from Proposition 16.33 that $\partial\phi^*(0) = \operatorname{Argmin} \phi$. Therefore the result follows from (24.2), Example 13.26, and Proposition 24.27 (applied to ϕ^*). □

24.4 Functions on the Real Line

As evidenced by the results in the previous sections, proximity operators of functions in $\Gamma_0(\mathbb{R})$ play a prominent role in the computation of various proximity operators on $\mathcal{H}$.

Proposition 24.31 *Let $\varrho\colon \mathbb{R} \to \mathbb{R}$. Then ϱ is the proximity operator of a function in $\Gamma_0(\mathbb{R})$ if and only if ϱ is nonexpansive and increasing.*

Proof. Let ξ and η be real numbers. First, suppose that $\varrho = \operatorname{Prox}_\phi$, where $\phi \in \Gamma_0(\mathbb{R})$. Then it follows from Proposition 12.28 that ϱ is nonexpansive and that $0 \leqslant |\varrho(\xi)-\varrho(\eta)|^2 \leqslant (\xi-\eta)(\varrho(\xi)-\varrho(\eta))$, which shows that ϱ is increasing. Conversely, suppose that ϱ is nonexpansive and increasing and, without loss of generality, that $\xi < \eta$. Then $0 \leqslant \varrho(\xi) - \varrho(\eta) \leqslant \xi - \eta$ and therefore $|\varrho(\xi)-\varrho(\eta)|^2 \leqslant (\xi-\eta)(\varrho(\xi)-\varrho(\eta))$. Thus, ϱ is firmly nonexpansive. However, by, Corollary 23.9, every firmly nonexpansive operator $R\colon \mathbb{R} \to \mathbb{R}$ is of the form $R = (\mathrm{Id} + A)^{-1}$, where $A\colon \mathbb{R} \to 2^{\mathbb{R}}$ is a maximally monotone operator. Since, by Corollary 22.23, the only maximally monotone operators in $\mathbb{R}$ are subdifferentials of functions in $\Gamma_0(\mathbb{R})$, we have $\varrho = (\mathrm{Id} + \partial\phi)^{-1} = \operatorname{Prox}_\phi$ for some $\phi \in \Gamma_0(\mathbb{R})$. ☐

Proposition 24.32 *Let $\phi \in \Gamma_0(\mathbb{R})$. Suppose that 0 is a minimizer of ϕ, as is the case when ϕ is even. Then*

$$(\forall \xi \in \mathbb{R}) \quad \begin{cases} 0 \leqslant \operatorname{Prox}_\phi \xi \leqslant \xi, & \text{if } \xi > 0;\\ \operatorname{Prox}_\phi \xi = 0, & \text{if } \xi = 0;\\ \xi \leqslant \operatorname{Prox}_\phi \xi \leqslant 0, & \text{if } \xi < 0, \end{cases} \tag{24.58}$$

that is, $\operatorname{Prox}_\phi \xi \in \operatorname{conv}\{0, \xi\}$.

Proof. Since $0 \in \operatorname{Argmin} \phi$, Proposition 12.29 yields $\operatorname{Prox}_\phi 0 = 0$. In turn, since Prox_ϕ is nonexpansive by Proposition 12.28, we have $(\forall \xi \in \mathbb{R})$ $|\operatorname{Prox}_\phi \xi| = |\operatorname{Prox}_\phi \xi - \operatorname{Prox}_\phi 0| \leqslant |\xi - 0| = |\xi|$. Altogether, since Proposition 24.31 asserts that Prox_ϕ is increasing, we obtain (24.58). The last assertion follows from Proposition 11.7(i). ☐

Proposition 24.33 *Let $\psi \in \Gamma_0(\mathbb{R})$, and let ρ and θ be real numbers in $\mathbb{R}_{++}$ such that the following hold:*

(i) $\psi \geqslant \psi(0) = 0$.
(ii) *ψ is differentiable at 0.*
(iii) *ψ is twice differentiable on $[-\rho, \rho] \smallsetminus \{0\}$ and $\inf_{0<|\xi|\leqslant\rho} \psi''(\xi) \geqslant \theta$.*

Then $(\forall \xi \in [-\rho, \rho])(\forall \eta \in [-\rho, \rho])$ $|\operatorname{Prox}_\psi \xi - \operatorname{Prox}_\psi \eta| \leqslant |\xi - \eta|/(1 + \theta)$.

Proof. Set $R = [-\rho, \rho] \smallsetminus \{0\}$ and $\varphi \colon R \to \mathbb{R} \colon \zeta \mapsto \zeta + \psi'(\zeta)$. We first infer from (iii) that

$$(\forall \zeta \in R) \quad \varphi'(\zeta) = 1 + \psi''(\zeta) \geqslant 1 + \theta. \tag{24.59}$$

Moreover, (24.2) yields $(\forall \zeta \in R)$ $\operatorname{Prox}_\psi \zeta = \varphi^{-1}(\zeta)$. Note also that, in the light of (24.2), (ii), and (i), we have $(\forall \zeta \in \mathbb{R})$ $\operatorname{Prox}_\psi \zeta = 0 \Leftrightarrow \zeta \in \partial\psi(0) = \{\psi'(0)\} = \{0\}$. Hence, Prox_ψ vanishes only at 0 and we derive from Proposition 12.28 that

$$(\forall \zeta \in R) \quad 0 < |\varphi^{-1}(\zeta)| = |\operatorname{Prox}_\psi \zeta - \operatorname{Prox}_\psi 0| \leqslant |\zeta - 0| \leqslant \rho. \tag{24.60}$$

In turn, we deduce from (24.59) that

$$\sup_{\zeta \in R} \operatorname{Prox}'_\psi \zeta = \frac{1}{\inf\limits_{\zeta \in R} \varphi'\big(\varphi^{-1}(\zeta)\big)} \leqslant \frac{1}{\inf\limits_{\zeta \in R} \varphi'(\zeta)} \leqslant \frac{1}{1+\theta}. \tag{24.61}$$

Now fix ξ and η in R. First, let us assume that either $\xi < \eta < 0$ or $0 < \xi < \eta$. Then, since Prox_ψ is increasing by Proposition 24.31, it follows from the mean value theorem and (24.61) that there exists $\mu \in]\xi, \eta[$ such that

$$0 \leqslant \operatorname{Prox}_\psi \eta - \operatorname{Prox}_\psi \xi = (\eta - \xi) \operatorname{Prox}'_\psi \mu \leqslant \frac{\eta - \xi}{1+\theta}. \tag{24.62}$$

Next, let us assume that $\xi < 0 < \eta$. Then the mean value theorem asserts that there exist $\mu \in]\xi, 0[$ and $\nu \in]0, \eta[$ such that

$$\begin{cases} \operatorname{Prox}_\psi 0 - \operatorname{Prox}_\psi \xi = -\xi \operatorname{Prox}'_\psi \mu, \\ \operatorname{Prox}_\psi \eta - \operatorname{Prox}_\psi 0 = \eta \operatorname{Prox}'_\psi \nu. \end{cases} \tag{24.63}$$

Since Prox_ψ is increasing and $\operatorname{Prox}_\psi 0 = 0$, we obtain

$$\begin{aligned} 0 &\leqslant \operatorname{Prox}_\psi \eta - \operatorname{Prox}_\psi \xi \\ &= \eta \operatorname{Prox}'_\psi \nu - \xi \operatorname{Prox}'_\psi \mu \\ &\leqslant (\eta - \xi) \sup_{\zeta \in R} \operatorname{Prox}'_\psi \zeta \\ &\leqslant \frac{\eta - \xi}{1+\theta}. \end{aligned} \tag{24.64}$$

Altogether, we have shown that, for every ξ and η in R, $|\operatorname{Prox}_\psi \xi - \operatorname{Prox}_\psi \eta| \leqslant |\xi - \eta|/(1+\theta)$. We conclude by observing that, due to the continuity of Prox_ψ, this inequality holds for every ξ and η in $[-\rho, \rho]$. ☐

Next, we provide examples of proximity operators on the real line. The first one is instrumental is the area of compressive sensing.

Example 24.34 Let Ω be a nonempty closed interval in $\mathbb{R}$, set $\underline{\omega} = \inf \Omega$, set $\overline{\omega} = \sup \Omega$, and let $\xi \in \mathbb{R}$. Then the following hold:

(i) $\operatorname{Prox}_{\iota_\Omega} \xi = P_\Omega \xi = \begin{cases} \underline{\omega}, & \text{if } \xi < \underline{\omega}; \\ \xi, & \text{if } \xi \in \Omega; \\ \overline{\omega}, & \text{if } \xi > \overline{\omega}. \end{cases}$

(ii) $\operatorname{Prox}_{\sigma_\Omega} \xi = \operatorname{soft}_\Omega \xi$, where

$$\operatorname{soft}_\Omega \colon \mathbb{R} \to \mathbb{R} \colon \eta \mapsto \begin{cases} \eta - \underline{\omega}, & \text{if } \eta < \underline{\omega}; \\ 0, & \text{if } \eta \in \Omega; \\ \eta - \overline{\omega}, & \text{if } \eta > \overline{\omega} \end{cases} \tag{24.65}$$

is the soft thresholder on Ω.

(iii) Let $\omega \in \mathbb{R}_{++}$. Then

$$\operatorname{Prox}_{\omega|\cdot|} \xi = \operatorname{sign}(\xi) \max\big\{|\xi| - \omega, 0\big\} = \begin{cases} \xi + \omega, & \text{if } \xi < -\omega; \\ 0, & \text{if } |\xi| \leqslant \omega; \\ \xi - \omega, & \text{if } \xi > \omega. \end{cases} \tag{24.66}$$

Proof. (i): Clear.

(ii): Since $\sigma_\Omega^* = \iota_\Omega$, this follows from (i) and (24.4).

(iii): Apply (ii) with $\Omega = [-\omega, \omega]$. ☐

Example 24.35 Let $\phi\colon \mathbb{R} \to \mathbb{R}\colon \eta \mapsto \max\{0, \eta\}$ be the positive part function, let $\xi \in \mathbb{R}$, and let $\gamma \in \mathbb{R}_{++}$. Then it follows from Example 7.9 that $\gamma\phi = \sigma_{[0,\gamma]}$, and we therefore derive from Example 24.34(ii) that

$$\operatorname{Prox}_{\gamma\phi} \xi = \begin{cases} \xi, & \text{if } \xi < 0; \\ 0, & \text{if } 0 \leqslant \xi \leqslant \gamma; \\ \xi - \gamma, & \text{if } \xi > \gamma. \end{cases} \tag{24.67}$$

Alternatively, note that $\phi = d_C$, where $C = \left]-\infty, 0\right]$, and use (24.54).

The following example is of interest in the area of support vector machines.

Example 24.36 Let $\phi\colon \mathbb{R} \to \mathbb{R}\colon \eta \mapsto \max\{0, 1 - \eta\}$ be the *hinge loss* function, let $\gamma \in \mathbb{R}_{++}$, and let $\xi \in \mathbb{R}$. Then $\phi = d_C$, where $C = \left[1, +\infty\right[$, and we therefore derive from (24.54) that

$$\operatorname{Prox}_{\gamma\phi} \xi = \begin{cases} \xi + \gamma, & \text{if } \xi < 1 - \gamma; \\ 1, & \text{if } 1 - \gamma \leqslant \xi \leqslant 1; \\ \xi, & \text{if } \xi > 1. \end{cases} \tag{24.68}$$

Alternatively, observe that $\gamma\phi = \tau_1 \sigma_\Omega$, where $\Omega = [-\gamma, 0]$ and $\tau_1\colon \eta \mapsto \eta - 1$. Hence, (24.68) follows from Proposition 24.8(ii) and Example 24.34(ii).

As seen in Section 24.2, proximity operators of functions in $\Gamma_0(\mathbb{R})$ are basic building blocks in proximity operators of more complicated functions. Here is an illustration.

Example 24.37 Suppose that $u \in \mathcal{H} \smallsetminus \{0\}$ and set

$$f\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \max\{0, 1 - \langle x \mid u\rangle\}. \tag{24.69}$$

Then

$$\operatorname{Prox}_f\colon x \mapsto \begin{cases} x + u, & \text{if } \ \langle x \mid u\rangle < 1 - \|u\|^2;\\ x + \dfrac{1 - \langle x \mid u\rangle}{\|u\|^2} u, & \text{if } \ 1 - \|u\|^2 \leqslant \langle x \mid u\rangle \leqslant 1;\\ x, & \text{if } \ \langle x \mid u\rangle > 1. \end{cases} \tag{24.70}$$

Proof. Let $x \in \mathcal{H}$ and set $\phi\colon \xi \mapsto \max\{0, 1 - \xi\}$. Combining Example 24.36 and Corollary 24.15 yields

$$\begin{aligned} \operatorname{Prox}_f x &= x + \frac{\operatorname{Prox}_{\|u\|^2\phi} \langle x \mid u\rangle - \langle x \mid u\rangle}{\|u\|^2} u \\ &= x + \frac{u}{\|u\|^2} \cdot \begin{cases} \|u\|^2, & \text{if } \ \langle x \mid u\rangle < 1 - \|u\|^2;\\ 1 - \langle x \mid u\rangle, & \text{if } \ 1 - \|u\|^2 \leqslant \langle x \mid u\rangle \leqslant 1;\\ 0, & \text{if } \ \langle x \mid u\rangle > 1, \end{cases} \end{aligned} \tag{24.71}$$

and we obtain (24.70). ☐

Example 24.38 Let $p \in \left]1, +\infty\right[$, let $\gamma \in \mathbb{R}_{++}$, let $\phi\colon \mathbb{R} \to \mathbb{R}\colon \eta \mapsto |\eta|^p$, and let $\xi \in \mathbb{R}$. Then $\operatorname{Prox}_{\gamma\phi} \xi = \operatorname{sign}(\xi)\varrho$, where ϱ is the unique solution in $\mathbb{R}_+$ to

$$\varrho + p\gamma\varrho^{p-1} = |\xi|. \tag{24.72}$$

In particular, the following hold:

(i) Suppose that $p = 4/3$ and set $\rho = \sqrt{\xi^2 + 256\gamma^3/729}$. Then

$$\operatorname{Prox}_{\gamma\phi} \xi = \xi + \frac{4\gamma}{3\sqrt[3]{2}}\Big(\sqrt[3]{\rho - \xi} - \sqrt[3]{\rho + \xi}\Big).$$

(ii) Suppose that $p = 3/2$. Then

$$\operatorname{Prox}_{\gamma\phi} \xi = \xi + \frac{9\gamma^2}{8} \operatorname{sign}(\xi)\Big(1 - \sqrt{1 + 16|\xi|/(9\gamma^2)}\Big).$$

(iii) Suppose that $p = 2$. Then $\operatorname{Prox}_{\gamma\phi} \xi = \xi/(1 + 2\gamma)$.

(iv) Suppose that $p = 3$. Then $\operatorname{Prox}_{\gamma\phi} \xi = \operatorname{sign}(\xi)\big(\sqrt{1 + 12\gamma|\xi|} - 1\big)/(6\gamma)$.

(v) Suppose that $p = 4$ and set $\rho = \sqrt{\xi^2 + 1/(27\gamma)}$. Then

$$\operatorname{Prox}_{\gamma\phi}\xi = \sqrt[3]{\frac{\rho+\xi}{8\gamma}} - \sqrt[3]{\frac{\rho-\xi}{8\gamma}}.$$

Proof. The main conclusion follows from Proposition 24.1, and the special cases follow by solving the polynomial equation (24.72) in $\mathbb{R}_+$ and using the duality relation (24.4) via Example 13.2(i). □

Example 24.39 Consider the negative Boltzmann–Shannon entropy

$$\phi\colon \mathbb{R} \to \left]-\infty,+\infty\right] : \xi \mapsto \begin{cases} \xi\ln(\xi)-\xi, & \text{if } \xi>0;\\ 0, & \text{if } \xi=0;\\ +\infty, & \text{if } \xi<0. \end{cases} \tag{24.73}$$

It follows from Example 13.2(v) that $\phi^* = \exp$. Hence, Proposition 24.1 and (24.4) yield $(\forall\xi\in\mathbb{R})$ $\operatorname{Prox}_\phi\xi = W(e^\xi)$ and $\operatorname{Prox}_{\exp}\xi = \xi - W(e^\xi)$, where W denotes the Lambert W-function.

Example 24.40 Let $\gamma\in\mathbb{R}_{++}$ and consider the negative Burg entropy

$$\phi\colon \mathbb{R} \to \left]-\infty,+\infty\right] : \xi \mapsto \begin{cases} -\ln\xi, & \text{if } \xi>0;\\ +\infty, & \text{if } \xi\leqslant 0. \end{cases} \tag{24.74}$$

Then Proposition 24.1 yields $(\forall\xi\in\mathbb{R})$ $\operatorname{Prox}_{\gamma\phi}\xi = (\xi+\sqrt{\xi^2+4\gamma})/2$.

Example 24.41 Let $\underline{\omega}\in\mathbb{R}_{--}$, let $\overline{\omega}\in\mathbb{R}_{++}$, and set

$$\phi\colon \mathbb{R} \to \left]-\infty,+\infty\right] : \xi \mapsto \begin{cases} -\ln(\xi-\underline{\omega})+\ln(-\underline{\omega}), & \text{if } \xi\in\left]\underline{\omega},0\right];\\ -\ln(\overline{\omega}-\xi)+\ln(\overline{\omega}), & \text{if } \xi\in\left]0,\overline{\omega}\right[;\\ +\infty, & \text{otherwise.} \end{cases} \tag{24.75}$$

Then, for every $\xi\in\mathbb{R}$,

$$\operatorname{Prox}_\phi\xi = \begin{cases} \dfrac{\xi+\underline{\omega}+\sqrt{|\xi-\underline{\omega}|^2+4}}{2}, & \text{if } \xi<1/\underline{\omega};\\ \dfrac{\xi+\overline{\omega}-\sqrt{|\xi-\overline{\omega}|^2+4}}{2}, & \text{if } \xi>1/\overline{\omega};\\ 0, & \text{otherwise.} \end{cases} \tag{24.76}$$

Proof. Let $\xi\in\mathbb{R}$ and set $\pi = \operatorname{Prox}_\phi\xi$. We have $\partial\phi(0) = [1/\underline{\omega}, 1/\overline{\omega}]$. Therefore (24.2) yields

$$\pi = 0 \quad\Leftrightarrow\quad \xi\in[1/\underline{\omega},1/\overline{\omega}]. \tag{24.77}$$

Now consider the case when $\xi>1/\overline{\omega}$. Since $0\in\operatorname{Argmin}\phi$, it follows from Proposition 24.32 and (24.77) that $\pi\in\left]0,\xi\right]$. Hence, we derive from (24.2) that π is the only solution in $\left]0,\xi\right]$ to $\pi+1/(\overline{\omega}-\pi)=\xi$, i.e., $\pi = (\xi+\overline{\omega}-$

$\sqrt{|\xi-\overline{\omega}|^2+4})/2$. Likewise, if $\xi < 1/\underline{\omega}$, then π is the only solution in $[\xi, 0[$ to $\pi - 1/(\pi-\underline{\omega}) = \xi$, which yields $\pi = (\xi + \underline{\omega} + \sqrt{|\xi-\underline{\omega}|^2+4})/2$. ☐

Example 24.42 Let $\omega \in \mathbb{R}_{++}$ and set

$$\phi\colon \mathbb{R} \to \left]-\infty,+\infty\right] : \xi \mapsto \omega|\xi| - \ln\big(1+\omega|\xi|\big). \tag{24.78}$$

Then, for every $\xi \in \mathbb{R}$,

$$\operatorname{Prox}_\phi \xi = \operatorname{sign}(\xi)\frac{\omega|\xi| - \omega^2 - 1 + \sqrt{\big|\omega|\xi| - \omega^2 - 1\big|^2 + 4\omega|\xi|}}{2\omega}. \tag{24.79}$$

Proof. In view of Proposition 24.10, since ϕ is even, we focus on the case when $\xi \geqslant 0$. As $\operatorname{argmin}\phi = 0$, using Proposition 11.7(i), Proposition 24.32 yields $\pi = \operatorname{Prox}_\phi \xi \geqslant 0$. We deduce from Proposition 24.1 that π is the unique solution in $\mathbb{R}_+$ to the equation $\omega\pi^2 + (\omega^2 + 1 - \omega\xi)\pi - \xi = 0$, which leads to (24.79). ☐

The proximity operators in the following examples are sigmoid functions and they arise in particular in neural networks. These formulas follow from Proposition 24.1 and Proposition 24.32 (alternatively, see Corollary 24.5).

Example 24.43 Set

$$\phi\colon \mathbb{R} \to \left]-\infty,+\infty\right] : \xi \mapsto \begin{cases} -\sqrt{1-\xi^2} - \frac{1}{2}\xi^2, & \text{if } |\xi| \leqslant 1;\\ +\infty, & \text{if } |\xi| > 1. \end{cases} \tag{24.80}$$

Then $(\forall \xi \in \mathbb{R})$ $\operatorname{Prox}_\phi \xi = \xi/\sqrt{1+\xi^2}$.

Example 24.44 Set

$$\begin{aligned}\phi\colon &\mathbb{R} \to \left]-\infty,+\infty\right]\\ &\xi \mapsto \begin{cases} \dfrac{(1+\xi)\ln(1+\xi) + (1-\xi)\ln(1-\xi) - \xi^2}{2}, & \text{if } |\xi| < 1;\\ \ln(2) - 1/2, & \text{if } |\xi| = 1;\\ +\infty, & \text{if } |\xi| > 1. \end{cases}\end{aligned} \tag{24.81}$$

Then $\operatorname{Prox}_\phi = \tanh$.

Example 24.45 Set

$$\begin{aligned}\phi\colon &\mathbb{R} \to \left]-\infty,+\infty\right]\\ &\xi \mapsto \begin{cases} -\frac{2}{\pi}\ln\left(\cos\dfrac{\pi}{2}\xi\right) - \frac{1}{2}\xi^2, & \text{if } |\xi| < 1;\\ +\infty, & \text{if } |\xi| \geqslant 1. \end{cases}\end{aligned} \tag{24.82}$$

Then $\operatorname{Prox}_\phi = (2/\pi)\arctan$.

Example 24.46 Set

$$\phi\colon \mathbb{R} \to \left]-\infty,+\infty\right]$$
$$\xi \mapsto \begin{cases} \xi \ln \xi + (1-\xi)\ln(1-\xi) - \frac{1}{2}\xi^2, & \text{if } 0<\xi<1; \\ 0, & \text{if } \xi \in \{0,1\}; \\ +\infty, & \text{otherwise.} \end{cases} \tag{24.83}$$

Then Prox_ϕ is the *logistic cumulative function*, i.e., $(\forall \xi \in \mathbb{R})$ $\operatorname{Prox}_\phi \xi = 1/(1+\exp(-\xi))$.

Further proximity operators can be obtained via the following sum rule.

Proposition 24.47 *Let $\phi \in \Gamma_0(\mathbb{R})$ and let Ω be a closed interval of $\mathbb{R}$ such that $\Omega \cap \operatorname{dom}\phi \neq \varnothing$. Then* $\operatorname{Prox}_{\iota_\Omega+\phi} = P_\Omega \circ \operatorname{Prox}_\phi$.

Proof. Fix $\xi \in \mathbb{R}$, and set $\psi\colon \eta \mapsto \phi(\eta) + (1/2)|\xi-\eta|^2$ and $\pi = \operatorname{Prox}_\phi \xi$. In view of Example 24.34(i), we must show that

$$\operatorname{Prox}_{\phi+\iota_\Omega} \xi = \begin{cases} \inf \Omega, & \text{if } \pi < \inf \Omega; \\ \pi, & \text{if } \pi \in \Omega; \\ \sup \Omega, & \text{if } \pi > \sup \Omega. \end{cases} \tag{24.84}$$

By (24.1), π minimizes ψ. Therefore, if $\pi \in \Omega$, it also minimizes $\iota_\Omega + \psi$, hence $\pi = \operatorname{Prox}_{\iota_\Omega+\phi}\xi$. Now suppose that $\pi < \inf\Omega$. Since ψ is strictly convex and has π as its unique minimizer, it increases strictly over the interval $\left[\pi,+\infty\right[\cap \operatorname{dom}\phi$. Since $\Omega \cap \operatorname{dom}\psi \neq \varnothing$, $\inf\Omega \in \Omega\cap\operatorname{dom}\psi$ and $\inf\Omega$ therefore minimizes $\iota_\Omega + \psi$, which shows that $\operatorname{Prox}_{\iota_\Omega+\phi}\xi = \inf\Omega$. The case when $\pi > \sup\Omega$ is treated analogously. ☐

24.5 Proximal Thresholding

In this section, we provide a detailed analysis of thresholding operators in the context of proximity operators.

Definition 24.48 Let Ω be a nonempty closed convex subset of $\mathcal{H}$ and let $T\colon \mathcal{H} \to \mathcal{H}$. Then T is a *proximal thresholder* on Ω if there exists a function $f \in \Gamma_0(\mathcal{H})$ such that

$$T = \operatorname{Prox}_f \quad \text{and} \quad \operatorname{zer} T = \Omega. \tag{24.85}$$

The next proposition provides characterizations of proximal thresholders.

Proposition 24.49 *Let Ω be a nonempty closed convex subset of $\mathcal{H}$ and let $f \in \Gamma_0(\mathcal{H})$. Then the following are equivalent:*

(i) Prox_f *is a proximal thresholder on* Ω.
(ii) $\partial f(0) = \Omega$.
(iii) $\operatorname{Fix}\operatorname{Prox}_{f^*} = \Omega$.
(iv) $\operatorname{Argmin} f^* = \Omega$.

Moreover, (i)–(iv) hold when

(v) $f = g + \sigma_\Omega$, *where the function* $g \in \Gamma_0(\mathcal{H})$ *is Gâteaux differentiable at* 0 *and* $\nabla g(0) = 0$.

Proof. The equivalences of (i)–(iv) are obtained by applying Proposition 17.5 to f^* and invoking Corollary 13.38.

(v)⇒(ii): Since (v) implies that $0 \in \operatorname{core}\operatorname{dom} g$, we have $0 \in (\operatorname{core}\operatorname{dom} g) \cap \operatorname{dom}\sigma_\Omega$ and it follows from Corollary 16.48 and Example 16.34 that

$$\partial f(0) = \partial(g + \sigma_\Omega)(0) = \partial g(0) + \partial\sigma_\Omega(0) = \partial g(0) + \Omega. \tag{24.86}$$

However, since $\partial g(0) = \{\nabla g(0)\} = \{0\}$, we obtain $\partial f(0) = \Omega$, and (ii) is therefore satisfied. ☐

Example 24.50 Suppose that $\mathcal{H} = \mathbb{R}^N$, set $I = \{1, \ldots, N\}$, and let $f \in \Gamma_0(\mathcal{H})$ be the negative Burg entropy, i.e.,

$$f\colon \mathcal{H} \to \left]-\infty, +\infty\right] : x = (\xi_i)_{i \in I} \mapsto \begin{cases} -\displaystyle\sum_{i \in I} \ln \xi_i, & \text{if } x \in \mathbb{R}^N_{++}; \\ +\infty, & \text{otherwise.} \end{cases} \tag{24.87}$$

Let $x \in \mathcal{H}$ and let $\mu \in \mathbb{R}_{++}$. Then

$$\operatorname{Prox}_{f + \mu\|\cdot\|_1} x = \left(\frac{\xi_i - \mu + \sqrt{\xi_i^2 - 2\mu\xi_i + \mu^2 + 4}}{2} \right)_{i \in I}. \tag{24.88}$$

Note that $\mu\|\cdot\|_1 = \sigma_\Omega$, where $\Omega = [-\mu, \mu]^N$. However, $\operatorname{Prox}_{f+\mu\|\cdot\|_1}$ is not a proximal thresholder on Ω (compare with Proposition 24.49(v)).

Proof. This follows from Proposition 24.17, Proposition 24.12(ii), and Example 24.40. ☐

When $\phi = \rho|\cdot|$, the following example reduces to Example 14.5.

Example 24.51 Let $\phi\colon \mathbb{R} \to \mathbb{R}$ be an even convex function. Suppose that ϕ is differentiable on $\mathbb{R} \smallsetminus \{0\}$, and set $f = \phi \circ \|\cdot\|$. Then

$$\operatorname{Prox}_f\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \begin{cases} x - \dfrac{\operatorname{Prox}_{\phi^*}\|x\|}{\|x\|} x, & \text{if } \|x\| > \max \partial\phi(0); \\ 0, & \text{if } \|x\| \leqslant \max \partial\phi(0). \end{cases} \tag{24.89}$$

If $\mathcal{H} = \mathbb{R}$, then Prox_f is a proximal thresholder on $[-\phi'(0;1), \phi'(0;1)]$.

Proof. It follows from Example 17.37, Theorem 17.18, and Proposition 24.49 that Prox_f is a proximal thresholder on $B(0; \max \partial\phi(0)) = [-\phi'(0;1), \phi'(0;1)]$. Finally, (24.89) is obtained by setting $C = \{0\}$ in Proposition 24.27. ☐

The following theorem is a significant refinement of a result of Proposition 24.49 in the case when $\mathcal{H} = \mathbb{R}$. It characterizes those functions in $\Gamma_0(\mathbb{R})$ the proximity operator of which is a proximal thresholder.

Theorem 24.52 *Let Ω be a nonempty closed interval in $\mathbb{R}$ and let $\phi \in \Gamma_0(\mathbb{R})$. Then the following are equivalent:*

(i) Prox_ϕ *is a proximal thresholder on Ω.*
(ii) $\phi = \psi + \sigma_\Omega$, *where $\psi \in \Gamma_0(\mathbb{R})$ is differentiable at 0 and $\psi'(0) = 0$.*

Proof. In view of Proposition 24.49, it is enough to show that $\partial\phi(0) = \Omega \Rightarrow$ (ii). To this end, assume that $\partial\phi(0) = \Omega$, and set $\underline{\omega} = \inf \Omega = \phi'_-(0)$ and $\overline{\omega} = \sup \Omega = \phi'_+(0)$ (see Proposition 17.16(ii)). Let us now identify the function ψ of (ii) in each of the four conceivable cases.

(a) $\underline{\omega} = -\infty$ and $\overline{\omega} = +\infty$: Then $\mathrm{dom}\,\phi = \{0\}$, $\sigma_\Omega = \iota_{\{0\}}$, and we set $\psi \equiv \phi(0)$.

(b) $\underline{\omega} = -\infty$ and $\overline{\omega} < +\infty$: Then $\{0\} \neq \mathrm{dom}\,\phi \subset \mathbb{R}_+$, $\overline{\omega} = \phi'_+(0)$, and we set

$$\psi\colon \mathbb{R} \to \left]-\infty, +\infty\right] : \xi \mapsto \begin{cases} \phi(0), & \text{if } \xi \leqslant 0; \\ \phi(\xi) - \overline{\omega}\xi, & \text{if } \xi > 0. \end{cases} \tag{24.90}$$

Observe that

$$(\forall \xi \in \mathrm{dom}\,\psi) \quad \psi'_+(\xi) = \begin{cases} 0, & \text{if } \xi < 0; \\ \phi'_+(\xi) - \overline{\omega}, & \text{if } \xi \geqslant 0, \end{cases} \tag{24.91}$$

and

$$(\forall \xi \in \mathrm{dom}\,\psi) \quad \psi'_-(\xi) = \begin{cases} 0, & \text{if } \xi \leqslant 0; \\ \phi'_-(\xi) - \overline{\omega}, & \text{if } \xi > 0. \end{cases} \tag{24.92}$$

Since $\overline{\omega} = \phi'_+(0)$, it follows that $\psi'_+(0) = \psi'_-(0) = 0$. Thus, $\psi'(0) = 0$. Hence ψ is lower semicontinuous and, by Corollary 17.30, convex.

(c) $\underline{\omega} > -\infty$ and $\overline{\omega} = +\infty$: This case is similar to (b). We set

$$\psi\colon \mathbb{R} \to \left]-\infty, +\infty\right] : \xi \mapsto \begin{cases} \phi(0), & \text{if } \xi \geqslant 0; \\ \phi(\xi) - \underline{\omega}\xi, & \text{if } \xi > 0. \end{cases} \tag{24.93}$$

(d) $\underline{\omega} > -\infty$ and $\overline{\omega} < +\infty$: Then $0 \in \mathrm{int\,dom}\,\phi$, and we set

$$\psi\colon \mathbb{R} \to \left]-\infty, +\infty\right] : \xi \mapsto \begin{cases} \phi(\xi) - \underline{\omega}\xi, & \text{if } \xi < 0; \\ \phi(0), & \text{if } \xi = 0; \\ \phi(\xi) - \overline{\omega}\xi, & \text{if } \xi > 0. \end{cases} \tag{24.94}$$

Observe that

$$(\forall \xi \in \operatorname{dom} \psi) \quad \psi'_+(\xi) = \begin{cases} \phi'_+(\xi) - \underline{\omega}, & \text{if } \xi < 0; \\ \phi'_+(0) - \overline{\omega} = 0, & \text{if } \xi = 0; \\ \phi'_+(\xi) - \overline{\omega}, & \text{if } \xi > 0, \end{cases} \tag{24.95}$$

and

$$(\forall \xi \in \operatorname{dom} \psi) \quad \psi'_-(\xi) = \begin{cases} \phi'_-(\xi) - \underline{\omega}, & \text{if } \xi < 0; \\ \phi'_-(0) - \underline{\omega} = 0, & \text{if } \xi = 0; \\ \phi'_-(\xi) - \overline{\omega}, & \text{if } \xi > 0. \end{cases} \tag{24.96}$$

Thus $\psi'_+(0) = \psi'_-(0) = 0$ and hence $\psi'(0) = 0$. Therefore ψ is lower semicontinuous, while its convexity follows from Corollary 17.30. ☐

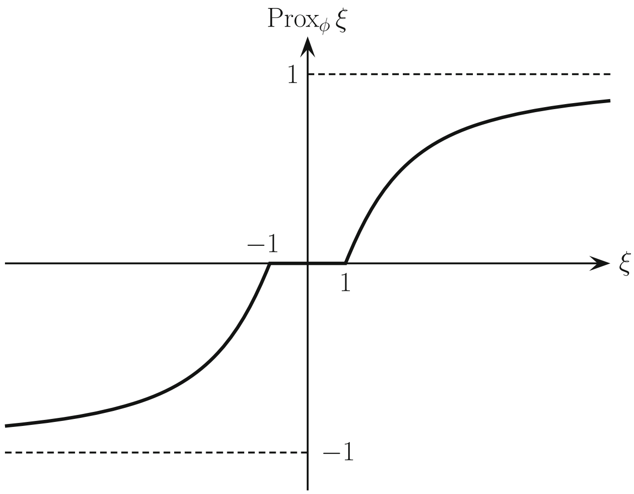

Fig. 24.1 Example of a proximal thresholder: the proximity operator of the function ϕ of (24.97) for $\omega = 1$.

Example 24.53 Let $\omega \in \mathbb{R}_{++}$ and set

$$\phi\colon \mathbb{R} \to \left]-\infty, +\infty\right] : \xi \mapsto \begin{cases} \ln(\omega) - \ln\big(\omega - |\xi|\big), & \text{if } |\xi| < \omega; \\ +\infty, & \text{otherwise.} \end{cases} \tag{24.97}$$

Furthermore, set

$$\psi\colon \mathbb{R} \to \left]-\infty, +\infty\right] : \xi \mapsto \begin{cases} \ln(\omega) - \ln\big(\omega - |\xi|\big) - |\xi|/\omega, & \text{if } |\xi| < \omega; \\ +\infty, & \text{otherwise} \end{cases} \tag{24.98}$$

and $\Omega = [-1/\omega, 1/\omega]$. Then $\psi \in \Gamma_0(\mathbb{R})$ is differentiable at 0, $\psi'(0) = 0$, and $\phi = \psi + \sigma_\Omega$. Therefore, Theorem 24.52 asserts that Prox_ϕ is a proximal thresholder on $[-1/\omega, 1/\omega]$. More explicitly, (see Figure 24.1), it follows from Example 24.41 that, for every $\xi \in \mathbb{R}$,

$$\operatorname{Prox}_\phi \xi = \begin{cases} \operatorname{sign}(\xi)\, \dfrac{|\xi| + \omega - \sqrt{\big||\xi| - \omega\big|^2 + 4}}{2}, & \text{if } |\xi| > 1/\omega; \\ 0, & \text{otherwise.} \end{cases} \tag{24.99}$$

Next, we provide a convenient decomposition rule for implementing proximal thresholders.

Proposition 24.54 *Let Ω be a nonempty closed interval in $\mathbb{R}$, let $\psi \in \Gamma_0(\mathbb{R})$ be differentiable at* 0 *and such that $\psi'(0) = 0$, and set $\phi = \psi + \sigma_\Omega$. Then* $\operatorname{Prox}_\phi = \operatorname{Prox}_\psi \circ \operatorname{soft}_\Omega$.

Proof. Set $A = \partial\psi$. Example 16.14 asserts that $0 \in \operatorname{dom} \partial\sigma_\Omega = \operatorname{dom} \partial\iota_C^* = \operatorname{dom}(\partial\iota_C)^{-1} = \operatorname{ran} \partial\iota_C = \operatorname{ran} N_C$. Furthermore, since ψ is differentiable at 0, we derive from Proposition 16.27 that $0 \in \operatorname{int} \operatorname{dom} \psi = \operatorname{int} \operatorname{dom} \partial\psi$. As a result, we have $\operatorname{cone}(\operatorname{dom} A - \operatorname{ran} N_C) = \overline{\operatorname{span}}\,(\operatorname{dom} A - \operatorname{ran} N_C)$. Now let $\xi \in \mathbb{R}$. If $\xi = 0$, then $0 + A0 = 0 + \{\psi'(0)\} = \{0\} = \mathbb{R}\, 0$. On the other hand, by monotonicity of A, if $\xi > 0$, then $\xi + A\xi \subset \mathbb{R}_{++} \subset \mathbb{R}_+ \xi$. Similarly, if $\xi < 0$, then $\xi + A\xi \subset \mathbb{R}_{--} \subset \mathbb{R}_- \xi$. Altogether, the result follows from Corollary 23.33 and Example 24.34(ii). ☐

Example 24.55 (Berhu function) Let $\rho \in \mathbb{R}_{++}$. The *reverse Huber*, or *Berhu*, function is obtained by swapping the roles played by $|\cdot|$ and $|\cdot|^2$ in (8.47) in the sense that the former is now used for small values and the latter for large ones, namely,

$$\phi\colon \mathbb{R} \to \mathbb{R}\colon \xi \mapsto \begin{cases} \dfrac{\xi^2 + \rho^2}{2\rho}, & \text{if } |\xi| > \rho; \\ |\xi|, & \text{if } |\xi| \leqslant \rho. \end{cases} \tag{24.100}$$

Now set $\Omega = [-1, 1]$, $C = [-\rho, \rho]$, and $\psi = d_C^2/(2\rho)$. Then $\phi = \psi + |\cdot| = \psi + \sigma_\Omega$, where $\psi = {}^{\rho}\iota_C$ (Example 12.21). Hence, $\psi \in \Gamma_0(\mathbb{R})$ is differentiable (Corollary 12.31) with $\psi'(0) = 0$. Thus Proposition 24.8(vii) yields

$$\operatorname{Prox}_\psi = \operatorname{Id} + \frac{1}{\rho + 1}\big(P_C - \operatorname{Id}\big) \tag{24.101}$$

and, since $\phi = \psi + \sigma_\Omega$, we derive from Proposition 24.54 and Example 24.34 that (see Figure 24.2)

$$\operatorname{Prox}_\phi = \operatorname{Prox}_\psi \circ \operatorname{soft}_\Omega \colon \mathbb{R} \to \mathbb{R}$$
$$\xi \mapsto \begin{cases} \dfrac{\rho\xi}{\rho+1}, & \text{if } |\xi| > \rho+1; \\ \xi - \operatorname{sign}(\xi), & \text{if } 1 < |\xi| \leqslant \rho+1; \\ 0, & \text{if } |\xi| \leqslant 1. \end{cases} \tag{24.102}$$

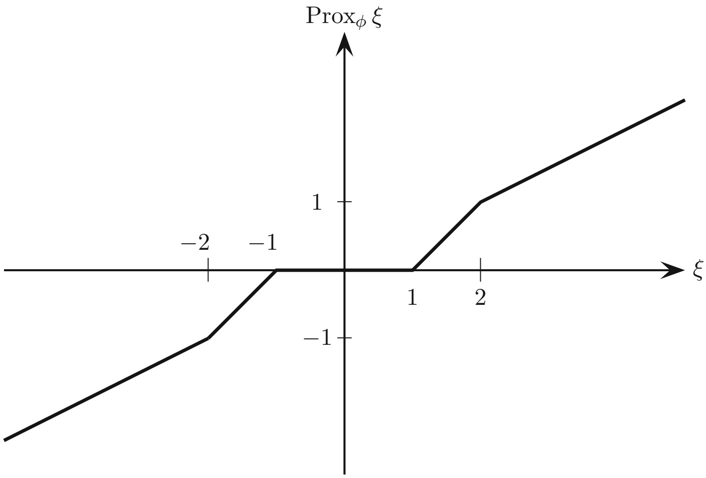

Fig. 24.2 Example of a proximal thresholder: the proximity operator of the Berhu function ϕ of (24.100) for $\rho = 1$.

24.6 Perspective Functions

Perspective functions were introduced in Proposition 8.25 and their lower semicontinuous convex envelopes in Proposition 9.42. We now determine the proximity operator of these envelopes.

Proposition 24.56 *Let $\varphi \in \Gamma_0(\mathcal{H})$ be such that* $\operatorname{dom} \varphi^* = \mathcal{H}$*, define*

$$f \colon \mathbb{R} \times \mathcal{H} \colon (\xi, x) \mapsto \begin{cases} \xi\varphi(x/\xi), & \text{if } \xi > 0; \\ (\operatorname{rec} \varphi)(x), & \text{if } \xi = 0; \\ +\infty, & \text{if } \xi < 0, \end{cases} \tag{24.103}$$

let $(\xi, x) \in \mathbb{R} \times \mathcal{H}$, and let $\gamma \in \mathbb{R}_{++}$. Then $f \in \Gamma_0(\mathbb{R} \oplus \mathcal{H})$ and the following hold:

(i) *Suppose that $\xi + \gamma\varphi^*(x/\gamma) \leqslant 0$. Then* $\operatorname{Prox}_{\gamma f}(\xi, x) = (0, 0)$.

(ii) *Suppose that* $\xi + \gamma\varphi^*(x/\gamma) > 0$. *Then*

$$\operatorname{Prox}_{\gamma f}(\xi, x) = \big(\xi + \gamma\varphi^*(p), x - \gamma p\big), \tag{24.104}$$

where p *is the unique solution to the inclusion*

$$x \in \gamma p + \big(\xi + \gamma\varphi^*(p)\big)\partial\varphi^*(p). \tag{24.105}$$

If φ^* *is differentiable at* p, *then* p *is the unique solution to the equation* $x = \gamma p + (\xi + \gamma\varphi^*(p))\nabla\varphi^*(p)$.

Proof. As we saw in Proposition 9.42, $f \in \Gamma_0(\mathbb{R} \oplus \mathcal{H})$ and f is the lower semicontinuous convex envelope of the perspective function (8.18) of φ. Hence, it follows from Proposition 13.16(iv) and Example 13.9 that

$$f^* = \iota_C, \quad \text{where} \quad C = \big\{(\mu, u) \in \mathbb{R} \times \mathcal{H} \;\big|\; \mu + \varphi^*(u) \leqslant 0\big\}. \tag{24.106}$$

Since $\varphi^* \in \Gamma_0(\mathcal{H})$, C is a nonempty closed convex set. It follows from Theorem 14.3(ii) that

$$\begin{aligned}\operatorname{Prox}_{\gamma f}(\xi, x) &= (\xi, x) - \gamma \operatorname{Prox}_{\gamma^{-1} f^*}\big(\xi/\gamma, x/\gamma\big)\\ &= (\xi, x) - \gamma P_C\big(\xi/\gamma, x/\gamma\big).\end{aligned} \tag{24.107}$$

Now set $(\pi, p) = P_C(\xi/\gamma, x/\gamma)$ and recall that, by Proposition 6.47, (π, p) is characterized by

$$(\xi/\gamma - \pi, x/\gamma - p) \in N_C(\pi, p). \tag{24.108}$$

(i): We have $(\xi/\gamma, x/\gamma) \in C$. Hence, $(\pi, p) = (\xi/\gamma, x/\gamma)$ and (24.107) yields $\operatorname{Prox}_{\gamma f}(\xi, x) = (0, 0)$.

(ii): Set $g\colon \mathbb{R} \times \mathcal{H} \to \left]-\infty, +\infty\right] : (\mu, u) \mapsto \mu + \varphi^*(u)$. Then $C = \operatorname{lev}_{\leqslant 0} g$. Now let $z \in \mathcal{H}$ and let $\zeta \in \left]-\infty, -\varphi^*(z)\right[$. Then $g(\zeta, z) < 0$. Therefore, we derive from Lemma 27.20 below and Proposition 16.9 that

$$N_C(\pi, p) = \begin{cases} \operatorname{cone} \partial g(\pi, p), & \text{if } \pi + \varphi^*(p) = 0;\\ \{(0,0)\}, & \text{if } \pi + \varphi^*(p) < 0 \end{cases} \tag{24.109}$$

$$= \begin{cases} \operatorname{cone}\big(\{1\} \times \partial\varphi^*(p)\big), & \text{if } \pi = -\varphi^*(p);\\ \{(0,0)\}, & \text{if } \pi < -\varphi^*(p). \end{cases} \tag{24.110}$$

Hence, if $\pi < -\varphi^*(p)$, then (24.108) yields $(\xi/\gamma, x/\gamma) = (\pi, p) \in C$, which is impossible since $(\xi/\gamma, x/\gamma) \notin C$. Thus, the characterization (24.108) becomes

$$\begin{cases} \pi = -\varphi^*(p),\\ (\exists\, \nu \in \mathbb{R}_{++})(\exists\, w \in \partial\varphi^*(p)) \quad \big(\xi + \gamma\varphi^*(p), x - \gamma p\big) = \nu(1, w), \end{cases} \tag{24.111}$$

which yields (24.105). For the last assertion, see Proposition 17.31(i). $\square$

Example 24.57 Consider the special case of Proposition 24.56 in which $\varphi = (1/2)\|\cdot\|^2$. Then (24.103) becomes

$$f\colon \mathbb{R} \oplus \mathcal{H} \to \left]-\infty, +\infty\right] : (\xi, x) \mapsto \begin{cases} \dfrac{\|x\|^2}{2\xi}, & \text{if } \xi > 0; \\ 0, & \text{if } x = 0 \text{ and } \xi = 0; \\ +\infty, & \text{otherwise,} \end{cases} \tag{24.112}$$

and, by Example 13.6, $\varphi^* = \varphi$. Let $\gamma \in \mathbb{R}_{++}$, and take $\xi \in \mathbb{R}$ and $x \in \mathcal{H}$. If $2\gamma\xi + \|x\|^2 \leqslant 0$, then Proposition 24.56(i) yields $\operatorname{Prox}_{\gamma f}(\xi, x) = (0, 0)$. Now suppose that $2\gamma\xi + \|x\|^2 > 0$. According to Proposition 24.56(ii), since $\nabla\varphi^* = \mathrm{Id}$, we need to find $p \in \mathcal{H}$ such that

$$x = \big(\gamma + \xi + \gamma\|p\|^2/2\big)p. \tag{24.113}$$

Note that, since (24.111) implies that $\xi + \gamma\varphi^*(p) > 0$, we have $\gamma + \xi + \gamma\|p\|^2/2 > 0$. Hence, taking the norm on both sides in (24.113) yields

$$\|x\| = \big(\gamma + \xi + \gamma\|p\|^2/2\big)\|p\|. \tag{24.114}$$

If $x = 0$, then $\operatorname{Prox}_{\gamma g}(\xi, x) = (\xi, 0)$. Otherwise, $\|p\|$ is the unique solution in $\mathbb{R}_{++}$ to the depressed cubic equation

$$s^3 + \frac{2(\xi + \gamma)}{\gamma}s - \frac{2\|x\|}{\gamma} = 0, \tag{24.115}$$

which can be obtained explicitly via Cardano's formulas. We then derive from Proposition 24.56(ii) and (24.113) that

$$\operatorname{Prox}_{\gamma f}(\xi, x) = \left(\xi + \frac{\gamma\|p\|^2}{2}, \left(1 - \frac{\gamma\|p\|}{\|x\|}\right)x\right). \tag{24.116}$$

24.7 Symmetric Functions of Symmetric Matrices

In this section, N is a strictly positive integer.

A function $\varphi\colon \mathbb{R}^N \to \left]-\infty, +\infty\right]$ is *symmetric* if $\varphi \circ P = \varphi$ for every permutation matrix $P \in \mathbb{R}^{N \times N}$. Given $x \in \mathbb{R}^N$, the vector $x_\downarrow \in \mathbb{R}^N$ has the same entries as x, with the same multiplicities, but ordered decreasingly. Given a subset S of $\mathbb{R}^N$, we use the notation $S_\downarrow = \big\{x_\downarrow \;\big|\; x \in S\big\}$.

Proposition 24.58 *Let $\varphi\colon \mathbb{R}^N \to \,]-\infty,+\infty]$ be proper and symmetric, and let x and y be in $\mathbb{R}^N$. Then the following hold:*

(i) *φ^* is symmetric.*
(ii) *The following are equivalent:*

  (a) *$y \in \partial\varphi(x)$.*
  (b) *$y_\downarrow \in \partial\varphi(x_\downarrow)$ and $\langle x \mid y\rangle = \langle x_\downarrow \mid y_\downarrow\rangle$.*
  (c) *$y_\downarrow \in \partial\varphi(x_\downarrow)$ and there exists a permutation matrix $P \in \mathbb{R}^{N\times N}$ such that $Px = x_\downarrow$ and $Py = y_\downarrow$.*

(iii) *$(\partial\varphi(x))_\downarrow \subset \partial\varphi(x_\downarrow)$.*
(iv) *φ is differentiable at x if and only if φ is differentiable at $x_\downarrow$, in which case $\nabla\varphi(x_\downarrow) = (\nabla\varphi(x))_\downarrow$.*

Proof. (i): Let $P \in \mathbb{R}^{N\times N}$ be a permutation matrix. Using Proposition 13.23(iv), we obtain $\varphi^* = (\varphi\circ P)^* = \varphi^*\circ P^{*-1} = \varphi^*\circ P$.

(ii): Using (i), Proposition 16.10, Fact 2.18, Proposition 13.15, and Fact 2.18, we obtain (a) $\Leftrightarrow$ $\varphi(x_\downarrow) + \varphi^*(y_\downarrow) = \varphi(x) + \varphi^*(y) = \langle x \mid y\rangle \leqslant \langle x_\downarrow \mid y_\downarrow\rangle \leqslant \varphi(x_\downarrow)+\varphi^*(y_\downarrow) \Leftrightarrow \varphi(x_\downarrow)+\varphi^*(y_\downarrow) = \langle x_\downarrow \mid y_\downarrow\rangle$ and $\langle x \mid y\rangle = \langle x_\downarrow \mid y_\downarrow\rangle$ $\Leftrightarrow$ (b) $\Leftrightarrow$ (c).

(iii): Clear from (ii).

(iv): Let $P \in \mathbb{R}^{N\times N}$ be a permutation matrix such that $Px = x_\downarrow$. By Corollary 16.53(i), $\partial\varphi(x) = \partial(\varphi\circ P)(x) = P^{\mathsf{T}}\partial\varphi(Px)$ and hence $P\partial\varphi(x) = \partial\varphi(x_\downarrow)$. The conclusion follows from Proposition 17.45 and (iii). ☐

In the remainder of this section, $\mathbb{S}^N$ is the Hilbert space of symmetric matrices of Example 2.5. We set $I = \{1,\ldots,N\}$ and let $\mathbb{U}^N = \big\{U \in \mathbb{R}^{N\times N} \;\big|\; UU^{\mathsf{T}} = \mathrm{Id}\big\}$ be the set of orthogonal matrices of size $N\times N$. Furthermore, $\mathrm{Diag}\colon \mathbb{R}^N \to \mathbb{S}^N$ maps a vector $(\xi_1,\ldots,\xi_N)$ to the diagonal matrix with diagonal elements $\xi_1, \ldots, \xi_N$. Given $A \in \mathbb{S}^N$, the vector formed by the N (not necessarily distinct) eigenvalues of A, ordered decreasingly, is denoted by $\lambda(A) = (\lambda_1(A),\ldots,\lambda_N(A))$; this defines a function

$$\lambda\colon \mathbb{S}^N \to \mathbb{R}^N\colon A \mapsto \big(\lambda_1(A),\ldots,\lambda_N(A)\big). \tag{24.117}$$

Let $A \in \mathbb{S}^N$, $U \in \mathbb{U}^N$, and let $x \in \mathbb{R}^N$. Then

$$\|UAU^{\mathsf{T}}\|_{\mathsf{F}} = \|A\|_{\mathsf{F}} = \sqrt{\langle A \mid A\rangle} = \sqrt{\langle \lambda(A) \mid \lambda(A)\rangle} = \|\lambda(A)\| \tag{24.118}$$

and

$$\|U(\mathrm{Diag}\,x)U^{\mathsf{T}}\|_{\mathsf{F}} = \|\mathrm{Diag}\,x\|_{\mathsf{F}} = \|x\|. \tag{24.119}$$

We start with a refinement of the Cauchy–Schwarz inequality in $\mathbb{S}^N$.

Fact 24.59 (Theobald) (See [344]) Let A and B be in $\mathbb{S}^N$. Then the following hold:

(i) $\langle A \mid B\rangle \leqslant \langle \lambda(A) \mid \lambda(B)\rangle \leqslant \|\lambda(A)\|\,\|\lambda(B)\| = \|A\|_{\mathsf{F}}\|B\|_{\mathsf{F}}$.

(ii) $\langle A \mid B\rangle = \langle \lambda(A) \mid \lambda(B)\rangle \Leftrightarrow (\exists\, U \in \mathbb{U}^N)\ A = U(\operatorname{Diag}\lambda(A))U^{\mathsf{T}}$ and $B = U(\operatorname{Diag}\lambda(B))U^{\mathsf{T}}$.

Proposition 24.60 (Lewis) *Let* $\varphi\colon \mathbb{R}^N \to\,]-\infty,+\infty]$ *be symmetric and let* λ *be as in* (24.117). *Then*

$$(\varphi \circ \lambda)^* = \varphi^* \circ \lambda. \tag{24.120}$$

Proof. Let $B \in \mathbb{S}^N$. On the one hand, using Fact 24.59(i),

$$\begin{aligned}(\varphi\circ\lambda)^*(B) &= \sup_{A\in\mathbb{S}^N} \big(\langle A \mid B\rangle - (\varphi\circ\lambda)(A)\big)\\ &\leqslant \sup_{A\in\mathbb{S}^N} \big(\langle \lambda(A) \mid \lambda(B)\rangle - \varphi\big(\lambda(A)\big)\big)\\ &\leqslant \sup_{x\in\mathbb{R}^N} \big(\langle x \mid \lambda(B)\rangle - \varphi(x)\big)\\ &= \varphi^*\big(\lambda(B)\big). \end{aligned}\tag{24.121}$$

On the other hand, there exists $V \in \mathbb{U}^N$ such that $B = V(\operatorname{Diag}\lambda(B))V^{\mathsf{T}}$. For every $x \in \mathbb{R}^N$, $\lambda(V(\operatorname{Diag}x)V^{\mathsf{T}}) = x_{\downarrow}$, which implies that

$$\varphi\big(\lambda(V(\operatorname{Diag}x)V^{\mathsf{T}})\big) = \varphi(x_{\downarrow}) = \varphi(x) \tag{24.122}$$

and $\langle x \mid \lambda(B)\rangle = \operatorname{tra}((\operatorname{Diag}x)(\operatorname{Diag}\lambda(B))) = \operatorname{tra}((\operatorname{Diag}x)(V^{\mathsf{T}}BV)) = \operatorname{tra}(V(\operatorname{Diag}x)V^{\mathsf{T}}B) = \big\langle V(\operatorname{Diag}x)V^{\mathsf{T}} \mid B\big\rangle$. Thus

$$\begin{aligned}\varphi^*\big(\lambda(B)\big) &= \sup_{x\in\mathbb{R}^N} \big(\langle x \mid \lambda(B)\rangle - \varphi(x)\big)\\ &= \sup_{x\in\mathbb{R}^N} \big(\big\langle V(\operatorname{Diag}x)V^{\mathsf{T}} \mid B\big\rangle - \varphi(\lambda(V(\operatorname{Diag}x)V^{\mathsf{T}}))\big)\\ &\leqslant \sup_{A\in\mathbb{S}^N} \big(\langle A \mid B\rangle - (\varphi\circ\lambda)(A)\big)\\ &= (\varphi\circ\lambda)^*(B). \end{aligned}\tag{24.123}$$

Combining (24.121) and (24.123), we obtain the result. ☐

Corollary 24.61 (Davis) *Let* $\varphi\colon \mathbb{R}^N \to\,]-\infty,+\infty]$ *be proper and symmetric. Then* $\varphi\circ\lambda \in \Gamma_0(\mathbb{S}^N)$ *if and only if* $\varphi \in \Gamma_0(\mathbb{R}^N)$.

Proof. By Proposition 24.58(i), φ^{**} is symmetric. Using Theorem 13.37 and Proposition 24.60, we obtain $\varphi\circ\lambda \in \Gamma_0(\mathbb{S}^N) \Leftrightarrow (\varphi\circ\lambda)^{**} = \varphi\circ\lambda \Leftrightarrow \varphi^{**}\circ\lambda = \varphi\circ\lambda \Leftrightarrow \varphi^{**}|_{\mathbb{R}^N_{\downarrow}} = \varphi|_{\mathbb{R}^N_{\downarrow}} \Leftrightarrow \varphi^{**} = \varphi \Leftrightarrow \varphi \in \Gamma_0(\mathbb{R}^N)$. ☐

Corollary 24.62 *Let* $\varphi \in \Gamma_0(\mathbb{R}^N)$ *be symmetric, let* $\gamma \in \mathbb{R}_{++}$, *and let* $A \in \mathbb{S}^N$. *Then* $\operatorname{Prox}_{\gamma\varphi\circ\lambda} A = A - \gamma \operatorname{Prox}_{\gamma^{-1}\varphi^*\circ\lambda}(\gamma^{-1}A)$.

Proof. By Corollary 24.61, $\varphi\circ\lambda \in \Gamma_0(\mathbb{S}^N)$. Now combine Theorem 14.3(ii) and Proposition 24.60. ☐

Proposition 24.63 (Lewis) *Let $\varphi \in \Gamma_0(\mathbb{R}^N)$ be symmetric, and let A and B be in $\mathbb{S}^N$. Then the following hold:*

(i) $(\varphi \circ \lambda)(A) + (\varphi \circ \lambda)^*(B) = \varphi(\lambda(A)) + \varphi^*(\lambda(B)) \geqslant \langle \lambda(A) \mid \lambda(B)\rangle \geqslant \langle A \mid B\rangle$.

(ii) $B \in \partial(\varphi \circ \lambda)(A) \Leftrightarrow \big[\, \lambda(B) \in \partial\varphi(\lambda(A))$ *and there exists* $U \in \mathbb{U}^N$ *such that* $A = U(\operatorname{Diag}\lambda(A))U^{\mathsf{T}}$ *and* $B = U(\operatorname{Diag}\lambda(B))U^{\mathsf{T}}\,\big]$.

(iii) $\partial(\varphi \circ \lambda)(A) = \big\{U(\operatorname{Diag} y)U^{\mathsf{T}} \;\big|\; U \in \mathbb{U}^N,\ y \in \partial\varphi(\lambda(A)),$ *and* $A = U(\operatorname{Diag}\lambda(A))U^{\mathsf{T}}\big\}$.

(iv) $\varphi \circ \lambda$ *is differentiable at* A *if and only if* φ *is differentiable at* $\lambda(A)$*, in which case, for every* $U \in \mathbb{U}^N$ *such that* $A = U(\operatorname{Diag}\lambda(A))U^{\mathsf{T}}$*, we have*

$$\nabla(\varphi \circ \lambda)(A) = U\big(\operatorname{Diag}\nabla\varphi(\lambda(A))\big)U^{\mathsf{T}}. \tag{24.124}$$

Proof. Corollary 24.61 yields $\varphi \circ \lambda \in \Gamma_0(\mathbb{S}^N)$.

(i): This follows from Proposition 24.60, Fenchel–Young (Proposition 13.15), and Fact 24.59(i).

(ii): In view of Proposition 16.10, (i), and Fact 24.59(ii), we obtain the equivalences $B \in \partial(\varphi \circ \lambda)(A) \Leftrightarrow (\varphi \circ \lambda)(A) + (\varphi \circ \lambda)^*(B) = \langle A \mid B\rangle \Leftrightarrow [\varphi(\lambda(A)) + \varphi^*(\lambda(B)) = \langle \lambda(A) \mid \lambda(B)\rangle$ and $\langle \lambda(A) \mid \lambda(B)\rangle = \langle A \mid B\rangle] \Leftrightarrow [\lambda(B) \in \partial\varphi(\lambda(A))$ and $(\exists U \in \mathbb{U}^N)$ $A = U(\operatorname{Diag}\lambda(A))U^{\mathsf{T}}$ and $B = U(\operatorname{Diag}\lambda(B))U^{\mathsf{T}}]$.

(iii): Denote the set on the right-hand side of (iii) by S. By (ii), $\partial(\varphi \circ \lambda)(A) \subset S$. Conversely, suppose that $B \in S$ and set $x = \lambda(A)$. Then $x_{\downarrow} = x$ and there exist $y \in \partial\varphi(x)$ and $U \in \mathbb{U}^N$ such that $A = U(\operatorname{Diag} x)U^{\mathsf{T}}$ and $B = U(\operatorname{Diag} y)U^{\mathsf{T}}$. By Proposition 24.58(ii), $\lambda(B) = y_{\downarrow} \in \partial\varphi(x_{\downarrow}) = \partial\varphi(x) = \partial\varphi(\lambda(A))$ and there exists a permutation matrix $P \in \mathbb{R}^{N\times N}$ such that $Px = x_{\downarrow} = x$ and $Py = y_{\downarrow}$. Hence $P(\operatorname{Diag} x)P^{\mathsf{T}} = \operatorname{Diag}(Px) = \operatorname{Diag}(x_{\downarrow}) = \operatorname{Diag} x = \operatorname{Diag}\lambda(A)$ and $P(\operatorname{Diag} y)P^{\mathsf{T}} = \operatorname{Diag}(Py) = \operatorname{Diag} y_{\downarrow} = \operatorname{Diag}\lambda(B)$. Now set $V = UP^{\mathsf{T}}$. Then $V \in \mathbb{U}^N$, $V(\operatorname{Diag}\lambda(A))V^{\mathsf{T}} = U(P^{\mathsf{T}}(\operatorname{Diag}\lambda(A))P)U^{\mathsf{T}} = U(\operatorname{Diag} x)U^{\mathsf{T}} = A$, and $V(\operatorname{Diag}\lambda(B))V^{\mathsf{T}} = U(P^{\mathsf{T}}(\operatorname{Diag}\lambda(B))P)U^{\mathsf{T}} = U(\operatorname{Diag} y)U^{\mathsf{T}} = B$. Therefore, $B \in \partial(\varphi \circ \lambda)(A)$ by (ii).

(iv): We use Proposition 17.45 repeatedly. First, consider the case when $\varphi \circ \lambda$ is differentiable at A, say $\partial(\varphi \circ \lambda)(A) = \{B\}$. Let $U \in \mathbb{U}^N$ be such that $A = U(\operatorname{Diag}\lambda(A))U$, and let $y \in \partial\varphi(\lambda(A))$. By (iii), $B = U(\operatorname{Diag} y)U^{\mathsf{T}}$. Therefore, $\|B\|_{\mathsf{F}}^2 = \langle B \mid B\rangle = \operatorname{tra}(BB) = \operatorname{tra}(U(\operatorname{Diag} y)U^{\mathsf{T}}U(\operatorname{Diag} y)U^{\mathsf{T}}) = \operatorname{tra}((\operatorname{Diag} y)(\operatorname{Diag} y)) = \|y\|^2$. It follows that $\|\partial\varphi(\lambda(A))\| = \{\|B\|_{\mathsf{F}}\}$. Hence, by Proposition 16.4(iii) and Proposition 3.7, $\partial\varphi(\lambda(A))$ is a singleton. Thus φ is differentiable at $\lambda(A)$. Conversely, assume that φ is differentiable at $x = \lambda(A)$. Set $y = \nabla\varphi(x)$ and let $B \in \partial(\varphi \circ \lambda)(A)$. By (iii), there exists $U \in \mathbb{U}^M$ such that $B = U(\operatorname{Diag} y)U^{\mathsf{T}}$. As before, $\|B\|_{\mathsf{F}} = \|y\|$ and hence $\|\partial(\varphi \circ \lambda)(A)\| = \{\|y\|\}$. By Proposition 16.4(iii) and Proposition 3.7, $\partial(\varphi \circ$

$\lambda)(A)$ is a singleton. Thus $\varphi \circ \lambda$ is differentiable at A. The identity (24.124) follows from (iii). ☐

Example 24.64 Let $k \in \{1, \dots, N\}$. Then $f: \mathbb{S}^N \to \mathbb{R}: A \mapsto \sum_{i=1}^{k} \lambda_i(A)$ is convex and Lipschitz continuous with constant $\sqrt{k}$.

Proof. Set $I = \{1, \dots, N\}$ and

$$\varphi: \mathbb{R}^N \to \mathbb{R}: (\xi_i)_{i \in I} \mapsto \max \left\{ \sum_{i \in J} \xi_i \;\middle|\; J \subset I,\ \operatorname{card} J = k \right\}. \tag{24.125}$$

Then $\varphi \in \Gamma_0(\mathbb{R}^N)$ and, by Corollary 24.61, $f = \varphi \circ \lambda$ is convex. Now denote the standard unit vectors of $\mathbb{R}^N$ by $(e_i)_{i \in I}$. By Theorem 18.5,

$$\operatorname{ran} \partial \varphi \subset \operatorname{conv} \left\{ \sum_{i \in J} e_i \;\middle|\; J \subset I,\ \operatorname{card} J = k \right\} \subset B\big(0; \sqrt{k}\,\big) \subset \mathbb{R}^N. \tag{24.126}$$

In turn, (24.119) and Proposition 24.63(iii) yield $\operatorname{ran} \partial f \subset B(0; \sqrt{k}) \subset \mathbb{S}^N$. The conclusion therefore follows from Corollary 16.57. ☐

Corollary 24.65 *Let $\varphi \in \Gamma_0(\mathbb{R}^N)$ be symmetric, let $A \in \mathbb{S}^N$, and let $U \in \mathbb{U}^N$ be such that $A = U(\operatorname{Diag} \lambda(A))U^{\mathsf{T}}$. Then*

$$\operatorname{Prox}_{\varphi \circ \lambda} A = U \operatorname{Diag}\big(\operatorname{Prox}_{\varphi} \lambda(A)\big) U^{\mathsf{T}}. \tag{24.127}$$

Proof. Set $f = \varphi \circ \lambda$ and denote by q and ϱ the halved squared norms of the Hilbert spaces $\mathbb{S}^N$ and $\mathbb{R}^N$, respectively. Note that $q = \varrho \circ \lambda$ by Fact 24.59(i). Hence $f + q = (\varphi + \varrho) \circ \lambda$, which yields $(f^* \mathbin{\square} q) = (f + q)^* = ((\varphi + \varrho) \circ \lambda)^* = (\varphi + \varrho)^* \circ \lambda = (\varphi^* \mathbin{\square} \varrho) \circ \lambda$ by (14.1) and Proposition 24.60. Using (14.7) and Proposition 24.63(iv), we deduce that $\operatorname{Prox}_f A = \nabla(f^* \mathbin{\square} q)(A) = \nabla\big((\varphi^* \mathbin{\square} \varrho) \circ \lambda\big)(A) = U(\operatorname{Diag} \nabla(\varphi^* \mathbin{\square} \varrho)(\lambda(A)))U^{\mathsf{T}} = U(\operatorname{Diag} \operatorname{Prox}_{\varphi}(\lambda(A)))U^{\mathsf{T}}$. ☐

Example 24.66 Set $I = \{1, \dots, N\}$ and

$$f: \mathbb{S}^N \to \left]-\infty, +\infty\right] : A \mapsto \begin{cases} -\ln(\det A), & \text{if } A \succ 0; \\ +\infty, & \text{otherwise.} \end{cases} \tag{24.128}$$

Then $f \in \Gamma_0(\mathbb{S}^N)$. Now let $A \in \mathbb{S}^N$ and let $U \in \mathbb{U}^N$ be such that $A = U(\operatorname{Diag} \lambda(A))U^{\mathsf{T}}$. Then

$$\operatorname{Prox}_f A = \tfrac{1}{2} U \operatorname{Diag}\Big(\lambda_i(A) + \sqrt{\lambda_i^2(A) + 4}\Big)_{i \in I} U^{\mathsf{T}}. \tag{24.129}$$

Moreover, if $A \succ 0$, then f is differentiable at A and $\nabla f(A) = -A^{-1}$.

Proof. Set

$$\varphi\colon \mathbb{R}^N \to \left]-\infty,+\infty\right] : x = (\xi_i)_{i\in I} \mapsto \begin{cases} -\displaystyle\sum_{i\in I} \ln \xi_i, & \text{if } x \in \mathbb{R}^N_{++}; \\ +\infty, & \text{otherwise.} \end{cases} \tag{24.130}$$

Then φ is symmetric and, by Example 9.36(viii) and Proposition 8.6, $\varphi \in \Gamma_0(\mathbb{R}^N)$. Then, if $A \succ 0$, $f(A) = -\ln(\prod_{i\in I} \lambda_i(A)) = -\sum_{i\in I} \ln(\lambda_i(A)) = (\varphi \circ \lambda)(A)$. The conclusions therefore follow from Corollary 24.61, Proposition 24.63(iv), Example 24.40, Proposition 24.11, and Corollary 24.65. ☐

24.8 Functions of Rectangular Matrices

In this section, $\mathbb{R}^{M\times N}$ is the Hilbert matrix space of Example 2.4. Let $m = \min\{M,N\}$. Then $\mathrm{Diag}\colon \mathbb{R}^m \to \mathbb{R}^{M\times N}$ maps a vector $(\xi_1,\ldots,\xi_m)$ to the diagonal matrix the diagonal elements of which, starting from the upper left corner, are $\xi_1, \ldots, \xi_m$. The singular values of a matrix $A \in \mathbb{R}^{M\times N}$ of rank r are denoted by $(\sigma_i(A))_{1\leqslant i\leqslant m}$ with the convention that they are listed in decreasing order, i.e.,

$$\sigma_1(A) \geqslant \cdots \geqslant \sigma_r(A) > \sigma_{r+1}(A) = \cdots = \sigma_m(A) = 0. \tag{24.131}$$

Fact 24.67 (von Neumann trace inequality) (See [235, Theorem 4.6]) Let A and B be in $\mathbb{R}^{M\times N}$, set $m = \min\{M,N\}$, and set $I = \{1,\ldots,m\}$. Then $|\operatorname{tra}(A^{\mathsf{T}}B)| = |\langle A \mid B\rangle| \leqslant \sum_{i\in I} \sigma_i(A)\sigma_i(B)$.

Proposition 24.68 *Suppose that $m = \min\{M,N\} \geqslant 2$, set $I = \{1,\ldots,m\}$, and let $\phi \in \Gamma_0(\mathbb{R})$ be even. Set*

$$f\colon \mathbb{R}^{M\times N} \to \left]-\infty,+\infty\right] : A \mapsto \sum_{i\in I} \phi\big(\sigma_i(A)\big). \tag{24.132}$$

Now let $A \in \mathbb{R}^{M\times N}$, set $r = \operatorname{rank} A$, let $U \in \mathbb{U}^M$, let $\Sigma \in \mathbb{R}^{M\times N}$, and let $V \in \mathbb{U}^N$ be such that $A = U\Sigma V^{\mathsf{T}}$ is a singular value decomposition of A. Set

$$\Sigma_f = \mathrm{Diag}\left(\mathrm{Prox}_\phi \sigma_1(A),\ldots,\mathrm{Prox}_\phi \sigma_r(A),0,\ldots,0\right) \in \mathbb{R}^{M\times N}. \tag{24.133}$$

Then

$$\mathrm{Prox}_f A = U\Sigma_f V^{\mathsf{T}}. \tag{24.134}$$

Proof. We deduce from Lemma 1.28 and [235, Proposition 6.1] that $f \in \Gamma_0(\mathbb{R}^{M\times N})$. Furthermore, $\Sigma = \mathrm{Diag}(\sigma_i(A))_{i\in I}$. Now set

$$\big(\forall B \in \mathbb{R}^{M\times N}\big) \quad \begin{cases} g(B) = f(B) + \frac{1}{2}\|\Sigma - B\|_{\mathsf{F}}^2, \\ h(B) = \sum_{i\in I} \Big(\phi\big(\sigma_i(B)\big) + \frac{1}{2}|\sigma_i(A) - \sigma_i(B)|^2\Big). \end{cases} \tag{24.135}$$

It follows from Fact 24.67 and from the invariance of $\|\cdot\|_{\mathsf{F}}$ under orthogonal transformations that, for every $B \in \mathbb{R}^{M\times N}$,

$$\begin{aligned} g(B) &= f(B) + \tfrac{1}{2}\|\Sigma\|_{\mathsf{F}}^2 - \langle \Sigma \mid B\rangle + \tfrac{1}{2}\|B\|_{\mathsf{F}}^2 \\ &\geqslant f(B) + \tfrac{1}{2}\|\Sigma\|_{\mathsf{F}}^2 - \langle \Sigma \mid \mathrm{Diag}(\sigma_i(B))_{i\in I}\rangle + \tfrac{1}{2}\|\mathrm{Diag}(\sigma_i(B))_{i\in I}\|_{\mathsf{F}}^2 \\ &= f(B) + \tfrac{1}{2}\|\Sigma - \mathrm{Diag}(\sigma_i(B))_{i\in I}\|_{\mathsf{F}}^2 \\ &= h(B). \end{aligned} \tag{24.136}$$

On the other hand, it follows from (24.1), (24.135), and (24.133) that Σ_f minimizes h. Hence, since $g \geqslant h$ and $g(\Sigma_f) = h(\Sigma_f)$, we have $\Sigma_f = \operatorname{argmin} g = \mathrm{Prox}_f \Sigma$. Now set $L\colon \mathbb{R}^{M\times N} \to \mathbb{R}^{M\times N}\colon B \mapsto U^{\mathsf{T}}BV$. Then $L^{-1} = L^*\colon B \mapsto UBV^{\mathsf{T}}$. Therefore L preserves the singular values and hence $f = f \circ L$. Altogether, it follows from Proposition 24.8(iv) that $\mathrm{Prox}_f A = \mathrm{Prox}_{f\circ L} A = L^* \mathrm{Prox}_f(LA) = L^* \mathrm{Prox}_f \Sigma = L^*\Sigma_f = U\Sigma_f V^{\mathsf{T}}$. Finally, Proposition 24.10 yields $(\forall i \in \{r+1,\ldots,m\})\ \mathrm{Prox}_\phi \sigma_i(A) = 0$. ☐

Example 24.69 Suppose that $\min\{M,N\} \geqslant 2$, let $A \in \mathbb{R}^{M\times N}$, set $r = \operatorname{rank} A$, let $U \in \mathbb{U}^M$, let $\Sigma \in \mathbb{R}^{M\times N}$, and let $V \in \mathbb{U}^N$ be such that $A = U\Sigma V^{\mathsf{T}}$ is a singular value decomposition of A. Then the *nuclear norm* of A is

$$\|A\|_{\mathsf{nuc}} = \sum_{i=1}^{r} \sigma_i(A). \tag{24.137}$$

Now let $\gamma \in \mathbb{R}_{++}$ and define $\Sigma_{\mathsf{nuc}} \in \mathbb{R}^{M\times N}$ by

$$\Sigma_{\mathsf{nuc}} = \mathrm{Diag}\,\big(\max\{\sigma_1(A)-\gamma,0\},\ldots,\max\{\sigma_r(A)-\gamma,0\},0,\ldots,0\big). \tag{24.138}$$

Then applying Proposition 24.68 with $\phi = \gamma|\cdot|$ and using (24.19) yields

$$\mathrm{Prox}_{\gamma\|\cdot\|_{\mathsf{nuc}}} A = U\Sigma_{\mathsf{nuc}}V^{\mathsf{T}}. \tag{24.139}$$

Example 24.70 Let $p \in \left]1,+\infty\right[$, suppose that $\min\{M,N\} \geqslant 2$, let $A \in \mathbb{R}^{M\times N}$, and set $r = \operatorname{rank} A$. Then the *Schatten p-norm* of $A \in \mathbb{R}^{M\times N}$ is

$$\|A\|_{\mathsf{Sch}} = \left(\sum_{i=1}^{r} |\sigma_i(A)|^p\right)^{1/p}. \tag{24.140}$$

The proximity operator of $\gamma\|\cdot\|_{\mathsf{Sch}}^p$ can be deduced by applying Proposition 24.68 with $\phi = \gamma|\cdot|^p$ and using Example 24.38.

Exercises

Exercise 24.1 Give the details of the proof of Corollary 24.15.

Exercise 24.2 Let C be a nonempty closed convex subset of $\mathcal{H}$ and let $\gamma \in \mathbb{R}_{++}$. Compute $\operatorname{Prox}_{(\iota_{B(0;1)}+\gamma\sigma_C)}$.

Exercise 24.3 Determine the proximity operator of the hinge loss function ϕ (see Example 24.36) graphically: first plot $\partial\phi$, then $\operatorname{Id} + \gamma\partial\phi$, and finally $\operatorname{Prox}_{\gamma\phi} = (\operatorname{Id} + \gamma\partial\phi)^{-1}$.

Exercise 24.4 Let $\lambda \in \left]0, 1\right]$, let $f \in \Gamma_0(\mathcal{H})$, and set $T = \operatorname{Id} + \lambda(\operatorname{Prox}_f - \operatorname{Id})$. Use Corollary 18.20 to show that T is a proximity operator.

Exercise 24.5 Provide a direct proof of Example 24.28 using the characterization (24.2).

Exercise 24.6 Let $f \colon \mathbb{R} \to \mathbb{R}$ be convex and twice continuously differentiable. Show that $(\forall x \in \mathbb{R})\ \operatorname{Prox}_f' x = 1/(1 + f''(\operatorname{Prox}_f x))$.

Exercise 24.7 Let f and g be functions in $\Gamma_0(\mathcal{H})$. Show that the following are equivalent:

(i) $(\exists\, \gamma \in \mathbb{R})\ f = g + \gamma$.
(ii) $\operatorname{Prox}_f = \operatorname{Prox}_g$.
(iii) $\partial f = \partial g$.
(iv) $\operatorname{gra} \partial f \subset \operatorname{gra} \partial g$.

Exercise 24.8 Find functions f and g in $\Gamma_0(\mathbb{R})$ such that $\operatorname{Prox}_{f+g} \neq \operatorname{Prox}_f \circ \operatorname{Prox}_g$.

Exercise 24.9 Give the details of the derivations in Example 24.38.

Exercise 24.10 Give the details of the derivations in Example 24.43, Example 24.44, Example 24.45, and Example 24.46.

Exercise 24.11 Show that Proposition 24.47 fails if $\mathcal{H}$ has dimension at least 2.

Chapter 25
Sums of Monotone Operators

The sum of two monotone operators is monotone. However, maximal monotonicity of the sum of two maximally monotone operators is not automatic and requires additional assumptions. In this chapter, we provide flexible sufficient conditions for maximality. Under additional assumptions, conclusions can be drawn about the range of the sum, which is important for deriving surjectivity results. We also consider the parallel sum of two set-valued operators.

Throughout this chapter, F_A designates the Fitzpatrick function of a monotone operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ (see Definition 20.51).

25.1 Maximal Monotonicity of Sums and Composite Sums

As the following example shows, the sum of two maximally monotone operators need not be maximally monotone.

Example 25.1 Suppose that x and y are linearly independent vectors in $\mathcal{H}$, and set $C = B(x; \|x\|)$ and $D = -C$. Then N_C and N_D are maximally monotone by Example 20.26, with $\operatorname{dom}(\mathrm{Id} + N_C + N_D) = (\operatorname{dom} N_C) \cap (\operatorname{dom} N_D) = \{0\}$. Hence, by using Example 6.39, $\operatorname{ran}(\mathrm{Id} + N_C + N_D) = N_C 0 + N_D 0 = \mathbb{R}_+ (-x) + \mathbb{R}_+ x = \mathbb{R}\, x \subset \mathcal{H} \smallsetminus \{y\}$ and it follows from Theorem 21.1 that $N_C + N_D$ is not maximally monotone.

Theorem 25.2 *Set $Q_1\colon \mathcal{H} \times \mathcal{H} \to \mathcal{H}\colon (x, u) \mapsto x$, and let A and B be maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$ such that*

$$0 \in \operatorname{sri} Q_1\big(\operatorname{dom} F_A - \operatorname{dom} F_B\big). \tag{25.1}$$

Then $A + B$ is maximally monotone.

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_25

Proof. Proposition 20.58 and Proposition 20.56(iv) imply that, for every (x, u_1, u_2) in $\mathcal{H}^3$,

$$\langle x \mid u_1 + u_2\rangle \leqslant F_A(x, u_1) + F_B(x, u_2) \leqslant F_A^*(u_1, x) + F_B^*(u_2, x). \tag{25.2}$$

Thus the function

$$F\colon \mathcal{H}\times\mathcal{H} \to \left]-\infty,+\infty\right] : (x,u) \mapsto \big(F_A(x,\cdot)\,\square\, F_B(x,\cdot)\big)(u) \tag{25.3}$$

is proper and convex, and it satisfies

$$F \geqslant \langle\cdot \mid \cdot\rangle\,. \tag{25.4}$$

Corollary 15.8 and (25.2) yield

$$\begin{aligned}\big(\forall (u,x)\in\mathcal{H}\times\mathcal{H}\big)\quad F^*(u,x) &= \big(F_A^*(\cdot,x)\,\boxdot\, F_B^*(\cdot,x)\big)(u)\\ &\geqslant F(x,u)\\ &\geqslant \langle x \mid u\rangle\,.\end{aligned} \tag{25.5}$$

Now fix $(x,u)\in\mathcal{H}\times\mathcal{H}$ and assume that $F^*(u,x) = \langle x \mid u\rangle$. Then (25.5) and Proposition 20.61(iii) guarantee the existence of u_1 and u_2 in $\mathcal{H}$ such that $u_1 + u_2 = u$ and $\langle x \mid u\rangle = F^*(u,x) = F_A^*(u_1,x) + F_B^*(u_2,x) \geqslant \langle x \mid u_1\rangle + \langle x \mid u_2\rangle = \langle x \mid u\rangle$. It follows that $F_A^*(u_1,x) = \langle x \mid u_1\rangle$ and that $F_B^*(u_2,x) = \langle x \mid u_2\rangle$. By (20.51), $(x,u) = (x, u_1+u_2) \in \operatorname{gra}(A+B)$. Now assume that $(x,u)\in\operatorname{gra}(A+B)$. Then there exist $u_1\in Ax$ and $u_2\in Bx$ such that $u = u_1+u_2$. Hence $F_A^*(u_1,x) = \langle x \mid u_1\rangle$ and $F_B^*(u_2,x) = \langle x \mid u_2\rangle$, by (20.51). This and (25.5) imply that $\langle x \mid u\rangle = \langle x \mid u_1\rangle + \langle x \mid u_2\rangle = F_A^*(u_1,x) + F_B^*(u_2,x) \geqslant \big(F_A^*(\cdot,x)\,\boxdot\, F_B^*(\cdot,x)\big)(u) = F^*(u,x) \geqslant \langle x \mid u\rangle$. Therefore, $F^*(u,x) = \langle x \mid u\rangle$. Altogether, we have shown that

$$\big\{(x,u)\in\mathcal{H}\times\mathcal{H} \;\big|\; F^*(u,x) = \langle x \mid u\rangle\big\} = \operatorname{gra}(A+B). \tag{25.6}$$

Combining (25.5), (25.6), (25.4), and Theorem 20.46(ii), we conclude that $A+B$ is maximally monotone. ☐

Theorem 25.3 *Let $\mathcal{K}$ be a real Hilbert space, let $A\colon \mathcal{H}\to 2^{\mathcal{H}}$ and $B\colon \mathcal{K}\to 2^{\mathcal{K}}$ be maximally monotone, let $L\in\mathcal{B}(\mathcal{H},\mathcal{K})$, and suppose that*

$$\operatorname{cone}\big(\operatorname{dom} B - L(\operatorname{dom} A)\big) = \overline{\operatorname{span}}\,\big(\operatorname{dom} B - L(\operatorname{dom} A)\big). \tag{25.7}$$

Then $A + L^*BL$ is maximally monotone.

Proof. Set $Q_1^{\mathcal{H}}\colon \mathcal{H}\times\mathcal{H}\to\mathcal{H}\colon (x,u)\mapsto x$, and $Q_1^{\mathcal{K}}\colon \mathcal{K}\times\mathcal{K}\to\mathcal{K}\colon (y,v)\mapsto y$. It follows from Proposition 21.12, Corollary 21.14, and Proposition 3.46 that $\operatorname{dom} A \subset Q_1^{\mathcal{H}}(\operatorname{dom} F_A) \subset \overline{\operatorname{dom}}\,A \subset \overline{\operatorname{conv}}\operatorname{dom} A$ and $\operatorname{dom} B \subset Q_1^{\mathcal{K}}(\operatorname{dom} F_B) \subset \overline{\operatorname{conv}}\operatorname{dom} B$. Hence, in view of Proposition 6.21(ii), we obtain

$$0 \in \operatorname{sri}\Big(\big((Q_1^{\mathcal{H}}\operatorname{dom} F_A)\times(Q_1^{\mathcal{K}}\operatorname{dom} F_B)\big) - \operatorname{gra} L\Big). \tag{25.8}$$

Turning to the maximal monotonicity of $A + L^*BL$, we work in $\boldsymbol{\mathcal{H}} = \mathcal{H} \oplus \mathcal{K}$ and we set $Q_1^{\boldsymbol{\mathcal{H}}} \colon \boldsymbol{\mathcal{H}} \times \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} \colon ((x, y), (u, v)) \mapsto (x, y)$. Using Proposition 20.23 and Example 20.26, we see that the operators $\boldsymbol{A} \colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}} \colon (x, y) \mapsto Ax \times By$ and $\boldsymbol{B} = N_{\operatorname{gra} L}$ are maximally monotone with $\operatorname{dom} \boldsymbol{A} = \operatorname{dom} A \times \operatorname{dom} B$ and $\operatorname{dom} \boldsymbol{B} = \operatorname{gra} L$. In turn, Proposition 20.57(ii) and Example 21.13 imply that

$$\begin{aligned} Q_1^{\boldsymbol{\mathcal{H}}}(\operatorname{dom} F_{\boldsymbol{A}} - \operatorname{dom} F_{\boldsymbol{B}}) &= Q_1^{\boldsymbol{\mathcal{H}}}(\operatorname{dom} F_{\boldsymbol{A}}) - Q_1^{\boldsymbol{\mathcal{H}}}(\operatorname{dom} F_{\boldsymbol{B}}) \\ &= (Q_1^{\mathcal{H}} \operatorname{dom} F_A) \times (Q_1^{\mathcal{K}} \operatorname{dom} F_B) - \operatorname{gra} L. \end{aligned} \tag{25.9}$$

Hence, by (25.8) and Theorem 25.2,

$$\boldsymbol{A} + \boldsymbol{B} \text{ is maximally monotone.} \tag{25.10}$$

Now let $(x, u) \in \mathcal{H} \times \mathcal{H}$ and $v \in \mathcal{K}$. Then $(\boldsymbol{A} + \boldsymbol{B})(x, Lx) = (Ax \times B(Lx)) + N_{\operatorname{gra} L}(x, Lx)$ and hence

$$\begin{aligned} (u, v) \in (\boldsymbol{A} + \boldsymbol{B})(x, Lx) &\Leftrightarrow (u, v) \in \big(Ax \times B(Lx)\big) + N_{\operatorname{gra} L}(x, Lx) \\ &\Leftrightarrow (u, v) \in \big(Ax \times B(Lx)\big) + \big\{(L^*w, -w) \bigm| w \in \mathcal{K}\big\} \\ &\Leftrightarrow (\exists\, w \in \mathcal{K}) \quad u \in L^*w + Ax \text{ and } v + w \in B(Lx) \\ &\Leftrightarrow u + L^*v \in \big(A + L^*BL\big)x. \end{aligned} \tag{25.11}$$

Next, let $(z, w) \in \boldsymbol{\mathcal{H}}$ be monotonically related to $\operatorname{gra}(A + L^*BL)$, i.e.,

$$(\forall y \in \mathcal{H}) \quad \inf \langle y - z \mid (Ay + L^*BLy) - w\rangle \geqslant 0, \tag{25.12}$$

and suppose that $((x, Lx), (u, v)) \in \operatorname{gra}(\boldsymbol{A} + \boldsymbol{B})$. Then, by (25.11), $u + L^*v \in (A + L^*BL)x$. Moreover, (25.12) yields

$$\begin{aligned} 0 &\leqslant \langle x - z \mid u + L^*v - w\rangle \\ &= \langle x - z \mid u - w\rangle + \langle Lx - Lz \mid v\rangle \\ &= \langle (x, Lx) - (z, Lz) \mid (u, v) - (w, 0)\rangle . \end{aligned} \tag{25.13}$$

Hence, $((z, Lz), (w, 0)) \in \operatorname{gra}(\boldsymbol{A} + \boldsymbol{B})$ by (25.10). In view of (25.11), this is equivalent to $w \in (A + L^*BL)z$. Appealing to (20.16), we conclude that $A + L^*BL$ is maximally monotone. ☐

Corollary 25.4 *Let A and B be maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$ such that*

$$\operatorname{cone}(\operatorname{dom} A - \operatorname{dom} B) = \overline{\operatorname{span}}\,(\operatorname{dom} A - \operatorname{dom} B). \tag{25.14}$$

Then $A + B$ is maximally monotone.

Proof. This follows from Theorem 25.3. ☐

Corollary 25.5 *Let A and B be maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$ such that one of the following holds:*

(i) $\operatorname{dom} B = \mathcal{H}$.
(ii) $\operatorname{dom} A \cap \operatorname{int} \operatorname{dom} B \neq \varnothing$.
(iii) $0 \in \operatorname{int}(\operatorname{dom} A - \operatorname{dom} B)$.
(iv) $(\forall x \in \mathcal{H})(\exists\, \varepsilon \in \mathbb{R}_{++})\ [0, \varepsilon x] \subset \operatorname{dom} A - \operatorname{dom} B$.
(v) $\operatorname{dom} A$ *and* $\operatorname{dom} B$ *are convex and*

$$0 \in \operatorname{sri}(\operatorname{dom} A - \operatorname{dom} B). \tag{25.15}$$

Then $A + B$ is maximally monotone.

Proof. We derive from (6.11) that (i)⇒(ii)⇒(iii)⇒(iv)⇒(v). The conclusion thus follows from Corollary 25.4. ☐

Corollary 25.6 *Let $\mathcal{K}$ be a real Hilbert space, let $A\colon \mathcal{K} \to 2^{\mathcal{K}}$ be maximally monotone, let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, and suppose that* $\operatorname{cone}(\operatorname{ran} L - \operatorname{dom} A) = \overline{\operatorname{span}}\,(\operatorname{ran} L - \operatorname{dom} A)$. *Then L^*AL is maximally monotone.*

Proof. This follows from Theorem 25.3. ☐

25.2 Local Maximal Monotonicity

Definition 25.7 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be monotone, and suppose that U is an open convex subset of $\mathcal{H}$ such that $U \cap \operatorname{ran} A \neq \varnothing$. Then A is *locally maximally monotone with respect to* U if, for every $(x, u) \in (\mathcal{H} \times U) \smallsetminus \operatorname{gra} A$, there exists $(y, v) \in (\mathcal{H} \times U) \cap \operatorname{gra} A$ such that $\langle x - y \mid u - v \rangle < 0$. Moreover, A is *locally maximally monotone* if A is locally maximally monotone with respect to every open convex subset of $\mathcal{H}$ that has a nonempty intersection with $\operatorname{ran} A$.

As the next result illustrates, local maximal monotonicity is closely related to maximal monotonicity.

Proposition 25.8 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be monotone. Then the following hold:*

(i) *A is locally maximally monotone with respect to $\mathcal{H}$ if and only if A is maximally monotone.*
(ii) *A is locally maximally monotone if and only if, for every bounded closed convex subset C of $\mathcal{H}$ such that* $\operatorname{ran} A \cap \operatorname{int} C \neq \varnothing$ *and for every* $(x, u) \in (\mathcal{H} \times \operatorname{int} C) \smallsetminus \operatorname{gra} A$, *there exists* $(y, v) \in (\mathcal{H} \times C) \cap \operatorname{gra} A$ *such that* $\langle x - y \mid u - v \rangle < 0$.

Proof. (i): This is clear from the definition.

(ii): Assume first that A is locally maximally monotone, that C is a bounded closed convex subset of $\mathcal{H}$ such that $\operatorname{ran} A \cap \operatorname{int} C \neq \varnothing$, and that

$(x,u) \in (\mathcal{H} \times \operatorname{int} C) \smallsetminus \operatorname{gra} A$. Then the existence of (y,v) with the desired properties is guaranteed by the local maximal monotonicity of A with respect to $\operatorname{int} C$.

To prove the other implication, let U be an open convex subset of $\mathcal{H}$ such that $U \cap \operatorname{ran} A \neq \varnothing$, let $(z,w) \in (\mathcal{H} \times U) \cap \operatorname{gra} A$, and let $(x,u) \in (\mathcal{H} \times U) \smallsetminus \operatorname{gra} A$. Then there exists $\varepsilon \in \mathbb{R}_{++}$ such that $B(u;\varepsilon) \cap B(w;\varepsilon) \subset U$. Set $C = [u,w] + B(0;\varepsilon)$. Then C is a bounded closed convex subset of U such that $u \in \operatorname{int} C$ and $w \in C \cap \operatorname{ran} A$. By assumption, there exists $(y,v) \in (\mathcal{H} \times C) \cap \operatorname{gra} A \subset (\mathcal{H} \times U) \cap \operatorname{gra} A$ such that $\langle x - y \mid u - v \rangle < 0$. ☐

Proposition 25.9 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone. Then A is locally maximally monotone.*

Proof. Let C be a bounded closed convex subset of $\mathcal{H}$ such that $\operatorname{ran} A \cap \operatorname{int} C \neq \varnothing$, and let $(x,u) \in (\mathcal{H} \times \operatorname{int} C) \smallsetminus \operatorname{gra} A$. Set $B = A^{-1}$. Then $\operatorname{ran} A \cap \operatorname{int} C = \operatorname{dom} B \cap \operatorname{int} \operatorname{dom} N_C \neq \varnothing$ and $x \notin Bu = (B + N_C)u$. By Corollary 25.5(ii), $B + N_C$ is maximally monotone. Since $(u,x) \notin \operatorname{gra}(B + N_C)$, there exist $(v,y) \in \operatorname{gra} B$ and $(v,z) \in \operatorname{gra} N_C$ such that $\langle u - v \mid x - (y+z) \rangle < 0$. Hence $(y,v) \in (\mathcal{H} \times C) \cap \operatorname{gra} A$ and $\langle u - v \mid z \rangle \leqslant 0$, and thus $\langle u - v \mid x - y \rangle \leqslant \langle u - v \mid x - y \rangle - \langle u - v \mid z \rangle = \langle u - v \mid x - (y+z) \rangle < 0$. Therefore, the conclusion follows from Proposition 25.8(ii). ☐

25.3 3* Monotone Operators of Brézis and Haraux

Definition 25.10 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be monotone. Then A is 3^* *monotone* if

$$\operatorname{dom} A \times \operatorname{ran} A \subset \operatorname{dom} F_A. \tag{25.16}$$

The next result explains why 3^* monotone operators are sometimes called rectangular.

Proposition 25.11 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and 3^* monotone. Then $\overline{\operatorname{dom}}\, F_A = \overline{\operatorname{dom}}\, A \times \overline{\operatorname{ran}}\, A$.*

Proof. Combine (25.16) and Corollary 21.14. ☐

Proposition 25.12 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be 3-cyclically monotone. Then A is 3^* monotone.*

Proof. Let (x,u) and (z,w) be in $\operatorname{gra} A$. Then

$$\big(\forall (y,v) \in \operatorname{gra} A\big) \quad \langle y - x \mid u \rangle + \langle z - y \mid v \rangle + \langle x - z \mid w \rangle \leqslant 0. \tag{25.17}$$

It follows that $(\forall (y,v) \in \operatorname{gra} A)$ $\langle z \mid v \rangle + \langle y \mid u \rangle - \langle y \mid v \rangle \leqslant \langle x \mid u \rangle + \langle x - z \mid w \rangle$. Hence $F_A(z,u) \leqslant \langle x \mid u \rangle + \langle x - z \mid w \rangle$ and therefore $(z,u) \in \operatorname{dom} F_A$. ☐

Example 25.13 Let $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ be proper. Then ∂f is 3^* monotone.

Proof. Combine Proposition 22.14 and Proposition 25.12. ☐

Example 25.14 Let C be a nonempty closed convex subset of $\mathcal{H}$. Then N_C is 3^* monotone.

Proof. Apply Example 25.13 to $f = \iota_C$ and use Example 16.13. ☐

Example 25.15 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be monotone. Then $\operatorname{dom} A \times \mathcal{H} \subset \operatorname{dom} F_A$ and A is 3^* monotone in each of the following cases:

(i) $(\forall x \in \operatorname{dom} A)\ \displaystyle\lim_{\rho\to+\infty} \inf_{\substack{(z,w)\in\operatorname{gra} A\\ \|z\|\geqslant\rho}} \frac{\langle z - x \mid w\rangle}{\|z\|} = +\infty.$

(ii) $\operatorname{dom} A$ is bounded.

(iii) A is uniformly monotone with a supercoercive modulus.

(iv) A is strongly monotone.

Proof. We suppose that $\operatorname{gra} A \neq \varnothing$.

(i): Let $(x, v) \in \operatorname{dom} A \times \mathcal{H}$, and set $\mu = \|v\| + 1$. Then there exists $\rho \in \mathbb{R}_{++}$ such that, for every $(z, w) \in \operatorname{gra} A$ with $\|z\| \geqslant \rho$, we have $\langle z - x \mid w\rangle \geqslant \mu\|z\|$. Thus,

$$\big(\forall (z,w) \in \operatorname{gra} A\big) \quad \|z\| \geqslant \rho \ \Rightarrow\ \langle x - z \mid w\rangle \leqslant -\mu\|z\|. \tag{25.18}$$

Take $u \in Ax$ and $(z, w) \in \operatorname{gra} A$. Then, by monotonicity of A,

$$\langle x - z \mid w\rangle \leqslant \langle x - z \mid u\rangle\,. \tag{25.19}$$

Hence, if $\|z\| < \rho$, then $\langle x - z \mid w\rangle + \langle z \mid v\rangle \leqslant \langle x - z \mid u\rangle + \langle z \mid v\rangle \leqslant \|x - z\|\,\|u\| + \|z\|\,\|v\| \leqslant (\|x\| + \|z\|)\|u\| + \|z\|\,\|v\| \leqslant (\|x\| + \rho)\|u\| + \rho\|v\|$. On the other hand, if $\|z\| \geqslant \rho$, then (25.18) implies that $\langle x - z \mid w\rangle + \langle z \mid v\rangle \leqslant -\mu\|z\| + \|z\|\,\|v\| = \|z\|\,(\|v\| - \mu) = -\|z\| \leqslant 0$. Altogether,

$$F_A(x, v) = \sup_{(z,w)\in\operatorname{gra} A} \big(\langle x - z \mid w\rangle + \langle z \mid v\rangle\big) \leqslant \big(\|x\| + \rho\big)\|u\| + \rho\|v\|, \tag{25.20}$$

which shows that $(x, v) \in \operatorname{dom} F_A$.

(ii): Clear from (i).

(iii): Denote by ϕ the modulus of uniform monotonicity of A, fix $(x, v) \in \operatorname{dom} A \times \mathcal{H}$, fix $u \in Ax$, and let $(z, w) \in \operatorname{gra} A$. Then $\langle z - x \mid w - u\rangle \geqslant \phi(\|z - x\|)$ and therefore, using Cauchy–Schwarz, we obtain

$$\langle z - x \mid w\rangle \geqslant \phi(\|x - z\|) + \langle x - z \mid u\rangle \geqslant \phi(\|z - x\|) - \|z\|\,\|u\| + \langle x \mid u\rangle\,. \tag{25.21}$$

Thus, since ϕ is supercoercive, we obtain (i).

(iv): Clear from (iii). ☐

Proposition 25.16 *Let $A \in \mathcal{B}(\mathcal{H})$ be monotone. Then the following are equivalent for some $\beta \in \mathbb{R}_{++}$:*

(i) *A is 3^* monotone.*
(ii) *A is β-cocoercive.*
(iii) *A^* is β-cocoercive.*
(iv) *A^* is 3^* monotone.*

Proof. (i)⇒(ii): Set

$$f\colon \mathcal{H} \to \left]-\infty,+\infty\right] : x \mapsto F_A(x,0). \tag{25.22}$$

Then $f \in \Gamma_0(\mathcal{H})$ by Proposition 20.56(i)&(ii). Since A is 3^* monotone, we see that $\mathcal{H} \times \{0\} \subset \operatorname{dom} A \times \operatorname{ran} A \subset \operatorname{dom} F_A$. Hence $\operatorname{dom} f = \mathcal{H}$. By Corollary 8.39(ii), $0 \in \operatorname{cont} f$ and thus there exist $\rho \in \mathbb{R}_{++}$ and $\mu \in \mathbb{R}_{++}$ such that $(\forall x \in B(0;\rho))\ f(x) = F_A(x,0) = \sup_{y\in\mathcal{H}}(\langle x \mid Ay\rangle - \langle y \mid Ay\rangle) \leqslant \mu$. Now let $x \in B(0;\rho)$ and $y \in \mathcal{H}$. Then

$$\begin{aligned}(\forall t \in \mathbb{R}) \quad 0 &\leqslant \mu + \langle ty \mid A(ty)\rangle - \langle x \mid A(ty)\rangle \\ &= \mu + t^2 \langle y \mid Ay\rangle - t\langle x \mid Ay\rangle .\end{aligned} \tag{25.23}$$

This implies that

$$(\forall x \in B(0;\rho))(\forall y \in \mathcal{H}) \quad \langle x \mid Ay\rangle^2 \leqslant 4\mu \langle y \mid Ay\rangle . \tag{25.24}$$

Now let $y \in \mathcal{H} \smallsetminus \ker A$ and set $x = (\rho/\|Ay\|)Ay \in B(0;\rho)$. By (25.24), we arrive at $4\mu\langle y \mid Ay\rangle \geqslant (\rho/\|Ay\|)^2\|Ay\|^4$, i.e., $\langle y \mid Ay\rangle \geqslant \|Ay\|^2\rho^2/(4\mu)$. The last inequality is also true when $y \in \ker A$; hence we set $\beta = \rho^2/(4\mu)$, and we deduce that, for every $y \in \mathcal{H}$, $\langle y \mid \beta Ay\rangle \geqslant \|\beta Ay\|^2$. Therefore, since A is linear, we deduce that A is β-cocoercive.

(ii)⇔(iii): Clear from Corollary 4.5.

(ii)⇒(i)&(iii)⇒(i): Fix x and y in $\mathcal{H}$, and take $z \in \mathcal{H}$. Then

$$\begin{aligned}&\langle x \mid Az\rangle + \langle z \mid Ay\rangle - \langle z \mid Az\rangle \\ &\qquad = \big(\langle x \mid Az\rangle - \tfrac{1}{2}\langle z \mid Az\rangle\big) + \big(\langle A^*z \mid y\rangle - \tfrac{1}{2}\langle z \mid A^*z\rangle\big) \\ &\qquad \leqslant \big(\|x\|\,\|Az\| - \tfrac{1}{2}\beta\|Az\|^2\big) + \big(\|A^*z\|\,\|y\| - \tfrac{1}{2}\beta\|A^*z\|^2\big) \\ &\qquad \leqslant \frac{1}{2\beta}\big(\|x\|^2 + \|y\|^2\big),\end{aligned} \tag{25.25}$$

where (25.25) was obtained by computing the maximum of the quadratic functions $t \mapsto \|x\|t-(1/2)\beta t^2$ and $t \mapsto \|y\|t-(1/2)\beta t^2$. It follows from (25.25) that

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad F_A(x,Ay) \leqslant \frac{1}{2\beta}\big(\|x\|^2 + \|y\|^2\big). \tag{25.26}$$

Hence $\mathcal{H} \times \operatorname{ran} A \subset \operatorname{dom} F_A$.

(iii)⇔(iv): Apply the already established equivalence (i)⇔(ii) to A^*. ☐

Example 25.17 Let $\mathcal{K}$ be a real Hilbert space and let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$. Then LL^* is 3^* monotone.

Proof. Combine Example 25.13, Example 2.60, and Proposition 17.31(i). Alternatively, combine Corollary 18.18 and Proposition 25.16. ☐

Example 25.18 Let N be a strictly positive integer, and let $R\colon \mathcal{H}^N \to \mathcal{H}^N \colon (x_1, x_2, \ldots, x_N) \mapsto (x_N, x_1, \ldots, x_{N-1})$ be the *cyclic right-shift operator* in $\mathcal{H}^N$. Then $\mathrm{Id} - R$ is 3^* monotone.

Proof. By Proposition 4.4 and Proposition 25.16, $-R$ is nonexpansive $\Leftrightarrow$ $2((1/2)(\mathrm{Id}-R))-\mathrm{Id}$ is nonexpansive $\Leftrightarrow$ $(1/2)(\mathrm{Id}-R)$ is firmly nonexpansive $\Rightarrow$ $\mathrm{Id} - R$ is 3^* monotone. ☐

Proposition 25.19 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be monotone and let $\gamma \in \mathbb{R}_{++}$. Then the following hold:*

(i) *A is 3^* monotone if and only if A^{-1} is 3^* monotone.*
(ii) *A is 3^* monotone if and only if γA is 3^* monotone.*

Proof. (i): By Proposition 20.56(vi), A is 3^* monotone $\Leftrightarrow$ $\mathrm{dom}\, A \times \mathrm{ran}\, A \subset \mathrm{dom}\, F_A$ $\Leftrightarrow$ $\mathrm{ran}\, A \times \mathrm{dom}\, A \subset \mathrm{dom}\, F_A^{\intercal}$ $\Leftrightarrow$ $\mathrm{dom}\, A^{-1} \times \mathrm{ran}\, A^{-1} \subset \mathrm{dom}\, F_{A^{-1}}$ $\Leftrightarrow$ A^{-1} is 3^* monotone.

(ii): By Proposition 20.56(vii), A is 3^* monotone $\Leftrightarrow$ $\mathrm{dom}\, A \times \mathrm{ran}\, A \subset \mathrm{dom}\, F_A$ $\Leftrightarrow$ $\mathrm{dom}\, A \times \gamma\, \mathrm{ran}\, A \subset (\mathrm{Id} \times \gamma \mathrm{Id})(\mathrm{dom}\, F_A)$ $\Leftrightarrow$ $\mathrm{dom}(\gamma A) \times \mathrm{ran}(\gamma A) \subset \mathrm{dom}\, F_{\gamma A}$ $\Leftrightarrow$ γA is 3^* monotone. ☐

Example 25.20 Let D be a nonempty subset of $\mathcal{H}$, let $T\colon D \to \mathcal{H}$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be monotone, and let β and γ be in $\mathbb{R}_{++}$. Suppose that one of the following holds:

(i) T is β-cocoercive.
(ii) T is firmly nonexpansive.
(iii) $T = J_{\gamma A}$.
(iv) $T = {}^{\gamma}A$.
(v) $\mathrm{Id} - T$ is nonexpansive.

Then T is 3^* monotone.

Proof. (i): T^{-1} is β–strongly monotone by Example 22.7, hence 3^* monotone by Example 25.15(iv). In turn, Proposition 25.19(i) implies that T is 3^* monotone.

(ii): This corresponds to the case when $\beta = 1$ in (i).

(iii): This follows from (ii) and Proposition 23.8(i)&(ii).

(iv): Combine (i) with Proposition 23.21(i).

(v): Combine Proposition 4.11 with (i). ☐

Proposition 25.21 *Let A and B be monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$. Then $(\forall x \in \mathcal{H})(\forall u \in \mathcal{H})$ $F_{A+B}(x, u) \leqslant (F_A(x, \cdot) \,\square\, F_B(x, \cdot))(u)$.*

Proof. Fix $(x,u) \in \mathcal{H} \times \mathcal{H}$ and suppose that $(u_1,u_2) \in \mathcal{H} \times \mathcal{H}$ satisfies $u = u_1 + u_2$. Take $(y,v_1) \in \operatorname{gra} A$ and $(y,v_2) \in \operatorname{gra} B$. Then $\langle x \mid v_1 + v_2\rangle + \langle y \mid u\rangle - \langle y \mid v_1 + v_2\rangle = (\langle x \mid v_1\rangle + \langle y \mid u_1\rangle - \langle y \mid v_1\rangle) + (\langle x \mid v_2\rangle + \langle y \mid u_2\rangle - \langle y \mid v_2\rangle) \leqslant F_A(x,u_1) + F_B(x,u_2)$ and hence $F_{A+B}(x,u) \leqslant F_A(x,u_1) + F_B(x,u_2)$. In turn, this implies that $F_{A+B}(x,u) \leqslant \inf_{u_1+u_2=u} F_A(x,u_1) + F_B(x,u_2) = (F_A(x,\cdot) \,\square\, F_B(x,\cdot))(u)$. ☐

Proposition 25.22 *Let A and B be monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$. Then $A + B$ is 3^* monotone in each of the following cases:*

(i) $(\operatorname{dom} A \cap \operatorname{dom} B) \times \mathcal{H} \subset \operatorname{dom} F_B$.
(ii) *A and B are 3^* monotone.*

Proof. Suppose that $\operatorname{gra}(A + B) \neq \varnothing$.

(i): Let $(x,v) \in \operatorname{dom}(A + B) \times \mathcal{H} = (\operatorname{dom} A \cap \operatorname{dom} B) \times \mathcal{H}$ and take $u \in Ax$. By Proposition 20.56(i), $F_A(x,u) = \langle x \mid u\rangle < +\infty$. Thus, by Proposition 25.21, $F_{A+B}(x,v) \leqslant (F_A(x,\cdot) \,\square\, F_B(x,\cdot))(v) \leqslant F_A(x,u)+F_B(x,v-u) < +\infty$. Hence $\operatorname{dom}(A+B) \times \operatorname{ran}(A+B) \subset (\operatorname{dom} A \cap \operatorname{dom} B) \times \mathcal{H} \subset \operatorname{dom} F_{A+B}$, and therefore $A + B$ is 3^* monotone.

(ii): Let $x \in \operatorname{dom}(A + B) = \operatorname{dom} A \cap \operatorname{dom} B$ and let $w \in \operatorname{ran}(A + B)$. Then there exist $u \in \operatorname{ran} A$ and $v \in \operatorname{ran} B$ such that $w = u + v$. Observe that $(x,u) \in \operatorname{dom} A \times \operatorname{ran} A \subset \operatorname{dom} F_A$ and $(x,v) \in \operatorname{dom} B \times \operatorname{ran} B \subset \operatorname{dom} F_B$. Using Proposition 25.21, we obtain $F_{A+B}(x,w) \leqslant F_A(x,u)+F_B(x,v) < +\infty$. Therefore, $\operatorname{dom}(A + B) \times \operatorname{ran}(A + B) \subset \operatorname{dom} F_{A+B}$. ☐

25.4 The Brézis–Haraux Theorem

Theorem 25.23 (Simons) *Let A and B be monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$ such that $A + B$ is maximally monotone. Suppose that*

$$(\forall u \in \operatorname{ran} A)(\forall v \in \operatorname{ran} B)(\exists\, x \in \mathcal{H}) \quad (x,u) \in \operatorname{dom} F_A \text{ and } (x,v) \in \operatorname{dom} F_B. \tag{25.27}$$

Then $\overline{\operatorname{ran}}\,(A + B) = \overline{\operatorname{ran} A + \operatorname{ran} B}$ *and* $\operatorname{int} \operatorname{ran}(A + B) = \operatorname{int}(\operatorname{ran} A + \operatorname{ran} B)$.

Proof. We set $Q_2 \colon \mathcal{H} \times \mathcal{H} \to \mathcal{H} \colon (z,w) \mapsto w$. Fix $u \in \operatorname{ran} A$ and $v \in \operatorname{ran} B$, and let $x \in \mathcal{H}$ be such that $(x,u) \in \operatorname{dom} F_A$ and $(x,v) \in \operatorname{dom} F_B$. Proposition 25.21 implies that

$$F_{A+B}(x,u+v) \leqslant F_A(x,u) + F_B(x,v) < +\infty, \tag{25.28}$$

and thus that $u+v \in Q_2(\operatorname{dom} F_{A+B})$. Hence $\operatorname{ran} A+\operatorname{ran} B \subset Q_2(\operatorname{dom} F_{A+B})$ and therefore

$$\overline{\operatorname{ran} A + \operatorname{ran} B} \subset \overline{Q_2(\operatorname{dom} F_{A+B})} \quad \text{and}$$

$$\operatorname{int}(\operatorname{ran} A + \operatorname{ran} B) \subset \operatorname{int} Q_2(\operatorname{dom} F_{A+B}). \quad (25.29)$$

Since $A+B$ is maximally monotone, Corollary 21.14 yields

$$\overline{Q_2(\operatorname{dom} F_{A+B})} = \overline{\operatorname{ran}}\,(A+B) \quad \text{and} \quad \operatorname{int} Q_2(\operatorname{dom} F_{A+B}) = \operatorname{int}\operatorname{ran}(A+B). \quad (25.30)$$

Altogether,

$$\overline{\operatorname{ran} A + \operatorname{ran} B} \subset \overline{\operatorname{ran}}\,(A+B) \quad \text{and} \quad \operatorname{int}(\operatorname{ran} A + \operatorname{ran} B) \subset \operatorname{int}\operatorname{ran}(A+B). \quad (25.31)$$

The identities follow since the reverse inclusions in (25.31) always hold. ☐

Theorem 25.24 (Brézis–Haraux) *Let A and B be monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$ such that $A+B$ is maximally monotone and one of the following conditions is satisfied:*

(i) *A and B are 3^* monotone.*
(ii) *$\operatorname{dom} A \subset \operatorname{dom} B$ and B is 3^* monotone.*

Then $\overline{\operatorname{ran}}\,(A+B) = \overline{\operatorname{ran} A + \operatorname{ran} B}$ and $\operatorname{int}\operatorname{ran}(A+B) = \operatorname{int}(\operatorname{ran} A + \operatorname{ran} B)$.

Proof. Let $u \in \operatorname{ran} A$ and $v \in \operatorname{ran} B$. In view of Theorem 25.23, it suffices to show that there exists $x \in \mathcal{H}$ such that $(x,u) \in \operatorname{dom} F_A$ and $(x,v) \in \operatorname{dom} F_B$.

(i): For every $x \in \operatorname{dom} A \cap \operatorname{dom} B$, $(x,u) \in \operatorname{dom} A \times \operatorname{ran} A \subset \operatorname{dom} F_A$ and $(x,v) \in \operatorname{dom} B \times \operatorname{ran} B \subset \operatorname{dom} F_B$.

(ii): Since $u \in \operatorname{ran} A$, there exists $x \in \operatorname{dom} A$ such that $(x,u) \in \operatorname{gra} A$. Proposition 20.56(i) yields $(x,u) \in \operatorname{dom} F_A$. Furthermore, since $x \in \operatorname{dom} A$ and $\operatorname{dom} A \subset \operatorname{dom} B$, we have $(x,v) \in \operatorname{dom} B \times \operatorname{ran} B \subset \operatorname{dom} F_B$. ☐

Example 25.25 It follows from Example 25.14 that, in Example 25.1, N_C and N_D are 3^* monotone. However, $\overline{\operatorname{ran}}\,(N_C + N_D) = \mathbb{R} \times \{0\} \neq \mathbb{R}^2 = \overline{\operatorname{ran} N_C + \operatorname{ran} N_D}$. The assumption that $A+B$ be maximally monotone in Theorem 25.24 is therefore critical.

Example 25.26 Suppose that $\mathcal{H} = \mathbb{R}^2$ and that $f \in \Gamma_0(\mathcal{H})$ is as in Remark 16.28. Set $A = B = \partial f^*$. Then A and B are 3^* monotone, $A+B$ is maximally monotone, and $\operatorname{ran}(A+B) \neq \operatorname{ran} A + \operatorname{ran} B$. The closure or interior operation in the conclusion of Theorem 25.24 must therefore not be omitted.

Corollary 25.27 *Let A and B be monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$ such that $A+B$ is maximally monotone, A or B is surjective, and one of the following conditions is satisfied:*

(i) *A and B are 3^* monotone.*
(ii) *$\operatorname{dom} A \subset \operatorname{dom} B$ and B is 3^* monotone.*

Then $A+B$ is surjective.

Corollary 25.28 *Let A and B be maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$ such that $A+B$ is maximally monotone and B is uniformly monotone with a supercoercive modulus. Suppose, in addition, that A is 3^* monotone or that* $\operatorname{dom} A \subset \operatorname{dom} B$. *Then the following hold:*

(i) $\operatorname{ran}(A+B) = \mathcal{H}$.
(ii) $\operatorname{zer}(A+B)$ *is a singleton.*

Proof. (i): Proposition 22.11(i) and Example 25.15(iii) imply that B is surjective and 3^* monotone. We then deduce from Corollary 25.27 that $\operatorname{ran}(A+B) = \mathcal{H}$.

(ii): Since A is monotone and B is strictly monotone, $A+B$ is strictly monotone. Hence, the inclusion $0 \in Ax + Bx$ has at most one solution by Proposition 23.35. Existence follows from (i). □

25.5 Parallel Sum

Definition 25.29 Let A and B be operators from $\mathcal{H}$ to $2^{\mathcal{H}}$. The *parallel sum* of A and B is

$$A \,\Box\, B = \big(A^{-1} + B^{-1}\big)^{-1}. \tag{25.32}$$

Some elementary properties are collected in the next result.

Proposition 25.30 *Let A and B be operators from $\mathcal{H}$ to $2^{\mathcal{H}}$, and let x and u be in $\mathcal{H}$. Then the following hold:*

(i) $(A \,\Box\, B)x = \bigcup_{y \in \mathcal{H}} \big(Ay \cap B(x-y)\big)$.
(ii) $u \in (A \,\Box\, B)x \Leftrightarrow (\exists\, y \in \mathcal{H})\ y \in A^{-1}u$ *and* $x - y \in B^{-1}u$.
(iii) $\operatorname{dom}(A \,\Box\, B) = \operatorname{ran}(A^{-1} + B^{-1})$.
(iv) $\operatorname{ran}(A \,\Box\, B) = \operatorname{ran} A \cap \operatorname{ran} B$.
(v) *Suppose that A and B are monotone. Then $A \,\Box\, B$ is monotone.*

Proposition 25.31 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be at most single-valued and let $B\colon \mathcal{H} \to \mathcal{H}$ be linear. Then*

$$A \,\Box\, B = A(A+B)^{-1}B. \tag{25.33}$$

Proof. Let $(x,u) \in \mathcal{H} \times \mathcal{H}$. Proposition 25.30(ii) yields

$$\begin{aligned}
(x,u) \in \operatorname{gra}(A \,\Box\, B) &\Leftrightarrow (\exists\, y \in \mathcal{H}) \quad y \in A^{-1}u \ \text{ and } \ x - y \in B^{-1}u \\
&\Leftrightarrow (\exists\, y \in \mathcal{H}) \quad u = Ay = Bx - By \\
&\Leftrightarrow (\exists\, y \in \mathcal{H}) \quad u = Ay \ \text{ and } \ (A+B)y = Bx \\
&\Leftrightarrow (\exists\, y \in \mathcal{H}) \quad u = Ay \ \text{ and } \ y \in (A+B)^{-1}Bx \\
&\Leftrightarrow (x,u) \in \operatorname{gra}\big(A(A+B)^{-1}B\big).
\end{aligned} \tag{25.34}$$

Hence (25.33) holds. □

Proposition 25.32 *Let $f \in \Gamma_0(\mathcal{H})$, let $g \in \Gamma_0(\mathcal{H})$, and suppose that $0 \in \operatorname{sri}(\operatorname{dom} f^* - \operatorname{dom} g^*)$. Then*

$$(\partial f) \,\Box\, (\partial g) = \partial(f \boxdot g). \tag{25.35}$$

Proof. Using (25.32), Corollary 16.30, Corollary 16.48(i), Proposition 15.1, and Proposition 15.7(i), we obtain $(\partial f) \,\Box\, (\partial g) = ((\partial f)^{-1} + (\partial g)^{-1})^{-1} = (\partial f^* + \partial g^*)^{-1} = (\partial(f^* + g^*))^{-1} = \partial(f^* + g^*)^* = \partial(f \,\Box\, g)^{\smallsmile} = \partial(f \boxdot g)$. ☐

Proposition 25.33 *Let A and B be monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$. Then*

$$J_A \,\Box\, J_B = J_{(1/2)(A+B)} \circ \tfrac{1}{2}\mathrm{Id}. \tag{25.36}$$

Proof. We have

$$\begin{aligned} J_A \,\Box\, J_B &= \big((\mathrm{Id} + A) + (\mathrm{Id} + B)\big)^{-1} \\ &= \Big(2\big(\mathrm{Id} + \tfrac{1}{2}(A + B)\big)\Big)^{-1} \\ &= \Big(\mathrm{Id} + \tfrac{1}{2}(A + B)\Big)^{-1} \circ \tfrac{1}{2}\mathrm{Id} \\ &= J_{(1/2)(A+B)} \circ \tfrac{1}{2}\mathrm{Id}, \end{aligned} \tag{25.37}$$

as claimed. ☐

Corollary 25.34 *Let A and B be monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$. Then $J_{A+B} = (J_{2A} \,\Box\, J_{2B}) \circ 2\mathrm{Id}$.*

Corollary 25.35 *Let f and g be functions in $\Gamma_0(\mathcal{H})$ such that $\operatorname{dom} f \cap \operatorname{dom} g \neq \varnothing$ and $\partial f + \partial g = \partial(f + g)$. Then*

$$\mathrm{Prox}_{f+g} = (\mathrm{Prox}_{2f} \,\Box\, \mathrm{Prox}_{2g}) \circ 2\mathrm{Id}. \tag{25.38}$$

Corollary 25.36 *Let C and D be closed convex subsets of $\mathcal{H}$ such that $C \cap D \neq \varnothing$ and $N_C + N_D = N_{C \cap D}$. Then*

$$P_{C\cap D} = \big(P_C \,\Box\, P_D\big) \circ 2\mathrm{Id}. \tag{25.39}$$

Proposition 25.37 *Let C and D be closed linear subspaces of $\mathcal{H}$ such that $C + D$ is closed. Then the following hold:*

(i) $P_{C\cap D} = 2P_C(P_C + P_D)^{-1}P_D$.
(ii) *Suppose that* $\operatorname{ran}(P_C + P_D)$ *is closed. Then* $P_{C\cap D} = 2P_C(P_C + P_D)^{\dagger}P_D$.

Proof. Since $C + D$ is closed, it follows from Corollary 15.35 that $C^\perp + D^\perp = \overline{C^\perp + D^\perp} = (C \cap D)^\perp$. Hence, $N_C + N_D = N_{C\cap D}$.

(i): Using Corollary 25.36, the linearity of $P_{C\cap D}$, and Proposition 25.31, we see that $P_{C\cap D} = 2(P_C \,\Box\, P_D) = 2P_D(P_C + P_D)^{-1}P_D$.

(ii): Since, by Corollary 3.24 and Corollary 20.28 the projectors P_C and P_D are continuous, maximally monotone, and self-adjoint, so is their sum $P_C + P_D$. Furthermore, by Proposition 4.16 and Proposition 25.16, P_C and P_D are 3^* monotone. Theorem 25.24 therefore yields

$$\operatorname{ran}(P_C + P_D) = \overline{\operatorname{ran}}\,(P_C + P_D) = \overline{\operatorname{ran} P_C + \operatorname{ran} P_D} = \overline{C + D} = C + D. \tag{25.40}$$

In turn, Exercise 3.14 implies that

$$\begin{aligned}(P_C + P_D)^\dagger &= P_{\operatorname{ran}(P_C+P_D)^*} \circ (P_C + P_D)^{-1} \circ P_{\operatorname{ran}(P_C+P_D)}\\ &= P_{C+D}(P_C + P_D)^{-1} P_{C+D}.\end{aligned} \tag{25.41}$$

On the other hand, by Exercise 4.12, we have $P_C P_{C+D} = P_C$ and $P_{C+D} P_D = P_D$. Altogether,

$$\begin{aligned}P_C(P_C + P_D)^\dagger P_D &= P_C\big(P_{C+D}(P_C + P_D)^{-1} P_{C+D}\big)P_D\\ &= \big(P_C P_{C+D}\big)(P_C + P_D)^{-1}\big(P_{C+D} P_D\big)\\ &= P_C(P_C + P_D)^{-1} P_D.\end{aligned} \tag{25.42}$$

In view of (i), the proof is therefore complete. ☐

Corollary 25.38 (Anderson–Duffin) *Suppose that $\mathcal{H}$ is finite-dimensional, and let C and D be linear subspaces of $\mathcal{H}$. Then*

$$P_{C\cap D} = 2P_C(P_C + P_D)^\dagger P_D. \tag{25.43}$$

25.6 Parallel Composition

Definition 25.39 Let $\mathcal{K}$ be a real Hilbert space, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$. The *parallel composition* of A by L is the operator from $\mathcal{K}$ to $2^{\mathcal{K}}$ given by

$$L \rhd A = \big(L \circ A^{-1} \circ L^*\big)^{-1}. \tag{25.44}$$

First, we justify the terminology via the following connection with the parallel sum operation discussed in Section 25.5.

Example 25.40 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and let $B\colon \mathcal{H} \to 2^{\mathcal{H}}$. Set $\boldsymbol{\mathcal{H}} = \mathcal{H} \oplus \mathcal{H}$, $\boldsymbol{A}\colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}\colon (x,y) \mapsto Ax \times By$, and $L\colon \boldsymbol{\mathcal{H}} \to \mathcal{H}\colon (x,y) \mapsto x + y$. Then $L \rhd \boldsymbol{A} = A \,\square\, B$.

Proof. Since $L^*\colon \mathcal{H} \to \boldsymbol{\mathcal{H}}\colon x \mapsto (x,x)$, the identity is an immediate consequence of (25.32) and (25.44). ☐

We now explore some basic properties of the parallel composition operation.

Proposition 25.41 *Let $\mathcal{K}$ be a real Hilbert space, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$, let $B\colon \mathcal{K} \to 2^{\mathcal{K}}$, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$. Then the following hold:*

(i) $((L \rhd A) \,\Box\, B)^{-1} = L \circ A^{-1} \circ L^* + B^{-1}$.
(ii) *Suppose that A and B are monotone. Then $(L \rhd A) \,\Box\, B$ is monotone.*
(iii) *Suppose that A and B are maximally monotone and that*

$$\operatorname{cone}\big(\operatorname{ran} A - L^*(\operatorname{ran} B)\big) = \overline{\operatorname{span}}\big(\operatorname{ran} A - L^*(\operatorname{ran} B)\big). \tag{25.45}$$

Then $(L \rhd A) \,\Box\, B$ is maximally monotone.
(iv) *Suppose that A is maximally monotone and that* $\operatorname{cone}(\operatorname{ran} A - \operatorname{ran} L^*) = \overline{\operatorname{span}}\,(\operatorname{ran} A - \operatorname{ran} L^*)$. *Then $L \rhd A$ is maximally monotone.*

Proof. (i): This follows from (25.32) and (25.44).
(ii): Combine (i) and Proposition 20.10.
(iii): The operators A^{-1} and B^{-1} are maximally monotone by Proposition 20.22, and $L^*(\operatorname{ran} B) - \operatorname{ran} A = L^*(\operatorname{dom} B^{-1}) - \operatorname{dom} A^{-1}$. Hence, $L \circ A^{-1} \circ L^* + B^{-1}$ is maximally monotone by Theorem 25.3, and so is its inverse which, in view of (i), is $(L \rhd A) \,\Box\, B$.
(iv): Apply (iii) with $\operatorname{gra} B = \{0\} \times \mathcal{K}$. $\square$

Proposition 25.42 *Let $\mathcal{G}$ and $\mathcal{K}$ be real Hilbert spaces, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$, let $B\colon \mathcal{H} \to 2^{\mathcal{H}}$, let $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$, and let $M \in \mathcal{B}(\mathcal{G},\mathcal{K})$. Then the following hold:*

(i) $L \rhd (A \,\Box\, B) = (L \rhd A) \,\Box\, (L \rhd B)$.
(ii) $M \rhd (L \rhd A) = (M \circ L) \rhd A$.

Proof. (i): It follows from (25.32) and (25.44) that

$$\begin{aligned} L \rhd (A \,\Box\, B) &= \big(L \circ (A \,\Box\, B)^{-1} \circ L^*\big)^{-1} \\ &= \big(L \circ (A^{-1} + B^{-1}) \circ L^*\big)^{-1} \\ &= \big(L \circ A^{-1} \circ L^* + L \circ B^{-1} \circ L^*\big)^{-1} \\ &= \big((L \rhd A)^{-1} + (L \rhd B)^{-1}\big)^{-1} \\ &= (L \rhd A) \,\Box\, (L \rhd B). \end{aligned} \tag{25.46}$$

(ii): $M \rhd (L \rhd A) = (M \circ (L \rhd A)^{-1} \circ M^*)^{-1} = (M \circ L \circ A^{-1} \circ L^* \circ M^*)^{-1} = (M \circ L) \rhd A$. $\square$

In the next proposition we make connections with the infimal convolution and postcomposition operations of (12.1) and (12.36).

Proposition 25.43 *Let $\mathcal{K}$ be a real Hilbert space, let $f \in \Gamma_0(\mathcal{H})$, let $g \in \Gamma_0(\mathcal{K})$, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$ be such that $0 \in \operatorname{sri}(\operatorname{dom} f^* - L^*(\operatorname{dom} g^*))$. Then the following hold:*

(i) $(L \rhd f) \,\Box\, g \in \Gamma_0(\mathcal{K})$.
(ii) $\partial((L \rhd f) \,\Box\, g) = (L \rhd \partial f) \,\Box\, \partial g$.

Proof. (i): Since $0 \in L^*(\operatorname{dom} g^*) - \operatorname{dom} f^*$ and, by Corollary 13.38, $f^* \in \Gamma_0(\mathcal{H})$ and $g^* \in \Gamma_0(\mathcal{K})$, we have $f^* \circ L^* + g^* \in \Gamma_0(\mathcal{K})$. Hence $(f^* \circ L^* + g^*)^* \in \Gamma_0(\mathcal{K})$. However, in view of Theorem 15.27(i), the assumptions also imply that $(f^* \circ L^* + g^*)^* = (L \triangleright f) \,\square\, g$.

(ii): Let y and v be in $\mathcal{K}$. Then (i), Corollary 16.30, Proposition 13.24(i) &(iv), and Theorem 16.47(i) enable us to write

$$\begin{aligned} v \in \partial\big((L \triangleright f) \,\square\, g\big)(y) &\Leftrightarrow y \in \Big(\partial\big((L \triangleright f) \,\square\, g\big)\Big)^{-1}(v) \\ &\Leftrightarrow y \in \partial\big((L \triangleright f) \,\square\, g\big)^*(v) \\ &\Leftrightarrow y \in \partial(f^* \circ L^* + g^*)(v) \\ &\Leftrightarrow y \in \big(L \circ (\partial f^*) \circ L^* + \partial g^*\big)(v) \\ &\Leftrightarrow y \in \big(L \circ (\partial f)^{-1} \circ L^* + (\partial g)^{-1}\big)(v) \\ &\Leftrightarrow v \in \big((L \triangleright \partial f) \,\square\, \partial g\big) y, \end{aligned} \tag{25.47}$$

which establishes the announced identity. ☐

Corollary 25.44 *Let $\mathcal{K}$ be a real Hilbert space, let $f \in \Gamma_0(\mathcal{H})$, and let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ be such that $0 \in \operatorname{sri}(\operatorname{ran} L^* - \operatorname{dom} f^*)$. Then the following hold:*

(i) $L \triangleright f \in \Gamma_0(\mathcal{K})$.
(ii) $\partial(L \triangleright f) = L \triangleright \partial f$.

Proof. Set $g = \iota_{\{0\}}$ in Proposition 25.43. ☐

Exercises

Exercise 25.1 Show that Corollary 25.6 fails if $\operatorname{cone}(\operatorname{ran} L - \operatorname{dom} A) \neq \overline{\operatorname{span}}\,(\operatorname{ran} L - \operatorname{dom} A)$.

Exercise 25.2 Is the upper bound for F_{A+B} provided in Proposition 25.21 sharp when A and B are operators in $\mathcal{B}(\mathcal{H})$ such that $A^* = -A$ and $B^* = -B$?

Exercise 25.3 Is the upper bound for F_{A+B} provided in Proposition 25.21 sharp when A and B are operators in $\mathcal{B}(\mathcal{H})$ such that A is self-adjoint and monotone, and $B^* = -B$?

Exercise 25.4 Is the upper bound for F_{A+B} provided in Proposition 25.21 sharp when V is a closed linear subspace of $\mathcal{H}$, $A = P_V$, and $B = P_{V^\perp}$?

Exercise 25.5 Is the upper bound for F_{A+B} provided in Proposition 25.21 sharp when $\mathcal{H} = \mathbb{R}$, $A = P_K$, and $B = P_{K^\perp}$, where $K = \mathbb{R}_+$?

Exercise 25.6 An operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ is *angle bounded* with constant $\beta \in \mathbb{R}_{++}$ if, for all (x,u), (y,v), and (z,w) in $\mathcal{H} \times \mathcal{H}$,

$$\left.\begin{array}{l}(x,u) \in \operatorname{gra} A,\\ (y,v) \in \operatorname{gra} A,\\ (z,w) \in \operatorname{gra} A\end{array}\right\} \quad\Rightarrow\quad \langle y-z \mid w-u\rangle \leqslant \beta\,\langle x-y \mid u-v\rangle\,. \tag{25.48}$$

Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be 3-cyclically monotone. Show that A is angle bounded with constant 1.

Exercise 25.7 Show that every angle bounded monotone operator is 3^* monotone.

Exercise 25.8 Provide an example of a maximally monotone operator that is not 3^* monotone.

Exercise 25.9 Provide an example of maximally monotone operators A and B such that $A+B$ is maximally monotone, $\operatorname{ran}(A+B) = \{0\}$, and $\operatorname{ran} A = \operatorname{ran} B = \mathcal{H}$. Conclude that the assumptions (i) and (ii) of Theorem 25.24 are critical.

Exercise 25.10 Check Example 25.26.

Exercise 25.11 Let R be as in Example 25.18. Prove that

$$\ker(\mathrm{Id}-R) = \big\{(x,\ldots,x) \in \mathcal{H}^N \;\big|\; x \in \mathcal{H}\big\} \tag{25.49}$$

and that

$$\operatorname{ran}(\mathrm{Id}-R) = \big(\ker(\mathrm{Id}-R)\big)^{\perp} = \left\{(x_1,\ldots,x_N) \in \mathcal{H}^N \;\middle|\; \sum_{i=1}^{N} x_i = 0\right\}. \tag{25.50}$$

Exercise 25.12 Let $A \in \mathcal{B}(\mathcal{H})$ be monotone and self-adjoint, and let $B \in \mathcal{B}(\mathcal{H})$ be such that $B^* = -B$. Suppose that $\operatorname{ran} A$ is closed and set $C = A+B$. Show that C is 3^* monotone if and only if $\operatorname{ran} B \subset \operatorname{ran} A$.

Exercise 25.13 Let $(T_i)_{i\in I}$ be a finite family of firmly nonexpansive mappings from $\mathcal{H}$ to $\mathcal{H}$, let $(\omega_i)_{i\in I}$ be real numbers in $]0,1]$, and set $T = \sum_{i\in I} \omega_i T_i$. Show that if, for some $i \in I$, T_i is surjective, then T is likewise.

Exercise 25.14 Let $(A_i)_{\in I}$ be a finite family of maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$, let $(\omega_i)_{i\in I}$ be real numbers in $]0,1]$ such that $\sum_{i\in I} \omega_i = 1$, and consider the corresponding resolvent average $A = \big(\sum_{i\in I} \omega_i J_{A_i}\big)^{-1} - \mathrm{Id}$. Show the following:

(i) If, for some $i \in I$, A_i has full domain, then A is likewise.
(ii) If, for some $i \in I$, A_i is surjective, then A is likewise.

Exercise 25.15 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $z \in \mathcal{H}$, and set

$$B\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x \mapsto \begin{cases} \mathcal{H}, & \text{if } x = z; \\ \varnothing, & \text{otherwise.} \end{cases} \tag{25.51}$$

Determine $A \,\Box\, B$ and $A \,\Box\, B^{-1}$.

Exercise 25.16 Let A and B be strictly monotone self-adjoint surjective operators in $\mathcal{B}(\mathcal{H})$. Set $q_A\colon x \mapsto (1/2)\,\langle x \mid Ax\rangle$. Show that $A+B$ is surjective and observe that the surjectivity assumption in Exercise 12.17 is therefore superfluous. Furthermore, use Exercise 12.17 to deduce that $q_{A \,\Box\, B} = q_A \,\boxdot\, q_B$.

Exercise 25.17 Let A and B be monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$ such that $\operatorname{dom}(A \,\Box\, B) \neq \varnothing$. Let (x, u_1) and (x, u_2) be in $\operatorname{gra}(A \,\Box\, B)$, and let y_1 and y_2 be such that $u_1 \in Ay_1 \cap B(x - y_1)$ and $u_2 \in Ay_2 \cap B(x - y_2)$. Show that $\langle y_1 - y_2 \mid u_1 - u_2\rangle = 0$.

Exercise 25.18 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, and let $\gamma \in \mathbb{R}_{++}$. Show that $A \,\Box\, (\gamma^{-1}\mathrm{Id}) = {}^{\gamma}\!A$.

Exercise 25.19 Let A and B be maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$. Use Lemma 2.14(i) and Exercise 25.17 to show that, for every $x \in \mathcal{H}$, $(A \,\Box\, B)x$ is convex.

Exercise 25.20 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ be paramonotone operators. Show that $A \,\Box\, B$ is paramonotone.

Chapter 26
Zeros of Sums of Monotone Operators

Properties of the zeros of a single monotone operator were discussed in Section 23.4. In this chapter, we first characterize the zeros of sums of monotone operators and then present basic algorithms to construct such zeros iteratively. Duality for monotone inclusion problems is also discussed. In the case of two operators A and B such that $A + B$ is maximally monotone, a point in $\operatorname{zer}(A + B)$ could in principle be constructed via Theorem 23.41. However, this approach is numerically viable only when it is easy to compute $J_{\gamma(A+B)}$. A more widely applicable alternative is to devise an *operator splitting* algorithm, in which the operators A and B are employed in separate steps. Various such algorithms are discussed in this chapter. Splitting methods for finding a zero of composite monotone operators of the form $A+L^*BL$, where L is a linear operator, are also presented.

26.1 Duality and Characterizations

Numerous problems in various areas of nonlinear analysis can be reduced to finding a zero of the sum of two set-valued operators $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{H} \to 2^{\mathcal{H}}$, i.e.,

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + Bx. \tag{26.1}$$

It will be useful to associate with (26.1) the dual inclusion problem

$$\text{find } u \in \mathcal{H} \text{ such that } 0 \in -A^{-1}(-u) + B^{-1}u. \tag{26.2}$$

Note that this dual problem depends on the ordered pair (A, B). As was seen in Chapter 15, in convex analysis, duality is expressed through con-

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-48311-5_31

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_26

jugate functions. More generally, in view of Corollary 16.30, it is natural to express duality in monotone operator theory in terms of inverses and to pair (26.1) with (26.2). Indeed, if $A = \partial f$ and $B = \partial g$ for some functions f and g in $\Gamma_0(\mathcal{H})$ such that $0 \in \operatorname{sri}(\operatorname{dom} f - \operatorname{dom} g)$, then it follows from Theorem 16.3 and Corollary 16.48(i) that (26.1) yields the primal problem (15.22) and (26.2) corresponds to the Fenchel dual problem (15.23). Using the notation (23.12), we characterize the sets of solutions $\mathscr{P} = \operatorname{zer}(A+B)$ of (26.1) and $\mathscr{D} = \operatorname{zer}(-A^{-1} \circ (-\mathrm{Id}) + B^{-1})$ of (26.2) as follows.

Proposition 26.1 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$, let $B\colon \mathcal{H} \to 2^{\mathcal{H}}$, let $\gamma \in \mathbb{R}_{++}$, and denote the sets of solutions to (26.1) and (26.2) by $\mathscr{P}$ and $\mathscr{D}$, respectively. Then the following hold:*

(i) *We have*

$$\mathscr{P} = \big\{x \in \mathcal{H} \;\big|\; (\exists\, u \in \mathscr{D}) \;\; -u \in Ax \;\; \text{and} \;\; u \in Bx\big\} \tag{26.3}$$

and

$$\mathscr{D} = \big\{u \in \mathcal{H} \;\big|\; (\exists\, x \in \mathscr{P}) \;\; x \in A^{-1}(-u) \;\; \text{and} \;\; x \in B^{-1}u\big\}. \tag{26.4}$$

(ii) *Let C be a closed affine subspace of $\mathcal{H}$, set $V = C - C$, and suppose that $A = N_C$. Then*

$$\mathscr{P} = \big\{x \in C \;\big|\; V^{\perp} \cap Bx \neq \varnothing\big\} \;\; \text{and} \;\; \mathscr{D} = \big\{u \in V^{\perp} \;\big|\; C \cap B^{-1}u \neq \varnothing\big\}. \tag{26.5}$$

(iii) *Suppose that A and B are maximally monotone and let*

$$T_{A,B} = J_A R_B + \mathrm{Id} - J_B \tag{26.6}$$

be the associated Douglas–Rachford splitting operator. Then the following hold:

(a) *$T_{\gamma A,\gamma B}$ is firmly nonexpansive.*
(b) $\mathscr{P} = J_{\gamma B}\big(\operatorname{Fix} R_{\gamma A} R_{\gamma B}\big) = J_{\gamma B}\big(\operatorname{Fix} T_{\gamma A,\gamma B}\big)$.
(c) $\mathscr{D} = {}^{\gamma}\!B\big(\operatorname{Fix} R_{\gamma A} R_{\gamma B}\big) = {}^{\gamma}\!B\big(\operatorname{Fix} T_{\gamma A,\gamma B}\big)$.
(d) $\mathscr{P} \neq \varnothing \Leftrightarrow \mathscr{D} \neq \varnothing \Leftrightarrow \operatorname{Fix} T_{\gamma A,\gamma B} \neq \varnothing$.

(iv) *Suppose that A is maximally monotone. Let $B\colon \mathcal{H} \to \mathcal{H}$ and set*

$$T_{A,B} = J_A \circ (\mathrm{Id} - B). \tag{26.7}$$

Then the following hold:

(a) $\mathscr{P} = \operatorname{Fix} T_{\gamma A,\gamma B}$.
(b) $\mathscr{D} = B(\mathscr{P})$.
(c) $\mathscr{P} \neq \varnothing \Leftrightarrow \mathscr{D} \neq \varnothing \Leftrightarrow \operatorname{Fix} T_{\gamma A,\gamma B} \neq \varnothing$.

(d) *Suppose that B is β-cocoercive for some $\beta \in \mathbb{R}_{++}$ and that $\gamma \in \left]0, 2\beta\right[$, and set $\alpha = 2\beta/(4\beta - \gamma)$. Then the associated* forward-backward splitting operator $T_{\gamma A,\gamma B}$ *is α-averaged.*

(v) *Suppose that A is maximally monotone and that $B^{-1} \colon \mathcal{H} \to \mathcal{H}$. Then*

$$\mathscr{D} = \operatorname{Fix}\big(- {}^{\gamma}A \circ (B^{-1} - \gamma \mathrm{Id})\big) \quad \textit{and} \quad \mathscr{P} = B^{-1}(\mathscr{D}). \tag{26.8}$$

Proof. (i): Let $x \in \mathcal{H}$. It follows from (1.7) that

$$\begin{aligned} x \in \mathscr{P} &\Leftrightarrow 0 \in Ax + Bx \\ &\Leftrightarrow (\exists\, u \in \mathcal{H}) \ \ -u \in Ax \ \text{ and } \ u \in Bx \\ &\Leftrightarrow (\exists\, u \in \mathcal{H}) \ \ -x \in -A^{-1}(-u) \ \text{ and } \ x \in B^{-1}u \\ &\Leftrightarrow (\exists\, u \in \mathscr{D}) \ \ -u \in Ax \ \text{ and } \ u \in Bx, \end{aligned} \tag{26.9}$$

which establishes (26.3). An analogous argument yields (26.4).

(ii): Let $x \in \mathcal{H}$. Then it follows from Example 6.43 that $0 \in N_C x + Bx \Leftrightarrow (\exists\, u \in Bx)\ -u \in N_C x \Leftrightarrow [x \in C \text{ and } (\exists\, u \in Bx)\ u \in V^{\perp}]$. Likewise, let $u \in \mathcal{H}$. Then $0 \in -(N_C)^{-1}(-u) + B^{-1}u \Leftrightarrow (\exists\, x \in B^{-1}u)\ x \in (N_C)^{-1}(-u) \Leftrightarrow (\exists\, x \in B^{-1}u)\ -u \in N_C x \Leftrightarrow [u \in V^{\perp} \text{ and } (\exists\, x \in B^{-1}u)\ x \in C]$.

(iii): Corollary 23.9 and Proposition 4.31(ii) yield (iii)(a). Now let $x \in \mathcal{H}$. Then

$$\begin{aligned} x \in \mathscr{P} &\Leftrightarrow (\exists\, y \in \mathcal{H}) \ \ x - y \in \gamma Ax \ \text{ and } \ y - x \in \gamma Bx \\ &\Leftrightarrow (\exists\, y \in \mathcal{H}) \ \ 2x - y \in (\mathrm{Id} + \gamma A)x \ \text{ and } \ x = J_{\gamma B} y \\ &\Leftrightarrow (\exists\, y \in \mathcal{H}) \ \ x = J_{\gamma A}(R_{\gamma B} y) \ \text{ and } \ x = J_{\gamma B} y \\ &\Leftrightarrow (\exists\, y \in \mathcal{H}) \ \ y = 2x - R_{\gamma B} y = R_{\gamma A}(R_{\gamma B} y) \ \text{ and } \ x = J_{\gamma B} y \\ &\Leftrightarrow (\exists\, y \in \operatorname{Fix} R_{\gamma A} R_{\gamma B}) \ \ x = J_{\gamma B} y, \end{aligned} \tag{26.10}$$

which proves the first equality in (iii)(b). For the second equality in (iii)(b) and (iii)(c), use Proposition 4.31(iii). Next, let $u \in \mathcal{H}$. Then it follows from Proposition 23.2(iii) that

$$\begin{aligned} u \in \mathscr{D} &\Leftrightarrow (\exists\, v \in \mathcal{H}) \ \ v - \gamma u \in A^{-1}(-u) \ \text{ and } \ v - \gamma u \in B^{-1}u \\ &\Leftrightarrow (\exists\, v \in \mathcal{H}) \ \ -u \in A(v - \gamma u) \ \text{ and } \ u \in B(v - \gamma u) \\ &\Leftrightarrow (\exists\, v \in \mathcal{H}) \ \ v - 2\gamma u \in (\mathrm{Id} + \gamma A)(v - \gamma u) \ \text{ and } \ u = {}^{\gamma}Bv \\ &\Leftrightarrow (\exists\, v \in \mathcal{H}) \ \ v - \gamma u = J_{\gamma A}(v - 2\gamma u) \ \text{ and } \ u = {}^{\gamma}Bv \\ &\Leftrightarrow (\exists\, v \in \mathcal{H}) \ \ J_{\gamma B} v = J_{\gamma A}(R_{\gamma B} v) \ \text{ and } \ u = {}^{\gamma}Bv \\ &\Leftrightarrow (\exists\, v \in \mathcal{H}) \ \ v = 2J_{\gamma B} v - R_{\gamma B} v = 2J_{\gamma A}(R_{\gamma B} v) - R_{\gamma B} v = R_{\gamma A}(R_{\gamma B} v) \\ &\qquad\qquad\qquad \text{and } \ u = {}^{\gamma}Bv \\ &\Leftrightarrow (\exists\, v \in \operatorname{Fix} R_{\gamma A} R_{\gamma B}) \ \ u = {}^{\gamma}Bv. \end{aligned} \tag{26.11}$$

Finally, (iii)(b) and (iii)(c) imply (iii)(d).

(iv)(a): Let $x \in \mathcal{H}$. We have $x \in \mathscr{P} \Leftrightarrow -Bx \in Ax \Leftrightarrow x - \gamma Bx \in x + \gamma Ax \Leftrightarrow x \in ((\mathrm{Id} + \gamma A)^{-1} \circ (\mathrm{Id} - \gamma B))x \Leftrightarrow x = (J_{\gamma A} \circ (\mathrm{Id} - \gamma B))x$.

(iv)(b): This follows from (26.4).

(iv)(c): This follows from (iv)(a) and (iv)(b).

(iv)(d): On the one hand, Corollary 23.9 and Remark 4.34(iii) imply that $J_{\gamma A}$ is $1/2$-averaged. On the other hand, Proposition 4.39 implies that $\mathrm{Id} - \gamma B$ is $\gamma/(2\beta)$-averaged. Hence, Proposition 4.44 implies that $T_{\gamma A, \gamma B}$ is α-averaged.

(v): Let $u \in \mathcal{H}$. Then it follows from Proposition 23.7(ii) that $u \in \mathscr{D} \Leftrightarrow B^{-1}u \in A^{-1}(-u) \Leftrightarrow -u + \gamma^{-1}B^{-1}u \in (\mathrm{Id} + \gamma^{-1}A^{-1})(-u) \Leftrightarrow -u = J_{\gamma^{-1}A^{-1}}(\gamma^{-1}(B^{-1}u - \gamma u)) \Leftrightarrow u = -({}^{\gamma}A \circ (B^{-1} - \gamma \mathrm{Id}))u$, which yields the formula for $\mathscr{D}$. The formula for $\mathscr{P}$ then follows from (26.3). ☐

Remark 26.2 Items (iii)(b) and (iii)(c) above lead to the nicely symmetric identities $\mathscr{P} = J_B(\operatorname{Fix} R_A R_B) = J_B(\operatorname{Fix} T_{A,B})$ and $\mathscr{D} = (\mathrm{Id} - J_B)(\operatorname{Fix} R_A R_B) = (\mathrm{Id} - J_B)(\operatorname{Fix} T_{A,B})$.

Corollary 26.3 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $B\colon \mathcal{H} \to \mathcal{H}$, and let $\gamma \in \mathbb{R}_{++}$. Then* $\operatorname{Fix} J_A \circ (\mathrm{Id} + \gamma(B - \mathrm{Id})) = \operatorname{Fix} J_{\gamma^{-1}A} \circ B$.

Proof. Proposition 26.1(iv)(a) asserts that $\operatorname{zer}(A + B) = \operatorname{Fix} J_A \circ (\mathrm{Id} - B) = \operatorname{Fix} J_{\gamma A} \circ (\mathrm{Id} - \gamma B)$. Applying the second identity to $\gamma^{-1}A$ instead of A, and to $\mathrm{Id} - B$ instead of B yields the result. ☐

In order to study the zeros of an arbitrary finite sum of maximally monotone operators, we need a few technical facts.

Proposition 26.4 *Let m be an integer such that $m \geqslant 2$, set $I = \{1, \ldots, m\}$, and let $(A_i)_{i \in I}$ be maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$. Furthermore, set*

$$\begin{cases} \boldsymbol{\mathcal{H}} = \bigoplus_{i \in I} \mathcal{H}, \\ \boldsymbol{D} = \big\{(x, \ldots, x) \in \boldsymbol{\mathcal{H}} \bigm| x \in \mathcal{H}\big\}, \\ \boldsymbol{j}\colon \mathcal{H} \to \boldsymbol{D}\colon x \mapsto (x, \ldots, x), \\ \boldsymbol{A}\colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}\colon (x_i)_{i \in I} \mapsto \bigtimes_{i \in I} A_i x_i. \end{cases} \tag{26.12}$$

Then the following hold for every $\boldsymbol{x} = (x_i)_{i \in I} \in \boldsymbol{\mathcal{H}}$:

(i) $\boldsymbol{D}^{\perp} = \big\{\boldsymbol{u} \in \boldsymbol{\mathcal{H}} \bigm| \sum_{i \in I} u_i = 0\big\}$.

(ii) $N_{\boldsymbol{D}}\boldsymbol{x} = \begin{cases} \big\{\boldsymbol{u} \in \boldsymbol{\mathcal{H}} \bigm| \sum_{i \in I} u_i = 0\big\}, & \text{if } \boldsymbol{x} \in \boldsymbol{D}; \\ \varnothing, & \text{otherwise.} \end{cases}$

(iii) $P_{\boldsymbol{D}}\boldsymbol{x} = \boldsymbol{j}\big((1/m) \sum_{i \in I} x_i\big)$.

(iv) $P_{\boldsymbol{D}^{\perp}}\boldsymbol{x} = \big(x_i - (1/m) \sum_{j \in I} x_j\big)_{i \in I}$.

(v) $\boldsymbol{A}$ *is maximally monotone.*
(vi) $J_{\gamma\boldsymbol{A}}\boldsymbol{x} = \big(J_{\gamma A_i}x_i\big)_{i\in I}$.
(vii) $\boldsymbol{j}\big(\operatorname{zer}\sum_{i\in I} A_i\big) = \operatorname{zer}\big(N_{\boldsymbol{D}} + \boldsymbol{A}\big)$.
(viii) *Let* $x \in \mathcal{H}$ *and set* $\boldsymbol{x} = \boldsymbol{j}(x)$*. Then*

$$x \in \operatorname{zer}\Bigg(\sum_{i\in I} A_i\Bigg) \quad\Leftrightarrow\quad \big(\exists\, \boldsymbol{u} \in \boldsymbol{D}^{\perp}\big) \quad \boldsymbol{u} \in \boldsymbol{A}\boldsymbol{x}. \tag{26.13}$$

Proof. (i): This follows from (2.6).

(ii): Combine (i) and Example 6.43.

(iii): Set $p = (1/m)\sum_{i\in I} x_i$ and $\boldsymbol{p} = \boldsymbol{j}(p)$, and let $\boldsymbol{y} = \boldsymbol{j}(y)$, where $y \in \mathcal{H}$. Then $\boldsymbol{p} \in \boldsymbol{D}$, $\boldsymbol{y}$ is an arbitrary point in $\boldsymbol{D}$, and $\langle \boldsymbol{x} - \boldsymbol{p} \mid \boldsymbol{y}\rangle = \sum_{i\in I}\langle x_i - p \mid y\rangle = \big\langle \sum_{i\in I} x_i - mp \mid y\big\rangle = 0$. Hence, by Corollary 3.24(i), $\boldsymbol{p} = P_{\boldsymbol{D}}\boldsymbol{x}$.

(iv): Corollary 3.24(v).

(v)&(vi): Proposition 23.18.

(vii): Let $x \in \mathcal{H}$. Then (ii) implies that

$$\begin{aligned}
0 \in \sum_{i\in I} A_i x &\Leftrightarrow \Bigg(\exists\,(u_i)_{i\in I} \in \bigtimes_{i\in I} A_i x\Bigg) \ \sum_{i\in I} u_i = 0 \\
&\Leftrightarrow \big(\exists\, \boldsymbol{u} \in \boldsymbol{A}\boldsymbol{j}(x)\big) \quad -\boldsymbol{u} \in \boldsymbol{D}^{\perp} = N_{\boldsymbol{D}}\,\boldsymbol{j}(x) \qquad (26.14)\\
&\Leftrightarrow \boldsymbol{j}(x) \in \operatorname{zer}(N_{\boldsymbol{D}} + \boldsymbol{A}) \subset \boldsymbol{D}, \qquad (26.15)
\end{aligned}$$

and we obtain the announced identity.

(viii): See (26.14). ☐

Proposition 26.5 *Let* $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ *be maximally monotone, let* $\mathcal{K}$ *be a real Hilbert space, let* $B\colon \mathcal{K} \to 2^{\mathcal{K}}$ *be maximally monotone, and let* $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$*. Let* $(x_n,u_n)_{n\in\mathbb{N}}$ *be a sequence in* $\operatorname{gra} A$*, let* $(y_n,v_n)_{n\in\mathbb{N}}$ *be a sequence in* $\operatorname{gra} B$*, let* $x \in \mathcal{H}$*, and let* $v \in \mathcal{K}$*. Suppose that* $x_n \rightharpoonup x$*,* $v_n \rightharpoonup v$*,* $Lx_n - y_n \to 0$*, and* $u_n + L^*v_n \to 0$*. Then the following hold:*

(i) $(x, -L^*v) \in \operatorname{gra} A$ *and* $(Lx, v) \in \operatorname{gra} B$.
(ii) $x \in \operatorname{zer}(A + L^*\circ B\circ L)$ *and* $v \in \operatorname{zer}(-L\circ A^{-1}\circ(-L^*) + B^{-1})$.

Proof. (i): Set $\boldsymbol{\mathcal{H}} = \mathcal{H}\oplus\mathcal{K}$, $\boldsymbol{M}\colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}\colon (x,v) \mapsto Ax \times B^{-1}v$ and $\boldsymbol{S}\colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}\colon (x,v) \mapsto (L^*v, -Lx)$. Since A and B^{-1} are maximally monotone, so is $\boldsymbol{M}$ on account of Proposition 20.22 and Proposition 20.23. On the other hand, $\boldsymbol{S}$ is linear, bounded, and monotone since

$$(\forall (x,v) \in \boldsymbol{\mathcal{H}}) \quad \langle \boldsymbol{S}(x,v) \mid (x,v)\rangle = \langle x \mid L^*v\rangle + \langle -Lx \mid v\rangle = 0. \tag{26.16}$$

Thus, Example 20.34 asserts that $\boldsymbol{S}$ is maximally monotone with $\operatorname{dom}\boldsymbol{S} = \boldsymbol{\mathcal{H}}$. In turn, we derive from Corollary 25.5(i) that

$$\boldsymbol{M} + \boldsymbol{S} \text{ is maximally monotone.} \tag{26.17}$$

Now set $(\forall n \in \mathbb{N})$ $\boldsymbol{x}_n = (x_n, v_n)$ and $\boldsymbol{u}_n = (u_n + L^* v_n, y_n - Lx_n)$. Then $\boldsymbol{x}_n \rightharpoonup (x, v)$, $\boldsymbol{u}_n \to \mathbf{0}$, and $(\forall n \in \mathbb{N})$ $(\boldsymbol{x}_n, \boldsymbol{u}_n) \in \operatorname{gra}(\boldsymbol{M} + \boldsymbol{S})$. Hence, it follows from (26.17) and Proposition 20.38(ii) that $\mathbf{0} \in (\boldsymbol{M} + \boldsymbol{S})(x, v)$, which yields the announced inclusions.

(ii): By (i), $(-L^* v, v) \in Ax \times B(Lx)$ and therefore $(-L^* v, L^* v) \in Ax \times L^*(B(Lx))$, which yields $0 \in Ax + L^*(B(Lx))$. Likewise, since $(x, Lx) \in A^{-1}(-L^* v) \times B^{-1} v$, we have $(-Lx, Lx) \in -L(A^{-1}(-L^* v)) \times B^{-1} v$ and therefore $0 \in -L(A^{-1}(-L^* v)) + B^{-1} v$. ☐

Corollary 26.6 *Let A and B be maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$, let $(x_n, u_n)_{n\in\mathbb{N}}$ be a sequence in $\operatorname{gra} A$, let $(y_n, v_n)_{n\in\mathbb{N}}$ be a sequence in $\operatorname{gra} B$, let $x \in \mathcal{H}$, and let $u \in \mathcal{H}$. Suppose that*

$$x_n \rightharpoonup x, \quad u_n \rightharpoonup u, \quad x_n - y_n \to 0, \quad \textit{and} \quad u_n + v_n \to 0. \tag{26.18}$$

Then $x \in \operatorname{zer}(A + B)$, $-u \in \operatorname{zer}(-A^{-1} \circ (-\mathrm{Id}) + B^{-1})$, $(x, u) \in \operatorname{gra} A$, and $(x, -u) \in \operatorname{gra} B$.

Proof. Apply Proposition 26.5 with $\mathcal{K} = \mathcal{H}$ and $L = \mathrm{Id}$. ☐

The following asymptotic result complements Proposition 26.1(ii).

Example 26.7 Let V be a closed linear subspace of $\mathcal{H}$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $(x_n, u_n)_{n\in\mathbb{N}}$ be a sequence in $\operatorname{gra} A$, and let $(x, u) \in \mathcal{H} \times \mathcal{H}$. Suppose that

$$x_n \rightharpoonup x, \quad u_n \rightharpoonup u, \quad P_{V^\perp} x_n \to 0, \quad \text{and} \quad P_V u_n \to 0. \tag{26.19}$$

Then the following hold:

(i) $x \in \operatorname{zer}(N_V + A)$.
(ii) $(x, u) \in (V \times V^\perp) \cap \operatorname{gra} A$.
(iii) $\langle x_n \mid u_n \rangle \to \langle x \mid u \rangle = 0$.

Proof. (i): In view of Example 6.43, this is an application of Corollary 26.6 with $B = N_V$ and $(\forall n \in \mathbb{N})$ $y_n = P_V x_n$ and $v_n = P_V u_n - u_n$.

(ii)&(iii): This follows from Proposition 20.60. ☐

Corollary 26.8 *Let m be an integer such that $m \geqslant 2$, set $I = \{1, \ldots, m\}$, and let $(A_i)_{i\in I}$ be maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$. For every $i \in I$, let $(x_{i,n}, u_{i,n})_{n\in\mathbb{N}}$ be a sequence in $\operatorname{gra} A_i$ and let $(x_i, u_i) \in \mathcal{H} \times \mathcal{H}$. Suppose that*

$$\sum_{i\in I} u_{i,n} \to 0 \quad \textit{and} \quad (\forall i \in I) \begin{cases} x_{i,n} \rightharpoonup x_i, \\ u_{i,n} \rightharpoonup u_i, \\ m x_{i,n} - \displaystyle\sum_{j\in I} x_{j,n} \to 0. \end{cases} \tag{26.20}$$

Then there exists $x \in \operatorname{zer} \sum_{i \in I} A_i$ *such that the following hold:*

(i) $x = x_1 = \cdots = x_m$.

(ii) $\sum_{i \in I} u_i = 0$.

(iii) $(\forall i \in I)\ (x, u_i) \in \operatorname{gra} A_i$.

(iv) $\sum_{i \in I} \langle x_{i,n} \mid u_{i,n} \rangle \to \big\langle x \mid \sum_{i \in I} u_i \big\rangle = 0$.

Proof. Define $\boldsymbol{\mathcal{H}}$, $\boldsymbol{D}$, and $\boldsymbol{A}$ as in Proposition 26.4, set $\boldsymbol{x} = (x_i)_{i \in I}$ and $\boldsymbol{u} = (u_i)_{i \in I}$, and note that, by Proposition 23.18, $\boldsymbol{A}$ is maximally monotone. Now set $(\forall n \in \mathbb{N})$ $\boldsymbol{x}_n = (x_{i,n})_{i \in I}$ and $\boldsymbol{u}_n = (u_{i,n})_{i \in I}$. Then $(\boldsymbol{x}_n, \boldsymbol{u}_n)_{n \in \mathbb{N}}$ lies in $\operatorname{gra} \boldsymbol{A}$ and we derive from (26.20), Proposition 26.4(iii), and Proposition 26.4(iv), that $\boldsymbol{x}_n \rightharpoonup \boldsymbol{x}$, $\boldsymbol{u}_n \rightharpoonup \boldsymbol{u}$, $P_{\boldsymbol{D}^\perp} \boldsymbol{x}_n \to 0$, and $P_{\boldsymbol{D}} \boldsymbol{u}_n \to 0$. The assertions hence follow from Example 26.7 and Proposition 26.4(i)&(vii). ☐

26.2 Spingarn's Method of Partial Inverses

The method of partial inverses (see Section 20.3) is an application of the proximal-point algorithm.

Theorem 26.9 (Spingarn) *Let* $A \colon \mathcal{H} \to 2^{\mathcal{H}}$ *be maximally monotone and let* V *be a closed linear subspace of* $\mathcal{H}$*. Suppose that the problem*

$$\text{find } x \in V \text{ and } u \in V^\perp \text{ such that } u \in Ax \tag{26.21}$$

has at least one solution, let $x_0 \in V$*, let* $u_0 \in V^\perp$*, and set*

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} y_n = J_A(x_n + u_n), \\ v_n = x_n + u_n - y_n, \\ (x_{n+1}, u_{n+1}) = (P_V y_n, P_{V^\perp} v_n). \end{array} \right. \end{array} \tag{26.22}$$

Then there exists a solution (x, u) *to* (26.21) *such that* $x_n \rightharpoonup x$ *and* $u_n \rightharpoonup u$.

Proof. Set $(\forall n \in \mathbb{N})$ $z_n = x_n + u_n$. Then, since $(x_n)_{n \in \mathbb{N}}$ lies in V and $(u_n)_{n \in \mathbb{N}}$ lies in $V^\perp$, (26.22) yields

$$\begin{aligned} (\forall n \in \mathbb{N}) \quad & P_V z_{n+1} + P_{V^\perp}(z_n - z_{n+1}) \\ &= x_{n+1} + P_{V^\perp} u_n - u_{n+1} \\ &= P_V y_n + P_{V^\perp}(v_n - x_n + y_n) - P_{V^\perp} v_n \\ &= P_V y_n + P_{V^\perp} v_n + P_{V^\perp} y_n - P_{V^\perp} v_n \\ &= J_A z_n. \end{aligned} \tag{26.23}$$

Hence, it follows from Proposition 23.30(ii) that (26.22) reduces to the proximal-point iterations $(\forall n \in \mathbb{N})$ $z_{n+1} = J_{A_V} z_n$. Therefore, we derive from

Proposition 20.44(vi) and Example 23.40(i) that $(z_n)_{n\in\mathbb{N}}$ converges weakly to some $z \in \operatorname{zer} A_V$ and that $(x,u) = (P_V z, P_{V^\perp} z)$ solves (26.21). Finally, Lemma 2.41 yields $x_n = P_V z_n \rightharpoonup P_V z$ and $u_n = P_{V^\perp} z_n \rightharpoonup P_{V^\perp} z$. ☐

The following application of the method of partial inverses leads to a parallel splitting method for finding a zero of a finite sum of maximally monotone operators.

Proposition 26.10 *Let m be an integer such that $m \geqslant 2$, set $I = \{1,\ldots,m\}$, let $(A_i)_{i\in I}$ be maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$ such that $\operatorname{zer}\sum_{i\in I} A_i \neq \varnothing$, let $x_0 \in \mathcal{H}$, and let $(u_{i,0})_{i\in I} \in \mathcal{H}^m$ be such that $\sum_{i\in I} u_{i,0} = 0$. Set*

$$
\begin{array}{l}
\textit{for } n = 0,1,\ldots \\
\left\lfloor
\begin{array}{l}
\textit{for every } i \in I \\
\left\lfloor
\begin{array}{l}
y_{i,n} = J_{A_i}(x_n + u_{i,n}), \\
v_{i,n} = x_n + u_{i,n} - y_{i,n},
\end{array}
\right. \\
x_{n+1} = \dfrac{1}{m}\displaystyle\sum_{i\in I} y_{i,n}, \\
\textit{for every } i \in I \\
\left\lfloor
\begin{array}{l}
u_{i,n+1} = v_{i,n} - \dfrac{1}{m}\displaystyle\sum_{j\in I} v_{j,n}.
\end{array}
\right.
\end{array}
\right.
\end{array}
\tag{26.24}
$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{zer}\sum_{i\in I} A_i$.

Proof. Define $\boldsymbol{\mathcal{H}}$, $\boldsymbol{D}$, $\boldsymbol{j}$, and $\boldsymbol{A}$ as in (26.12). Then $\boldsymbol{A}$ is maximally monotone by Proposition 26.4(v) and it follows from Proposition 26.4(viii) that a point $x \in \operatorname{zer}(\sum_{i\in I} A_i)$ can be found by solving

$$\text{find } \boldsymbol{x} \in \boldsymbol{D} \text{ and } \boldsymbol{u} \in \boldsymbol{D}^\perp \text{ such that } \boldsymbol{u} \in \boldsymbol{A}\boldsymbol{x} \tag{26.25}$$

since, if $(\boldsymbol{x},\boldsymbol{u})$ solves (26.25), then $x = \boldsymbol{j}^{-1}(\boldsymbol{x}) \in \operatorname{zer}(\sum_{i\in I} A_i)$. As seen in Theorem 26.9, given $\boldsymbol{x}_0 \in \boldsymbol{D}$ and $\boldsymbol{u}_0 \in \boldsymbol{D}^\perp$, the algorithm

$$
\begin{array}{l}
\text{for } n = 0,1,\ldots \\
\left\lfloor
\begin{array}{l}
\boldsymbol{y}_n = J_{\boldsymbol{A}}(\boldsymbol{x}_n + \boldsymbol{u}_n), \\
\boldsymbol{v}_n = \boldsymbol{x}_n + \boldsymbol{u}_n - \boldsymbol{y}_n, \\
(\boldsymbol{x}_{n+1}, \boldsymbol{u}_{n+1}) = (P_{\boldsymbol{D}}\boldsymbol{y}_n, P_{\boldsymbol{D}^\perp}\boldsymbol{v}_n),
\end{array}
\right.
\end{array}
\tag{26.26}
$$

produces sequences $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ and $(\boldsymbol{u}_n)_{n\in\mathbb{N}}$ such that $\boldsymbol{x}_n \rightharpoonup \boldsymbol{x}$ and $\boldsymbol{u}_n \rightharpoonup \boldsymbol{u}$, where $(\boldsymbol{x},\boldsymbol{u})$ solves (26.25). Now write $(\forall n \in \mathbb{N})$ $\boldsymbol{x}_n = \boldsymbol{j}(x_n)$, $\boldsymbol{y}_n = (y_{i,n})_{i\in I}$, $\boldsymbol{u}_n = (u_{i,n})_{i\in I}$, and $\boldsymbol{v}_n = (v_{i,n})_{i\in I}$. Then, using items (iii), (iv), and (vi) in Proposition 26.4, we can rewrite (26.26) as (26.24). Altogether, $x_n = \boldsymbol{j}^{-1}(\boldsymbol{x}_n) \rightharpoonup \boldsymbol{j}^{-1}(\boldsymbol{x}_n) = x \in \operatorname{zer}(\sum_{i\in I} A_i)$. ☐

26.3 Douglas–Rachford Splitting Algorithm

The following splitting algorithm for solving (26.1) is referred to as the *Douglas–Rachford* algorithm because, in a special affine case, it is akin to a method proposed by Douglas and Rachford for solving certain matrix equations.

Theorem 26.11 (Douglas–Rachford algorithm) *Let A and B be maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$ such that $\operatorname{zer}(A+B) \neq \varnothing$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0,2]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$, and let $\gamma \in \mathbb{R}_{++}$. Let $\mathscr{P}$ be the set of solutions to the problem*

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + Bx, \tag{26.27}$$

and let $\mathscr{D}$ be the set of solutions to the dual problem

$$\text{find } u \in \mathcal{H} \text{ such that } 0 \in -A^{-1}(-u) + B^{-1}u. \tag{26.28}$$

Let $y_0 \in \mathcal{H}$, and set

$$\begin{array}{l} \text{for } n = 0,1,\ldots \\ \left\lfloor \begin{array}{l} x_n = J_{\gamma B} y_n, \\ u_n = \gamma^{-1}(y_n - x_n), \\ z_n = J_{\gamma A}(2x_n - y_n), \\ y_{n+1} = y_n + \lambda_n(z_n - x_n). \end{array} \right. \end{array} \tag{26.29}$$

Then there exists $y \in \operatorname{Fix} R_{\gamma A} R_{\gamma B}$ such that $y_n \rightharpoonup y$. Now set $x = J_{\gamma B} y$ and $u = {}^{\gamma}B y$. Then the following hold:

(i) *$x \in \mathscr{P}$ and $u \in \mathscr{D}$.*
(ii) *$x_n - z_n \to 0$.*
(iii) *$x_n \rightharpoonup x$ and $z_n \rightharpoonup x$.*
(iv) *$u_n \rightharpoonup u$.*
(v) *Suppose that $A = N_C$, where C is a closed affine subspace of $\mathcal{H}$. Then $P_C y_n \rightharpoonup x$.*
(vi) *Suppose that one of the following holds (see Remark 22.3):*
   (a) *A is uniformly monotone on every nonempty bounded subset of $\operatorname{dom} A$.*
   (b) *B is uniformly monotone on every nonempty bounded subset of $\operatorname{dom} B$.*

   Then $\mathscr{P} = \{x\}$, $x_n \to x$, and $z_n \to x$.

Proof. Set $T = J_{\gamma A} R_{\gamma B} + \mathrm{Id} - J_{\gamma B}$. Then it follows from Corollary 23.11(i) and Proposition 4.31(i)–(iii) that $\operatorname{Fix} T = \operatorname{Fix} R_{\gamma A} R_{\gamma B}$ and that T is firmly nonexpansive. In addition, we derive from (26.29) that

$$(\forall n \in \mathbb{N}) \quad y_{n+1} = y_n + \lambda_n\big(Ty_n - y_n\big) \tag{26.30}$$

and from Proposition 26.1(iii)(d) that $\operatorname{Fix} T \neq \varnothing$. Therefore, Proposition 5.4(i) and Corollary 5.17 assert that $(y_n)_{n\in\mathbb{N}}$ is a bounded sequence, that

$$z_n - x_n = J_{\gamma A}(2x_n - y_n) - x_n = Ty_n - y_n \to 0, \tag{26.31}$$

and that $(y_n)_{n\in\mathbb{N}}$ converges weakly to a point $y \in \operatorname{Fix} T$. Proposition 23.2(iii) entails that

$$(x, u) = (y - \gamma u, u) \in \operatorname{gra} B. \tag{26.32}$$

It will be convenient to set

$$(\forall n \in \mathbb{N}) \quad w_n = \gamma^{-1}(2x_n - y_n - z_n), \tag{26.33}$$

and to observe that (26.29) and Proposition 23.2(ii) yield

$$(\forall n \in \mathbb{N}) \quad \begin{cases} (z_n, w_n) \in \operatorname{gra} A, \\ (x_n, u_n) \in \operatorname{gra} B, \\ x_n - z_n = \gamma(u_n + w_n). \end{cases} \tag{26.34}$$

(i): Proposition 26.1(iii)(b)–(iii)(c).

(ii): See (26.31).

(iii): By Corollary 23.11(i), $(\forall n \in \mathbb{N})$ $\|x_n - x_0\| = \|J_{\gamma B} y_n - J_{\gamma B} y_0\| \leqslant \|y_n - y_0\|$. Hence, since $(y_n)_{n\in\mathbb{N}}$ is bounded, so is $(x_n)_{n\in\mathbb{N}}$. Now let z be a weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$, say $x_{k_n} \rightharpoonup z$. Then, since $y_{k_n} \rightharpoonup y$, it follows from (ii), (26.33), and (26.34) that

$$\begin{cases} z_{k_n} \rightharpoonup z, \\ w_{k_n} \rightharpoonup \gamma^{-1}(z - y), \\ z_n - x_n \to 0, \\ w_n + u_n = \gamma^{-1}(x_n - z_n) \to 0. \end{cases} \tag{26.35}$$

In turn, Corollary 26.6 yields $z \in \operatorname{zer}(A + B) = \mathscr{P}$,

$$\big(z, \gamma^{-1}(z - y)\big) \in \operatorname{gra} A, \quad \text{and} \quad \big(z, \gamma^{-1}(y - z)\big) \in \operatorname{gra} B. \tag{26.36}$$

Hence, Proposition 23.2(ii) implies that

$$z = J_{\gamma B} y \quad \text{and} \quad z \in \operatorname{dom} A. \tag{26.37}$$

Thus, $x = J_{\gamma B} y$ is the unique weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ and, appealing to Lemma 2.46, we obtain $x_n \rightharpoonup x$ and then, in view of (ii), $z_n \rightharpoonup x$. Moreover, we derive from (26.36) that

$$(x, -u) = \big(z, \gamma^{-1}(z - y)\big) \in \operatorname{gra} A. \tag{26.38}$$

(iv): We have $y_n \rightharpoonup y$ and, by (iii), $x_n \rightharpoonup x$. Hence,

$$u_n = \gamma^{-1}(y_n - x_n) \rightharpoonup \frac{1}{\gamma}(y - x) = \frac{1}{\gamma}(y - J_{\gamma B} y) = {}^{\gamma}B y = u. \tag{26.39}$$

(v): On the one hand, since $y_n \rightharpoonup y$, Proposition 4.19(i) and Example 23.4 yield $P_C y_n \rightharpoonup P_C y = J_{\gamma A} y$. On the other hand, by Proposition 4.31(iv), $P_C y = J_{\gamma B} y$. Altogether, $P_C y_n \rightharpoonup J_{\gamma B} y = x$.

(vi): We have $y_n \rightharpoonup y$ and, as seen in (i), $x \in \mathscr{P}$. Since A or B is strictly monotone, so is $A+B$. Hence Proposition 23.35 yields $\mathscr{P} = \operatorname{zer}(A+B) = \{x\}$. It follows from (iii), (iv), and (26.33) that

$$z_n \rightharpoonup x, \quad w_n \rightharpoonup -u, \quad x_n \rightharpoonup x, \quad \text{and} \quad u_n \rightharpoonup u. \tag{26.40}$$

(vi)(a): Set $C = \{x\} \cup \{z_n\}_{n\in\mathbb{N}}$. Since (iii) and (26.34) imply that $(z_n)_{n\in\mathbb{N}}$ is a bounded sequence in $\operatorname{dom} A$, and since $x \in \mathscr{P} \subset \operatorname{dom} A$, C is a bounded subset of $\operatorname{dom} A$. Hence, it follows from (26.34), (26.38), and (22.7) that there exists an increasing function $\phi_A \colon \mathbb{R}_+ \to [0, +\infty]$ that vanishes only at 0 such that

$$(\forall n \in \mathbb{N}) \quad \phi_A\big(\|z_n - x\|\big) \leqslant \langle z_n - x \mid w_n + u\rangle . \tag{26.41}$$

On the other hand, by (26.32), (26.34), and the monotonicity of B,

$$(\forall n \in \mathbb{N}) \quad 0 \leqslant \langle x_n - x \mid u_n - u\rangle . \tag{26.42}$$

Thus, using (26.34), we obtain

$$\begin{aligned}
(\forall n \in \mathbb{N}) \quad \phi_A\big(\|z_n - x\|\big) &\leqslant \langle z_n - x \mid w_n + u\rangle + \langle x_n - x \mid u_n - u\rangle \\
&= \langle z_n - x_n \mid w_n + u\rangle + \langle x_n - x \mid w_n + u\rangle \\
&\quad + \langle x_n - x \mid u_n - u\rangle \\
&= \langle z_n - x_n \mid w_n + u\rangle + \langle x_n - x \mid w_n + u_n\rangle \\
&= \langle z_n - x_n \mid w_n + u\rangle + \gamma^{-1} \langle x_n - x \mid x_n - z_n\rangle \\
&= \big\langle z_n - x_n \mid w_n + \gamma^{-1}(y - x_n)\big\rangle .
\end{aligned} \tag{26.43}$$

However, it follows from (ii) that $z_n - x_n \to 0$ and from (26.40) that $w_n + \gamma^{-1}(y - x_n) \rightharpoonup 0$. Thus, it follows from Lemma 2.51(iii) that $\phi_A(\|z_n - x\|) \to 0$ and therefore $z_n \to x$. In turn, (ii) yields $x_n \to x$.

(vi)(b): We argue as in (vi)(a), except that the roles of A and B are interchanged. Consequently, (26.43) becomes

$$\begin{aligned}
(\forall n \in \mathbb{N}) \quad \phi_B\big(\|x_n - x\|\big) &\leqslant \langle x_n - x \mid u_n - u\rangle + \langle z_n - x \mid w_n + u\rangle \\
&\leqslant \langle x_n - z_n \mid u_n - u\rangle + \langle z_n - x \mid u_n - u\rangle \\
&\quad + \langle z_n - x \mid w_n + u\rangle \\
&= \langle x_n - z_n \mid u_n - u\rangle + \langle z_n - x \mid u_n + w_n\rangle \\
&= \langle x_n - z_n \mid u_n - u\rangle + \gamma^{-1} \langle z_n - x \mid x_n - z_n\rangle \\
&= \big\langle x_n - z_n \mid u_n + \gamma^{-1}(z_n - y)\big\rangle .
\end{aligned} \tag{26.44}$$

Thus, since $x_n - z_n \to 0$ and $u_n + \gamma^{-1}(z_n - y) \rightharpoonup 0$, we get $\phi_B(\|x_n - x\|) \to 0$, hence $x_n \to x$ and $z_n \to x$. ☐

By recasting the Douglas–Rachford algorithm (26.29) in a product space, we obtain a parallel splitting algorithm for finding a zero of a finite sum of maximally monotone operators.

Proposition 26.12 (Parallel splitting algorithm) *Let m be an integer such that $m \geqslant 2$, set $I = \{1,\ldots,m\}$, let $(A_i)_{i\in I}$ be maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$ such that $\operatorname{zer}\sum_{i\in I} A_i \neq \varnothing$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0,2]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$, let $\gamma \in \mathbb{R}_{++}$, and let $(y_{i,0})_{i\in I} \in \mathcal{H}^m$. Set*

$$
\begin{array}{l}
\text{for } n = 0,1,\ldots \\
\left\lfloor
\begin{array}{l}
p_n = \dfrac{1}{m}\displaystyle\sum_{i\in I} y_{i,n}, \\
(\forall i \in I)\ \ x_{i,n} = J_{\gamma A_i} y_{i,n}, \\
q_n = \dfrac{1}{m}\displaystyle\sum_{i\in I} x_{i,n}, \\
(\forall i \in I)\ \ y_{i,n+1} = y_{i,n} + \lambda_n\big(2q_n - p_n - x_{i,n}\big).
\end{array}
\right.
\end{array}
\tag{26.45}
$$

Then $(p_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{zer}\sum_{i\in I} A_i$.

Proof. Define $\boldsymbol{\mathcal{H}}$, $\boldsymbol{D}$, and $\boldsymbol{j}$ as in (26.12), and set $\boldsymbol{A} = N_{\boldsymbol{D}}$ and $\boldsymbol{B}\colon (x_i)_{i\in I} \mapsto \bigtimes_{i\in I} A_i x_i$. It follows from Example 20.26, Example 23.4, and items (iii), (v), and (vi) in Proposition 26.4 that $\boldsymbol{A}$ and $\boldsymbol{B}$ are maximally monotone, with

$$
(\forall \boldsymbol{x} \in \boldsymbol{\mathcal{H}}) \quad J_{\gamma \boldsymbol{A}}\boldsymbol{x} = \boldsymbol{j}\bigg(\frac{1}{m}\sum_{i\in I} x_i\bigg) \quad \text{and} \quad J_{\gamma \boldsymbol{B}}\boldsymbol{x} = \big(J_{\gamma A_i} x_i\big)_{i\in I}. \tag{26.46}
$$

Moreover, Proposition 26.4(vii) yields

$$
\boldsymbol{j}\bigg(\operatorname{zer}\sum_{i\in I} A_i\bigg) = \operatorname{zer}\big(\boldsymbol{A} + \boldsymbol{B}\big). \tag{26.47}
$$

Now write $(\forall n \in \mathbb{N})$ $\boldsymbol{x}_n = (x_{i,n})_{i\in I}$, $\boldsymbol{p}_n = \boldsymbol{j}(p_n)$, $\boldsymbol{y}_n = (y_{i,n})_{i\in I}$, and $\boldsymbol{q}_n = \boldsymbol{j}(q_n)$. Then it follows from (26.46) that (26.45) can be rewritten as

$$
\begin{array}{l}
\text{for } n = 0,1,\ldots \\
\left\lfloor
\begin{array}{l}
\boldsymbol{p}_n = P_{\boldsymbol{D}}\boldsymbol{y}_n, \\
\boldsymbol{x}_n = J_{\gamma \boldsymbol{B}}\boldsymbol{y}_n, \\
\boldsymbol{q}_n = P_{\boldsymbol{D}}\boldsymbol{x}_n, \\
\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + \lambda_n\big(2\boldsymbol{q}_n - \boldsymbol{p}_n - \boldsymbol{x}_n\big).
\end{array}
\right.
\end{array}
\tag{26.48}
$$

In turn, since $J_{\gamma \boldsymbol{A}} = P_{\boldsymbol{D}}$ is linear, this is equivalent to

$$
\begin{array}{l}
\text{for } n = 0,1,\ldots \\
\left\lfloor
\begin{array}{l}
\boldsymbol{p}_n = J_{\gamma \boldsymbol{A}}\boldsymbol{y}_n, \\
\boldsymbol{x}_n = J_{\gamma \boldsymbol{B}}\boldsymbol{y}_n, \\
\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + \lambda_n\big(J_{\gamma \boldsymbol{A}}(2\boldsymbol{x}_n - \boldsymbol{y}_n) - \boldsymbol{x}_n\big).
\end{array}
\right.
\end{array}
\tag{26.49}
$$

Hence, we derive from Theorem 26.11(v) and (26.47) that $(\boldsymbol{p}_n)_{n\in\mathbb{N}}$ converges weakly to a point $\boldsymbol{j}(x)$, where $x \in \operatorname{zer}\sum_{i\in I} A_i$. Thus, $p_n = \boldsymbol{j}^{-1}(\boldsymbol{p}_n) \rightharpoonup x$. ☐

26.4 Peaceman–Rachford Splitting Algorithm

The Peaceman–Rachford algorithm can be regarded as a limiting case of the Douglas–Rachford algorithm (26.29) in which $(\forall n \in \mathbb{N})$ $\lambda_n = 2$.

Proposition 26.13 (Peaceman–Rachford algorithm) *Let A and B be maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$ such that $\operatorname{zer}(A+B) \neq \varnothing$ and B is uniformly monotone, let $\gamma \in \mathbb{R}_{++}$, and let $\bar{x}$ be the solution to the problem*

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + Bx. \tag{26.50}$$

Let $y_0 \in \mathcal{H}$, and set

$$\begin{array}{l} \text{for } n = 0,1,\ldots \\ \left\lfloor \begin{array}{l} x_n = J_{\gamma B} y_n, \\ z_n = J_{\gamma A}(2x_n - y_n), \\ y_{n+1} = y_n + 2(z_n - x_n). \end{array} \right. \end{array} \tag{26.51}$$

Then $x_n \to \bar{x}$.

Proof. Since B is uniformly monotone, so is $A+B$, and it follows from Proposition 23.35 that $\operatorname{zer}(A+B)$ is a singleton. Hence $\operatorname{zer}(A+B) = \{\bar{x}\}$. Now set $T = R_{\gamma A} R_{\gamma B}$. Then Proposition 26.1(iii)(d) implies that $\operatorname{Fix} T \neq \varnothing$. In addition, we derive from (26.51) that $(\forall n \in \mathbb{N})$ $y_{n+1} = Ty_n$. Next, let $y \in \operatorname{Fix} T$ and let ϕ be the modulus of uniform monotonicity of B. Then Proposition 26.1(iii)(b) yields $\bar{x} = J_{\gamma B} y$ and therefore we derive from Corollary 23.11(ii), Proposition 23.12, and (26.51) that

$$\begin{aligned}
(\forall n \in \mathbb{N}) \quad \|y_{n+1} - y\|^2 &= \|R_{\gamma A}(R_{\gamma B} y_n) - R_{\gamma A}(R_{\gamma B} y)\|^2 \\
&\leqslant \|R_{\gamma B} y_n - R_{\gamma B} y\|^2 \\
&= \|y_n - y\|^2 - 4\left\langle J_{\gamma B} y_n - J_{\gamma B} y \mid y_n - y\right\rangle \\
&\quad + 4\|J_{\gamma B} y_n - J_{\gamma B} y\|^2 \\
&\leqslant \|y_n - y\|^2 - 4\gamma\phi(\|J_{\gamma B} y_n - J_{\gamma B} y\|) \\
&= \|y_n - y\|^2 - 4\gamma\phi(\|x_n - \bar{x}\|).
\end{aligned} \tag{26.52}$$

Hence $\phi(\|x_n - \bar{x}\|) \to 0$ and, in turn, $\|x_n - \bar{x}\| \to 0$. ☐

26.5 Forward-Backward Splitting Algorithm

We focus on the case when B is single-valued in (26.1). The algorithm described next is known as a *forward-backward* algorithm. It alternates an explicit step using the operator B with an implicit resolvent step involving the operator A.

Theorem 26.14 (Forward-backward algorithm) *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $\beta \in \mathbb{R}_{++}$, let $B\colon \mathcal{H} \to \mathcal{H}$ be β-cocoercive, let $\gamma \in \left]0, 2\beta\right[$, and set $\delta = 2 - \gamma/(2\beta)$. Furthermore, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0,\delta]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$, and let $x_0 \in \mathcal{H}$. Suppose that* $\operatorname{zer}(A+B) \neq \varnothing$ *and set*

$$\begin{array}{l} \text{for } n = 0,1,\ldots \\ \left\lfloor \begin{array}{l} y_n = x_n - \gamma B x_n, \\ x_{n+1} = x_n + \lambda_n\big(J_{\gamma A} y_n - x_n\big). \end{array} \right. \end{array} \tag{26.53}$$

Then the following hold:

(i) $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* $\operatorname{zer}(A+B)$.

(ii) *Let $x \in \operatorname{zer}(A+B)$. Then $(Bx_n)_{n\in\mathbb{N}}$ converges strongly to the unique dual solution Bx.*

(iii) *Suppose that one of the following holds:*

   (a) *A is uniformly monotone on every nonempty bounded subset of* $\operatorname{dom} A$.

   (b) *B is uniformly monotone on every nonempty bounded subset of $\mathcal{H}$.*

   Then $(x_n)_{n\in\mathbb{N}}$ converges strongly to the unique point in $\operatorname{zer}(A+B)$.

Proof. Set $T = J_{\gamma A} \circ (\mathrm{Id} - \gamma B)$ and $\alpha = 1/\delta$. It follows from Proposition 26.1(iv)(d) that T is α-averaged. Moreover, by Proposition 26.1(iv)(a), $\operatorname{Fix} T = \operatorname{zer}(A+B)$. We also deduce from (26.53) that $(x_n)_{n\in\mathbb{N}}$ is generated by (5.18).

(i): The claim follows from Proposition 5.16(iii).

(ii): By Example 22.7, B^{-1} is strongly monotone and $-A^{-1} \circ (-\mathrm{Id}) + B^{-1}$ is therefore likewise. In turn, Corollary 23.37(ii) asserts that the dual problem (26.2) has a unique solution which, by Proposition 26.1(iv)(b), is Bx. Set $y = x - \gamma Bx$ and let $n \in \mathbb{N}$. Then $Tx_n = J_{\gamma A} y_n$ and, by Proposition 26.1(iv)(a), $x = Tx = J_{\gamma A} y$. Hence, we derive from Corollary 23.9 and Proposition 4.4 that

$$\begin{aligned} &\langle Tx_n - x \mid x_n - Tx_n - \gamma(Bx_n - Bx)\rangle \\ &\quad = \langle Tx_n - x \mid y_n - Tx_n + x - y\rangle \\ &\quad = \langle J_{\gamma A} y_n - J_{\gamma A} y \mid (\mathrm{Id} - J_{\gamma A}) y_n - (\mathrm{Id} - J_{\gamma A}) y\rangle \\ &\quad \geqslant 0. \end{aligned} \tag{26.54}$$

Therefore, since T is nonexpansive and B is β-cocoercive,

$$\begin{aligned}
\|x_n - x\| \, \|Tx_n - x_n\| &\geqslant \|Tx_n - x\| \, \|Tx_n - x_n\| \\
&\geqslant \langle Tx_n - x \mid x_n - Tx_n \rangle \\
&\geqslant \gamma \langle Tx_n - x \mid Bx_n - Bx \rangle \\
&= \gamma\big(\langle Tx_n - x_n \mid Bx_n - Bx \rangle + \langle x_n - x \mid Bx_n - Bx \rangle\big) \\
&\geqslant -\gamma \|Tx_n - x_n\| \, \|Bx_n - Bx\| + \gamma\beta \|Bx_n - Bx\|^2 \\
&\geqslant -\frac{\gamma}{\beta} \|Tx_n - x_n\| \, \|x_n - x\| + \gamma\beta \|Bx_n - Bx\|^2
\end{aligned} \tag{26.55}$$

and hence Proposition 5.16(i) yields

$$\gamma\beta \|Bx_n - Bx\|^2 \leqslant 3\|x_n - x\| \, \|Tx_n - x_n\| \leqslant 3\|x_0 - x\| \, \|Tx_n - x_n\|. \tag{26.56}$$

The claim therefore follows from Proposition 5.16(ii).

(iii): As seen in (i), there exists $x \in \operatorname{zer}(A+B)$ such that $x_n \rightharpoonup x$.

(iii)(a): Set $(\forall n \in \mathbb{N})$ $z_n = J_{\gamma A} y_n$. Then

$$-\gamma Bx \in \gamma Ax \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x_n - \gamma Bx_n - z_n = y_n - z_n \in \gamma A z_n. \tag{26.57}$$

On the other hand, we derive from Proposition 5.16(ii) that

$$z_n - x_n = Tx_n - x_n \to 0. \tag{26.58}$$

Thus, $z_n \rightharpoonup x$ and hence $C = \{x\} \cup \{z_n\}_{n\in\mathbb{N}}$ is a bounded subset of $\operatorname{dom} A$. In turn, it follows from (22.7) and (26.57) that there exists an increasing function $\phi_A \colon \mathbb{R}_+ \to [0,+\infty]$ that vanishes only at 0 such that

$$(\forall n \in \mathbb{N}) \quad \langle z_n - x \mid x_n - z_n - \gamma(Bx_n - Bx) \rangle \geqslant \gamma \phi_A\big(\|z_n - x\|\big). \tag{26.59}$$

However, $z_n - x \rightharpoonup 0$, and it follows from (26.58) and (ii) that $x_n - z_n - \gamma(Bx_n - Bx) \to 0$. We thus derive from (26.59) that $\phi_A(\|z_n - x\|) \to 0$, which forces $z_n \to x$. In view of (26.58), we conclude that $x_n \to x$.

(iii)(b): Since $x_n \rightharpoonup x$, $C = \{x\} \cup \{x_n\}_{n\in\mathbb{N}}$ is a bounded subset of $\mathcal{H}$. Hence, it follows from (22.7) that there exists an increasing function $\phi_B \colon \mathbb{R}_+ \to [0,+\infty]$ that vanishes only at 0 such that

$$(\forall n \in \mathbb{N}) \quad \langle x_n - x \mid Bx_n - Bx \rangle \geqslant \phi_B\big(\|x_n - x\|\big). \tag{26.60}$$

Thus, (ii) yields $\phi_B(\|x_n - x\|) \to 0$ and therefore $x_n \to x$. ☐

Remark 26.15 For $\lambda_n = 1$, the updating rule in (26.53) can be written as $x_{n+1} = J_{\gamma A}(x_n - \gamma Bx_n)$, i.e., by Proposition 23.2(ii),

$$-\frac{1}{\gamma}(x_{n+1} - x_n) \in Ax_{n+1} + Bx_n. \tag{26.61}$$

This is a discrete-time version of the continuous-time dynamics

$$-x'(t) \in Ax(t) + Bx(t) \tag{26.62}$$

obtained using a time-step γ, a forward Euler discretization for the operator B, and a backward Euler discretization for the operator A.

We conclude this section with two instances of linear convergence of the forward-backward algorithm.

Proposition 26.16 *Let D be a nonempty closed subset of $\mathcal{H}$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and such that* $\operatorname{dom} A \subset D$*, let $B\colon D \to \mathcal{H}$, let $\alpha \in \mathbb{R}_{++}$, and let $\beta \in \mathbb{R}_{++}$. Suppose that one of the following holds:*

(i) *A is α-strongly monotone, B is β-cocoercive, and $\gamma \in \left]0, 2\beta\right[$.*
(ii) *$\alpha \leqslant \beta$, B is α-strongly monotone and β-Lipschitz continuous, and $\gamma \in \left]0, 2\alpha/\beta^2\right[$.*

Let $x_0 \in D$ and set

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} y_n = x_n - \gamma B x_n, \\ x_{n+1} = J_{\gamma A} y_n. \end{array} \right. \end{array} \tag{26.63}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges linearly to the unique point in $\operatorname{zer}(A + B)$.

Proof. Set $T = J_{\gamma A} \circ (\mathrm{Id} - \gamma B)$ and note that, by Proposition 26.1(iv)(a), $\operatorname{Fix} T = \operatorname{zer}(A + B)$. Since Proposition 23.2(i) asserts that $\operatorname{ran} J_{\gamma A} = \operatorname{dom} A$, T is a well-defined operator from D to D, and (26.63) reduces to $(\forall n \in \mathbb{N})$ $x_{n+1} = Tx_n$. Since D, as a closed subset of $\mathcal{H}$, is a complete metric space, in view of Theorem 1.50, it is enough to show that in both cases T is Lipschitz continuous with a constant in $\left[0, 1\right[$.

(i): Set $\tau = 1/(\alpha\gamma+1)$. On the one hand, we derive from Proposition 23.13 that $J_{\gamma A}$ is Lipschitz continuous with constant τ. On the other hand, we derive from Proposition 4.39 and Remark 4.34(i) that $\mathrm{Id}-\gamma B$ is nonexpansive. Altogether, T is Lipschitz continuous with constant $\tau \in \left]0, 1\right[$.

(ii): Observe that $\gamma(2\alpha - \gamma\beta^2) \in \left]0, 1\right]$ and set $\tau = \sqrt{1 - \gamma(2\alpha - \gamma\beta^2)}$. On the one hand, we derive from Corollary 23.11(i) that $J_{\gamma A}$ is nonexpansive. On the other hand, for every x and y in D, we have

$$\begin{aligned} \|(\mathrm{Id} - \gamma B)x - (\mathrm{Id} - \gamma B)y\|^2 &= \|x - y\|^2 - 2\gamma \langle x - y \mid Bx - By\rangle \\ &\quad + \gamma^2\|Bx - By\|^2 \\ &\leqslant \|x - y\|^2 - 2\gamma\alpha\|x - y\|^2 + \gamma^2\beta^2\|x - y\|^2 \\ &= \big(1 - \gamma\big(2\alpha - \gamma\beta^2\big)\big)\|x - y\|^2. \end{aligned} \tag{26.64}$$

Altogether, T is Lipschitz continuous with constant $\tau \in \left[0, 1\right[$. ☐

26.6 Tseng's Splitting Algorithm

In this section, we once again consider the case when B is single-valued in (26.1), but relax the cocoercivity condition imposed by the forward-backward algorithm in Theorem 26.14 at the expense of additional computations. More precisely, the algorithm presented below involves at each iteration two forward steps, a backward step, and a projection step.

Theorem 26.17 (Tseng's algorithm) *Let D be a nonempty subset of $\mathcal{H}$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and such that $\operatorname{dom} A \subset D$, let $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a monotone operator that is single-valued on D, and let $\beta \in \mathbb{R}_{++}$. Suppose that $A+B$ is maximally monotone, and that C is a closed convex subset of D such that $C \cap \operatorname{zer}(A+B) \neq \varnothing$ and B is $1/\beta$-Lipschitz continuous relative to $C \cup \operatorname{dom} A$. Let $x_0 \in C$, let $\gamma \in \left]0,\beta\right[$, and set*

$$
\begin{array}{l}
\text{for } n = 0,1,\ldots\\
\left\lfloor
\begin{array}{l}
y_n = x_n - \gamma B x_n,\\
z_n = J_{\gamma A} y_n,\\
r_n = z_n - \gamma B z_n,\\
x_{n+1} = P_C\big(x_n - y_n + r_n\big).
\end{array}
\right.
\end{array}
\tag{26.65}
$$

Then the following hold:

(i) $(x_n - z_n)_{n\in\mathbb{N}}$ *converges strongly to* 0.
(ii) $(x_n)_{n\in\mathbb{N}}$ *and* $(z_n)_{n\in\mathbb{N}}$ *converge weakly to a point in* $C \cap \operatorname{zer}(A+B)$.
(iii) *Suppose that A or B is uniformly monotone on every nonempty bounded subset of* $\operatorname{dom} A$. *Then $(x_n)_{n\in\mathbb{N}}$ and $(z_n)_{n\in\mathbb{N}}$ converge strongly to the unique point in* $C \cap \operatorname{zer}(A+B)$.

Proof. Suppose that, for some $n \in \mathbb{N}$, $x_n \in C$. Then $x_n \in D$ and y_n is therefore well defined. In turn, we derive from Proposition 23.2(i) that $z_n \in \operatorname{ran} J_{\gamma A} = \operatorname{dom} A \subset D$, which makes r_n well defined. Finally, $x_{n+1} = P_C(r_n + x_n - y_n) \in C$. This shows by induction that the sequences $(x_n)_{n\in\mathbb{N}}$, $(y_n)_{n\in\mathbb{N}}$, $(z_n)_{n\in\mathbb{N}}$, and $(r_n)_{n\in\mathbb{N}}$ are well defined. Let us set

$$
(\forall n \in \mathbb{N}) \quad u_n = \gamma^{-1}(x_n - z_n) + Bz_n - Bx_n. \tag{26.66}
$$

Note that (26.65) yields

$$
(\forall n \in \mathbb{N}) \quad u_n = \gamma^{-1}(y_n - z_n) + Bz_n \in Az_n + Bz_n. \tag{26.67}
$$

Now let $z \in C \cap \operatorname{zer}(A+B)$ and let $n \in \mathbb{N}$. We first observe that

$$
z = P_C z \quad \text{and} \quad (z, -\gamma Bz) \in \operatorname{gra} \gamma A. \tag{26.68}
$$

On the other hand, by Proposition 23.2(ii) and (26.65), $(z_n, y_n - z_n) \in \operatorname{gra} \gamma A$. Hence, by (26.68) and monotonicity of γA, $\langle z_n - z \mid z_n - y_n - \gamma Bz\rangle \leqslant 0$.

However, by monotonicity of B, $\langle z_n - z \mid \gamma Bz - \gamma Bz_n\rangle \leqslant 0$. Upon adding these two inequalities, we obtain

$$\langle z_n - z \mid z_n - y_n - \gamma Bz_n\rangle \leqslant 0. \tag{26.69}$$

In turn, we derive from (26.65) that

$$\begin{aligned}
2\gamma \langle z_n - z \mid Bx_n - Bz_n\rangle &= 2\langle z_n - z \mid z_n - y_n - \gamma Bz_n\rangle \\
&\quad + 2\langle z_n - z \mid \gamma Bx_n + y_n - z_n\rangle \\
&\leqslant 2\langle z_n - z \mid \gamma Bx_n + y_n - z_n\rangle \\
&= 2\langle z_n - z \mid x_n - z_n\rangle \\
&= \|x_n - z\|^2 - \|z_n - z\|^2 - \|x_n - z_n\|^2
\end{aligned} \tag{26.70}$$

and, therefore, from Proposition 4.16 that

$$\begin{aligned}
\|x_{n+1} - z\|^2 &= \|P_C(r_n + x_n - y_n) - P_C z\|^2 \\
&\leqslant \|r_n + x_n - y_n - z\|^2 \\
&= \|(z_n - z) + \gamma(Bx_n - Bz_n)\|^2 \\
&= \|z_n - z\|^2 + 2\gamma\langle z_n - z \mid Bx_n - Bz_n\rangle + \gamma^2\|Bx_n - Bz_n\|^2 \\
&\leqslant \|x_n - z\|^2 - \|x_n - z_n\|^2 + \gamma^2\|Bx_n - Bz_n\|^2 \\
&\leqslant \|x_n - z\|^2 - (1 - \gamma^2/\beta^2)\|x_n - z_n\|^2.
\end{aligned} \tag{26.71}$$

This shows that

$$(x_n)_{n\in\mathbb{N}} \text{ is Fejér monotone with respect to } C \cap \operatorname{zer}(A+B). \tag{26.72}$$

(i): An immediate consequence of (26.71).

(ii): It follows from (i), the relative Lipschitz continuity of B, and (26.66) that

$$Bz_n - Bx_n \to 0 \quad \text{and} \quad u_n \to 0. \tag{26.73}$$

Now let x be a weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$, say $x_{k_n} \rightharpoonup x$. Let us show that $x \in C \cap \operatorname{zer}(A+B)$. Since $(x_n)_{n\in\mathbb{N}}$ lies in C, Corollary 3.35 asserts that $x \in C$ and it remains to show that $(x, 0) \in \operatorname{gra}(A+B)$. It follows from (i) that $z_{k_n} \rightharpoonup x$, and from (26.73) that $u_{k_n} \to 0$. Altogether, $(z_{k_n}, u_{k_n})_{n\in\mathbb{N}}$ lies in $\operatorname{gra}(A+B)$ by (26.67), and it satisfies

$$z_{k_n} \rightharpoonup x \quad \text{and} \quad u_{k_n} \to 0. \tag{26.74}$$

Since $A+B$ is maximally monotone, it follows from Proposition 20.38(ii) that $(x, 0) \in \operatorname{gra}(A+B)$. In view of (26.72), Theorem 5.5, and (i), the assertions are proved.

(iii): Since $A+B$ is strictly monotone, it follows from Proposition 23.35 that $\operatorname{zer}(A+B)$ is a singleton. As shown in (ii), there exists $x \in C \cap \operatorname{zer}(A+B) \subset \operatorname{dom} A$ such that

$$z_n \rightharpoonup x. \tag{26.75}$$

The assumptions imply that $A+B$ is uniformly monotone on $\{x\}\cup\{z_n\}_{n\in\mathbb{N}} \subset \operatorname{dom} A$. Hence, since $0 \in (A+B)x$, it follows from (26.67) and (22.7) that there exists an increasing function $\phi\colon \mathbb{R}_+ \to [0,+\infty]$ that vanishes only at 0 such that

$$(\forall n \in \mathbb{N}) \quad \langle z_n - x \mid u_n\rangle \geqslant \phi\big(\|z_n - x\|\big). \tag{26.76}$$

We therefore deduce from (26.75) and (26.73) that $\phi(\|z_n - x\|) \to 0$, which implies that $z_n \to x$. In turn, (i) yields $x_n \to x$. ☐

Remark 26.18 Here are a few observations concerning Theorem 26.17.

(i) Sufficient conditions for the maximal monotonicity of $A+B$ are discussed in Corollary 25.4 and Corollary 25.5.
(ii) The set C can be used to impose constraints on the zeros of $A+B$.
(iii) If $\operatorname{dom} A$ is closed, then it follows from Corollary 21.14 that it is convex and we can therefore choose $C = \operatorname{dom} A$.
(iv) If $\operatorname{dom} B = \mathcal{H}$, we can choose $C = \mathcal{H}$. In this case, it follows that B is continuous and single-valued on $\mathcal{H}$, hence maximally monotone by Corollary 20.28. In turn, Corollary 25.5(i) implies that $A+B$ is maximally monotone and (26.65) reduces to the *forward-backward-forward algorithm*

$$\begin{array}{l}\text{for } n = 0,1,\ldots\\ \left\lfloor\begin{array}{l} y_n = x_n - \gamma B x_n,\\ z_n = J_{\gamma A} y_n,\\ x_{n+1} = x_n - y_n + z_n - \gamma B z_n.\end{array}\right.\end{array} \tag{26.77}$$

26.7 Variational Inequalities

Definition 26.19 Let $f \in \Gamma_0(\mathcal{H})$ and let $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone. The associated *variational inequality* problem is to

$$\text{find } x \in \mathcal{H} \text{ such that } (\exists\, u \in Bx)(\forall y \in \mathcal{H})\ \langle x - y \mid u\rangle + f(x) \leqslant f(y). \tag{26.78}$$

Here are a few examples of variational inequalities (additional examples will arise in the context of minimization problems in Section 27.1 and Section 27.2).

Example 26.20 In Definition 26.19, let $z \in \mathcal{H}$ and set $B\colon x \mapsto x - z$. Then we obtain the variational inequality problem

$$\text{find } x \in \mathcal{H} \text{ such that } (\forall y \in \mathcal{H})\ \langle x - y \mid x - z\rangle + f(x) \leqslant f(y). \tag{26.79}$$

As seen in Proposition 12.26, the unique solution to this problem is $\operatorname{Prox}_f z$.

Example 26.21 In Definition 26.19, set $f = \iota_C$, where C is a nonempty closed convex subset of $\mathcal{H}$, and let $B\colon \mathcal{H} \to \mathcal{H}$ be maximally monotone. Then we obtain the classical variational inequality problem

$$\text{find } x \in C \text{ such that } (\forall y \in C)\ \langle x - y \mid Bx\rangle \leqslant 0. \tag{26.80}$$

In particular, if $B\colon x \mapsto x - z$, where $z \in \mathcal{H}$, we recover the variational inequality that characterizes the projection of z onto C (see Theorem 3.16).

Example 26.22 (Complementarity problem) In Example 26.21, set $C = K$, where K is a nonempty closed convex cone in $\mathcal{H}$. Then we obtain the *complementarity problem*

$$\text{find } x \in K \text{ such that } x \perp Bx \text{ and } Bx \in K^{\oplus}. \tag{26.81}$$

Proof. If $x \in K$, then $\{x/2, 2x\} \subset K$ and the condition $\sup_{y\in K} \langle x - y \mid Bx\rangle \leqslant 0$ implies that $\langle x - x/2 \mid Bx\rangle \leqslant 0$ and $\langle x - 2x \mid Bx\rangle \leqslant 0$, hence $\langle x \mid Bx\rangle = 0$. It therefore reduces to $\sup_{y\in K} \langle -y \mid Bx\rangle \leqslant 0$, i.e., by Definition 6.22, to $Bx \in K^{\oplus}$. ☐

Remark 26.23 In view of (16.1), the variational inequality problem (26.78) can be recast as that of finding a point in $\operatorname{zer}(A + B)$, where $A = \partial f$ is maximally monotone by Theorem 20.25.

It follows from Remark 26.23 that we can apply the splitting algorithms considered earlier in this chapter to solve variational inequalities. We start with an application of the Douglas–Rachford algorithm.

Proposition 26.24 (Douglas–Rachford algorithm) *Let $f \in \Gamma_0(\mathcal{H})$, let $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, and suppose that the variational inequality*

$$\textit{find } x \in \mathcal{H} \textit{ such that } (\exists\, u \in Bx)(\forall y \in \mathcal{H})\ \langle x - y \mid u\rangle + f(x) \leqslant f(y) \tag{26.82}$$

has at least one solution. Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0, 2]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2 - \lambda_n) = +\infty$, let $\gamma \in \mathbb{R}_{++}$, and let $y_0 \in \mathcal{H}$. Set

$$\begin{array}{l}\textit{for } n = 0, 1, \ldots\\ \left\lfloor \begin{array}{l} x_n = J_{\gamma B} y_n,\\ z_n = \operatorname{Prox}_{\gamma f}(2x_n - y_n),\\ y_{n+1} = y_n + \lambda_n(z_n - x_n). \end{array}\right.\end{array} \tag{26.83}$$

Then there exists $y \in \mathcal{H}$ such that the following hold:

(i) *$J_{\gamma B} y$ is a solution to (26.82).*
(ii) *$(y_n)_{n\in\mathbb{N}}$ converges weakly to y.*
(iii) *$(x_n)_{n\in\mathbb{N}}$ and $(z_n)_{n\in\mathbb{N}}$ converge weakly to $J_{\gamma B} y$.*
(iv) *Suppose that one of the following holds:*

(a) *f is uniformly convex on every nonempty bounded subset of* $\operatorname{dom}\partial f$.
(b) *B is uniformly monotone on every nonempty bounded subset of* $\operatorname{dom} B$.

Then $(x_n)_{n\in\mathbb{N}}$ and $(z_n)_{n\in\mathbb{N}}$ converge strongly to the unique solution to (26.82).

Proof. Apply items (i), (iii), and (vi) of Theorem 26.11 to $A=\partial f$, and use Example 23.3 and Example 22.5. ☐

Next, we consider the case when B is single-valued, and present an application of the forward-backward algorithm (linear convergence results can be derived from Proposition 26.16 in a similar fashion).

Proposition 26.25 (Forward-backward algorithm) *Let $f\in\Gamma_0(\mathcal{H})$, let $\beta\in\mathbb{R}_{++}$, let $B\colon\mathcal{H}\to\mathcal{H}$ be β-cocoercive, let $\gamma\in\left]0,2\beta\right[$, and set $\delta=2-\gamma/(2\beta)$. Suppose that the variational inequality*

$$\text{find } x\in\mathcal{H} \text{ such that } (\forall y\in\mathcal{H})\ \langle x-y \mid Bx\rangle + f(x)\leqslant f(y) \tag{26.84}$$

possesses at least one solution. Furthermore, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0,\delta]$ such that $\sum_{n\in\mathbb{N}}\lambda_n(\delta-\lambda_n)=+\infty$, let $x_0\in\mathcal{H}$, and set

$$\begin{array}{l}\text{for } n=0,1,\ldots\\ \left\lfloor\begin{array}{l} y_n=x_n-\gamma Bx_n,\\ x_{n+1}=x_n+\lambda_n\big(\operatorname{Prox}_{\gamma f} y_n-x_n\big).\end{array}\right.\end{array} \tag{26.85}$$

Then the following hold:

(i) *$(x_n)_{n\in\mathbb{N}}$ converges weakly to a solution to (26.84).*
(ii) *Let x be a solution to (26.84). Then $(Bx_n)_{n\in\mathbb{N}}$ converges strongly to Bx.*
(iii) *Suppose that one of the following holds:*

(a) *f is uniformly convex on every nonempty bounded subset of* $\operatorname{dom}\partial f$.
(b) *B is uniformly monotone on every nonempty bounded subset of $\mathcal{H}$.*

Then $(x_n)_{n\in\mathbb{N}}$ converges strongly to the unique solution to (26.84).

Proof. This is an application of Theorem 26.14 to $A=\partial f$, Example 23.3, and Example 22.5. ☐

Example 26.26 Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\beta\in\mathbb{R}_{++}$, and let $B\colon\mathcal{H}\to\mathcal{H}$ be β-cocoercive. Suppose that the variational inequality

$$\text{find } x\in C \text{ such that } (\forall y\in C)\ \langle x-y\mid Bx\rangle\leqslant 0 \tag{26.86}$$

has at least one solution. Let $x_0\in\mathcal{H}$ and set

$$(\forall n\in\mathbb{N})\quad x_{n+1}=P_C(x_n-\beta Bx_n). \tag{26.87}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a solution x to (26.86) and, moreover, $(Bx_n)_{n\in\mathbb{N}}$ converges strongly to Bx.

Proof. Apply Proposition 26.25 to $f = \iota_C$, $\gamma = \beta$, and $\lambda_n \equiv 1$, and use Example 12.25. ☐

In instances in which B is not cocoercive on $\mathcal{H}$, we can turn to Tseng's splitting algorithm of Section 26.6 as we illustrate next.

Example 26.27 Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\beta \in \mathbb{R}_{++}$, and let $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator that is single-valued and β-Lipschitz continuous relative to C. Suppose that

$$\operatorname{cone}(C - \operatorname{dom} B) = \overline{\operatorname{span}}\,(C - \operatorname{dom} B), \tag{26.88}$$

and that the variational inequality

$$\text{find } x \in C \ \text{ such that } \ (\forall y \in C)\ \langle x - y \mid Bx\rangle \leqslant 0 \tag{26.89}$$

possesses at least one solution. Let $x_0 \in C$, let $\gamma \in \left]0, 1/\beta\right[$, and set

$$\begin{array}{l}\text{for } n = 0, 1, \ldots\\ \left\lfloor\begin{array}{l} y_n = x_n - \gamma B x_n,\\ z_n = P_C y_n,\\ r_n = z_n - \gamma B z_n,\\ x_{n+1} = P_C\big(x_n - y_n + r_n\big).\end{array}\right.\end{array} \tag{26.90}$$

Then the following hold:

(i) $(x_n - z_n)_{n\in\mathbb{N}}$ converges strongly to 0.
(ii) $(x_n)_{n\in\mathbb{N}}$ and $(z_n)_{n\in\mathbb{N}}$ converge weakly to a solution to (26.89).
(iii) Suppose that B is uniformly monotone on every nonempty bounded subset of C. Then $(x_n)_{n\in\mathbb{N}}$ and $(z_n)_{n\in\mathbb{N}}$ converge strongly to the unique solution to (26.89).

Proof. It follows from Example 20.26, (26.88), and Corollary 25.4 that the operator $N_C + B$ is maximally monotone. Hence, the results follow from Theorem 26.17 applied to $D = C$ and $A = N_C$, and by invoking Example 20.26 and Example 23.4. ☐

26.8 Composite Inclusion Problems

We investigate a primal-dual composite inclusion problem involving two monotone operators and a linear operator.

Problem 26.28 Let $z \in \mathcal{H}$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $\mathcal{K}$ be a real Hilbert space, let $r \in \mathcal{K}$, let $B\colon \mathcal{K} \to 2^{\mathcal{K}}$ be maximally monotone, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$. Solve the primal inclusion

$$\text{find } x \in \mathcal{H} \text{ such that } z \in Ax + L^*\big(B(Lx - r)\big), \tag{26.91}$$

together with the dual inclusion

$$\text{find } v \in \mathcal{K} \text{ such that } -r \in -L\big(A^{-1}(z - L^*v)\big) + B^{-1}v. \tag{26.92}$$

The sets of solutions to (26.91) and (26.92) are denoted by $\mathscr{P}$ and $\mathscr{D}$, respectively.

Remark 26.29 As discussed in Section 26.1, the duality at work in (26.1)–(26.2) can be regarded as an extension of Fenchel duality to set-valued operators. Likewise, the operator duality presented in Problem 26.28 can be regarded as an extension of the Fenchel–Rockafellar duality framework of Section 15.3. Indeed, if $A = \partial f$ and $B = \partial g$ for some functions $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{K})$ such that $0 \in \operatorname{sri}(\operatorname{dom} g - L(\operatorname{dom} f))$, $z = 0$, and $r = 0$, then it follows from Theorem 16.3 and Theorem 16.47 that (26.91) yields the primal problem (19.32) and (26.92) corresponds to the dual problem (19.33).

To analyze and solve Problem 26.28, it will be convenient to introduce the following auxiliary problem.

Problem 26.30 In the setting of Problem 26.28, let $\boldsymbol{\mathcal{H}} = \mathcal{H} \oplus \mathcal{K}$ and set

$$\begin{cases} \boldsymbol{M}\colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}\colon (x,v) \mapsto (-z + Ax) \times (r + B^{-1}v), \\ \boldsymbol{S}\colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}\colon (x,v) \mapsto (L^*v, -Lx). \end{cases} \tag{26.93}$$

The problem is to

$$\text{find } \boldsymbol{x} \in \boldsymbol{\mathcal{H}} \text{ such that } \boldsymbol{0} \in \boldsymbol{M}\boldsymbol{x} + \boldsymbol{S}\boldsymbol{x}. \tag{26.94}$$

We denote by $\boldsymbol{Z}$ its set of solutions, i.e.,

$$\begin{aligned} \boldsymbol{Z} &= \operatorname{zer}(\boldsymbol{M} + \boldsymbol{S}) \\ &= \big\{(x,v) \in \mathcal{H} \oplus \mathcal{K} \;\big|\; z - L^*v \in Ax \text{ and } Lx - r \in B^{-1}v\big\}, \end{aligned} \tag{26.95}$$

and we call it the set of *Kuhn–Tucker points* associated with Problem 26.28.

Remark 26.31 Proposition 19.20(v) asserts that, in the context of Remark 26.29, the set $\boldsymbol{Z}$ introduced in (26.95) is the set of saddle points of the Lagrangian (19.34).

Proposition 26.32 *Consider Problem 26.28 and Problem 26.30. Then the following hold:*

(i) $\boldsymbol{M}$ *is maximally monotone.*
(ii) $\boldsymbol{S} \in \mathcal{B}(\boldsymbol{\mathcal{H}})$*,* $\boldsymbol{S}^* = -\boldsymbol{S}$*, and* $\|\boldsymbol{S}\| = \|L\|$*.*

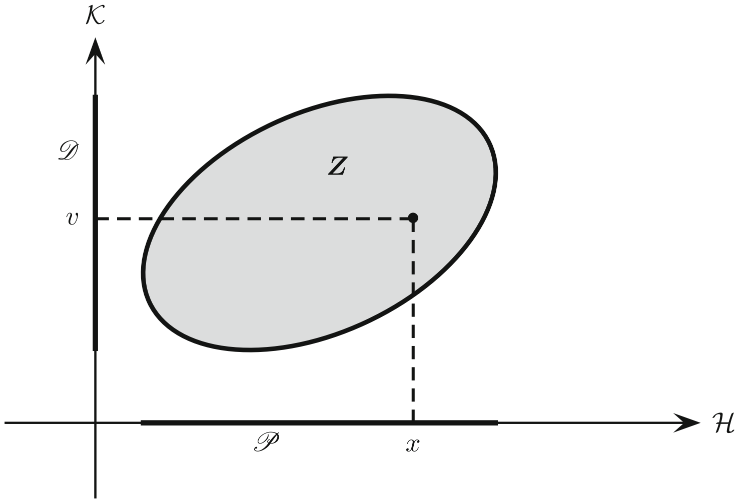

Fig. 26.1 The Kuhn–Tucker set $\boldsymbol{Z}$ of (26.95) is a closed convex subset of $\mathcal{H} \oplus \mathcal{K}$ and, for every $(x, v) \in \boldsymbol{Z}$, x is a solution to the primal problem (26.91) and v is a solution to the dual problem (26.92).

(iii) *$\boldsymbol{M} + \boldsymbol{S}$ is maximally monotone.*
(iv) *Let $(x, v) \in \boldsymbol{\mathcal{H}}$ and let $\gamma \in \mathbb{R}_{++}$. Then*

$$J_{\gamma \boldsymbol{M}}(x, v) = \Big(J_{\gamma A}(x + \gamma z)\,, \, v - \gamma\big(r + J_{\gamma^{-1} B}(\gamma^{-1} v - r)\big)\Big). \qquad (26.96)$$

Proof. (i): Since A and B are maximally monotone, it follows from Proposition 20.22 and Proposition 20.23 that $\boldsymbol{M}$ is likewise.

(ii): The first two assertions are clear. Now let $(x, v) \in \boldsymbol{\mathcal{H}}$. Then, using Fact 2.25(ii), $\|\boldsymbol{S}(x, v)\|^2 = \|(L^* v, -Lx)\|^2 = \|L^* v\|^2 + \|Lx\|^2 \leqslant \|L\|^2(\|v\|^2 + \|x\|^2) = \|L\|^2 \|(x, v)\|^2$. Thus, $\|\boldsymbol{S}\| \leqslant \|L\|$. On the other hand, $\|x\| \leqslant 1 \Rightarrow \|(x, 0)\| \leqslant 1 \Rightarrow \|Lx\| = \|\boldsymbol{S}(x, 0)\| \leqslant \|\boldsymbol{S}\|$. Hence, $\|L\| \leqslant \|\boldsymbol{S}\|$.

(iii): By (i), $\boldsymbol{M}$ is maximally monotone. On the other hand, it follows from (ii) that $\boldsymbol{S}$ is monotone and continuous, and hence maximally monotone by Example 20.35. Altogether, since $\operatorname{dom} \boldsymbol{S} = \boldsymbol{\mathcal{H}}$, it follows from Corollary 25.5(i) that $\boldsymbol{M} + \boldsymbol{S}$ is maximally monotone.

(iv): This follows from Proposition 23.17(ii), Proposition 23.18, and Proposition 23.20. ☐

Let us describe the interplay between Problem 26.28 and Problem 26.30.

Proposition 26.33 *Consider Problem 26.28 and Problem 26.30. Then the following hold (see Figure 26.1):*

(i) *$\boldsymbol{Z}$ is a closed convex subset of $\mathscr{P} \times \mathscr{D}$.*

(ii) *Set* $Q_{\mathcal{H}}\colon \mathcal{H}\oplus\mathcal{K}\to\mathcal{H}\colon (x,v)\mapsto x$ *and* $Q_{\mathcal{K}}\colon \mathcal{H}\oplus\mathcal{K}\to\mathcal{K}\colon (x,v)\mapsto v$. *Then* $\mathscr{P}=Q_{\mathcal{H}}(\boldsymbol{Z})$ *and* $\mathscr{D}=Q_{\mathcal{K}}(\boldsymbol{Z})$.

(iii) $\mathscr{P}\neq\varnothing \Leftrightarrow \boldsymbol{Z}\neq\varnothing \Leftrightarrow \mathscr{D}\neq\varnothing$.

Proof. Let $(x,v)\in\mathcal{H}\oplus\mathcal{K}$.

(i): We deduce from (26.95) that $(x,v)\in\boldsymbol{Z} \Leftrightarrow [0\in -z+Ax+L^*v$ and $0\in r+B^{-1}v-Lx] \Leftrightarrow [z-L^*v\in Ax$ and $Lx-r\in B^{-1}v] \Leftrightarrow [z-L^*v\in Ax$ and $v\in B(Lx-r)] \Rightarrow [z-L^*v\in Ax$ and $L^*v\in L^*(B(Lx-r))] \Rightarrow z\in Ax+L^*(B(Lx-r)) \Leftrightarrow x\in\mathscr{P}$. Similarly, $(x,v)\in\boldsymbol{Z} \Leftrightarrow [x\in A^{-1}(z-L^*v)$ and $r-Lx\in -B^{-1}v] \Rightarrow [Lx\in L(A^{-1}(z-L^*v))$ and $r-Lx\in -B^{-1}v] \Rightarrow r\in L(A^{-1}(z-L^*v))-B^{-1}v \Leftrightarrow v\in\mathscr{D}$. Thus $\boldsymbol{Z}\subset\mathscr{P}\times\mathscr{D}$. We conclude by invoking Proposition 26.32(iii) and Proposition 23.39.

(ii): $x\in\mathscr{P} \Leftrightarrow z\in Ax+L^*(B(Lx-r)) \Leftrightarrow (\exists\, w\in\mathcal{K})\ [z-L^*w\in Ax$ and $w\in B(Lx-r)] \Leftrightarrow (\exists\, w\in\mathcal{K})\ (x,w)\in\boldsymbol{Z}$. Hence $\mathscr{P}\neq\varnothing \Leftrightarrow \boldsymbol{Z}\neq\varnothing$. Likewise, $v\in\mathscr{D} \Leftrightarrow -r\in -LA^{-1}(z-L^*v)+B^{-1}v \Leftrightarrow (\exists\, y\in\mathcal{H})\ [y\in A^{-1}(z-L^*v)$ and $-r\in -Ly+B^{-1}v] \Leftrightarrow (\exists\, y\in\mathcal{H})\ [z-L^*v\in Ay$ and $Ly-r\in B^{-1}v] \Leftrightarrow (\exists\, y\in\mathcal{H})\ (y,v)\in\boldsymbol{Z}$.

(iii): Clear from (ii). ☐

We now introduce a splitting algorithm to solve Problem 26.28.

Theorem 26.34 *In Problem 26.28, suppose that* $L\neq 0$ *and that*

$$z\in\operatorname{ran}\big(A+L^*\circ B\circ(L\cdot -r)\big). \tag{26.97}$$

Let $x_0\in\mathcal{H}$, *let* $v_0\in\mathcal{K}$, *let* $\gamma\in\,]0,1/\|L\|[$, *and set*

$$\begin{array}{l}\text{for } n=0,1,\ldots\\ \left\lfloor\begin{array}{l} y_{1,n}=x_n-\gamma L^*v_n,\\ y_{2,n}=v_n+\gamma Lx_n,\\ p_{1,n}=J_{\gamma A}(y_{1,n}+\gamma z),\\ p_{2,n}=y_{2,n}-\gamma\big(r+J_{\gamma^{-1}B}(\gamma^{-1}y_{2,n}-r)\big),\\ q_{1,n}=p_{1,n}-\gamma L^*p_{2,n},\\ q_{2,n}=p_{2,n}+\gamma Lp_{1,n},\\ x_{n+1}=x_n-y_{1,n}+q_{1,n},\\ v_{n+1}=v_n-y_{2,n}+q_{2,n}.\end{array}\right.\end{array} \tag{26.98}$$

Then the following hold:

(i) $x_n-p_{1,n}\to 0$ *and* $v_n-p_{2,n}\to 0$.

(ii) *There exist a solution* $\bar{x}$ *to* (26.91) *and a solution* $\bar{v}$ *to* (26.92) *such that* $z-L^*\bar{v}\in A\bar{x}$, $L\bar{x}-r\in B^{-1}\bar{v}$, $x_n\rightharpoonup\bar{x}$, *and* $v_n\rightharpoonup\bar{v}$.

Proof. Consider the setting of Problem 26.30. It follows from Proposition 26.32(i)&(ii) that $\boldsymbol{M}$ is maximally monotone and that $\boldsymbol{S}\in\mathcal{B}(\boldsymbol{\mathcal{H}})$ is monotone and $\|L\|$–Lipschitz continuous. Moreover, (26.97) and Proposition 26.33 yield $\varnothing\neq\operatorname{zer}(\boldsymbol{M}+\boldsymbol{S})=\boldsymbol{Z}\subset\mathscr{P}\times\mathscr{D}$. Now set $(\forall n\in\mathbb{N})$ $\boldsymbol{x}_n=(x_n,v_n)$, $\boldsymbol{y}_n=(y_{1,n},y_{2,n})$, $\boldsymbol{p}_n=(p_{1,n},p_{2,n})$, and $\boldsymbol{q}_n=(q_{1,n},q_{2,n})$. Using (26.93) and Proposition 26.32(iv), (26.98) can be written in $\boldsymbol{\mathcal{H}}$ as

$$\begin{array}{l}\text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} \boldsymbol{y}_n = \boldsymbol{x}_n - \gamma \boldsymbol{S}\boldsymbol{x}_n \\ \boldsymbol{p}_n = J_{\gamma \boldsymbol{M}}\, \boldsymbol{y}_n \\ \boldsymbol{q}_n = \boldsymbol{p}_n - \gamma \boldsymbol{S}\boldsymbol{p}_n \\ \boldsymbol{x}_{n+1} = \boldsymbol{x}_n - \boldsymbol{y}_n + \boldsymbol{q}_n, \end{array} \right. \end{array} \tag{26.99}$$

which is precisely the form of (26.65) with $C = \boldsymbol{\mathcal{H}}$. Altogether, the assertions follow from Theorem 26.17(i)&(ii) applied to $\boldsymbol{M}$ and $\boldsymbol{S}$ in $\boldsymbol{\mathcal{H}}$. ☐

Next, we present an application to a multiple operator composite inclusion problem.

Problem 26.35 Let $z \in \mathcal{H}$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, and let I be a nonempty finite set. For every $i \in I$, let $\mathcal{K}_i$ be a real Hilbert space, let $r_i \in \mathcal{K}_i$, let $B_i\colon \mathcal{K}_i \to 2^{\mathcal{K}_i}$ be maximally monotone, and suppose that $L_i \in \mathcal{B}(\mathcal{H}, \mathcal{K}_i) \smallsetminus \{0\}$. Solve the primal inclusion

$$\text{find } \ x \in \mathcal{H} \ \text{ such that } \ z \in Ax + \sum_{i\in I} L_i^*\big(B_i(L_i x - r_i)\big), \tag{26.100}$$

together with the dual inclusion

$$\text{find } \ (v_i)_{i\in I} \in \bigoplus_{i\in I} \mathcal{K}_i \ \text{ such that} \qquad (\forall i \in I) \ \ -r_i \in -L_i A^{-1}\bigg(z - \sum_{j\in I} L_j^* v_j\bigg) + B_i^{-1} v_i. \tag{26.101}$$

Corollary 26.36 *In Problem 26.35, suppose that*

$$z \in \operatorname{ran}\bigg(A + \sum_{i\in I} L_i^* \circ B_i \circ (L_i \cdot - r_i)\bigg). \tag{26.102}$$

Let $x_0 \in \mathcal{H}$*, let* $(v_{i,0})_{i\in I} \in \bigoplus_{i\in I} \mathcal{K}_i$*, let* $\gamma \in \big]0, 1/\sqrt{\sum_{i\in I}\|L_i\|^2}\big[$*, and set*

$$\begin{array}{l}\textit{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} y_{1,n} = x_n - \gamma \sum_{i\in I} L_i^* v_{i,n}, \\ p_{1,n} = J_{\gamma A}(y_{1,n} + \gamma z), \\ \textit{for every } i \in I \\ \left\lfloor \begin{array}{l} y_{2,i,n} = v_{i,n} + \gamma L_i x_n, \\ p_{2,i,n} = y_{2,n} - \gamma\big(r_i + J_{\gamma^{-1}B_i}(\gamma^{-1} y_{2,i,n} - r_i)\big), \\ q_{2,i,n} = p_{2,i,n} + \gamma L_i p_{1,n}, \\ v_{i,n+1} = v_{i,n} - y_{2,i,n} + q_{2,i,n}, \end{array} \right. \\ q_{1,n} = p_{1,n} - \gamma \sum_{i\in I} L_i^* p_{2,i,n}, \\ x_{n+1} = x_n - y_{1,n} + q_{1,n}. \end{array} \right. \end{array} \tag{26.103}$$

Then the following hold:

(i) $x_n - p_{1,n} \to 0$ *and* $(\forall i \in I)$ $v_{i,n} - p_{2,i,n} \to 0$.

(ii) *There exist a solution* $\overline{x}$ *to* (26.100) *and a solution* $(\overline{v}_i)_{i\in I}$ *to* (26.101) *such that* $z - \sum_{i\in I} L_i^* \overline{v}_i \in A\overline{x}$, $x_n \rightharpoonup \overline{x}$, *and* $(\forall i \in I)$ $L_i \overline{x} - r_i \in B_i^{-1} \overline{v}_i$ *and* $v_{i,n} \rightharpoonup \overline{v}_i$.

Proof. Problem 26.35 is an instance of Problem 26.28 in which $\mathcal{K} = \bigoplus_{i\in I} \mathcal{K}_i$, $L\colon \mathcal{H} \to \mathcal{K}\colon x \mapsto (L_i x)_{i\in I}$ (hence $L^*\colon \mathcal{K} \to \mathcal{H}\colon (y_i)_{i\in I} \mapsto \sum_{i\in I} L_i^* y_i$), and $B\colon \mathcal{K} \to 2^{\mathcal{K}}\colon (y_i)_{i\in I} \mapsto \bigtimes_{i\in I} B_i y_i$. Now set $(\forall n \in \mathbb{N})$ $q_{2,n} = (q_{2,i,n})_{i\in I}$, $p_{2,n} = (p_{2,i,n})_{i\in I}$, $y_{2,n} = (y_{2,i,n})_{i\in I}$, and $v_n = (v_{i,n})_{i\in I}$. In this setting, it follows from (23.14) that (26.98) coincides with (26.103) and the claims therefore follow from Theorem 26.34. ☐

In the remainder of this section, we propose an alternative approach to solving Problem 26.28 which does not require the knowledge of $\|L\|$ as in Theorem 26.34. This approach draws on the following analogy between Problem 26.28 and the primal-dual problem (26.1)–(26.2).

Proposition 26.37 *Consider the setting of Problem 26.28. Set* $\boldsymbol{V} = \operatorname{gra} L$, $\boldsymbol{\mathcal{H}} = \mathcal{H} \oplus \mathcal{K}$, $\boldsymbol{A} = N_{\boldsymbol{V}}$, *and* $\boldsymbol{B}\colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}\colon (x, y) \mapsto (-z + Ax) \times B(y - r)$. *Consider the primal inclusion*

$$\text{find } \boldsymbol{x} \in \boldsymbol{\mathcal{H}} \text{ such that } \boldsymbol{0} \in \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{x} \tag{26.104}$$

and the dual inclusion

$$\text{find } \boldsymbol{u} \in \boldsymbol{\mathcal{H}} \text{ such that } \boldsymbol{0} \in -\boldsymbol{A}^{-1}(-\boldsymbol{u}) + \boldsymbol{B}^{-1}\boldsymbol{u}. \tag{26.105}$$

Then the following hold:

(i) *Let* $\boldsymbol{x} = (x, y) \in \boldsymbol{\mathcal{H}}$. *Then* $\boldsymbol{x}$ *solves* (26.104) *if and only if* $y = Lx$ *and* x *solves* (26.91).

(ii) *Let* $\boldsymbol{u} = (u, v) \in \boldsymbol{\mathcal{H}}$. *Then* $\boldsymbol{u}$ *solves* (26.105) *if and only if* $u = -L^* v$ *and* v *solves* (26.92).

Proof. Let $\boldsymbol{Z}$ be the set of Kuhn–Tucker points defined in (26.95).

(i): We derive from Example 6.43 and Fact 2.25(vii) that

$$\begin{aligned}
\boldsymbol{0} \in \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{x} &\Leftrightarrow (0,0) \in N_{\boldsymbol{V}}(x,y) + \boldsymbol{B}(x,y) \\
&\Leftrightarrow \begin{cases} (x,y) \in \boldsymbol{V} \\ (0,0) \in \boldsymbol{V}^{\perp} + \big((-z + Ax) \times B(y - r)\big) \end{cases} \\
&\Leftrightarrow \begin{cases} Lx = y \\ \big(\exists\, u \in -z + Ax\big)\big(\exists\, v \in B(y - r)\big) \;\; u = -L^* v \end{cases} \\
&\Leftrightarrow \begin{cases} Lx = y \\ (\exists\, v \in \mathcal{K}) \;\; z - L^* v \in Ax \;\text{ and }\; Lx - r \in B^{-1} v \end{cases} \\
&\Leftrightarrow \begin{cases} Lx = y \\ (\exists\, v \in \mathcal{K}) \;\; (x, v) \in \boldsymbol{Z}, \end{cases}
\end{aligned} \tag{26.106}$$

and it therefore follows from Proposition 26.33(ii) that x solves (26.91). Conversely, if x solves (26.91), Proposition 26.33(ii) implies that there exists $v \in \mathcal{K}$ such that $(x, v) \in \boldsymbol{Z}$. In turn, it follows from (26.106) that (x, Lx) solves (26.104).

(ii): By Example 6.43 and Fact 2.25(vii),

$$\begin{aligned}
\boldsymbol{0} \in -\boldsymbol{A}^{-1}(-\boldsymbol{u}) + \boldsymbol{B}^{-1}\boldsymbol{u}
&\Leftrightarrow (0,0) \in -N_{\boldsymbol{V}^\perp}(u,v) + \boldsymbol{B}^{-1}(u,v) \\
&\Leftrightarrow \begin{cases} (u,v) \in \boldsymbol{V}^\perp \\ (0,0) \in \boldsymbol{V} + \big(A^{-1}(u+z) \times (r + B^{-1}v)\big) \end{cases} \\
&\Leftrightarrow \begin{cases} u = -L^*v \\ \big(\exists\, x \in A^{-1}(u+z)\big)\big(\exists\, y \in r + B^{-1}v\big) \;\; y = Lx \end{cases} \\
&\Leftrightarrow \begin{cases} u = -L^*v \\ (\exists\, x \in \mathcal{H}) \;\; z - L^*v \in Ax \;\text{ and }\; Lx - r \in B^{-1}v \end{cases} \\
&\Leftrightarrow \begin{cases} u = -L^*v \\ (\exists\, x \in \mathcal{H}) \;\; (x,v) \in \boldsymbol{Z}, \end{cases}
\end{aligned} \tag{26.107}$$

and we deduce from Proposition 26.33(ii) that v solves (26.92). Finally, if v solves (26.92), Proposition 26.33(ii) asserts that there exists $x \in \mathcal{H}$ such that $(x, v) \in \boldsymbol{Z}$, and (26.107) implies that $(-L^*v, v)$ solves (26.105). ☐

Proposition 26.38 *Consider the setting of Problem 26.28 and suppose that $\mathscr{P} \neq \varnothing$. Let $s_0 \in \mathcal{H}$, let $t_0 \in \mathcal{K}$, and let $\gamma \in \mathbb{R}_{++}$. Set $Q = (\mathrm{Id} + L^*L)^{-1}$ and*

$$\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor \begin{array}{l}
x_n = J_{\gamma A}(s_n + \gamma z), \\
y_n = r + J_{\gamma B}(t_n - r), \\
u_n = \gamma^{-1}(s_n - x_n), \\
v_n = \gamma^{-1}(t_n - y_n), \\
p_n = Q\big(2x_n - s_n + L^*(2y_n - t_n)\big), \\
q_n = Lp_n, \\
s_{n+1} = s_n + \lambda_n(p_n - x_n), \\
t_{n+1} = t_n + \lambda_n(q_n - y_n).
\end{array} \right.
\end{array} \tag{26.108}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\mathscr{P}$ and $(v_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\mathscr{D}$.

Proof. Set $\boldsymbol{V} = \operatorname{gra} L$, $\boldsymbol{\mathcal{H}} = \mathcal{H} \oplus \mathcal{K}$, $\boldsymbol{A} = N_{\boldsymbol{V}}$, and $\boldsymbol{B} \colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}} \colon (x,y) \mapsto (-z + Ax) \times B(y - r)$. We derive from Example 29.19(ii) that

$$J_{\gamma \boldsymbol{A}} = P_{\boldsymbol{V}} \colon (x,y) \mapsto \big(Q(x + L^*y), LQ(x + L^*y)\big), \tag{26.109}$$

and from Proposition 23.18 and Proposition 23.17(ii)&(iii) that

$$J_{\gamma \boldsymbol{B}}\colon (x,y) \mapsto \big(J_{\gamma A}(x+\gamma z), r + J_{\gamma B}(y-r)\big). \tag{26.110}$$

In addition, set $(\forall n \in \mathbb{N})$ $\boldsymbol{x}_n = (x_n, y_n)$, $\boldsymbol{u}_n = (u_n, v_n)$, $\boldsymbol{y}_n = (s_n, t_n)$, and $\boldsymbol{z}_n = (p_n, q_n)$. Altogether, (26.108) reduces to the Douglas–Rachford splitting algorithm

$$\begin{array}{l} \text{for } n = 0,1,\ldots \\ \left\lfloor \begin{array}{l} \boldsymbol{x}_n = J_{\gamma \boldsymbol{B}} \boldsymbol{y}_n, \\ \boldsymbol{u}_n = \gamma^{-1}(\boldsymbol{y}_n - \boldsymbol{x}_n), \\ \boldsymbol{z}_n = J_{\gamma \boldsymbol{A}}(2\boldsymbol{x}_n - \boldsymbol{y}_n), \\ \boldsymbol{y}_{n+1} = \boldsymbol{y}_n + \lambda_n(\boldsymbol{z}_n - \boldsymbol{x}_n). \end{array} \right. \end{array} \tag{26.111}$$

Therefore the claims follow from Theorem 26.11 and Proposition 26.37. ☐

Remark 26.39 Set $R = L^*(\mathrm{Id} + LL^*)^{-1}$. Using Example 29.19(i) and the fact that $\operatorname{ran} J_{\boldsymbol{A}} = \boldsymbol{V}$, we can replace (26.109) by

$$\begin{aligned} J_{\boldsymbol{A}}\colon (x,y) &\mapsto \big(x - R(Lx-y), y + (\mathrm{Id}+LL^*)^{-1}(Lx-y)\big) \\ &= \big(x - R(Lx-y), L(x - R(Lx-y))\big). \end{aligned} \tag{26.112}$$

Thus, the conclusions of Proposition 26.38 remain valid by replacing the updating rule for p_n by

$$p_n = 2x_n - s_n - R\big(L(2x_n - s_n) - 2y_n + t_n\big). \tag{26.113}$$

Exercises

Exercise 26.1 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$, let $B\colon \mathcal{H} \to 2^{\mathcal{H}}$, and denote the sets of solutions to (26.1) and (26.2) by $\mathscr{P}$ and $\mathscr{D}$, respectively. Show that $\mathscr{P} = \operatorname{dom}(A \cap (-B))$ and $\mathscr{D} = \operatorname{dom}((A^{-1} \circ (-\mathrm{Id})) \cap B^{-1})$.

Exercise 26.2 Let C and D be nonempty closed convex subsets of $\mathcal{H}$. Show that $C \cap D = P_D(\operatorname{Fix}(2P_C - \mathrm{Id}) \circ (2P_D - \mathrm{Id}))$.

Exercise 26.3 Let A and B be maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$. Show that the graph of the Douglas–Rachford splitting operator $J_A R_B + \mathrm{Id} - J_B$ is

$$\big\{(y+v, x+v) \;\big|\; (x,u) \in \operatorname{gra} A,\ (y,v) \in \operatorname{gra} B,\ x+u = y-v\big\}. \tag{26.114}$$

Exercise 26.4 Let A and B be maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$, and let $T_{A,B} = J_A R_B + \mathrm{Id} - J_B$ be the associated Douglas–Rachford splitting operator. Show that $T_{A,B} + T_{A^{-1},B} = \mathrm{Id}$.

Exercise 26.5 Let A and B be maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$. Use Exercise 25.19 to show that $\operatorname{zer}(A+B)$ is convex.

Exercise 26.6 Let A and B be operators from $\mathcal{H}$ to $2^{\mathcal{H}}$.

(i) Show that $\operatorname{zer}(A+B) \neq \varnothing \Leftrightarrow \operatorname{zer}(A^{-1} - (B^{-1})^{\vee}) \neq \varnothing$.
(ii) Let f and g be functions in $\Gamma_0(\mathcal{H})$ such that $0 \in \operatorname{sri}(\operatorname{dom} f - \operatorname{dom} g)$, and set $A = \partial f$ and $B = \partial g$. Do we recover Proposition 15.13 from (i)?

Exercise 26.7 In the setting of Theorem 26.11, suppose that $\lambda_n \equiv \gamma = 1$ so that (26.29) can be written as

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} x_n = J_B y_n, \\ y_{n+1} = J_A(2x_n - y_n) + y_n - x_n. \end{array} \right. \end{array} \tag{26.115}$$

On the other hand, suppose that $y_0 \in \operatorname{zer} B$, set $b_0 = y_0$ and $u_0 = 0 \in \mathcal{H}$, and update via

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} a_{n+1} = J_A(b_n - u_n), \\ b_{n+1} = J_B(a_{n+1} + u_n), \\ u_{n+1} = u_n + a_{n+1} - b_{n+1}. \end{array} \right. \end{array} \tag{26.116}$$

Show that $(a_n + u_{n-1})_{n\geqslant 1} = (x_n)_{n\geqslant 1}$ and $(b_n)_{n\geqslant 1} = (y_n)_{n\geqslant 1}$. (The latter algorithm is related to a special case of the *alternating direction method of multipliers*.)

Exercise 26.8 In the setting of Theorem 26.11, suppose that $\mathcal{H}$ is finite-dimensional. Show that $(x_n)_{n\in\mathbb{N}}$ converges to a point in $\operatorname{zer}(A+B)$ without using items (iii) and (iv).

Exercise 26.9 In the setting of Theorem 26.11, suppose that B is affine and continuous. Show that $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{zer}(A+B)$ without using items (iii) and (iv).

Exercise 26.10 Let A and B be maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$, let $\gamma \in \mathbb{R}_{++}$, and suppose that the problem

$$\text{find} \quad x \in \mathcal{H} \quad \text{such that} \quad 0 \in Ax + {}^{\mu}Bx \tag{26.117}$$

has at least one solution. Devise an algorithm using the resolvent of A and B separately to solve (26.117), and establish a convergence result.

Exercise 26.11 Find the dual problem of (26.82) in the form of a variational inequality.

Exercise 26.12 Deduce the strong and weak convergence results of Example 23.40 from Theorem 26.11.

Exercise 26.13 In Theorem 26.11 make the additional assumption that the operators A and B are odd and that $(\forall n \in \mathbb{N})\ \lambda_n = 1$. Prove the following:

(i) $J_{\gamma A}$ and $J_{\gamma B}$ are odd.
(ii) The convergence of the sequences $(x_n)_{n\in\mathbb{N}}$, $(y_n)_{n\in\mathbb{N}}$, and $(z_n)_{n\in\mathbb{N}}$ defined in (26.29) is strong.

Exercise 26.14 Deduce the convergence results of Example 23.40 from Theorem 26.14 and Proposition 26.16.

Exercise 26.15 In Theorem 26.14(i), make the additional assumption that $\operatorname{int}\operatorname{zer}(A+B) \neq \varnothing$. Show that the sequence produced by the forward-backward algorithm (26.53) converges strongly to a zero of $A+B$.

Exercise 26.16 In the setting of Proposition 26.16(ii), show that B is cocoercive.

Exercise 26.17 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, and let I be a nonempty finite set. For every $i \in I$, let $\mathcal{K}_i$ be a real Hilbert space, let $B_i\colon \mathcal{K}_i \to 2^{\mathcal{K}_i}$ be maximally monotone, let $r_i \in \mathcal{K}_i$, and suppose that $L_i \in \mathcal{B}(\mathcal{H},\mathcal{K}_i) \smallsetminus \{0\}$. Set $\mathcal{K} = \bigoplus_{i\in I} \mathcal{K}_i$, $\mathcal{G} = \mathcal{H} \oplus \mathcal{K}$, $L\colon \mathcal{H} \to \mathcal{G}\colon x \mapsto (x, (L_i x)_{i\in I})$, and $B\colon \mathcal{G} \to 2^{\mathcal{G}}\colon (x,(y_i)_{i\in I}) \mapsto Ax \times \bigtimes_{i\in I} B_i(y_i - r_i)$. Show that

$$\operatorname{zer}(L^*BL) = \operatorname{zer}\left(A + \sum_{i\in I} L_i^* B_i L_i(\cdot - r_i)\right). \tag{26.118}$$

Specialize to the case when $A = 0$, $I = \{1,2\}$, $\mathcal{K}_1 = \mathcal{K}_2 = \mathcal{H}$, $L_1 = L_2 = \mathrm{Id}$, and $r_1 = r_2 = 0$.

Exercise 26.18 Let $\mathcal{K}$ be a real Hilbert space, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{K} \to 2^{\mathcal{K}}$ be maximally monotone, let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$, and set $\boldsymbol{V} = \big\{(x,y) \in \mathcal{H} \oplus \mathcal{K} \;\big|\; Lx = y\big\}$. Set $\boldsymbol{A}\colon (x,y) \mapsto Ax \times By$ and $\boldsymbol{B} = N_{\boldsymbol{V}}$. Show that

$$\operatorname{zer}(\boldsymbol{A} + \boldsymbol{B}) = \big\{(x, Lx) \;\big|\; x \in \operatorname{zer}(A + L^*BL)\big\}. \tag{26.119}$$

Chapter 27
Fermat's Rule in Convex Optimization

Fermat's rule (Theorem 16.3) provides a simple characterization of the minimizers of a function as the zeros of its subdifferential. This chapter explores various consequences of this fact.

Throughout, $\mathcal{K}$ is a real Hilbert space.

27.1 General Characterizations of Minimizers

Let us first provide characterizations of the minimizers of a function in $\Gamma_0(\mathcal{H})$.

Proposition 27.1 *Let* $f \in \Gamma_0(\mathcal{H})$. *Then*

$$\operatorname{Argmin} f = \operatorname{zer} \partial f = \partial f^*(0) = \operatorname{Fix} \operatorname{Prox}_f = \operatorname{zer} \operatorname{Prox}_{f^*} . \tag{27.1}$$

Proof. Let $x \in \mathcal{H}$. Then $x \in \operatorname{Argmin} f \Leftrightarrow 0 \in \partial f(x)$ (by Theorem 16.3) $\Leftrightarrow$ $x - 0 = x \in \partial f^*(0)$ (by Corollary 16.30) $\Leftrightarrow \operatorname{Prox}_{f^*} x = 0$ (by (16.37)) $\Leftrightarrow$ $\operatorname{Prox}_f x = x$ (by (14.6)). ☐

The next theorem is an application of Fermat's rule to the minimization of the sum of two convex functions satisfying a constraint qualification.

Theorem 27.2 *Let* $f \in \Gamma_0(\mathcal{H})$, *let* $g \in \Gamma_0(\mathcal{K})$, *let* $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, *and let* $\bar{x} \in \mathcal{H}$. *Suppose that one of the following holds:*

(a) $0 \in \operatorname{sri}(\operatorname{dom} g - L(\operatorname{dom} f))$ *(see Proposition 6.19 for special cases).*
(b) $\mathcal{K}$ *is finite-dimensional,* g *is polyhedral, and* $\operatorname{dom} g \cap \operatorname{ri} L(\operatorname{dom} f) \neq \varnothing$.
(c) $\mathcal{H}$ *and* $\mathcal{K}$ *are finite-dimensional,* f *and* g *are polyhedral, and* $\operatorname{dom} g \cap L(\operatorname{dom} f) \neq \varnothing$.

Then the following are equivalent:

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_27

(i) $\bar{x}$ *is a solution to the problem*

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; f(x) + g(Lx). \tag{27.2}$$

(ii) $\bar{x} \in \operatorname{zer}(\partial f + L^* \circ (\partial g) \circ L)$.
(iii) $(\exists\, v \in \partial g(L\bar{x}))\ -L^*v \in \partial f(\bar{x})$.
(iv) $(\exists\, v \in \partial g(L\bar{x}))(\forall y \in \mathcal{H})\ \langle \bar{x} - y \mid L^*v\rangle + f(\bar{x}) \leqslant f(y)$.

Moreover, if g is Gâteaux differentiable at $L\bar{x}$, each of items (i)–(iv) is also equivalent to each of the following:

(v) $-L^*(\nabla g(L\bar{x})) \in \partial f(\bar{x})$.
(vi) $(\forall y \in \mathcal{H})\ \langle \bar{x} - y \mid L^*(\nabla g(L\bar{x}))\rangle + f(\bar{x}) \leqslant f(y)$.
(vii) $(\forall \gamma \in \mathbb{R}_{++})\ \bar{x} = \operatorname{Prox}_{\gamma f}(\bar{x} - \gamma L^*(\nabla g(L\bar{x}))$.
(viii) $f(\bar{x}) + f^*(-L^*(\nabla g(L\bar{x}))) + \langle L\bar{x} \mid \nabla g(L\bar{x})\rangle = 0$.

Proof. (i)⇔(ii): It follows from Proposition 27.1 and Theorem 16.47 that $\operatorname{Argmin}(f + g\circ L) = \operatorname{zer}\partial(f + g\circ L) = \operatorname{zer}(\partial f + L^* \circ (\partial g)\circ L)$.

(ii)⇔(iii): $\bar{x} \in \operatorname{zer}(\partial f + L^* \circ (\partial g) \circ L) \Leftrightarrow 0 \in \partial f(\bar{x}) + L^*(\partial g(L\bar{x})) \Leftrightarrow (\exists\, v \in \partial g(L\bar{x}))\ -L^*v \in \partial f(\bar{x})$.

(iii)⇔(iv): Definition 16.1.

Now assume that g is Gâteaux differentiable at $L\bar{x}$.

(iii)⇔(v): Proposition 17.31(i).

(iv)⇔(vi): Proposition 17.31(i).

(v)⇔(vii): Let $\gamma \in \mathbb{R}_{++}$. Then (16.37) yields $-L^*(\nabla g(L\bar{x})) \in \partial f(\bar{x}) \Leftrightarrow (\bar{x} - \gamma L^*(\nabla g(L\bar{x}))) - \bar{x} \in \partial(\gamma f)(\bar{x}) \Leftrightarrow \bar{x} = \operatorname{Prox}_{\gamma f}(\bar{x} - \gamma L^*(\nabla g(L\bar{x})))$.

(v)⇔(viii): Proposition 16.10. ☐

Corollary 27.3 *Let f and g be functions in $\Gamma_0(\mathcal{H})$, let $\bar{x} \in \mathcal{H}$, and let $\gamma \in \mathbb{R}_{++}$. Suppose that one of the following holds:*

(a) $0 \in \operatorname{sri}(\operatorname{dom} g - \operatorname{dom} f)$ *(see Proposition 6.19 for special cases).*
(b) $\mathcal{H}$ *is finite-dimensional, g is polyhedral, and* $\operatorname{dom} g \cap \operatorname{ri}\operatorname{dom} f \neq \varnothing$.
(c) $\mathcal{H}$ *is finite-dimensional, f and g are polyhedral, and* $\operatorname{dom} f \cap \operatorname{dom} g \neq \varnothing$.

Then the following are equivalent:

(i) $\bar{x}$ *is a solution to the problem*

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; f(x) + g(x). \tag{27.3}$$

(ii) $\bar{x} \in \operatorname{zer}(\partial f + \partial g)$.
(iii) $\bar{x} \in \operatorname{Prox}_{\gamma g}(\operatorname{Fix}(2\operatorname{Prox}_{\gamma f} - \operatorname{Id}) \circ (2\operatorname{Prox}_{\gamma g} - \operatorname{Id}))$.
(iv) $(\exists\, u \in \partial g(\bar{x}))\ -u \in \partial f(\bar{x})$.
(v) $(\exists\, u \in \partial g(\bar{x}))(\forall y \in \mathcal{H})\ \langle \bar{x} - y \mid u\rangle + f(\bar{x}) \leqslant f(y)$.

Moreover, if g is Gâteaux differentiable at $\bar{x}$, each of items (i)–(v) is also equivalent to each of the following:

(vi) $-\nabla g(\bar{x}) \in \partial f(\bar{x})$.
(vii) $(\forall y \in \mathcal{H})\ \langle \bar{x} - y \mid \nabla g(\bar{x})\rangle + f(\bar{x}) \leqslant f(y)$.
(viii) $\bar{x} = \operatorname{Prox}_{\gamma f}(\bar{x} - \gamma \nabla g(\bar{x}))$.
(ix) $f(\bar{x}) + f^*(-\nabla g(\bar{x})) + \langle \bar{x} \mid \nabla g(\bar{x})\rangle = 0$.

Proof. Applying Theorem 27.2 to $\mathcal{K} = \mathcal{H}$ and $L = \mathrm{Id}$ yields all the results, except the equivalences involving (iii). The equivalence (ii)⇔(iii) follows from Theorem 20.25, Proposition 26.1(iii), and Example 23.3. ☐

Remark 27.4 Condition (v) in Corollary 27.3 is an instance of the variational inequality featured in Definition 26.19.

Proposition 27.5 *Let $\mathcal{K}$ be a real Hilbert space, let $f \in \Gamma_0(\mathcal{H})$, let $g \in \Gamma_0(\mathcal{K})$, and let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$. Then the following hold:*

(i) $\operatorname{zer}(\partial f + L^* \circ (\partial g) \circ L) \subset \operatorname{Argmin}(f + g \circ L)$.
(ii) $\operatorname{zer}(-L \circ (\partial f^*) \circ (-L^*) + \partial g^*) \subset \operatorname{Argmin}(f^* \circ (-L^*) + g^*)$.
(iii) *Suppose that one of the following is satisfied:*
 (a) $\operatorname{Argmin}(f + g \circ L) \neq \varnothing$ *and one of the following holds:*
  1/ $0 \in \operatorname{sri}(\operatorname{dom} g - L(\operatorname{dom} f))$ *(see Proposition 6.19 for special cases).*
  2/ $\mathcal{K}$ *is finite-dimensional and* $(\operatorname{ri} \operatorname{dom} g) \cap (\operatorname{ri} L(\operatorname{dom} f)) \neq \varnothing$.
  3/ $\mathcal{K}$ *is finite-dimensional, g is polyhedral, and*
$$\operatorname{dom} g \cap \operatorname{ri} L(\operatorname{dom} f) \neq \varnothing. \tag{27.4}$$
  4/ $\mathcal{H}$ *and* $\mathcal{K}$ *are finite-dimensional, f and g are polyhedral, and* $\operatorname{dom} g \cap L(\operatorname{dom} f) \neq \varnothing$.
 (b) $\operatorname{Argmin}(f+g\circ L) \subset \operatorname{Argmin} f \cap \operatorname{Argmin} g\circ L \neq \varnothing$ *and* $0 \in \operatorname{sri}(\operatorname{dom} g - \operatorname{ran} L)$.
 (c) *C and D are closed convex subsets of $\mathcal{H}$ and $\mathcal{K}$, respectively, such that* $C \cap L^{-1}(D) \neq \varnothing$ *and* $0 \in \operatorname{sri}(D - \operatorname{ran} L)$, $f = \iota_C$, *and* $g = \iota_D$.

 Then
$$\operatorname{Argmin}(f + g \circ L) = \operatorname{zer}\big(\partial f + L^* \circ (\partial g) \circ L\big) \neq \varnothing. \tag{27.5}$$

Proof. (i)&(ii): By Proposition 16.6(ii) and Theorem 16.3,
$$\operatorname{zer}\big(\partial f + L^* \circ (\partial g) \circ L\big) \subset \operatorname{zer} \partial(f + g \circ L) = \operatorname{Argmin}(f + g \circ L). \tag{27.6}$$

This shows (i). We obtain (ii) similarly.

(iii)(a): It follows from Theorem 16.3 and Theorem 16.47 that
$$\varnothing \neq \operatorname{Argmin}(f + g \circ L) = \operatorname{zer} \partial(f + g \circ L) = \operatorname{zer}(\partial f + L^* \circ (\partial g) \circ L). \tag{27.7}$$

(iii)(b): Using Theorem 16.3, Corollary 16.53(i), and (i), we obtain

$$\begin{aligned}
\varnothing &\neq \operatorname{Argmin} f \cap \operatorname{Argmin}\big(g \circ L\big) \\
&= (\operatorname{zer} \partial f) \cap \operatorname{zer} \partial(g \circ L) \\
&= (\operatorname{zer} \partial f) \cap \operatorname{zer}\big(L^* \circ (\partial g) \circ L\big) \\
&\subset \operatorname{zer}\big(\partial f + L^* \circ (\partial g) \circ L\big) \\
&\subset \operatorname{Argmin}(f + g \circ L) \\
&\subset \operatorname{Argmin} f \cap \operatorname{Argmin}(g \circ L).
\end{aligned} \tag{27.8}$$

(iii)(c): Since $\operatorname{dom}(\iota_C + \iota_D \circ L) = C \cap L^{-1}(D)$,

$$\begin{aligned}
\operatorname{Argmin}\big(\iota_C + \iota_D \circ L\big) &= \operatorname{Argmin} \iota_{C \cap L^{-1}(D)} \\
&= C \cap L^{-1}(D) \\
&= \operatorname{Argmin} \iota_C \cap \operatorname{Argmin}(\iota_D \circ L) \\
&\neq \varnothing.
\end{aligned} \tag{27.9}$$

Applying (iii)(b) to $f = \iota_C$ and $g = \iota_D$ completes the proof. ☐

Corollary 27.6 *Let f and g be functions in $\Gamma_0(\mathcal{H})$ such that one of the following holds:*

(i) $\operatorname{Argmin}(f + g) \neq \varnothing$ *(see Corollary 11.16 for sufficient conditions) and one of the following is satisfied:*

   (a) $0 \in \operatorname{sri}(\operatorname{dom} g - \operatorname{dom} f)$.
   (b) $\mathcal{H}$ *is finite-dimensional and* $(\operatorname{ri}\operatorname{dom} g) \cap (\operatorname{ri}\operatorname{dom} f) \neq \varnothing$.
   (c) $\mathcal{H}$ *is finite-dimensional, g is polyhedral, and* $\operatorname{dom} g \cap \operatorname{ri}\operatorname{dom} f \neq \varnothing$.
   (d) $\mathcal{H}$ *is finite-dimensional, f and g are polyhedral, and* $\operatorname{dom} g \cap \operatorname{dom} f \neq \varnothing$.

(ii) $\operatorname{Argmin}(f + g) \subset \operatorname{Argmin} f \cap \operatorname{Argmin} g \neq \varnothing$.
(iii) *C and D are closed convex subsets of $\mathcal{H}$ such that $C \cap D \neq \varnothing$, $f = \iota_C$, and $g = \iota_D$.*

Then $\operatorname{Argmin}(f + g) = \operatorname{zer}(\partial f + \partial g) \neq \varnothing$.

Proof. Apply Proposition 27.5(iii) with $\mathcal{K} = \mathcal{H}$ and $L = \operatorname{Id}$. ☐

Remark 27.7 It follows from Example 16.51 that assumption (iii) in Corollary 27.6 may hold while (i)(a) fails.

27.2 Abstract Constrained Minimization Problems

The problem under consideration in this section is the characterization of the minimizers of a function $f \in \Gamma_0(\mathcal{H})$ over a closed convex set.

Proposition 27.8 *Let C be a closed convex subset of $\mathcal{H}$, let $f \in \Gamma_0(\mathcal{H})$, let $\bar{x} \in \mathcal{H}$, and let $\gamma \in \mathbb{R}_{++}$. Suppose that one of the following holds:*

(a) $0 \in \operatorname{sri}(C - \operatorname{dom} f)$ *(see Proposition 6.19 for special cases).*
(b) *$\mathcal{H}$ is finite-dimensional, C is polyhedral, and $C \cap \operatorname{ri}(\operatorname{dom} f) \neq \varnothing$.*
(c) *$\mathcal{H}$ is finite-dimensional, C is a polyhedral set, f is a polyhedral function, and $C \cap \operatorname{dom} f \neq \varnothing$.*

Then the following are equivalent:

(i) *$\bar{x}$ is a solution to the problem*

$$\underset{x \in C}{\operatorname{minimize}} \;\; f(x). \tag{27.10}$$

(ii) $\bar{x} \in \operatorname{zer}(N_C + \partial f)$.
(iii) $\bar{x} \in \operatorname{Prox}_{\gamma f}(\operatorname{Fix}(2P_C - \operatorname{Id}) \circ (2\operatorname{Prox}_{\gamma f} - \operatorname{Id}))$.
(iv) $(\exists\, u \in N_C \bar{x})\ -u \in \partial f(\bar{x})$.
(v) $(\exists\, u \in \partial f(\bar{x}))\ -u \in N_C \bar{x}$.
(vi) *$\bar{x} \in C$ and* $(\exists\, u \in \partial f(\bar{x}))(\forall y \in C)\ \langle \bar{x} - y \mid u \rangle \leqslant 0$.

Moreover, if f is Gâteaux differentiable at $\bar{x}$, each of items (i)–(vi) is also equivalent to each of the following:

(vi) $-\nabla f(\bar{x}) \in N_C \bar{x}$.
(vii) *$\bar{x} \in C$ and* $(\forall y \in C)\ \langle \bar{x} - y \mid \nabla f(\bar{x}) \rangle \leqslant 0$.
(viii) $\bar{x} = P_C(\bar{x} - \gamma \nabla f(\bar{x}))$.

Proof. Apply Corollary 27.3 to the functions ι_C and f, and use Example 16.13 and Example 12.25. ☐

Remark 27.9 Condition (vii) in Proposition 27.8 is an instance of the variational inequality considered in Example 26.21.

Example 27.10 Let C be a nonempty closed convex subset of $\mathcal{H}$, let $x \in \mathcal{H}$, let $p \in \mathcal{H}$, and set $f = (1/2)\|\cdot - x\|^2$. Then Example 2.60 yields $\nabla f \colon y \mapsto y - x$ and we deduce from Proposition 27.8 that $p = P_C x \Leftrightarrow x - p \in N_C p$. We thus recover Proposition 6.47.

The following two results have found many uses in the study of partial differential equations.

Example 27.11 (Stampacchia) Let $F \colon \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ be a bilinear form such that, for some α and β in $\mathbb{R}_{++}$,

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad |F(x,y)| \leqslant \beta \|x\|\,\|y\| \quad \text{and} \quad F(x,x) \geqslant \alpha \|x\|^2, \tag{27.11}$$

let C be a nonempty closed convex subset of $\mathcal{H}$, and let $\ell \in \mathcal{B}(\mathcal{H}, \mathbb{R})$. Then the following hold:

(i) There exists a unique point $\bar{x} \in \mathcal{H}$ such that

$$\bar{x} \in C \quad \text{and} \quad (\forall y \in C) \quad F(\bar{x}, y - \bar{x}) \geqslant \ell(y - \bar{x}). \tag{27.12}$$

(ii) Suppose that F is symmetric, let $\bar{x} \in \mathcal{H}$, and set $f \colon \mathcal{H} \to \mathbb{R} \colon x \mapsto (1/2)F(x,x) - \ell(x)$. Then the following are equivalent:

(a) $\bar{x}$ solves (27.12).
(b) $\operatorname{Argmin}_C f = \{\bar{x}\}$.

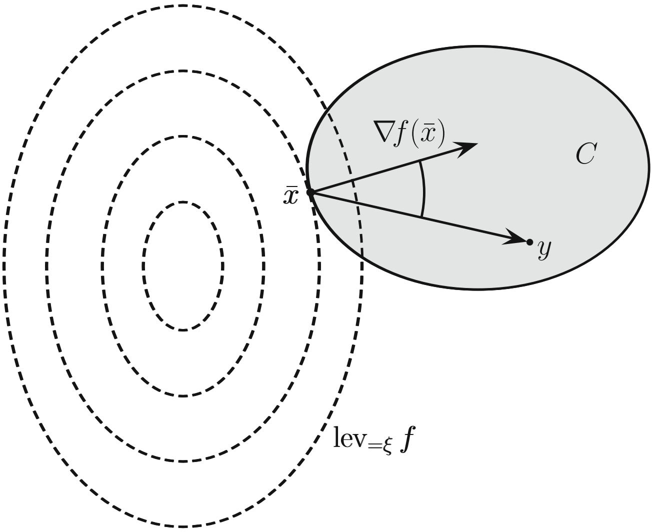

Fig. 27.1 Illustration of the equivalence (i)⇔(vii) in Proposition 27.8 when $\mathcal{H} = \mathbb{R}^2$: $\bar{x} \in C$ is a minimizer of f over C if and only if, for every $y \in C$, $\langle y - \bar{x} \mid \nabla f(\bar{x}) \rangle \geqslant 0$, i.e., the vectors $y - \bar{x}$ and $\nabla f(\bar{x})$ form an acute angle. Each dashed line represents a level line $\operatorname{lev}_{=\xi} f$.

Proof. (i): By Fact 2.24, there exists $u \in \mathcal{H}$ such that $\ell = \langle \cdot \mid u \rangle$. Similarly, for every $x \in \mathcal{H}$, $F(x, \cdot) \in \mathcal{B}(\mathcal{H}, \mathbb{R})$ and, therefore, there exists a vector in $\mathcal{H}$, which we denote by Lx, such that $F(x, \cdot) = \langle \cdot \mid Lx \rangle$. We derive from (27.11) and Example 20.34 that L is a strongly monotone, maximally monotone operator in $\mathcal{B}(\mathcal{H})$, and hence that $B \colon x \mapsto Lx - u$ is likewise. Furthermore, (6.35) implies that the set of solutions to (27.12) is $\operatorname{zer}(N_C + B)$, while Example 20.26 and Corollary 25.5(i) imply that $N_C + B$ is maximally monotone. Thus, since $N_C + B$ is strongly monotone, the claim follows from Corollary 23.37(ii).

(ii): It is clear that $f \in \Gamma_0(\mathcal{H})$ and that $\operatorname{dom} f = \mathcal{H}$. Moreover, by Example 2.59, $\mathsf{D}f(x) = F(x, \cdot) - \ell$. Hence, the result follows from (i) and the equivalence (i)⇔(vii) in Proposition 27.8. $\square$

Example 27.12 (Lax–Milgram) Let $F \colon \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ be a bilinear form such that, for some α and β in $\mathbb{R}_{++}$,

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad |F(x,y)| \leqslant \beta\|x\|\,\|y\| \quad \text{and} \quad F(x,x) \geqslant \alpha\|x\|^2, \quad (27.13)$$

and let $\ell \in \mathcal{B}(\mathcal{H}, \mathbb{R})$. Then the following hold:

(i) There exists a unique point $\bar{x} \in \mathcal{H}$ such that

$$(\forall y \in \mathcal{H}) \quad F(\bar{x}, y) = \ell(y). \qquad (27.14)$$

(ii) Suppose that F is symmetric, let $\bar{x} \in \mathcal{H}$, and set $f \colon \mathcal{H} \to \mathbb{R} \colon x \mapsto (1/2)F(x,x) - \ell(x)$. Then the following are equivalent:

(a) $\bar{x}$ solves (27.14).
(b) $\operatorname{Argmin} f = \{\bar{x}\}$.

Proof. Set $C = \mathcal{H}$ in Example 27.11. ☐

Proposition 27.13 *Let $f \in \Gamma_0(\mathcal{H})$ be such that* $\operatorname{int}\operatorname{dom} f \neq \varnothing$*, and let C be a nonempty convex subset of* $\operatorname{int}\operatorname{dom} f$*. Suppose that f is Gâteaux differentiable on C, and that x and y belong to* $\operatorname{Argmin}_C f$*. Then $\nabla f(x) = \nabla f(y)$.*

Proof. By the implication (i)⇒(vii) in Proposition 27.8, $\langle x - y \mid \nabla f(x)\rangle \leqslant 0$ and $\langle y - x \mid \nabla f(y)\rangle \leqslant 0$. Hence, $\langle x - y \mid \nabla f(x) - \nabla f(y)\rangle \leqslant 0$ and, by Proposition 17.7(iii), $\langle x - y \mid \nabla f(x) - \nabla f(y)\rangle = 0$. In turn, by Example 22.4(i) and Proposition 17.31(i), ∇f is paramonotone. Thus, $\nabla f(x) = \nabla f(y)$. ☐

27.3 Affine Constraints

We first revisit the setting of Proposition 19.21 in the light of Theorem 27.2.

Proposition 27.14 *Let $f \in \Gamma_0(\mathcal{H})$, let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, let $\bar{x} \in \mathcal{H}$, and suppose that*

$$r \in \operatorname{sri} L(\operatorname{dom} f). \qquad (27.15)$$

Consider the problem

$$\underset{\substack{x \in \mathcal{H} \\ Lx = r}}{\text{minimize}} \;\; f(x). \qquad (27.16)$$

Then $\bar{x}$ is a solution to (27.16) if and only if

$$L\bar{x} = r \quad \textit{and} \quad (\exists\, \bar{v} \in \mathcal{K}) \;\; -L^*\bar{v} \in \partial f(\bar{x}), \qquad (27.17)$$

in which case $\bar{v}$ is a Lagrange multiplier associated with $\bar{x}$, and $\bar{x}$ solves the problem

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x) + \langle x \mid L^*\bar{v}\rangle\,. \qquad (27.18)$$

Moreover, if f is Gâteaux differentiable at $\bar{x}$, (27.17) becomes

$$L\bar{x} = r \quad \textit{and} \quad (\exists\, \bar{v} \in \mathcal{K}) \;\; \nabla f(\bar{x}) = -L^*\bar{v}. \qquad (27.19)$$

Proof. Set $g = \iota_{\{r\}}$. Then Problem (27.16) is a special case of (27.2). Moreover, (27.15) implies that condition (a) in Theorem 27.2 is satisfied. Hence, the characterization (27.17) follows from the equivalence (i)⇔(iii) in Theorem 27.2, from which we deduce the characterization (27.19) via Proposition 17.31(i). Finally, it follows from Proposition 19.21(v), Theorem 15.23, and Remark 19.22 that $\bar{v}$ is a Lagrange multiplier associated with $\bar{x}$, and that $\bar{x}$ solves (27.18). ☐

In the next corollary, we revisit the setting of Corollary 19.23 using the tools of Proposition 27.14.

Corollary 27.15 *Let $f \in \Gamma_0(\mathcal{H})$, let m be a strictly positive integer, set $I = \{1, \ldots, m\}$, let $(u_i)_{i\in I} \in \mathcal{H}^m$, and suppose that*

$$(\rho_i)_{i\in I} \in \operatorname{ri}\big\{(\langle x \mid u_i\rangle)_{i\in I} \;\big|\; x \in \operatorname{dom} f\big\}. \tag{27.20}$$

Consider the problem

$$\underset{\substack{x\in\mathcal{H}\\ \langle x|u_1\rangle=\rho_1,\ldots,\langle x|u_m\rangle=\rho_m}}{\text{minimize}} \quad f(x), \tag{27.21}$$

and let $\bar{x} \in \mathcal{H}$. Then $\bar{x}$ is a solution to (27.21) if and only if

$$(\forall i \in I)\ \ \langle \bar{x} \mid u_i\rangle = \rho_i \quad \textit{and} \quad \big(\exists\, (\bar{\nu}_i)_{i\in I} \in \mathbb{R}^m\big) \ \ -\sum_{i\in I} \bar{\nu}_i u_i \in \partial f(\bar{x}), \tag{27.22}$$

in which case $(\bar{\nu}_i)_{i\in I}$ are Lagrange multipliers associated with $\bar{x}$, and $\bar{x}$ solves the problem

$$\underset{x\in\mathcal{H}}{\text{minimize}}\ \ f(x) + \sum_{i\in I} \bar{\nu}_i \langle x \mid u_i\rangle\,. \tag{27.23}$$

Moreover, if f is Gâteaux differentiable at $\bar{x}$, (27.22) becomes

$$(\forall i \in I)\ \ \langle \bar{x} \mid u_i\rangle = \rho_i \quad \textit{and} \quad \big(\exists\, (\bar{\nu}_i)_{i\in I} \in \mathbb{R}^m\big) \ \ \nabla f(\bar{x}) = -\sum_{i\in I} \bar{\nu}_i u_i. \tag{27.24}$$

Proof. Set $\mathcal{K} = \mathbb{R}^m$, $L\colon \mathcal{H} \to \mathcal{K}\colon x \mapsto (\langle x \mid u_i\rangle)_{i\in I}$, and $r = (\rho_i)_{i\in I}$. Then (27.21) appears as a special case of (27.16), $L^*\colon (\eta_i)_{i\in I} \mapsto \sum_{i\in I} \eta_i u_i$, and it follows from Fact 6.14(i) that (27.20) coincides with (27.15). Altogether, the results follow from Proposition 27.14. ☐

Next, we consider a more geometrical formulation by revisiting Proposition 27.8 in the case when C is a closed affine subspace.

Proposition 27.16 *Let $f \in \Gamma_0(\mathcal{H})$, let V be a closed linear subspace of $\mathcal{H}$, let $z \in \mathcal{H}$, and let $\bar{x} \in \mathcal{H}$. Suppose that*

$$V + \operatorname{cone}(z - \operatorname{dom} f)\ \ \textit{is a closed linear subspace}, \tag{27.25}$$

and consider the problem

$$\underset{x \in z+V}{\text{minimize}} \;\; f(x). \tag{27.26}$$

Then $\bar{x}$ is a solution to (27.26) if and only if

$$\bar{x} - z \in V \quad \textit{and} \quad \partial f(\bar{x}) \cap V^{\perp} \neq \varnothing. \tag{27.27}$$

Moreover, if f is Gâteaux differentiable at $\bar{x}$, (27.27) becomes

$$\bar{x} - z \in V \quad \textit{and} \quad \nabla f(\bar{x}) \perp V. \tag{27.28}$$

Proof. Set $C = z + V$. We have $C - \operatorname{dom} f = V + (z - \operatorname{dom} f)$. Hence, it follows from Proposition 6.19(iii) that $0 \in \operatorname{sri}(V + (z - \operatorname{dom} f)) = \operatorname{sri}(C - \operatorname{dom} f)$. Thus, (27.25) implies that (a) in Proposition 27.8 is satisfied. Therefore, the characterizations (27.27) and (27.28) follow from Example 6.43 and, respectively, from the equivalences (i)⇔(v) and (i)⇔(vi) in Proposition 27.8. ☐

27.4 Cone Constraints

We consider the problem of minimizing a convex function over the inverse image of a convex cone by a linear operator.

Proposition 27.17 *Let $f \in \Gamma_0(\mathcal{H})$, let K be a closed convex cone in $\mathcal{K}$, let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, and let $\bar{x} \in \mathcal{H}$. Suppose that one of the following holds:*

(a) $K - \operatorname{cone} L(\operatorname{dom} f)$ *is a closed linear subspace.*
(b) $\mathcal{K}$ *is finite-dimensional, K is polyhedral, and $K \cap \operatorname{ri} L(\operatorname{dom} f) \neq \varnothing$.*
(c) $\mathcal{H}$ *and $\mathcal{K}$ are finite-dimensional, f is a polyhedral function, K is polyhedral, and $K \cap L(\operatorname{dom} f) \neq \varnothing$.*

Consider the problem

$$\underset{Lx \in K}{\text{minimize}} \;\; f(x). \tag{27.29}$$

Then $\bar{x}$ is a solution to (27.29) if and only if

$$L\bar{x} \in K \quad \textit{and} \quad \big(\exists\, \bar{v} \in K^{\ominus}\big) \; \begin{cases} -L^*\bar{v} \in \partial f(\bar{x}), \\ \langle \bar{x} \mid L^*\bar{v} \rangle = 0, \end{cases} \tag{27.30}$$

in which case $\bar{v}$ is a Lagrange multiplier associated with $\bar{x}$, and $\bar{x}$ solves the problem

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x) + \langle x \mid L^*\bar{v} \rangle\,. \tag{27.31}$$

Moreover, if f is Gâteaux differentiable at $\bar{x}$, then (27.30) becomes

$$L\bar{x} \in K \quad \textit{and} \quad \big(\exists\, \bar{v} \in K^{\ominus}\big) \; \begin{cases} \nabla f(\bar{x}) = -L^*\bar{v}, \\ \langle \bar{x} \mid L^*\bar{v} \rangle = 0. \end{cases} \tag{27.32}$$

Proof. Set $g = \iota_K$. Then Problem (27.29) is a special case of (27.2). Using Proposition 6.19(iii) for (a), we note that conditions (a)–(c) imply their counterparts in Theorem 27.2. In turn, we derive from Example 16.13 and Example 6.40 that the characterization (27.30) follows from the equivalence (i)⇔(iii) in Theorem 27.2, from which we deduce the characterization (27.32) via Proposition 17.31(i). Finally, it follows from Proposition 27.1 that $-L^*\bar{v} \in \partial f(\bar{x}) \Rightarrow 0 \in \partial(f + \langle \cdot \mid L^*\bar{v}\rangle)(\bar{x}) \Rightarrow \bar{x}$ solves (27.31). Hence, if $\bar{x}$ solves (27.29), we derive from (27.30), Proposition 19.25(v), and Remark 19.26 that $\bar{v}$ is a Lagrange multiplier associated with $\bar{x}$. ☐

Example 27.18 Let M and N be strictly positive integers, let $A \in \mathbb{R}^{M\times N}$, let $f \in \Gamma_0(\mathbb{R}^N)$, and let $\bar{x} \in \mathbb{R}^N$. Suppose that $\mathbb{R}^M_+ \cap \operatorname{ri} A(\operatorname{dom} f) \neq \varnothing$ and consider the problem

$$\underset{Ax\in\mathbb{R}^M_+}{\text{minimize}} \ \ f(x). \tag{27.33}$$

Then $\bar{x}$ is a solution to (27.33) if and only if

$$A\bar{x} \in \mathbb{R}^M_+ \quad \text{and} \quad \big(\exists\, \bar{v} \in \mathbb{R}^M_-\big) \ \begin{cases} -A^{\mathsf{T}}\bar{v} \in \partial f(\bar{x}), \\ \big\langle \bar{x} \mid A^{\mathsf{T}}\bar{v}\big\rangle = 0, \end{cases} \tag{27.34}$$

in which case $\bar{v}$ is a Lagrange multiplier associated with $\bar{x}$, and $\bar{x}$ solves the problem

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \ \ f(x) + \big\langle x \mid A^{\mathsf{T}}\bar{v}\big\rangle\,. \tag{27.35}$$

Proof. Suppose that $\mathcal{H} = \mathbb{R}^N$, $K = \mathbb{R}^M_+$, and $L = A$. Then K is a closed convex polyhedral cone and condition (b) in Proposition 27.17 is satisfied. Hence, the result follows from Proposition 27.17. ☐

Corollary 27.19 *Let K be a closed convex cone in $\mathcal{H}$, let $f \in \Gamma_0(\mathcal{H})$, and let $\bar{x} \in \mathcal{H}$. Suppose that one of the following holds:*

(a) $K - \operatorname{cone}(\operatorname{dom} f)$ *is a closed linear subspace.*
(b) $\mathcal{H}$ *is finite-dimensional, K is polyhedral, and $K \cap \operatorname{ri} \operatorname{dom} f \neq \varnothing$.*
(c) $\mathcal{H}$ *is finite-dimensional, f is a polyhedral function, K is polyhedral, and $K \cap \operatorname{dom} f \neq \varnothing$.*

Consider the problem

$$\underset{x\in K}{\text{minimize}} \ \ f(x). \tag{27.36}$$

Then $\bar{x}$ is a solution to (27.36) if and only if

$$\bar{x} \in K \quad \textit{and} \quad \big(\exists\, u \in K^{\oplus} \cap \partial f(\bar{x})\big) \ \ \langle \bar{x} \mid u\rangle = 0. \tag{27.37}$$

Proof. Apply Proposition 27.17 to $\mathcal{K} = \mathcal{H}$ and $L = \mathrm{Id}$. ☐

27.5 Convex Inequality Constraints

In this section, we turn our attention to the minimization of a convex function subject to convex inequality constraints. The following result will be required.

Lemma 27.20 *Let* $g\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ *be a convex function such that* $\operatorname{lev}_{<0} g \neq \varnothing$*, set* $C = \operatorname{lev}_{\leqslant 0} g$*, and let* $x \in C$*. Then*

$$N_C x = \begin{cases} N_{\operatorname{dom} g}\, x \cup \operatorname{cone} \partial g(x), & \text{if } g(x) = 0;\\ N_{\operatorname{dom} g}\, x, & \text{if } g(x) < 0. \end{cases} \tag{27.38}$$

Proof. Clearly, $C \subset \operatorname{dom} g$. Hence,

$$\begin{aligned}(\forall u \in \mathcal{H}) \quad u \in N_{\operatorname{dom} g}\, x &\Leftrightarrow \sup \langle \operatorname{dom} g - x \mid u\rangle \leqslant 0\\ &\Rightarrow \sup \langle C - x \mid u\rangle \leqslant 0\\ &\Leftrightarrow u \in N_C x. \end{aligned} \tag{27.39}$$

Thus,

$$N_{\operatorname{dom} g}\, x \subset N_C x. \tag{27.40}$$

We consider two cases.

(a) $g(x) < 0$: Take $u \in N_C x$ and fix $y \in \operatorname{dom} g$. Then $\sup \langle C - x \mid u\rangle \leqslant 0$. Moreover, for $\alpha \in \left]0,1\right[$ sufficiently small, $g((1-\alpha)x + \alpha y) \leqslant (1-\alpha)g(x) + \alpha g(y) \leqslant 0$ and therefore $\alpha \langle y - x \mid u\rangle = \langle ((1-\alpha)x + \alpha y) - x \mid u\rangle \leqslant 0$. Thus $\langle y - x \mid u\rangle \leqslant 0$ and it follows that $u \in N_{\operatorname{dom} g}\, x$. In view of (27.40), we obtain $N_C x = N_{\operatorname{dom} g}\, x$.

(b) $g(x) = 0$: Set

$$K = \operatorname{cone}\big(\operatorname{lev}_{<0} g - x\big). \tag{27.41}$$

Then K is a nonempty convex cone and, using Proposition 6.24(iii) and Corollary 9.11, we deduce that

$$N_C x = (C - x)^{\ominus} = (\overline{C} - x)^{\ominus} = (\overline{\operatorname{lev}_{<0} g} - x)^{\ominus} = (\operatorname{lev}_{<0} g - x)^{\ominus} = K^{\ominus}. \tag{27.42}$$

We claim that

$$K = \big\{ y \in \mathcal{H} \;\big|\; g'(x;y) < 0 \big\}. \tag{27.43}$$

Let $y \in \mathcal{H}$. If $y \in K$, then $y = \alpha(z - x)$, for some $z \in \operatorname{lev}_{<0} g$ and some $\alpha \in \mathbb{R}_{++}$. Hence, $g'(x;y) = g'(x;\alpha(z-x)) = \alpha g'(x;z-x) \leqslant \alpha(g(z) - g(x)) = \alpha g(z) < 0$ by Proposition 17.2(ii)&(iv). Conversely, if $g'(x;y) < 0$ then, for some $\alpha \in \left]0,1\right[$ sufficiently small, we have $(g(x + \alpha y) - g(x))/\alpha = g(x + \alpha y)/\alpha < 0$ and hence $y = ((x + \alpha y) - x)/\alpha \in K$. This verifies (27.43). We now set

$$h = g'(x;\cdot). \tag{27.44}$$

Then h is sublinear and $h(0) = 0$ by Proposition 17.2(iv). Moreover, $h^* = \iota_{\partial g(x)}$ by Proposition 17.17. We consider two cases.

(b.1) $x \in \operatorname{dom} \partial g$: Then $\operatorname{dom} h^* = \partial g(x) \neq \varnothing$ and hence Proposition 13.45 implies that $h^{**} = \breve{h}$. Corollary 9.11 thus yields

$$\overline{K} = \overline{\operatorname{lev}_{<0} h} = \operatorname{lev}_{\leqslant 0} \breve{h} = \operatorname{lev}_{\leqslant 0} h^{**} = \operatorname{lev}_{\leqslant 0} \iota^*_{\partial g(x)} = \operatorname{lev}_{\leqslant 0} \sigma_{\partial g(x)} = \big(\partial g(x)\big)^{\ominus}. \tag{27.45}$$

In view of (27.42), Proposition 6.24(iii), and Proposition 6.33, we deduce that

$$N_C x = K^{\ominus} = \overline{K}^{\ominus} = \big(\partial g(x)\big)^{\ominus\ominus} = \overline{\operatorname{cone}}\, \partial g(x). \tag{27.46}$$

On the other hand, $\partial g(x)$ is nonempty by assumption, and closed and convex by Proposition 16.4(iii). Moreover, since $\operatorname{lev}_{<0} g \neq \varnothing$ and $g(x) = 0$ by assumption, $x \notin \operatorname{Argmin} g$ and Theorem 16.3 yields $0 \notin \partial g(x)$. Consequently, Corollary 6.53 implies that $\overline{\operatorname{cone}}\, \partial g(x) = (\operatorname{cone} \partial g(x)) \cup (\operatorname{rec} \partial g(x))$. However, using successively Proposition 16.5, Theorem 21.2, Proposition 21.17, Corollary 16.39, and Proposition 13.46(i), we obtain $\operatorname{rec} \partial g(x) = \operatorname{rec} \partial g^{**}(x) = N_{\overline{\operatorname{dom}}\, \partial g^{**}} x = N_{\overline{\operatorname{dom}}\, g^{**}} x = N_{\overline{\operatorname{dom}}\, g} x = N_{\operatorname{dom} g} x$. Altogether,

$$N_C x = \overline{\operatorname{cone}}\, \partial g(x) = \big(\operatorname{cone} \partial g(x)\big) \cup N_{\operatorname{dom} g} x, \tag{27.47}$$

as announced in (27.38).

(b.2) $x \notin \operatorname{dom} \partial g$: If $\breve{h}$ is proper, then so is $(\breve{h})^* = h^*$ by Proposition 13.16(iv) and Theorem 13.37. Hence, since $h^* = \iota_{\partial g(x)} = \iota_{\varnothing} \equiv +\infty$ is not proper, neither is $\breve{h}$. In view of Proposition 9.6, $\breve{h}$ therefore takes on only the values $-\infty$ and $+\infty$. Using Proposition 9.8(iv), we thus see that $\operatorname{lev}_{\leqslant 0} \breve{h} = \operatorname{dom} \breve{h} = \overline{\operatorname{dom}}\, h$. On the other hand, $\operatorname{dom} h = \operatorname{cone}(\operatorname{dom} g - x)$ by Proposition 17.2(v) and hence $\overline{\operatorname{dom}}\, h = \overline{\operatorname{cone}}\,(\operatorname{dom} g - x)$. It follows from (27.43), (27.44), and Corollary 9.11 that

$$\overline{K} = \overline{\operatorname{lev}_{<0} h} = \operatorname{lev}_{\leqslant 0} \breve{h} = \overline{\operatorname{dom}}\, h = \overline{\operatorname{cone}}\,(\operatorname{dom} g - x). \tag{27.48}$$

Therefore, using (27.42) and Proposition 6.24(iii), we conclude that

$$N_C x = K^{\ominus} = \big(\overline{\operatorname{cone}}\,(\operatorname{dom} g - x)\big)^{\ominus} = (\operatorname{dom} g - x)^{\ominus} = N_{\operatorname{dom} g} x. \tag{27.49}$$

The entire lemma is proven. ☐

The minimization of a convex function under convex inequality constraints was already examined in Corollary 19.30. We investigate it here in a new light, and provide in particular a sufficient condition for the existence of Lagrange multipliers.

Proposition 27.21 *Let $f \in \Gamma_0(\mathcal{H})$, let m be a strictly positive integer, set $I = \{1, \ldots, m\}$, suppose that $(g_i)_{i \in I}$ are functions in $\Gamma_0(\mathcal{H})$ such that the* Slater condition

$$\begin{cases} (\forall i \in I) \quad \operatorname{lev}_{\leqslant 0} g_i \subset \operatorname{int} \operatorname{dom} g_i, \\ \operatorname{dom} f \cap \bigcap_{i \in I} \operatorname{lev}_{<0} g_i \neq \varnothing, \end{cases} \tag{27.50}$$

is satisfied, and let $\bar{x} \in \mathcal{H}$. Consider the problem

$$\underset{\substack{x \in \mathcal{H} \\ g_1(x) \leqslant 0, \ldots, g_m(x) \leqslant 0}}{\operatorname{minimize}} \quad f(x). \tag{27.51}$$

Then $\bar{x}$ is a solution to (27.51) if and only if

$$\Big(\exists\, (\bar{\nu}_i)_{i \in I} \in \mathbb{R}_+^m\Big) \Big(\exists\, (u_i)_{i \in I} \in \bigtimes_{i \in I} \partial g_i(\bar{x})\Big)$$
$$-\sum_{i \in I} \bar{\nu}_i u_i \in \partial f(\bar{x}) \quad \text{and} \quad (\forall i \in I) \; \begin{cases} g_i(\bar{x}) \leqslant 0, \\ \bar{\nu}_i g_i(\bar{x}) = 0, \end{cases} \tag{27.52}$$

in which case $(\bar{\nu}_i)_{i \in I}$ are Lagrange multipliers associated with $\bar{x}$, and $\bar{x}$ solves the problem

$$\underset{x \in \mathcal{H}}{\operatorname{minimize}} \;\; f(x) + \sum_{i \in I} \bar{\nu}_i g_i(x). \tag{27.53}$$

Moreover, if the functions f and $(g_i)_{i \in I}$ are Gâteaux differentiable at $\bar{x}$, then (27.52) becomes

$$\big(\exists\, (\bar{\nu}_i)_{i \in I} \in \mathbb{R}_+^m\big) \quad \begin{cases} \nabla f(\bar{x}) = -\sum_{i \in I} \bar{\nu}_i \nabla g_i(\bar{x}), \\ (\forall i \in I) \quad \begin{cases} g_i(\bar{x}) \leqslant 0, \\ \bar{\nu}_i g_i(\bar{x}) = 0. \end{cases} \end{cases} \tag{27.54}$$

Proof. Set

$$C = \bigcap_{i \in I} C_i, \quad \text{where} \quad (\forall i \in I) \quad C_i = \operatorname{lev}_{\leqslant 0} g_i. \tag{27.55}$$

Corollary 8.39(ii) states that $(\forall i \in I)$ $\operatorname{int} \operatorname{dom} g_i = \operatorname{cont} g_i$. Hence, it follows from (27.50) and Corollary 8.47(i) that $\operatorname{int} C = \bigcap_{i \in I} \operatorname{int} C_i = \bigcap_{i \in I} \operatorname{lev}_{<0} g_i$ and that $0 \in \operatorname{int}(C - \operatorname{dom} f) \subset \operatorname{sri}(C - \operatorname{dom} f)$. Therefore, Proposition 27.8 asserts that

$$\bar{x} \in \operatorname{Argmin}_C f \quad \Leftrightarrow \quad (\exists\, u \in N_C \bar{x}) \quad -u \in \partial f(\bar{x}). \tag{27.56}$$

Now suppose that $\bar{x} \in C$. To determine $N_C \bar{x}$, set

$$I_+ = \big\{ i \in I \;\big|\; g_i(\bar{x}) = 0 \big\} \quad \text{and} \quad I_- = \big\{ i \in I \;\big|\; g_i(\bar{x}) < 0 \big\}. \tag{27.57}$$

As seen above,

$$\bigcap_{i\in I} \operatorname{int}\operatorname{dom}\iota_{C_i} = \bigcap_{i\in I} \operatorname{int} C_i = \bigcap_{i\in I} \operatorname{lev}_{<0} g_i \neq \varnothing. \tag{27.58}$$

On the other hand, (27.50) yields $\bar{x} \in \bigcap_{i\in I} \operatorname{int}\operatorname{dom} g_i$, which implies, by Proposition 6.44(ii), that

$$(\forall i \in I) \quad N_{\operatorname{dom} g_i} \bar{x} = \{0\}. \tag{27.59}$$

Hence, Example 16.13, (27.55), (27.58), Corollary 16.50(iv), and Lemma 27.20 yield

$$\begin{aligned} N_C\bar{x} &= \partial\iota_{\bigcap_{i\in I} C_i}(\bar{x}) \\ &= \partial\Big(\sum_{i\in I}\iota_{C_i}\Big)(\bar{x}) \\ &= \sum_{i\in I}\partial\iota_{C_i}(\bar{x}) \\ &= \sum_{i\in I_+} N_{C_i}\bar{x} + \sum_{i\in I_-} N_{C_i}\bar{x} \\ &= \sum_{i\in I_+}\bigcup_{\nu_i\in\mathbb{R}_{++}} \nu_i\partial g_i(\bar{x}). \end{aligned} \tag{27.60}$$

Thus, $N_C\bar{x}$ consists of all vectors of the form $\sum_{i\in I_+} \nu_i u_i$, where $(\forall i \in I_+)$ $\nu_i \in \mathbb{R}_{++}$ and $u_i \in \partial g_i(x)$. Therefore, (27.56) implies (27.52) and, in turn, (27.54) via Proposition 17.31(i). This provides the announced characterizations.

Next, suppose that $\bar{x}$ solves (27.51). Then we derive from (27.52), (27.50), (27.58), and Corollary 16.50(iv) that there exist $(\bar{\nu}_i)_{i\in I} \in \mathbb{R}^m_+$ such that

$$0 \in \partial f(\bar{x}) + \sum_{i\in I}\bar{\nu}_i\partial g_i(\bar{x}) = \partial\Big(f + \sum_{i\in I}\bar{\nu}_i g_i\Big)(\bar{x}). \tag{27.61}$$

In view of Theorem 16.3, this shows that $\bar{x}$ solves (27.53). In turn, we derive from (27.52), from Proposition 19.25(v) applied to $\mathcal{K} = \mathbb{R}^m$, $K = \mathbb{R}^m_-$, and $R\colon x \mapsto (g_i(x))_{i\in I}$, and from Remark 19.26 that $(\bar{\nu}_i)_{i\in I}$ are Lagrange multipliers associated with $\bar{x}$. ☐

Remark 27.22 The conditions in (27.52) are often referred to as the *Karush–Kuhn–Tucker conditions* associated with (27.51).

27.6 Regularization of Minimization Problems

Let $f \in \Gamma_0(\mathcal{H})$ and suppose that $\operatorname{Argmin} f \neq \varnothing$, i.e., the minimization problem

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x) \tag{27.62}$$

has at least one solution. In order to obtain a specific minimizer, one can introduce, for every $\varepsilon \in \left]0,1\right[$, the regularized problem

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x) + \varepsilon g(x), \tag{27.63}$$

where $g \in \Gamma_0(\mathcal{H})$. The objective is to choose the regularization function g such that (27.63) admits a unique solution x_ε and such that the net $(x_\varepsilon)_{\varepsilon \in]0,1[}$ converges to a specific point in $\operatorname{Argmin} f$. For instance, when $g = (1/2)\|\cdot\|^2$, we obtain the classical *Tykhonov regularization* framework. In this case, it follows from Proposition 27.8 that (27.63) is equivalent to

$$\text{find} \;\; x_\varepsilon \in \mathcal{H} \quad \text{such that} \quad 0 \in \partial f(x_\varepsilon) + \varepsilon x_\varepsilon, \tag{27.64}$$

which is a special case of (23.40). In turn, if we denote by x_0 the minimum norm minimizer of f, Theorem 23.44(i) asserts that $x_\varepsilon \to x_0$ as $\varepsilon \downarrow 0$. The next theorem explores the asymptotic behavior of the curve $(x_\varepsilon)_{\varepsilon \in]0,1[}$ for more general choices of the regularization function g.

Theorem 27.23 *Let f and g be in $\Gamma_0(\mathcal{H})$. Suppose that* $\operatorname{Argmin} f \cap \operatorname{dom} g \neq \varnothing$ *and that g is coercive and strictly convex. Then g admits a unique minimizer x_0 over* $\operatorname{Argmin} f$ *and, for every $\varepsilon \in \left]0,1\right[$, the regularized problem*

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x) + \varepsilon g(x) \tag{27.65}$$

admits a unique solution x_ε. Moreover, the following hold:

(i) $x_\varepsilon \rightharpoonup x_0$ *as* $\varepsilon \downarrow 0$.
(ii) $g(x_\varepsilon) \to g(x_0)$ *as* $\varepsilon \downarrow 0$.
(iii) *Suppose that g is uniformly convex on every closed ball in $\mathcal{H}$. Then* $x_\varepsilon \to x_0$ *as* $\varepsilon \downarrow 0$.

Proof. Set $S = \operatorname{Argmin} f$. Let $\varepsilon \in \left]0,1\right[$ and set $h_\varepsilon = f + \varepsilon g$. Then $h_\varepsilon \in \Gamma_0(\mathcal{H})$ and h_ε is strictly convex. Moreover, since $\inf f(\mathcal{H}) > -\infty$, existence and uniqueness of x_0 and x_ε follow from Corollary 11.16(ii) and Corollary 11.9. Now fix $z \in S \cap \operatorname{dom} g$, let $(\varepsilon_n)_{n \in \mathbb{N}}$ be an arbitrary sequence in $\left]0,1\right[$ such that $\varepsilon_n \downarrow 0$, and set

$$(\forall n \in \mathbb{N}) \;\; y_n = x_{\varepsilon_n}. \tag{27.66}$$

(i): We have

$$(\forall n \in \mathbb{N})\ \ f(y_n) + \varepsilon_n g(y_n) = \inf\big(f + \varepsilon_n g\big)(\mathcal{H}) \qquad$$
$$\leqslant f(z) + \varepsilon_n g(z) \tag{27.67}$$
$$\leqslant f(y_n) + \varepsilon_n g(z). \tag{27.68}$$

Therefore

$$(\forall n \in \mathbb{N})\ \ g(y_n) \leqslant g(z) < +\infty. \tag{27.69}$$

Accordingly, $(y_n)_{n\in\mathbb{N}}$ lies in $\operatorname{lev}_{\leqslant g(z)} g$ and it follows from Proposition 11.12 that $(y_n)_{n\in\mathbb{N}}$ is bounded. Furthermore, $\inf g(\mathcal{H}) > -\infty$. Thus, (27.67) implies that

$$(\forall n \in \mathbb{N}) \quad f(y_n) \leqslant \inf f(\mathcal{H}) + \varepsilon_n\big(g(z) - g(y_n)\big)$$
$$\leqslant \inf f(\mathcal{H}) + \varepsilon_n\big(g(z) - \inf g(\mathcal{H})\big). \tag{27.70}$$

Thus, since $\varepsilon_n \downarrow 0$, we get

$$\overline{\lim}\, f(y_n) \leqslant \inf f(\mathcal{H}). \tag{27.71}$$

Now, let x be a weak sequential cluster point of $(y_n)_{n\in\mathbb{N}}$, say $y_{k_n} \rightharpoonup x$. Since, by Theorem 9.1, f is weakly lower semicontinuous, (27.71) yields

$$\inf f(\mathcal{H}) \leqslant f(x) \leqslant \underline{\lim}\, f(y_{k_n}) \leqslant \overline{\lim}\, f(y_{k_n}) \leqslant \inf f(\mathcal{H}). \tag{27.72}$$

Consequently, $f(x) = \inf f(\mathcal{H})$, i.e., $x \in S$. Furthermore, it follows from (27.69) and the weak lower semicontinuity of g that $g(x) \leqslant \underline{\lim}\, g(y_{k_n}) \leqslant \inf g(S)$. In turn, since $x \in S$, we obtain $g(x) = \inf g(S)$, i.e., $x = x_0$. Altogether, $(y_n)_{n\in\mathbb{N}}$ is bounded and has x_0 as its unique weak sequential cluster point. We therefore deduce from Lemma 2.46 that $y_n \rightharpoonup x_0$. Finally, since $(\varepsilon_n)_{n\in\mathbb{N}}$ was chosen arbitrarily in $]0,1[$, we conclude that $x_\varepsilon \rightharpoonup x_0$ as $\varepsilon \downarrow 0$.

(ii): Since g is weakly lower semicontinuous, (i) implies that $g(x_0) \leqslant \underline{\lim}\, g(y_n) \leqslant \overline{\lim}\, g(y_n)$. However, (27.69) yields $\overline{\lim}\, g(y_n) \leqslant g(x_0)$. Hence $g(y_n) \to g(x_0)$ and, in turn, $g(x_\varepsilon) \to g(x_0)$.

(iii): Since $(y_n)_{n\in\mathbb{N}}$ is bounded, there exists an increasing function $\phi\colon \mathbb{R}_+ \to \mathbb{R}_+$ that vanishes only at 0 such that

$$(\forall n \in \mathbb{N})\ \ \frac{1}{4}\phi\big(\|y_n - x_0\|\big) \leqslant \frac{g(y_n) + g(x_0)}{2} - g\Big(\frac{y_n + x_0}{2}\Big). \tag{27.73}$$

Consequently, it follows from (ii) that

$$\frac{1}{4}\,\overline{\lim}\,\phi\big(\|y_n - x_0\|\big) \leqslant \overline{\lim}\ \frac{g(y_n) + g(x_0)}{2} - \underline{\lim}\, g\Big(\frac{y_n + x_0}{2}\Big)$$
$$= g(x_0) - \underline{\lim}\, g\Big(\frac{y_n + x_0}{2}\Big). \tag{27.74}$$

However, since $(y_n+x_0)/2 \rightharpoonup x_0$ by (ii), $g(x_0) \leqslant \underline{\lim}\, g((y_n+x_0)/2)$. Altogether $\overline{\lim}\, \phi(\|y_n - x_0\|) \leqslant 0$; hence $y_n \to x_0$. We conclude that $x_\varepsilon \to x_0$ as $\varepsilon \downarrow 0$. $\square$

Example 27.24 Let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$ be such that $\operatorname{ran} L$ is closed. Then the generalized inverse of L (see Definition 3.28) satisfies

$$(\forall y \in \mathcal{K}) \quad L^\dagger y = \lim_{\varepsilon\downarrow 0} \big(L^*L + \varepsilon \mathrm{Id}\big)^{-1} L^* y. \tag{27.75}$$

Proof. Let $y \in \mathcal{K}$ and let $\varepsilon \in \mathbb{R}_{++}$. The solution to

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; \frac{1}{2}\|Lx - y\|^2 + \frac{\varepsilon}{2}\|x\|^2 \tag{27.76}$$

is $x_\varepsilon = (L^*L+\varepsilon\mathrm{Id})^{-1}L^*y$. However, by applying Theorem 27.23(iii) to $f\colon x \mapsto (1/2)\|Lx - y\|^2$ and the strongly convex function $g = (1/2)\|\cdot\|^2$, we deduce that $x_\varepsilon \to x_0$ as $\varepsilon \downarrow 0$, where x_0 is the minimum norm minimizer of $x \mapsto \|Lx - y\|^2/2$. On the other hand, it follows from Definition 3.28 and Proposition 3.27 that $x_0 = L^\dagger y$. Altogether, $(L^*L + \varepsilon\mathrm{Id})^{-1}L^*y \to L^\dagger y$ as $\varepsilon \downarrow 0$. $\square$

Exercises

Exercise 27.1 Let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be proper and convex. Show that $\operatorname{Argmin} f \subset \partial f^*(0)$ and that $\operatorname{Argmin} f \neq \partial f^*(0)$ may occur. Compare to Proposition 27.1.

Exercise 27.2 Provide two functions f and g in $\Gamma_0(\mathcal{H})$ such that $\operatorname{dom} f \cap \operatorname{dom} g \neq \varnothing$ and for which the conclusion of Corollary 27.3 fails.

Exercise 27.3 Let $F\colon \mathcal{H}\times\mathcal{H} \to \mathbb{R}$ be a symmetric bilinear form that satisfies (27.11) for some α and β in $\mathbb{R}_{++}$. Derive the conclusions of the Stampacchia theorem (Example 27.11) from Theorem 3.16.

Exercise 27.4 Let C and D be closed convex subsets of $\mathcal{H}$ such that $0 \in \operatorname{sri}(C - D)$. Show that

$$C \cap D = P_D\big(\operatorname{Fix}(2P_C - \mathrm{Id}) \circ (2P_D - \mathrm{Id})\big). \tag{27.77}$$

Exercise 27.5 Let F be as in Example 27.11. Show that there exists a unique operator $L \in \mathcal{B}(\mathcal{H})$ such that $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ F(x,y) = \langle y \mid Lx\rangle$. Moreover, show that $\|L\| \leqslant \beta$, that L is invertible, and that $\|L^{-1}\| \leqslant \alpha^{-1}$.

Exercise 27.6 Let C and D be closed convex subsets of $\mathcal{H}$ such that $C \cap D \neq \varnothing$. Use Proposition 27.8 to obtain the identity

$$C \cap D = \operatorname{Fix}(P_C \circ P_D). \tag{27.78}$$

Exercise 27.7 Suppose that $u\in\mathcal{H}\smallsetminus\{0\}$, let $\eta\in\mathbb{R}$, set $C=\big\{x\in\mathcal{H}\;\big|\;\langle x\mid u\rangle=\eta\big\}$, and let $g\colon\mathcal{H}\to\mathbb{R}$ be lower semicontinuous, strictly convex, and Gâteaux differentiable on $\mathcal{H}$. Suppose that $\operatorname{dom}\partial g^*=\operatorname{int}\operatorname{dom}g^*$ and that $\operatorname{Argmin}_C g\neq\varnothing$. Show that there exists a unique $\nu\in\mathbb{R}$ such that $\langle\nabla g^*(\nu u)\mid u\rangle=\eta$ and that the unique element in $\operatorname{Argmin}_C g$ is $\nabla g^*(\nu u)$.

Exercise 27.8 Let $f\in\Gamma_0(\mathcal{H})$ and let $L\in\mathcal{B}(\mathcal{H},\mathcal{K})$ be such that $\operatorname{ran}L$ is closed. Suppose that $0\in\operatorname{sri}(L(\operatorname{dom}f))$. Consider the problem

$$\underset{x\in\ker L}{\text{minimize}}\;\; f(x),\tag{27.79}$$

and let $\bar{x}\in\mathcal{H}$. Derive from Proposition 27.14 that $\bar{x}$ is a solution to (27.79) if and only if

$$L\bar{x}=0\quad\text{and}\quad(\exists\,\bar{v}\in\mathcal{K})\quad -L^*\bar{v}\in\partial f(\bar{x}).\tag{27.80}$$

Exercise 27.9 Suppose that $u\in\mathcal{H}\smallsetminus\{0\}$ and let $\eta\in\mathbb{R}$. Set

$$C=\big\{x\in\mathcal{H}\;\big|\;\langle x\mid u\rangle=\eta\big\}\tag{27.81}$$

and let $z\in\mathcal{H}$. Use Corollary 27.15 to show that

$$P_Cz=z+\frac{\eta-\langle z\mid u\rangle}{\|u\|^2}u.\tag{27.82}$$

Exercise 27.10 Suppose that $\mathcal{H}=\mathbb{R}^N$ and let $f\colon\mathcal{H}\to\mathbb{R}$ be convex and Gâteaux differentiable on $\mathcal{H}$. Set $K=\mathbb{R}^N_+$ and let $x\in\mathcal{H}$. Show that $x\in\operatorname{Argmin}_K f$ if and only if $x\in K$, $\nabla f(x)\in K$, and $x\perp\nabla f(x)$.

Exercise 27.11 Provide an example of a proper and convex function $g\colon\mathcal{H}\to\left]-\infty,+\infty\right]$ such that $\operatorname{lev}_{<0}g=\varnothing$ and $N_Cx\neq N_{\operatorname{dom}g}x\cup\operatorname{cone}\partial g(x)$, where $C=\operatorname{lev}_{\leqslant 0}g$ and $x\in C$. Compare to Lemma 27.20.

Exercise 27.12 Let $(y_n)_{n\in\mathbb{N}}$ be the sequence defined in the proof of Theorem 27.23. Show that $(y_n)_{n\in\mathbb{N}}$ is a minimizing sequence of f.

Chapter 28
Proximal Minimization

We saw in Chapter 27 that the solutions to minimization problems can be characterized by fixed point equations involving proximity operators. Since proximity operators are firmly nonexpansive, they can be used to devise efficient operator splitting algorithms to solve minimization problems. Such algorithms, called proximal algorithms, are investigated in this chapter.

Throughout this chapter, $\mathcal{K}$ is a real Hilbert space.

28.1 The Proximal-Point Algorithm

As seen in Proposition 27.1, minimizing a function $f \in \Gamma_0(\mathcal{H})$ amounts to finding a zero of its subdifferential operator ∂f, which is a maximally monotone operator with resolvent $J_{\partial f} = \operatorname{Prox}_f$. Thus, a minimizer of f can be obtained via the proximal-point algorithm (23.36). In that vein, our first result is a refinement of Theorem 23.41 that features a more relaxed condition on the parameter sequence $(\gamma_n)_{n\in\mathbb{N}}$.

Theorem 28.1 (Proximal-point algorithm) *Let $f \in \Gamma_0(\mathcal{H})$ be such that* $\operatorname{Argmin} f \neq \varnothing$*, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}_{++}$ such that $\sum_{n\in\mathbb{N}} \gamma_n = +\infty$, and let $x_0 \in \mathcal{H}$. Set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \operatorname{Prox}_{\gamma_n f} x_n. \tag{28.1}$$

Then the following hold:

(i) $(x_n)_{n\in\mathbb{N}}$ *is a minimizing sequence of f; more precisely, $f(x_n) \downarrow \min f(\mathcal{H})$.*
(ii) $(\forall n \in \mathbb{N})$ $x_{n+1} = P_{C_{n+1}} x_n$*, where $C_{n+1} = \operatorname{lev}_{\leqslant f(x_{n+1})} f$.*

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_28

(iii) $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* $\operatorname{Argmin} f$.
(iv) *Suppose that* f *is uniformly convex on every nonempty bounded subset of* $\operatorname{dom}\partial f$. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges strongly to the unique point in* $\operatorname{Argmin} f$.

Proof. It follows from (28.1) and (16.37) that

$$(\forall n\in\mathbb{N})\quad x_n - x_{n+1} \in \gamma_n \partial f(x_{n+1}). \tag{28.2}$$

Hence, (16.1) yields

$$(\forall z\in\mathcal{H})(\forall n\in\mathbb{N})\quad \langle z - x_{n+1} \mid x_n - x_{n+1}\rangle \leqslant \gamma_n\big(f(z) - f(x_{n+1})\big). \tag{28.3}$$

(i): We derive from (28.3) that

$$(\forall n\in\mathbb{N})\quad 0 \leqslant \langle x_n - x_{n+1} \mid x_n - x_{n+1}\rangle/\gamma_n \leqslant f(x_n) - f(x_{n+1}). \tag{28.4}$$

Now let $z\in\operatorname{Argmin} f$. Then, for every $n\in\mathbb{N}$, (28.3) yields

$$\begin{aligned}\|x_{n+1}-z\|^2 &= \|x_n - z\|^2 + 2\langle x_n - z \mid x_{n+1}-x_n\rangle + \|x_{n+1}-x_n\|^2\\ &= \|x_n - z\|^2 - \|x_{n+1}-x_n\|^2 + 2\langle x_{n+1} - z \mid x_{n+1}-x_n\rangle\\ &\leqslant \|x_n - z\|^2 - 2\gamma_n\big(f(x_{n+1}) - \min f(\mathcal{H})\big).\end{aligned} \tag{28.5}$$

This shows that $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to $\operatorname{Argmin} f$ and that $\sum_{n\in\mathbb{N}}\gamma_n(f(x_{n+1}) - \min f(\mathcal{H})) < +\infty$. Hence, since $\sum_{n\in\mathbb{N}}\gamma_n = +\infty$, we have $\underline{\lim}\, f(x_n) = \min f(\mathcal{H})$ and it follows from (28.4) that $f(x_n)\downarrow \min f(\mathcal{H})$.

(ii): Let $n\in\mathbb{N}$. Then $x_{n+1}\in C_{n+1}$ and, furthermore, we derive from (28.3) that

$$(\forall z\in C_{n+1})\quad \langle z - x_{n+1} \mid x_n - x_{n+1}\rangle \leqslant 0. \tag{28.6}$$

In view of (3.10) the proof is complete.

(iii): Let x be a weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$. It follows from (i) and Proposition 11.21 that $x\in\operatorname{Argmin} f$. Now apply Theorem 5.5.

(iv): It follows from (iii) and Theorem 16.3 that there exists $x \in \operatorname{Argmin} f \subset \operatorname{dom}\partial f$ such that $x_n \rightharpoonup x$, and from (28.5) and (28.2) that $(x_{n+1})_{n\in\mathbb{N}}$ is a bounded sequence in $\operatorname{dom}\partial f$. Hence, $\{x\}\cup\{x_{n+1}\}_{n\in\mathbb{N}}$ is a bounded subset of $\operatorname{dom}\partial f$ and we derive from (10.3) that there exists an increasing function $\phi\colon \mathbb{R}_+\to[0,+\infty]$ vanishing only at 0 such that

$$(\forall n\in\mathbb{N}\smallsetminus\{0\})\quad \frac{1}{4}\phi\big(\|x_n - x\|\big) \leqslant \frac{f(x_n)+f(x)}{2} - f\bigg(\frac{x_n + x}{2}\bigg). \tag{28.7}$$

By (i), we have $f(x_n)\downarrow f(x)$. In addition, since, by Proposition 10.25, f is weakly sequentially lower semicontinuous and since $(x_n+x)/2\rightharpoonup x$, we have $f(x)\leqslant \underline{\lim}\, f((x_n+x)/2) \leqslant \overline{\lim}\, f((x_n+x)/2) \leqslant \overline{\lim}(f(x_n)+f(x))/2 = f(x)$. Altogether, $\phi(\|x_n - x\|)\to 0$ and we conclude that $x_n\to x$. ☐

28.2 Spingarn's Method of Partial Inverses

We describe an application of the method of partial inverses of Section 26.2 to linearly constrained minimization problems.

Proposition 28.2 *Let $f \in \Gamma_0(\mathcal{H})$ and let V be a closed linear subspace of $\mathcal{H}$. Suppose that $0 \in \operatorname{sri}(V - \operatorname{dom} f)$ and that the problem*

$$\underset{x \in V}{\text{minimize}} \;\; f(x) \tag{28.8}$$

has at least one solution. Let $x_0 \in V$, let $u_0 \in V^\perp$, and set

$$\begin{array}{l} \text{for } n = 0,1,\ldots \\ \left\lfloor \begin{array}{l} y_n = \operatorname{Prox}_f(x_n + u_n), \\ v_n = x_n + u_n - y_n, \\ (x_{n+1}, u_{n+1}) = (P_V y_n, P_{V^\perp} v_n). \end{array} \right. \end{array} \tag{28.9}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a solution to (28.8).

Proof. In view of Proposition 27.16, this is an application of Theorem 26.9 with $A = \partial f$. ☐

28.3 Douglas–Rachford Splitting Algorithm

In this section we apply the results of Theorem 26.11 on the Douglas–Rachford splitting algorithm to the minimization of the sum of two functions in $\Gamma_0(\mathcal{H})$.

Corollary 28.3 (Douglas–Rachford algorithm) *Let f and g be functions in $\Gamma_0(\mathcal{H})$ such that (see Corollary 27.6 for sufficient conditions)*

$$\operatorname{zer}\big(\partial f + \partial g\big) \neq \varnothing, \tag{28.10}$$

let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0,2]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$, and let $\gamma \in \mathbb{R}_{++}$. Let $\mathscr{P}$ be the set of solutions to the problem

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x) + g(x), \tag{28.11}$$

and let $\mathscr{D}$ be the set of solutions to the dual problem

$$\underset{u \in \mathcal{H}}{\text{minimize}} \;\; f^*(-u) + g^*(u). \tag{28.12}$$

Let $y_0 \in \mathcal{H}$ and set

$$\begin{array}{l} \text{for } n = 0,1,\ldots \\ \left\lfloor \begin{array}{l} x_n = \operatorname{Prox}_{\gamma g} y_n, \\ u_n = \gamma^{-1}(y_n - x_n), \\ z_n = \operatorname{Prox}_{\gamma f}(2x_n - y_n), \\ y_{n+1} = y_n + \lambda_n(z_n - x_n). \end{array} \right. \end{array} \tag{28.13}$$

Then there exists $y \in \mathcal{H}$ such that $y_n \rightharpoonup y$. Now set $x = \operatorname{Prox}_{\gamma g} y$ and $u = \gamma^{-1}(y - x)$. Then the following hold:

(i) $x \in \mathscr{P}$ *and* $u \in \mathscr{D}$.
(ii) $x_n - z_n \to 0$.
(iii) $x_n \rightharpoonup x$ *and* $z_n \rightharpoonup x$.
(iv) $u_n \rightharpoonup u$.
(v) *Suppose that one of the following holds:*

(a) f *is uniformly convex on every nonempty bounded subset of* $\operatorname{dom} \partial f$.
(b) g *is uniformly convex on every nonempty bounded subset of* $\operatorname{dom} \partial g$.

Then $\mathscr{P} = \{x\}$, $x_n \to x$, *and* $z_n \to x$.

Proof. Set $A = \partial f$ and $B = \partial g$. Then, by Theorem 20.25, A and B are maximally monotone. In addition, we deduce from Proposition 16.6(ii) that $\operatorname{zer}(A+B) \subset \operatorname{zer} \partial(f+g) = \operatorname{Argmin}(f+g)$. The results therefore follow from Example 23.3, Example 22.5, and Theorem 26.11(i)–(iv)&(vi). ☐

Corollary 28.4 *Let $f \in \Gamma_0(\mathcal{H})$ and $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$. Suppose that $LL^* = \mu \mathrm{Id}$ for some $\mu \in \mathbb{R}_{++}$, that*

$$r \in \operatorname{sri} L(\operatorname{dom} f), \tag{28.14}$$

and that the problem

$$\underset{\substack{x \in \mathcal{H} \\ Lx = r}}{\text{minimize}} \;\; f(x) \tag{28.15}$$

has at least one solution. Let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in $[0, 2]$ such that $\sum_{n \in \mathbb{N}} \lambda_n(2 - \lambda_n) = +\infty$, and let $\gamma \in \mathbb{R}_{++}$. Let $y_0 \in \mathcal{H}$ and set

$$\begin{array}{l} \textit{for}\; n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} p_n = y_n + \mu^{-1} L^*(r - L y_n), \\ x_n = \operatorname{Prox}_{\gamma f} y_n, \\ q_n = x_n + \mu^{-1} L^*(r - L x_n), \\ y_{n+1} = y_n + \lambda_n \big(2 q_n - p_n - x_n\big). \end{array} \right. \end{array} \tag{28.16}$$

Then $(p_n)_{n \in \mathbb{N}}$ converges weakly to a solution to (28.15).

Proof. Set $C = \big\{ x \in \mathcal{H} \;\big|\; Lx = r \big\}$, $A = N_C$, and $B = \partial f$. It follows from Example 20.26 and Theorem 20.25 that A and B are maximally monotone. Moreover, since $\operatorname{ran} LL^* = \operatorname{ran} \mu \mathrm{Id} = \mathcal{K}$, we derive from Fact 2.26 that $\operatorname{ran} L^*$ is closed. In turn, Example 6.43 and Fact 2.25(v) yield

$$(\forall x \in \mathcal{H}) \quad N_C x = \begin{cases} (\ker L)^\perp = \operatorname{ran} L^*, & \text{if } Lx = r; \\ \varnothing, & \text{if } Lx \neq r. \end{cases} \tag{28.17}$$

Next, we note that the set of solutions to (28.15) is $\operatorname{Argmin}_C f$. However, it follows from (28.14), Proposition 27.14, and (28.17) that

$$\operatorname{Argmin}_C f = \big\{ x \in C \;\big|\; (\exists\, v \in \mathcal{K}) \;\; -L^* v \in \partial f(x) \big\} = \operatorname{zer}(A + B). \tag{28.18}$$

On the other hand, we derive from Example 23.4 and Proposition 24.14 that

$$J_{\gamma A} = P_C = \operatorname{Prox}_{\iota_{\{r\}}\circ L} = \mathrm{Id} - \mu^{-1}L^* \circ (\mathrm{Id} - P_{\{r\}}) \circ L = \mathrm{Id} - \mu^{-1}L^* \circ (L - r). \tag{28.19}$$

Therefore, since $J_{\gamma B} = \operatorname{Prox}_{\gamma f}$ by Example 23.3 and since $J_{\gamma A} = P_C$ is affine by Corollary 3.22(ii), we derive from (28.16) that

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} p_n = J_{\gamma A} y_n, \\ x_n = J_{\gamma B} y_n, \\ q_n = J_{\gamma A} x_n, \\ y_{n+1} = y_n + \lambda_n\big(J_{\gamma A}(2x_n - y_n) - x_n\big). \end{array} \right. \end{array} \tag{28.20}$$

Thus, we recover (26.29), and the result follows from Theorem 26.11(v). $\square$

Example 28.5 Let m be an integer such that $m \geqslant 2$, set $I = \{1, \ldots, m\}$, let $(f_i)_{i\in I}$ be functions in $\Gamma_0(\mathcal{H})$, let $r \in \mathcal{H}$, and suppose that the problem

$$\underset{\substack{x_1\in\mathcal{H},\ldots,x_m\in\mathcal{H}\\ \sum_{i\in I} x_i = r}}{\operatorname{minimize}} \ \sum_{i\in I} f_i(x_i) \tag{28.21}$$

has at least one solution and that

$$r \in \operatorname{sri}\left\{ \sum_{i\in I} x_i \;\middle|\; (\forall i \in I)\ \ x_i \in \operatorname{dom} f_i \right\}. \tag{28.22}$$

Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0, 2]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2 - \lambda_n) = +\infty$, let $\gamma \in \mathbb{R}_{++}$, and let $(y_{i,0})_{i\in I} \in \mathcal{H}^m$. Set

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} u_n = \dfrac{1}{m}\Big(r - \displaystyle\sum_{j\in I} y_{j,n}\Big), \\[2ex] (\forall i \in I) \quad \left\lfloor \begin{array}{l} p_{i,n} = y_{i,n} + u_n, \\ x_{i,n} = \operatorname{Prox}_{\gamma f_i} y_{i,n}, \end{array} \right. \\[2ex] v_n = \dfrac{1}{m}\Big(r - \displaystyle\sum_{j\in I} x_{j,n}\Big), \\[2ex] (\forall i \in I)\ \ y_{i,n+1} = y_{i,n} + \lambda_n\big(x_{i,n} - y_{i,n} + 2v_n - u_n\big). \end{array} \right. \end{array} \tag{28.23}$$

Then, for every $i \in I$, $(p_{i,n})_{n\in\mathbb{N}}$ converges weakly to a point $p_i \in \mathcal{H}$, and $(p_i)_{i\in I}$ is a solution to (28.21).

Proof. Set $\boldsymbol{\mathcal{H}} = \bigoplus_{i\in I} \mathcal{H}$, $\boldsymbol{f} = \bigoplus_{i\in I} f_i$, $\boldsymbol{\mathcal{K}} = \mathcal{H}$, and $\boldsymbol{L} \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{K}} \colon \boldsymbol{x} \mapsto \sum_{i\in I} x_i$, where $\boldsymbol{x} = (x_i)_{i\in I}$ denotes a generic element in $\boldsymbol{\mathcal{H}}$. Then (28.21) becomes

$$\underset{\substack{\boldsymbol{x}\in\boldsymbol{\mathcal{H}}\\ \boldsymbol{L}\boldsymbol{x}=r}}{\operatorname{minimize}} \ \boldsymbol{f}(\boldsymbol{x}). \tag{28.24}$$

Moreover, the assumptions imply that this problem has at least one solution and that $r \in \operatorname{sri} \boldsymbol{L}(\operatorname{dom} \boldsymbol{f})$. On the other hand, $\boldsymbol{L}^* \colon \boldsymbol{\mathcal{K}} \to \boldsymbol{\mathcal{H}} \colon x \mapsto (x, \ldots, x)$ and Proposition 24.11 yields

$$\operatorname{Prox}_{\gamma \boldsymbol{f}} \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} \colon \boldsymbol{x} \mapsto \big(\operatorname{Prox}_{\gamma f_i} x_i\big)_{i \in I}. \tag{28.25}$$

Hence upon setting $\mu = m$ and, for every $n \in \mathbb{N}$, $\boldsymbol{x}_n = (x_{i,n})_{i\in I}$, $\boldsymbol{p}_n = (p_{i,n})_{i\in I}$, $\boldsymbol{y}_n = (y_{i,n})_{i\in I}$, and $\boldsymbol{q}_n = (x_{i,n} + v_n)_{i\in I}$, we can rewrite (28.23) as

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} \boldsymbol{p}_n = \boldsymbol{y}_n + \mu^{-1} \boldsymbol{L}^*(r - \boldsymbol{L}\boldsymbol{y}_n), \\ \boldsymbol{x}_n = \operatorname{Prox}_{\gamma \boldsymbol{f}} \boldsymbol{y}_n, \\ \boldsymbol{q}_n = \boldsymbol{x}_n + \mu^{-1} \boldsymbol{L}^*(r - \boldsymbol{L}\boldsymbol{x}_n), \\ \boldsymbol{y}_{n+1} = \boldsymbol{y}_n + \lambda_n\big(2\boldsymbol{q}_n - \boldsymbol{p}_n - \boldsymbol{x}_n\big), \end{array} \right. \end{array} \tag{28.26}$$

which is precisely (28.16). Since Corollary 28.4 asserts that $(\boldsymbol{p}_n)_{n\in\mathbb{N}}$ converges weakly to a solution to (28.15), we obtain the announced result. ☐

We now turn to an algorithm for robust principal component analysis (PCA).

Example 28.6 (Robust PCA) Let $\mathcal{H} = \mathbb{R}^{M\times N}$ be the Hilbert matrix space of Example 2.4, let $I = \{1, \ldots, M\}$ and $J = \{1, \ldots, N\}$, and suppose that $m = \min\{M, N\} \geqslant 2$. A problem which arises in robust principal component analysis (PCA) is to recover a low rank matrix $R \in \mathcal{H}$ from the observation $A = R + S$, where $S \in \mathcal{H}$ is sparse. The problem of recovering the low rank and sparse components of A can be modeled by the optimization problem

$$\underset{\substack{R\in\mathcal{H},\, S\in\mathcal{H} \\ R+S=A}}{\operatorname{minimize}} \;\; \|R\|_{\mathsf{nuc}} + \|S\|_1, \tag{28.27}$$

where $\|S\|_1$ denotes the sum of the absolute values of the entries of S. This problem is a special case of (28.21) with $m = 2$, $f_1 = \|\cdot\|_{\mathsf{nuc}}$, $f_2 = \|\cdot\|_1$, and $r = A$. Note that (28.22) is satisfied since $\operatorname{dom} f_1 = \operatorname{dom} f_2 = \mathcal{H}$. Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0, 2]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2 - \lambda_n) = +\infty$, let $\gamma \in \mathbb{R}_{++}$, let $Y_{1,0} \in \mathcal{H}$, and let $Y_{2,0} \in \mathcal{H}$. We denote the singular value decomposition of a matrix $Y_{1,n} \in \mathcal{H}$ by (see Section 24.8) $U_n \operatorname{Diag}(\sigma_1(Y_{1,n}), \ldots, \sigma_m(Y_{1,n}))V_n^{\mathsf{T}}$, where $U_n \in \mathbb{U}^M$ and $V_n \in \mathbb{U}^N$, and we express a matrix $Y_{2,n} \in \mathcal{H}$ in terms of its components as $Y_{2,n} = [\eta_{i,j,n}]_{i\in I, j\in J}$. Using Example 24.69 and Example 24.22, we reduce (28.23) to

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} C_n = \frac{1}{2}\big(A - Y_{1,n} - Y_{2,n}\big), \\ R_n = Y_{1,n} + C_n, \\ S_n = Y_{2,n} + C_n, \\ X_{1,n} = U_n \operatorname{Diag}\big(\max\{\sigma_1(Y_{1,n}) - \gamma, 0\}, \ldots, \max\{\sigma_m(Y_{1,n}) - \gamma, 0\}\big) V_n^{\mathsf{T}}, \\ X_{2,n} = \big[\operatorname{sign}(\eta_{i,j,n}) \max\{|\eta_{i,j,n}| - \gamma, 0\}\big]_{i\in I, j\in J}, \\ Y_{1,n+1} = Y_{1,n} + \lambda_n\big(A - X_{2,n} - Y_{1,n} - C_n\big), \\ Y_{2,n+1} = Y_{2,n} + \lambda_n\big(A - X_{1,n} - Y_{2,n} - C_n\big). \end{array} \right. \end{array} \tag{28.28}$$

It follows from Example 28.5 that $(R_n, S_n)_{n\in\mathbb{N}}$ converges to a solution (R, S) to (28.27).

We now describe a parallel splitting algorithm for minimizing a finite sum of functions in $\Gamma_0(\mathcal{H})$.

Proposition 28.7 (Parallel splitting algorithm) *Let m be an integer such that $m \geqslant 2$, set $I = \{1, \ldots, m\}$, and let $(f_i)_{i\in I}$ be functions in $\Gamma_0(\mathcal{H})$. Suppose that the problem*

$$\underset{x\in\mathcal{H}}{\text{minimize}} \quad \sum_{i\in I} f_i(x) \tag{28.29}$$

has at least one solution and that one of the following holds:

(i) $0 \in \bigcap_{i=2}^{m} \operatorname{sri}\big(\operatorname{dom} f_i - \bigcap_{j=1}^{i-1} \operatorname{dom} f_j\big)$.
(ii) $\operatorname{dom} f_1 \cap \bigcap_{i=2}^{m} \operatorname{int}\operatorname{dom} f_i \neq \varnothing$.
(iii) *$\mathcal{H}$ is finite-dimensional and* $\bigcap_{i\in I} \operatorname{ri}\operatorname{dom} f_i \neq \varnothing$.

Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0, 2]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2 - \lambda_n) = +\infty$, let $\gamma \in \mathbb{R}_{++}$, and let $(y_{i,0})_{i\in I} \in \mathcal{H}^m$. Set

$$\begin{array}{l} \textit{for}\ n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} p_n = \dfrac{1}{m}\displaystyle\sum_{i\in I} y_{i,n}, \\ (\forall i \in I)\ \ x_{i,n} = \operatorname{Prox}_{\gamma f_i} y_{i,n}, \\ q_n = \dfrac{1}{m}\displaystyle\sum_{i\in I} x_{i,n}, \\ (\forall i \in I)\ \ y_{i,n+1} = y_{i,n} + \lambda_n\big(2q_n - p_n - x_{i,n}\big). \end{array} \right. \end{array} \tag{28.30}$$

Then $(p_n)_{n\in\mathbb{N}}$ converges weakly to a solution to (28.29).

Proof. Set $(\forall i \in I)$ $A_i = \partial f_i$ in Proposition 26.12. Then it follows from Example 23.3 that $(p_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{zer}\sum_{i\in I} \partial f_i$. However, Corollary 16.50 and Theorem 16.3 yield $\operatorname{zer}\sum_{i\in I} \partial f_i = \operatorname{zer}\partial(\sum_{i\in I} f_i) = \operatorname{Argmin}\sum_{i\in I} f_i$. ☐

28.4 Peaceman–Rachford Splitting Algorithm

The next proposition is an application of Proposition 26.13 in the context of convex minimization.

Proposition 28.8 (Peaceman–Rachford algorithm) *Let f and g be functions in $\Gamma_0(\mathcal{H})$ such that $\operatorname{zer}\big(\partial f + \partial g\big) \neq \varnothing$ (see Corollary 27.6 for* *sufficient conditions) and g is uniformly convex, let $\gamma \in \mathbb{R}_{++}$, and let $\bar{x}$ be the unique solution to the problem*

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f(x)+g(x). \tag{28.31}$$

Let $y_0 \in \mathcal{H}$, and set

$$\begin{array}{l} \textit{for}\; n=0,1,\ldots \\ \left\lfloor \begin{array}{l} x_n = \operatorname{Prox}_{\gamma g} y_n, \\ z_n = \operatorname{Prox}_{\gamma f}(2x_n - y_n), \\ y_{n+1} = y_n + 2(z_n - x_n). \end{array} \right. \end{array} \tag{28.32}$$

Then $x_n \to \bar{x}$.

Proof. Apply Proposition 26.13 with $A = \partial f$ and $B = \partial g$, and use Example 22.4(iii). ☐

28.5 Forward-Backward Splitting Algorithm

To minimize the sum of two functions in $\Gamma_0(\mathcal{H})$ when one of them is smooth, we can use the following version of the forward-backward algorithm (26.53). This method is sometimes called the *proximal-gradient algorithm*.

Corollary 28.9 (Forward-backward algorithm) *Let $f \in \Gamma_0(\mathcal{H})$, let $\beta \in \mathbb{R}_{++}$, let $g\colon \mathcal{H} \to \mathbb{R}$ be convex and differentiable with a $1/\beta$-Lipschitz continuous gradient, let $\gamma \in \left]0, 2\beta\right[$, and set $\delta = 2 - \gamma/(2\beta)$. Furthermore, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0,\delta]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$, and let $x_0 \in \mathcal{H}$. Suppose that $\operatorname{Argmin}(f+g) \neq \varnothing$ and set*

$$\begin{array}{l} \textit{for}\; n=0,1,\ldots \\ \left\lfloor \begin{array}{l} y_n = x_n - \gamma\nabla g(x_n), \\ x_{n+1} = x_n + \lambda_n\big(\operatorname{Prox}_{\gamma f} y_n - x_n\big). \end{array} \right. \end{array} \tag{28.33}$$

Then the following hold:

(i) *$(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{Argmin}(f+g)$.*
(ii) *Let $x \in \operatorname{Argmin}(f+g)$. Then $(\nabla g(x_n))_{n\in\mathbb{N}}$ converges strongly to the unique dual solution $\nabla g(x)$.*
(iii) *Suppose that one of the following holds:*

  (a) *f is uniformly convex on every nonempty bounded subset of $\operatorname{dom}\partial f$.*
  (b) *g is uniformly convex on every nonempty bounded subset of $\mathcal{H}$.*

  Then $(x_n)_{n\in\mathbb{N}}$ converges strongly to the unique minimizer of $f+g$.

Proof. Set $A = \partial f$ and $B = \nabla g$. Then A and B are maximally monotone by Theorem 20.25 and, since $\operatorname{dom} g = \mathcal{H}$, Corollary 27.3 yields $\operatorname{Argmin}(f+g) = \operatorname{zer}(A+B)$. On the other hand, by Corollary 18.17, B is β-cocoercive. In view of Example 23.3 and Example 22.5, the claims therefore follow from Theorem 26.14. ☐

A special case of the proximal-gradient algorithm (28.33) is the *projection-gradient algorithm* described next (this algorithm is also called the *projected gradient algorithm*).

Corollary 28.10 (Projection-gradient algorithm) *Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\beta \in \mathbb{R}_{++}$, let $f\colon \mathcal{H} \to \mathbb{R}$ be convex and differentiable with a $1/\beta$-Lipschitz continuous gradient, let $\gamma \in \left]0, 2\beta\right[$, and set $\delta = 2 - \gamma/(2\beta)$. Furthermore, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0, \delta]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$, and let $x_0 \in \mathcal{H}$. Suppose that $\operatorname{Argmin}_C f \neq \varnothing$ and set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n\big(P_C\big(x_n - \gamma\nabla f(x_n)\big) - x_n\big). \tag{28.34}$$

Then the following hold:

(i) *$(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{Argmin}_C f$.*
(ii) *Let $x \in \operatorname{Argmin}_C f$. Then $(\nabla f(x_n))_{n\in\mathbb{N}}$ converges strongly to $\nabla f(x)$.*
(iii) *Suppose that f is uniformly convex on every nonempty bounded subset of $\mathcal{H}$. Then $(x_n)_{n\in\mathbb{N}}$ converges strongly to the unique minimizer of f over C.*

Proof. Apply Corollary 28.9 to ι_C and f, and use Example 12.25 and Example 22.5. ☐

The next result concerns the alternating projection method.

Example 28.11 Let C and D be nonempty closed convex subsets of $\mathcal{H}$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0, 3/2]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(3 - 2\lambda_n) = +\infty$, and let $x_0 \in \mathcal{H}$. Suppose that C or D is bounded and set

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n\big(P_C P_D x_n - x_n\big). \tag{28.35}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point $x \in C$ that is at minimal distance from D.

Proof. Set $f = (1/2)d_D^2$. Then f is coercive if D is bounded, and we therefore deduce from Proposition 11.15 that $\operatorname{Argmin}_C f \neq \varnothing$. Next, it follows from Corollary 12.31 and Corollary 4.18 that $\nabla f = \mathrm{Id} - P_D$ is nonexpansive. Hence, the result is an application of Corollary 28.10(i) with $\beta = 1$ and $\gamma = 1$. ☐

We now consider an example of linear convergence.

Example 28.12 Let $f \in \Gamma_0(\mathcal{H})$ be α-strongly convex for some $\alpha \in \mathbb{R}_{++}$, let $g\colon \mathcal{H} \to \mathbb{R}$ be convex and differentiable with a $1/\beta$-Lipschitz continuous gradient for some $\beta \in \mathbb{R}_{++}$, and let $\gamma \in \left]0, 2\beta\right[$. Let $x_0 \in \mathcal{H}$ and set

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} y_n = x_n - \gamma\nabla g(x_n), \\ x_{n+1} = \operatorname{Prox}_{\gamma f} y_n. \end{array} \right. \end{array} \tag{28.36}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges linearly to the unique minimizer of $f + g$.

Proof. Set $A = \partial f$ and $B = \nabla g$. Then Corollary 27.3 yields $\operatorname{Argmin}(f+g) = \operatorname{zer}(A+B)$. We also note that A is maximally monotone by Theorem 20.25 and α-strongly monotone by Example 22.4(iv). Furthermore, we derive from Corollary 18.17 that B is β-cocoercive. In view of Example 23.3, the claim therefore follows from Proposition 26.16(i). ☐

The above forward-backward methods are based on monotone operator splitting techniques from Chapter 26. To conclude this section, we consider an alternative forward-backward splitting method. The sequence produced by this algorithm is minimizing, with decreasing objective values.

Proposition 28.13 *Let $f \in \Gamma_0(\mathcal{H})$, let $\beta \in \mathbb{R}_{++}$, let $g\colon \mathcal{H} \to \mathbb{R}$ be convex and differentiable with a $1/\beta$-Lipschitz continuous gradient, let $\varepsilon \in \left]0, \min\{1,\beta\}\right[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, 2\beta-\varepsilon]$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, 1]$, and let $x_0 \in \operatorname{dom} f$. Suppose that $\operatorname{Argmin}(f+g) \neq \varnothing$ and set*

$$\begin{array}{l} \textit{for}\ n = 0,1,\ldots \\ \left\lfloor \begin{array}{l} y_n = x_n - \gamma_n \nabla g(x_n), \\ z_n = \operatorname{Prox}_{\gamma_n f} y_n, \\ x_{n+1} = x_n + \lambda_n(z_n - x_n). \end{array} \right. \end{array} \tag{28.37}$$

Then the following hold:

(i) *Let $x \in \operatorname{Argmin}(f+g)$. Then $\sum_{n\in\mathbb{N}} \|\nabla g(x_n) - \nabla g(x)\|^2 < +\infty$.*
(ii) *$\sum_{n\in\mathbb{N}} \|z_n - x_n\|^2 < +\infty$.*
(iii) *$(x_n)_{n\in\mathbb{N}}$ is a minimizing sequence; more precisely*

$$\sum_{n\in\mathbb{N}} \big((f+g)(x_n) - \inf(f+g)(\mathcal{H})\big)^2 < +\infty$$
$$\text{and} \quad (f+g)(x_n) \downarrow \inf(f+g)(\mathcal{H}). \tag{28.38}$$

(iv) *$(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{Argmin}(f+g)$.*
(v) *Suppose that $f+g$ is uniformly quasiconvex on every bounded subset of $\operatorname{dom} f$. Then $(x_n)_{n\in\mathbb{N}}$ converges strongly to the unique minimizer of $f+g$.*

Proof. Set $h = f+g$. By (24.3), $(\forall n \in \mathbb{N})$ $z_n \in \operatorname{dom}\partial f \subset \operatorname{dom} f = \operatorname{dom} h$. Since $x_0 \in \operatorname{dom} f = \operatorname{dom} h$ and $\operatorname{dom} h$ is convex, it therefore follows that $(\forall n \in \mathbb{N})$ $x_n \in \operatorname{dom} h$. Now let $n \in \mathbb{N}$ and $x \in \operatorname{Argmin}(f+g)$. Since ∇g is β-cocoercive by Corollary 18.17, Proposition 4.39 implies that $\operatorname{Id} - \gamma_n\nabla g$ is $\gamma_n/(2\beta)$-averaged. We therefore deduce from Corollary 27.3(viii), the firm nonexpansiveness of $\operatorname{Prox}_{\gamma_n f}$, Proposition 4.4, and Proposition 4.35 that

$$\begin{aligned} \|z_n - x\|^2 &= \|\operatorname{Prox}_{\gamma_n f}(x_n - \gamma_n\nabla g(x_n)) - \operatorname{Prox}_{\gamma_n f}(x - \gamma_n \nabla g(x))\|^2 \\ &\leqslant \|(\operatorname{Id} - \gamma_n\nabla g)x_n - (\operatorname{Id} - \gamma_n \nabla g)x\|^2 - \|y_n - z_n + \gamma_n\nabla g(x)\|^2 \\ &\leqslant \|x_n - x\|^2 - \gamma_n(2\beta - \gamma_n)\|\nabla g(x_n) - \nabla g(x)\|^2 \\ &\quad\ - \|y_n - z_n + \gamma_n \nabla g(x)\|^2. \end{aligned} \tag{28.39}$$

On the other hand, by convexity of $\|\cdot\|^2$, $\|x_{n+1}-x\|^2 \leqslant (1-\lambda_n)\|x_n-x\|^2 + \lambda_n\|z_n-x\|^2$. Hence,

$$\begin{aligned}\|x_{n+1}-x\|^2 &\leqslant \|x_n-x\|^2 - \varepsilon\gamma_n(2\beta-\gamma_n)\|\nabla g(x_n)-\nabla g(x)\|^2\\ &\qquad - \varepsilon\|y_n-z_n+\gamma_n\nabla g(x)\|^2 \qquad (28.40)\\ &\leqslant \|x_n-x\|^2 - \varepsilon^3\|\nabla g(x_n)-\nabla g(x)\|^2 \qquad (28.41)\\ &\leqslant \|x_n-x\|^2, \qquad (28.42)\end{aligned}$$

which implies that

$$(x_n)_{n\in\mathbb{N}} \text{ is Fejér monotone with respect to } \operatorname{Argmin} h. \tag{28.43}$$

(i): In view of (28.41), $\sum_{n\in\mathbb{N}}\|\nabla g(x_n)-\nabla g(x)\|^2 \leqslant \varepsilon^{-3}\|x_0-x\|^2$.

(ii): By (28.40), $\sum_{n\in\mathbb{N}}\|z_n-y_n-\gamma_n\nabla g(x)\|^2 \leqslant \varepsilon^{-1}\|x_0-x\|^2$. Consequently, $(z_n-y_n-\gamma_n\nabla g(x))_{n\in\mathbb{N}} \in \ell^2(\mathbb{N})$. However, (i) yields $(\gamma_n\nabla g(x)-\gamma_n\nabla g(x_n))_{n\in\mathbb{N}} \in \ell^2(\mathbb{N})$. Thus $(z_n-x_n)_{n\in\mathbb{N}} = (z_n-y_n-\gamma_n\nabla g(x))_{n\in\mathbb{N}} + (\gamma_n\nabla g(x)-\gamma_n\nabla g(x_n))_{n\in\mathbb{N}} \in \ell^2(\mathbb{N})$.

(iii): Let $z\in\mathcal{H}$ and let $n\in\mathbb{N}$. Since $z_n = \operatorname{Prox}_{\gamma_n f} y_n$, it follows from Proposition 12.26 that

$$\begin{aligned}\big\langle z-z_n \mid \gamma_n^{-1}(x_n-z_n)-\nabla g(x_n)\big\rangle &= \big\langle z-z_n \mid \gamma_n^{-1}(y_n-z_n)\big\rangle\\ &\leqslant f(z)-f(z_n). \qquad (28.44)\end{aligned}$$

On the other hand, Proposition 17.7 yields

$$\langle z-z_n \mid \nabla g(z_n)\rangle + g(z_n) \leqslant g(z). \tag{28.45}$$

Now let $\mu = \sup_{k\in\mathbb{N}}\|z_k-z\|$. Since (28.43) implies that $(x_k)_{k\in\mathbb{N}}$ is bounded, (ii) yields $\mu<+\infty$. In addition, upon adding (28.44) and (28.45), we obtain

$$\begin{aligned}&-\mu\|x_n-z_n\|(1/\varepsilon+1/\beta)+h(z_n)\\ &\quad\leqslant -\gamma_n^{-1}\|z_n-z\|\,\|x_n-z_n\| - \|z_n-z\|\,\|\nabla g(x_n)-\nabla g(z_n)\| + h(z_n)\\ &\quad\leqslant \big\langle z-z_n \mid \gamma_n^{-1}(x_n-z_n)-\big(\nabla g(x_n)-\nabla g(z_n)\big)\big\rangle + h(z_n)\\ &\quad\leqslant h(z). \qquad (28.46)\end{aligned}$$

Hence $h(z_n)-h(z) \leqslant \mu(1/\varepsilon+1/\beta)\|z_n-x_n\| \leqslant (2\mu/\varepsilon)\|z_n-x_n\|$ and taking $z\in\operatorname{Argmin} h$ yields

$$0 \leqslant h(z_n)-\inf h(\mathcal{H}) \leqslant \frac{2\mu}{\varepsilon}\|z_n-x_n\|. \tag{28.47}$$

Next, we derive from (28.44) that

$$f(z_n) \leqslant f(x_n) - \big\langle x_n-z_n \mid \gamma_n^{-1}(x_n-z_n)-\nabla g(x_n)\big\rangle \tag{28.48}$$

and from the descent lemma (Theorem 18.15) that

$$g(z_n) \leqslant g(x_n) - \langle x_n - z_n \mid \nabla g(x_n)\rangle + \frac{1}{2\beta}\|z_n - x_n\|^2. \tag{28.49}$$

Upon adding (28.48) and (28.49), we obtain

$$h(z_n) \leqslant h(x_n) - \left(\frac{1}{\gamma_n} - \frac{1}{2\beta}\right)\|z_n - x_n\|^2 \leqslant h(x_n). \tag{28.50}$$

Hence, by convexity of h,

$$\begin{aligned} h(x_{n+1}) &= h\big(\lambda_n z_n + (1-\lambda_n)x_n\big) \\ &\leqslant \lambda_n h(z_n) + (1-\lambda_n)h(x_n) \\ &= h(x_n) - \lambda_n\big(h(x_n) - h(z_n)\big) \\ &\leqslant h(x_n). \end{aligned} \tag{28.51}$$

This shows that $(h(x_n))_{n\in\mathbb{N}}$ is decreasing and that $0 \leqslant h(x_n) - h(z_n) \leqslant \varepsilon^{-1}(h(x_n) - h(x_{n+1}))$. In turn,

$$\sum_{n\in\mathbb{N}} h(x_n) - h(z_n) \leqslant \varepsilon^{-1}h(x_0) < +\infty \tag{28.52}$$

and, therefore, $(h(x_n) - h(z_n))_{n\in\mathbb{N}} \in \ell^1(\mathbb{N}) \subset \ell^2(\mathbb{N})$. On the other hand, appealing to (ii) and (28.47), we get $(h(z_n) - \inf h(\mathcal{H}))_{n\in\mathbb{N}} \in \ell^2(\mathbb{N})$. Altogether, $(h(x_n) - \inf h(\mathcal{H}))_{n\in\mathbb{N}} \in \ell^2(\mathbb{N})$.

(iv): Since (iii) asserts that $(x_n)_{n\in\mathbb{N}}$ is a minimizing sequence and since $h \in \Gamma_0(\mathcal{H})$, this follows from (28.43), Theorem 5.5, and Proposition 11.21.

(v): It follows from Proposition 11.8(i) that h has a unique minimizer, say x. Furthermore, as shown above, $(x_n)_{n\in\mathbb{N}}$ is a minimizing sequence of h and $C = \{x_n\}_{n\in\mathbb{N}} \cup \{x\} \subset \operatorname{dom} h$. Now let ϕ be the modulus of uniform quasiconvexity of h on C and let $\alpha \in\,]0,1[$. Then it follows from (10.27) that, for every $n \in \mathbb{N}$,

$$\begin{aligned} h(x) + \alpha(1-\alpha)\phi(\|x_n - x\|) &= \inf h(\mathcal{H}) + \alpha(1-\alpha)\phi(\|x_n - x\|) \\ &\leqslant h\big(\alpha x_n + (1-\alpha)x\big) + \alpha(1-\alpha)\phi(\|x_n - x\|) \\ &\leqslant \max\{h(x_n), h(x)\} \\ &= h(x_n). \end{aligned} \tag{28.53}$$

Thus, since $h(x_n) \to h(x)$, $\phi(\|x_n - x\|) \to 0$ and hence $\|x_n - x\| \to 0$. $\square$

28.6 Tseng's Splitting Algorithm

We study the convergence of a variant of (28.33) under less restrictive assumptions on the functions f and g than those imposed in Corollary 28.9.

Proposition 28.14 (Tseng's algorithm) *Let D be a nonempty subset of $\mathcal{H}$, let $f \in \Gamma_0(\mathcal{H})$ be such that $\operatorname{dom} \partial f \subset D$, let $g \in \Gamma_0(\mathcal{H})$ be Gâteaux differentiable on D, and let $\beta \in \mathbb{R}_{++}$. Suppose that C is a closed convex subset of D such that $C \cap \operatorname{Argmin}(f+g) \neq \varnothing$, and that ∇g is $1/\beta$-Lipschitz continuous relative to $C \cup \operatorname{dom} \partial f$. Let $x_0 \in C$, let $\gamma \in \left]0, \beta\right[$, and set*

$$\begin{array}{l} \textit{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} y_n = x_n - \gamma \nabla g(x_n), \\ z_n = \operatorname{Prox}_{\gamma f} y_n, \\ r_n = z_n - \gamma \nabla g(z_n), \\ x_{n+1} = P_C\big(x_n - y_n + r_n\big). \end{array} \right. \end{array} \tag{28.54}$$

Then the following hold:

(i) *$(x_n - z_n)_{n\in\mathbb{N}}$ converges strongly to 0.*
(ii) *$(x_n)_{n\in\mathbb{N}}$ and $(z_n)_{n\in\mathbb{N}}$ converge weakly to a point in $C \cap \operatorname{Argmin}(f+g)$.*
(iii) *Suppose that f or g is uniformly convex on every nonempty bounded subset of $\operatorname{dom} \partial f$. Then $(x_n)_{n\in\mathbb{N}}$ and $(z_n)_{n\in\mathbb{N}}$ converge strongly to the unique point in $C \cap \operatorname{Argmin}(f+g)$.*

Proof. Set $A = \partial f$ and $B = \partial g$. Then A and B are maximally monotone by Theorem 20.25. Moreover, it follows from Proposition 17.31(i) that B is single-valued on D, where it coincides with ∇g. Furthermore, Proposition 17.50 and Proposition 16.27 imply that $\operatorname{dom} \partial f \subset D \subset \operatorname{int} \operatorname{dom} g = \operatorname{int} \operatorname{dom} \partial g$ and hence, by Corollary 25.5(ii), $A + B = \partial f + \partial g$ is maximally monotone. In turn, it follows from Corollary 16.48(ii) that $\partial(f+g) = \partial f + \partial g$ and hence from Proposition 27.1 that $\operatorname{Argmin}(f+g) = \operatorname{zer}(A+B)$. Altogether, the results follow from Theorem 26.17, Example 23.3, and Example 22.5. ☐

As an example, we obtain the following variant of the projection-gradient algorithm.

Example 28.15 Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\beta \in \mathbb{R}_{++}$, and let $g \in \Gamma_0(\mathcal{H})$ be differentiable with a $1/\beta$-Lipschitz continuous gradient relative to C. Suppose that $\operatorname{Argmin}_C g \neq \varnothing$, let $x_0 \in C$, let $\gamma \in \left]0, \beta\right[$, and set

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} y_n = x_n - \gamma \nabla g(x_n), \\ z_n = P_C y_n, \\ r_n = z_n - \gamma \nabla g(z_n), \\ x_{n+1} = P_C\big(x_n - y_n + r_n\big). \end{array} \right. \end{array} \tag{28.55}$$

Then the following hold:

(i) $(x_n - z_n)_{n\in\mathbb{N}}$ converges strongly to 0.
(ii) $(x_n)_{n\in\mathbb{N}}$ and $(z_n)_{n\in\mathbb{N}}$ converge weakly to a point in $\operatorname{Argmin}_C g$.
(iii) Suppose that g is uniformly convex on every nonempty bounded subset of C. Then $(x_n)_{n\in\mathbb{N}}$ and $(z_n)_{n\in\mathbb{N}}$ converge strongly to the unique point in $\operatorname{Argmin}_C g$.

Proof. This is a special case of Proposition 28.14 applied to $D = C$ and $f = \iota_C$. $\square$

28.7 A Primal-Dual Algorithm

In this section we revisit the duality framework investigated in Section 19.1 and provide an algorithm to solve both the primal and the dual problems in the setting of Proposition 19.5.

Proposition 28.16 *Let $\varphi \in \Gamma_0(\mathcal{H})$, let $\psi \in \Gamma_0(\mathcal{K})$, let $z \in \mathcal{H}$, let $r \in \mathcal{K}$, and suppose that $L \in \mathcal{B}(\mathcal{H},\mathcal{K}) \smallsetminus \{0\}$ satisfies $r \in \operatorname{sri}(L(\operatorname{dom}\varphi) - \operatorname{dom}\psi)$. Consider the (primal) problem*

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; \varphi(x) + \psi(Lx - r) + \tfrac{1}{2}\|x - z\|^2, \tag{28.56}$$

together with the dual problem

$$\underset{v\in\mathcal{K}}{\text{minimize}}\;\; {}^{1}(\varphi^*)(z - L^*v) + \psi^*(v) + \langle v \mid r\rangle\,. \tag{28.57}$$

Let $\gamma \in \left]0, 2\|L\|^{-2}\right[$, set $\delta = 2 - \gamma\|L\|^2/2$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,\delta[$ such that $\inf_{n\in\mathbb{N}} \lambda_n > 0$ and $\sup_{n\in\mathbb{N}} \lambda_n < \delta$, and let $v_0 \in \mathcal{K}$. Set

$$\begin{array}{l} \text{for } n = 0,1,\ldots \\ \left\lfloor \begin{array}{l} x_n = \operatorname{Prox}_{\varphi}(z - L^*v_n), \\ v_{n+1} = v_n + \lambda_n\big(\operatorname{Prox}_{\gamma\psi^*}\big(v_n + \gamma(Lx_n - r)\big) - v_n\big), \end{array}\right. \end{array} \tag{28.58}$$

and let $\bar{x}$ be the unique solution to (28.56). Then the following hold:

(i) *$(v_n)_{n\in\mathbb{N}}$ converges weakly to a solution $\bar{v}$ to the dual problem (28.57) and $\bar{x} = \operatorname{Prox}_{\varphi}(z - L^*\bar{v})$.*
(ii) *$(x_n)_{n\in\mathbb{N}}$ converges strongly to $\bar{x}$.*

Proof. Set $\mathfrak{h}\colon \mathcal{H} \to \left]-\infty,+\infty\right] : x \mapsto \varphi(x) + (1/2)\|x - z\|^2$ and $\mathfrak{j}\colon \mathcal{K} \to \left]-\infty,+\infty\right] : y \mapsto \psi(y - r)$. Then (28.56) can be rewritten as

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; \mathfrak{h}(x) + \mathfrak{j}(Lx). \tag{28.59}$$

Moreover, the assumptions imply that $\mathfrak{h} \in \Gamma_0(\mathcal{H})$, $\mathfrak{j} \in \Gamma_0(\mathcal{K})$, and $0 \in \operatorname{sri}(L(\operatorname{dom}\mathfrak{h}) - \operatorname{dom}\mathfrak{j})$. Hence, it follows from Theorem 15.23 that the dual

$$\underset{v\in\mathcal{K}}{\text{minimize}}\ \ \mathfrak{h}^*(-L^*v) + \mathfrak{j}^*(v) \tag{28.60}$$

of (28.59) has at least one solution. We derive from Proposition 14.1 and Proposition 13.23(iii) that $\mathfrak{h}^*\colon u \mapsto {}^1(\varphi^*)(u+z) - (1/2)\|z\|^2$ and $\mathfrak{j}^*\colon v \mapsto \psi^*(v) + \langle v \mid r\rangle$. Therefore, (28.60) coincides with (28.57) up to an additive constant. We have thus shown that (28.57) is the dual of (28.56) (up to an additive constant), and that it has at least one solution. Now let us define two functions f and g on $\mathcal{K}$ by $f\colon v \mapsto \psi^*(v) + \langle v \mid r\rangle$ and $g\colon v \mapsto {}^1(\varphi^*)(z - L^*v)$. Then (28.57) amounts to minimizing $f+g$ on $\mathcal{K}$. By Corollary 13.38, f and g are in $\Gamma_0(\mathcal{K})$ and, as just observed, $\operatorname{Argmin}(f+g) \neq \varnothing$. Moreover, it follows from Proposition 12.30 and (14.6) that g is differentiable on $\mathcal{K}$ with gradient

$$\nabla g\colon v \mapsto -L\big(\operatorname{Prox}_\varphi(z - L^*v)\big). \tag{28.61}$$

Hence, Proposition 12.28 and Fact 2.25(ii) yield

$$\begin{aligned}
(\forall v \in \mathcal{K})(\forall w \in \mathcal{K})\quad & \|\nabla g(v) - \nabla g(w)\| \\
&\leqslant \|L\|\,\|\operatorname{Prox}_\varphi(z - L^*v) - \operatorname{Prox}_\varphi(z - L^*w)\| \\
&\leqslant \|L\|\,\|L^*v - L^*w\| \\
&\leqslant \|L\|^2\,\|v - w\|.
\end{aligned} \tag{28.62}$$

On the other hand, we derive from (28.58), (28.61), and Proposition 24.8(vi)&(i) that, for every $n \in \mathbb{N}$,

$$\begin{aligned}
v_{n+1} &= v_n + \lambda_n\big(\operatorname{Prox}_{\gamma\psi^*}\big(v_n + \gamma(Lx_n - r)\big) - v_n\big) \\
&= v_n + \lambda_n\big(\operatorname{Prox}_{\gamma\psi^*}\big(v_n - \gamma(\nabla g(v_n) + r)\big) - v_n\big) \\
&= v_n + \lambda_n\big(\operatorname{Prox}_{\gamma(\psi^* + \langle\cdot\mid r\rangle)}\big(v_n - \gamma\nabla g(v_n)\big) - v_n\big) \\
&= v_n + \lambda_n\big(\operatorname{Prox}_{\gamma f}\big(v_n - \gamma\nabla g(v_n)\big) - v_n\big).
\end{aligned} \tag{28.63}$$

We thus recover the forward-backward iteration (28.33).

(i): In view of the above, this follows from Corollary 28.9 and Proposition 19.5.

(ii): As seen in (i), $v_n \rightharpoonup \bar{v}$, where $\bar{v}$ is a solution to (28.57), and $\bar{x} = \operatorname{Prox}_\varphi(z - L^*\bar{v})$. Now set $\rho = \sup_{n\in\mathbb{N}} \|v_n - \bar{v}\|$. Then Lemma 2.46 yields $\rho < +\infty$. Hence, using Proposition 12.28 and (28.61), we obtain, for every $n \in \mathbb{N}$,

$$\begin{aligned}
\|x_n - \bar{x}\|^2 &= \|\operatorname{Prox}_\varphi(z - L^*v_n) - \operatorname{Prox}_\varphi(z - L^*\bar{v})\|^2 \\
&\leqslant \langle -L^*v_n + L^*\bar{v} \mid \operatorname{Prox}_\varphi(z - L^*v_n) - \operatorname{Prox}_\varphi(z - L^*\bar{v})\rangle \\
&= \big\langle v_n - \bar{v} \mid -L\big(\operatorname{Prox}_\varphi(z - L^*v_n)\big) + L\big(\operatorname{Prox}_\varphi(z - L^*\bar{v})\big)\big\rangle
\end{aligned}$$

$$= \langle v_n - \bar{v} \mid \nabla g(v_n) - \nabla g(\bar{v}) \rangle$$
$$\leqslant \rho \|\nabla g(v_n) - \nabla g(\bar{v})\|. \tag{28.64}$$

However, Corollary 28.9(ii) yields $\|\nabla g(v_n) - \nabla g(\bar{v})\| \to 0$. Hence, we derive from (28.64) that $x_n \to \bar{x}$. ☐

As an illustration, we revisit Example 19.8 and Example 19.10.

Example 28.17 Let K be a closed convex cone in $\mathcal{H}$, let $\psi \in \Gamma_0(\mathcal{K})$ be positively homogeneous, set $D = \partial\psi(0)$, let $z \in \mathcal{H}$, let $r \in \mathcal{K}$, and suppose that $L \in \mathcal{B}(\mathcal{H},\mathcal{K}) \smallsetminus \{0\}$ satisfies $r \in \operatorname{sri}(L(K) - \operatorname{dom}\psi)$. Consider the (primal) problem

$$\underset{x \in K}{\text{minimize}} \;\; \psi(Lx - r) + \tfrac{1}{2}\|x - z\|^2, \tag{28.65}$$

together with the dual problem

$$\underset{v \in D}{\text{minimize}} \;\; \tfrac{1}{2} d_{K^{\ominus}}^2(z - L^* v) + \langle v \mid r \rangle \,. \tag{28.66}$$

Let $\gamma \in \left]0, 2\|L\|^{-2}\right[$, set $\delta = 2 - \gamma\|L\|^2/2$, let $\lambda \in \left]0, \delta\right[$, and let $v_0 \in \mathcal{K}$. Set

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} x_n = P_K(z - L^* v_n), \\ v_{n+1} = v_n + \lambda\big(P_D\big(v_n + \gamma(Lx_n - r)\big) - v_n\big), \end{array} \right. \end{array} \tag{28.67}$$

and let $\bar{x}$ be the unique solution to (28.65). Then the following hold:

(i) $(v_n)_{n\in\mathbb{N}}$ converges weakly to a solution $\bar{v}$ to (28.66) and $\bar{x} = P_K(z - L^*\bar{v})$.
(ii) $(x_n)_{n\in\mathbb{N}}$ converges strongly to $\bar{x}$.

Proof. In view of Example 19.8, this is an application of Proposition 28.16 with $\varphi = \iota_K$ and $\lambda_n \equiv \lambda$ since, by Proposition 16.24, $\psi^* = \sigma_D^* = \iota_D$ and therefore $\operatorname{Prox}_{\gamma\psi^*} = P_D$. ☐

Example 28.18 Let C and D be closed convex subsets of $\mathcal{H}$, let $z \in \mathcal{H}$, and suppose that $L \in \mathcal{B}(\mathcal{H},\mathcal{K}) \smallsetminus \{0\}$ satisfies $0 \in \operatorname{sri}(L(C) - D)$. Consider the best approximation (primal) problem

$$\underset{\substack{x \in C \\ Lx \in D}}{\text{minimize}} \;\; \tfrac{1}{2}\|x - z\|^2, \tag{28.68}$$

together with the dual problem

$$\underset{v \in \mathcal{K}}{\text{minimize}} \;\; \tfrac{1}{2}\|z - L^* v\|^2 - \tfrac{1}{2} d_C^2(z - L^* v) + \sigma_D(v). \tag{28.69}$$

Let $\gamma \in \left]0, 2\|L\|^{-2}\right[$, set $\delta = 2 - \gamma\|L\|^2/2$, let $\lambda \in \left]0, \delta\right[$, and let $v_0 \in \mathcal{K}$. Set

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} x_n = P_C(z - L^* v_n), \\ v_{n+1} = v_n + \gamma\lambda\big(Lx_n - P_D(\gamma^{-1} v_n + Lx_n)\big), \end{array} \right. \end{array} \tag{28.70}$$

and let $\bar{x}$ be the unique solution to (28.68). Then the following hold:

(i) $(v_n)_{n\in\mathbb{N}}$ converges weakly to a solution $\bar{v}$ to (28.69) and $\bar{x} = P_C(z - L^*\bar{v})$.
(ii) $(x_n)_{n\in\mathbb{N}}$ converges strongly to $\bar{x}$.

Proof. In view of Example 19.10, this is an application of Proposition 28.16 with $\varphi = \iota_C$, $\psi = \iota_D$, $r = 0$, and $\lambda_n \equiv \lambda$, where we have used Theorem 14.3(ii) to derive (28.70) from (28.58). ☐

Example 28.19 Let C be a closed convex subset of $\mathcal{H}$, let $z \in \mathcal{H}$, let $r \in \mathcal{K}$, and suppose that $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$ satisfies $\|L\| = 1$ and that $r \in \operatorname{sri} L(C)$. Consider the best approximation (primal) problem

$$\underset{\substack{x\in C\\ Lx=r}}{\text{minimize}} \;\; \tfrac{1}{2}\|x - z\|^2, \tag{28.71}$$

together with the dual problem

$$\underset{v\in\mathcal{K}}{\text{minimize}} \;\; \tfrac{1}{2}\|z - L^*v\|^2 - \tfrac{1}{2}d_C^2(z - L^*v) + \langle v \mid r\rangle\,. \tag{28.72}$$

Let $v_0 \in \mathcal{K}$, set

$$\begin{array}{l} \text{for } n = 0,1,\ldots \\ \left\lfloor \begin{array}{l} x_n = P_C(z - L^*v_n), \\ v_{n+1} = v_n + Lx_n - r, \end{array}\right. \end{array} \tag{28.73}$$

and let $\bar{x}$ be the unique solution to (28.71). Then the following hold:

(i) $(v_n)_{n\in\mathbb{N}}$ converges weakly to a solution $\bar{v}$ to (28.72) and $\bar{x} = P_C(z - L^*\bar{v})$.
(ii) $(x_n)_{n\in\mathbb{N}}$ converges strongly to $\bar{x}$.

Proof. Apply Example 28.18 with $D = \{r\}$, $\gamma = 1$, and $\lambda = 1$. ☐

28.8 Composite Minimization Problems

The focus of this section is on solving the following primal-dual minimization problem, which is based on Example 19.3.

Problem 28.20 Let $z \in \mathcal{H}$, let $h \in \Gamma_0(\mathcal{H})$, and let I be a nonempty finite set. For every $i \in I$, let $\mathcal{K}_i$ be a real Hilbert space, let $r_i \in \mathcal{K}_i$, let $g_i \in \Gamma_0(\mathcal{K}_i)$, and suppose that $L_i \in \mathcal{B}(\mathcal{H},\mathcal{K}_i) \smallsetminus \{0\}$. Solve the primal problem

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; h(x) - \langle x \mid z\rangle + \sum_{i\in I} g_i(L_i x - r_i) \tag{28.74}$$

together with the dual problem

$$\underset{(v_i)_{i\in I}\in\bigoplus_{i\in I}\mathcal{K}_i}{\text{minimize}} \;\; h^*\Big(z - \sum_{i\in I} L_i^* v_i\Big) + \sum_{i\in I}\big(g_i^*(v_i) + \langle v_i \mid r_i\rangle\big). \tag{28.75}$$

We propose the following primal-dual algorithm to solve this problem.

Proposition 28.21 *In Problem 28.20, suppose that*

$$z \in \operatorname{ran}\Big(\partial h + \sum_{i\in I} L_i^* \circ (\partial g_i) \circ (L_i \cdot - r_i)\Big). \tag{28.76}$$

Let $x_0 \in \mathcal{H}$, *let* $(v_{i,0})_{i\in I} \in \bigoplus_{i\in I} \mathcal{K}_i$, *let* $\gamma \in \big]0, 1/\sqrt{\sum_{i\in I}\|L_i\|^2}\big[$, *and set*

$$\begin{array}{l}
\text{for } n = 0,1,\ldots\\
\left\lfloor\begin{array}{l}
y_{1,n} = x_n - \gamma\sum_{i\in I} L_i^* v_{i,n},\\
p_{1,n} = \operatorname{Prox}_{\gamma h}(y_{1,n} + \gamma z),\\
\text{for every } i \in I\\
\left\lfloor\begin{array}{l}
y_{2,i,n} = v_{i,n} + \gamma L_i x_n,\\
p_{2,i,n} = y_{2,n} - \gamma\big(r_i + \operatorname{Prox}_{\gamma^{-1} g_i}(\gamma^{-1} y_{2,i,n} - r_i)\big),\\
q_{2,i,n} = p_{2,i,n} + \gamma L_i p_{1,n},\\
v_{i,n+1} = v_{i,n} - y_{2,i,n} + q_{2,i,n},
\end{array}\right.\\
q_{1,n} = p_{1,n} - \gamma\sum_{i\in I} L_i^* p_{2,i,n},\\
x_{n+1} = x_n - y_{1,n} + q_{1,n}.
\end{array}\right.
\end{array} \tag{28.77}$$

Then the following hold:

(i) $x_n - p_{1,n} \to 0$ *and* $(\forall i \in I)$ $v_{i,n} - p_{2,i,n} \to 0$.

(ii) *There exist a solution* $\bar{x}$ *to* (28.74) *and a solution* $(\bar{v}_i)_{i\in I}$ *to* (28.75) *such that* $z - \sum_{i\in I} L_i^*\bar{v}_i \in \partial h(\bar{x})$, $x_n \rightharpoonup \bar{x}$, *and* $(\forall i \in I)$ $L_i\bar{x} - r_i \in \partial g_i^*(\bar{v}_i)$ *and* $v_{i,n} \rightharpoonup \bar{v}_i$.

Proof. We first establish a connection between Problem 28.20 and Problem 26.35. To this end, set

$$A = \partial h \quad \text{and} \quad (\forall i \in I) \quad B_i = \partial g_i. \tag{28.78}$$

It is clear that (28.76) coincides with (26.102) and, using (24.6), that (28.77) coincides with (26.103). Moreover, it follows from Theorem 20.25 that the operators A and $(B_i)_{i\in I}$ are maximally monotone. Altogether, we apply Corollary 16.30 and Corollary 26.36 to obtain (i) and the existence of $\bar{x} \in \mathcal{H}$ and $(\bar{v}_i)_{i\in I} \in \bigoplus_{i\in I}\mathcal{K}_i$ such that

$$\begin{cases}
x_n \rightharpoonup \bar{x},\\
z - \sum_{i\in I} L_i^*\bar{v}_i \in \partial h(\bar{x}),\\
z \in \partial h(\bar{x}) + \sum_{i\in I} L_i^*\big(\partial g_i(L_i\bar{x} - r_i)\big)
\end{cases} \tag{28.79}$$

and

$$(\forall i \in I) \quad \begin{cases}
v_{i,n} \rightharpoonup \bar{v}_i,\\
L_i\bar{x} - r_i \in \partial g_i^*(\bar{v}_i),\\
-r_i \in -L_i\Big(\partial h^*\Big(z - \sum_{j\in I} L_j^*\bar{v}_j\Big)\Big) + \partial g_i^*(\bar{v}_i).
\end{cases} \tag{28.80}$$

Now set $f = h - \langle\cdot \mid z\rangle$, $\mathcal{K} = \bigoplus_{i\in I}\mathcal{K}_i$, $g = \bigoplus_{i\in I} g_i(\cdot - r_i)$, $r = (r_i)_{i\in I}$, and $L\colon \mathcal{H} \to \mathcal{K}\colon x \mapsto (L_i x)_{i\in I}$. Then $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{K})$, $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$, and $L^*\colon \mathcal{K} \to \mathcal{H}\colon (y_i)_{i\in I} \mapsto \sum_{i\in I} L_i^* y_i$. Moreover, it follows from (28.79), Proposition 16.9, and Proposition 27.5(i), that

$$\bar{x} \in \operatorname{zer}\big(\partial f + L^* \circ (\partial g) \circ L\big) \subset \operatorname{Argmin}(f + g \circ L). \tag{28.81}$$

Likewise, it follows from (28.80) and Proposition 27.5(ii), that

$$(\bar{v}_i)_{i\in I} \in \operatorname{zer}\big(-L\circ(\partial f^*)\circ(-L^*) + \partial g^*\big) \subset \operatorname{Argmin}\big(f^*\circ(-L^*) + g^*\big). \tag{28.82}$$

However, as seen in Example 19.3, (28.74) amounts to minimizing $f + g\circ L$ and (28.75) amounts to minimizing $f^*\circ(-L^*) + g^*$. Hence, we conclude that $\bar{x}$ solves (28.74) and $(\bar{v}_i)_{i\in I}$ solves (28.75). ☐

Remark 28.22 In the setting of Proposition 28.21 and its proof, we have $\operatorname{Argmin}(f + g\circ L) \neq \varnothing$ and (28.76) reduces to $\operatorname{zer}\big(\partial f + L^*\circ(\partial g)\circ L\big) \neq \varnothing$. Conditions for this property to be satisfied can be derived from Proposition 27.5(iii).

Exercises

Exercise 28.1 Let C and D be closed convex subsets of $\mathcal{H}$ such that $C\cap D \neq \varnothing$, and let $y_0 \in \mathcal{H}$. Set

$$(\forall n \in \mathbb{N}) \quad y_{n+1} = y_n + P_C(2P_D y_n - y_n) - P_D y_n. \tag{28.83}$$

Show that $(y_n)_{n\in\mathbb{N}}$ converges weakly to a point $y \in \mathcal{H}$ and that $P_D y_n \rightharpoonup P_D y \in C\cap D$.

Exercise 28.2 Let $\varphi \in \Gamma_0(\mathcal{H})$, let $\psi \in \Gamma_0(\mathcal{H})$, let $z \in \mathcal{H}$, and suppose that $0 \in \operatorname{sri}(\operatorname{dom}\psi - \operatorname{dom}\varphi)$. Consider the problem of constructing $\operatorname{Prox}_{\varphi+\psi} z$, i.e., of finding the solution to the problem

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; \varphi(x) + \psi(x) + \tfrac{1}{2}\|x - z\|^2. \tag{28.84}$$

Use the following results to solve (28.84) (in each case provide an explicit algorithm and a strong convergence result):

(i) Corollary 28.3 and Theorem 26.11(vi).
(ii) Proposition 28.16.

Exercise 28.3 Let $f \in \Gamma_0(\mathcal{H})$, let $x_0 \in \mathcal{H}$, and set $(\forall n \in \mathbb{N})$ $x_{n+1} = \operatorname{Prox}_f x_n$. Suppose that $(x_n)_{n\in\mathbb{N}}$ converges strongly to $z \in \operatorname{Argmin} f$. Use Proposition 24.26 to show that $n(f(x_n) - f(z)) \to 0$.

Exercise 28.4 In Corollary 28.10, suppose that C is a compact set and that $\sup_{n\in\mathbb{N}} \lambda_n \leqslant 1$. Show that the sequence $(x_n)_{n\in\mathbb{N}}$ produced by the projection-gradient algorithm (28.34) with $x_0 \in C$ converges strongly to a minimizer of f over C.

Exercise 28.5 Let m be a strictly positive integer, set $I = \{1, \ldots, m\}$, and let C and $(C_i)_{i\in I}$ be nonempty closed convex subsets of $\mathcal{H}$, at least one of which is bounded. Consider the problem

$$\underset{x\in C}{\text{minimize}} \quad \frac{1}{2m} \sum_{i\in I} d_{C_i}^2(x). \tag{28.85}$$

(i) Show that (28.85) has at least one solution.
(ii) Use the projection-gradient algorithm to solve (28.85). Check that all the assumptions in Corollary 28.10 are satisfied; provide an algorithm and a weak convergence result.

Exercise 28.6 Suppose that $\mathcal{H}$ is infinite-dimensional and separable, let $(e_k)_{k\in\mathbb{N}}$ be an orthonormal basis of $\mathcal{H}$, and let $(\phi_k)_{k\in\mathbb{N}}$ be a family of functions in $\Gamma_0(\mathbb{R})$ such that $(\forall k \in \mathbb{N})\ \phi_k \geqslant \phi_k(0) = 0$. Moreover, let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K}) \smallsetminus \{0\}$, let $r \in \mathcal{K}$, and assume that the problem

$$\underset{x\in\mathcal{H}}{\text{minimize}} \quad \sum_{k\in\mathbb{N}} \phi_k\big(\langle x \mid e_k\rangle\big) + \frac{1}{2}\|Lx - r\|^2 \tag{28.86}$$

admits at least one solution.

(i) Use the Douglas–Rachford algorithm to solve (28.86). Check that all the assumptions in Corollary 28.3 are satisfied; provide an algorithm and a weak convergence result.
(ii) Use the forward-backward algorithm to solve (28.86). Check that all the assumptions in Corollary 28.9 are satisfied; provide an algorithm and a weak convergence result.

Exercise 28.7 Suppose that $(e_k)_{k\in\mathbb{N}}$ is an orthonormal basis of $\mathcal{H}$ and let $(\phi_k)_{k\in\mathbb{N}}$ be a family of functions in $\Gamma_0(\mathbb{R})$ such that $(\forall k \in \mathbb{N})\ \phi_k \geqslant \phi_k(0) = 0$. Moreover, let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K}) \smallsetminus \{0\}$, let $\psi \in \Gamma_0(\mathcal{K})$, suppose that $r \in \operatorname{sri} L(\operatorname{dom} \varphi)$, and consider the problem

$$\underset{Lx=r}{\text{minimize}} \quad \sum_{k\in\mathbb{N}} \Big(\phi_k\big(\langle x \mid e_k\rangle\big) + |\langle x \mid e_k\rangle|^2\Big). \tag{28.87}$$

(i) Show that (28.87) admits exactly one solution.
(ii) Use the primal-dual algorithm (28.58) to solve (28.87). Check that all the assumptions in Proposition 28.16 are satisfied; in addition, provide an algorithm and a strong convergence result.

Chapter 29
Projection Operators

Let C be a nonempty closed convex subset of $\mathcal{H}$. The projection $P_C x$ of a point $x \in \mathcal{H}$ onto C is characterized by (see Theorem 3.16)

$$P_C x \in C \quad \text{and} \quad (\forall y \in C) \quad \langle y - P_C x \mid x - P_C x \rangle \leqslant 0 \tag{29.1}$$

or, equivalently, by (see Proposition 6.47) $x - P_C x \in N_C(P_C x)$. In this chapter, we investigate further the properties of projectors and provide a variety of examples.

29.1 Basic Properties

As a special case of proximity operator, projectors inherit their properties. Here are some examples.

Proposition 29.1 *Let C be a nonempty closed convex subset of $\mathcal{H}$ and let $x \in \mathcal{H}$. Then the following hold:*

(i) *Set $D = z + C$, where $z \in \mathcal{H}$. Then $P_D x = z + P_C(x - z)$.*
(ii) *Set $D = \rho C$, where $\rho \in \mathbb{R} \smallsetminus \{0\}$. Then $P_D x = \rho P_C(\rho^{-1} x)$.*
(iii) *Set $D = -C$. Then $P_D x = -P_C(-x)$.*

Proof. These properties are obtained by setting $f = \iota_C$ in the following results:
(i): Proposition 24.8(ii) (see also Proposition 3.19).
(ii): Proposition 24.8(v).
(iii): Proposition 24.8(vi). Alternatively, set $\rho = -1$ in (ii). ☐

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-48311-5_31

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_29

Proposition 29.2 *Let $\mathcal{K}$ be a real Hilbert space, let C be a nonempty closed convex subset of $\mathcal{K}$, let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, let $x \in \mathcal{H}$, and set $D = L^{-1}(C)$. Then the following hold:*

(i) *Suppose that $LL^* = \gamma \mathrm{Id}$ for some $\gamma \in \mathbb{R}_{++}$. Then $P_D x = x + \gamma^{-1} L^*(P_C(Lx) - Lx)$.*
(ii) *Suppose that L is invertible, with $L^{-1} = L^*$. Then $P_D x = L^{-1} P_C(Lx)$.*

Proof. (i): Set $f = \iota_C$ in Proposition 24.14.
(ii): This follows from (i). ☐

Proposition 29.3 *Let $(\mathcal{H}_i)_{i \in I}$ be a finite family of real Hilbert spaces, set $\boldsymbol{\mathcal{H}} = \bigoplus_{i \in I} \mathcal{H}_i$, and, for every $i \in I$, let C_i be a nonempty closed convex subset of $\mathcal{H}_i$ and let $x_i \in \mathcal{H}_i$. Set $\boldsymbol{x} = (x_i)_{i \in I}$ and $\boldsymbol{C} = \bigtimes_{i \in I} C_i$. Then $P_{\boldsymbol{C}} \boldsymbol{x} = (P_{C_i} x_i)_{i \in I}$.*

Proof. It is clear that $\boldsymbol{C}$ is a nonempty closed convex subset of $\boldsymbol{\mathcal{H}}$. The result therefore follows from Proposition 24.11, where $(\forall i \in I)\ f_i = \iota_{C_i}$. ☐

Proposition 29.4 *Let $(C_i)_{i \in I}$ be a family of closed intervals of $\mathbb{R}$ containing 0, suppose that $\mathcal{H} = \ell^2(I)$, set $\boldsymbol{C} = \mathcal{H} \cap \bigtimes_{i \in I} C_i$, and let $x = (\xi_i)_{i \in I} \in \mathcal{H}$. Then $P_{\boldsymbol{C}}$ is weakly sequentially continuous and $P_{\boldsymbol{C}} x = (P_{C_i} \xi_i)_{i \in I}$.*

Proof. Apply Proposition 24.12 with $(\forall i \in I)\ \phi_i = \iota_{C_i}$. ☐

Proposition 29.5 *Let C and D be nonempty closed convex subsets of $\mathcal{H}$, let $x \in \mathcal{H}$, and suppose that $P_C x \in D$. Then $P_{C \cap D} x = P_C x$.*

Proof. On the one hand, $P_C x \in C \cap D$. On the other hand, for every $y \in C \cap D \subset C$, (29.1) yields $\langle y - P_C x \mid x - P_C x \rangle \leqslant 0$. Altogether, Theorem 3.16 yields $P_{C \cap D} x = P_C x$. ☐

Proposition 29.6 *Let C and D be nonempty closed convex subsets of $\mathcal{H}$ such that $C \perp D$. Then $C + D$ is closed and convex, and $P_{C+D} = P_C + P_D$.*

Proof. It is clear that $C+D$ is convex. To see that $C+D$ is closed, let $(x_n)_{n \in \mathbb{N}}$ and $(y_n)_{n \in \mathbb{N}}$ be sequences in C and D, respectively, such that $(x_n + y_n)_{n \in \mathbb{N}}$ converges. The assumption $C \perp D$ implies that, for every m and every n in $\mathbb{N}$, we have $\|(x_n+y_n)-(x_m+y_m)\|^2 = \|x_n-x_m\|^2+\|y_n-y_m\|^2$. Since $(x_n+y_n)_{n \in \mathbb{N}}$ is a Cauchy sequence, so are therefore $(x_n)_{n \in \mathbb{N}}$ and $(y_n)_{n \in \mathbb{N}}$. Thus, $(x_n)_{n \in \mathbb{N}}$ and $(y_n)_{n \in \mathbb{N}}$ converge, which implies that $\lim(x_n + y_n) \in C + D$. Now let $z \in \mathcal{H}$, $x \in C$, and $y \in D$. Since $\{x - P_C z, -P_C z\} \perp \{y - P_D z, -P_D z\}$, we deduce from (29.1) that

$$
\begin{aligned}
&\langle x + y - P_C z - P_D z \mid z - P_C z - P_D z \rangle \\
&\qquad = \langle x - P_C z \mid z - P_C z \rangle + \langle x - P_C z \mid -P_D z \rangle \\
&\qquad\quad + \langle y - P_D z \mid z - P_D z \rangle + \langle y - P_D z \mid -P_C z \rangle \\
&\qquad = \langle x - P_C z \mid z - P_C z \rangle + \langle y - P_D z \mid z - P_D z \rangle \\
&\qquad \leqslant 0 \qquad (29.2)
\end{aligned}
$$

and hence that $P_{C+D}z = P_C z + P_D z$. □

Next, we discuss some asymptotic properties.

Proposition 29.7 *Let $(C_n)_{n\in\mathbb{N}}$ be a sequence of nonempty closed convex subsets of $\mathcal{H}$ such that $(\forall n \in \mathbb{N})$ $C_n \subset C_{n+1}$. Set $C = \overline{\bigcup_{n\in\mathbb{N}} C_n}$ and let $x \in \mathcal{H}$. Then $P_{C_n}x \to P_C x$.*

Proof. It follows from Proposition 3.45(i) that C is a nonempty closed convex set. By assumption, there exists a sequence $(y_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ such that $y_n \to P_C x$ and $(\forall n \in \mathbb{N})$ $y_n \in C_n$. In turn, $(\forall n \in \mathbb{N})$ $\|x-P_C x\| \leqslant \|x-P_{C_n}x\| \leqslant \|x-y_n\|$. Hence,

$$\|x - P_{C_n}x\| \to \|x - P_C x\|. \tag{29.3}$$

Therefore, every weak sequential cluster point z of $(P_{C_n}x)_{n\in\mathbb{N}}$ lies in C by Theorem 3.34 and since, by Lemma 2.42 and (29.3), $\|x - P_C x\| \leqslant \|x - z\| \leqslant \underline{\lim} \|x - P_{C_n}x\|$, it satisfies $\|x - z\| = \|x - P_C x\|$. It follows that $P_C x$ is the only weak sequential cluster point of $(P_{C_n}x)_{n\in\mathbb{N}}$ and, in turn, that $x - P_{C_n}x \rightharpoonup x - P_C x$ by Lemma 2.46. In view of (29.3) and Corollary 2.52, we conclude that $x - P_{C_n}x \to x - P_C x$. □

Proposition 29.8 *Let $(C_n)_{n\in\mathbb{N}}$ be a sequence of closed convex subsets of $\mathcal{H}$ such that $C = \bigcap_{n\in\mathbb{N}} C_n \neq \varnothing$ and $(\forall n \in \mathbb{N})$ $C_{n+1} \subset C_n$, and let $x \in \mathcal{H}$. Then $P_{C_n}x \to P_C x$.*

Proof. Set $(\forall n \in \mathbb{N})$ $p_n = P_{C_n}x$. Then $(\forall n \in \mathbb{N})$ $\|x - p_n\| \leqslant \|x - p_{n+1}\| \leqslant \|x-P_C x\|$. Hence, $(p_n)_{n\in\mathbb{N}}$ is bounded and $(\|x-p_n\|)_{n\in\mathbb{N}}$ converges. For every m and n in $\mathbb{N}$ such that $m \leqslant n$, since $(p_m + p_n)/2 \in C_m$, Lemma 2.12(iv) yields

$$\begin{aligned}\|p_n - p_m\|^2 &= 2\big(\|p_n - x\|^2 + \|p_m - x\|^2\big) - 4\,\|(p_n + p_m)/2 - x\|^2\\ &\leqslant 2\big(\|p_n - x\|^2 - \|p_m - x\|^2\big)\\ &\to 0 \quad \text{as} \quad \min\{m, n\} \to +\infty. \end{aligned}\tag{29.4}$$

Hence, $(p_n)_{n\in\mathbb{N}}$ is a Cauchy sequence and, in turn, $p_n \to p$ for some $p \in \mathcal{H}$. For every $n \in \mathbb{N}$, $(p_k)_{k\geqslant n}$ lies in C_n and hence $p \in C_n$, since C_n is closed. Thus, $p \in C$ and therefore $\|x-P_C x\| \leqslant \|x-p\| = \lim \|x-p_n\| \leqslant \|x-P_C x\|$. Since C is a Chebyshev set, we deduce that $p = P_C x$. □

Remark 29.9 As seen in Proposition 3.20, the condition $\bigcap_{n\in\mathbb{N}} C_n \neq \varnothing$ in Proposition 29.8 holds if C_0 is bounded.

The proof of the following result is left as Exercise 29.7.

Proposition 29.10 *Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\varepsilon \in \mathbb{R}_{++}$, let $x \in \mathcal{H}$, and set $D = C + B(0; \varepsilon)$. Then D is a nonempty closed convex subset of $\mathcal{H}$ and*

$$P_D x = \begin{cases} x, & \text{if } \|x - P_C x\| \leqslant \varepsilon;\\ P_C x + \varepsilon \dfrac{x - P_C x}{\|x - P_C x\|}, & \text{otherwise.}\end{cases}\tag{29.5}$$

Proposition 29.11 *Let C be a nonempty closed convex subset of $\mathcal{H}$, let $x \in \mathcal{H}$, and let $\gamma \in \left]1, +\infty\right[$. Then $\|P_C(\gamma x)\| \leqslant \gamma \|P_C x\|$.*

Proof. Since the conclusion is clear when $P_C x = P_C(\gamma x)$, we assume that

$$u = P_C(\gamma x) - P_C x \neq 0. \tag{29.6}$$

It follows from (3.10), (29.6), and again (3.10) that

$$\begin{aligned}\langle x \mid P_C(\gamma x) - P_C x\rangle &\leqslant \langle P_C x \mid P_C(\gamma x) - P_C x\rangle \\ &< \langle P_C(\gamma x) \mid P_C(\gamma x) - P_C x\rangle \\ &\leqslant \langle \gamma x \mid P_C(\gamma x) - P_C x\rangle \\ &= \gamma \langle x \mid P_C(\gamma x) - P_C x\rangle .\end{aligned} \tag{29.7}$$

Hence, $\langle x \mid P_C(\gamma x) - P_C x\rangle < \gamma \langle x \mid P_C(\gamma x) - P_C x\rangle$, which implies that

$$\langle x \mid P_C(\gamma x) - P_C x\rangle > 0. \tag{29.8}$$

We derive from (29.8) and (29.7) the following facts: $x \neq 0$, $P_C x \neq 0$, $P_C(\gamma x) \neq 0$, and

$$\delta = \frac{\langle P_C(\gamma x) \mid u\rangle}{\langle P_C x \mid u\rangle} \in \left]1, \gamma\right]. \tag{29.9}$$

Now set

$$H = \big\{ z \in \mathcal{H} \bigm| \langle z \mid u\rangle = \langle P_C(\gamma x) \mid u\rangle \big\}. \tag{29.10}$$

Then $\delta P_C x \in H$ and (3.16) implies that $P_H P_C x = P_C(\gamma x)$. Hence, $(\delta - 1)\|P_C x\| = \|P_C x - \delta P_C x\| \geqslant d_H(P_C x) = \|P_C x - P_C(\gamma x)\|$ and therefore

$$\gamma \geqslant \delta = \frac{(\delta - 1)\|P_C x\| + \|P_C x\|}{\|P_C x\|} \geqslant \frac{\|P_C(\gamma x) - P_C x\| + \|P_C x\|}{\|P_C x\|} \geqslant \frac{\|P_C(\gamma x)\|}{\|P_C x\|}, \tag{29.11}$$

as announced. ☐

Corollary 29.12 *Let C be a nonempty closed convex subset of $\mathcal{H}$, let $x \in \mathcal{H}$, and let $y \in \mathcal{H}$. Then $\phi \colon \mathbb{R}_{++} \to \mathbb{R}_+ \colon \alpha \mapsto \alpha^{-1}\|P_C(y + \alpha x) - y\|$ is decreasing.*

Proof. Take α and β in $\mathbb{R}_{++}$ such that $\alpha < \beta$, set $\gamma = \beta/\alpha$, and set $D = -y + C$. Then it follows from Proposition 29.1(i) and Proposition 29.11 that $\phi(\beta) = \beta^{-1}\|P_D(\beta x)\| = \beta^{-1}\|P_D(\gamma\alpha x)\| \leqslant \beta^{-1}\gamma\|P_D(\alpha x)\| = \phi(\alpha)$. ☐

Proposition 29.13 *Suppose that $\mathcal{H} = \mathbb{R}^N$, set $I = \{1, \ldots, N\}$, and let C be a nonempty closed convex subset of $\mathcal{H}$ such that $(\forall (\xi_i)_{i\in I} \in C)(\forall i \in I)$ $(\xi_1, \ldots, \xi_{i-1}, 0, \xi_{i+1}, \ldots, \xi_N) \in C$. Let $x = (\xi_i)_{i\in I} \in \mathcal{H}$ and set $p = (\pi_i)_{i\in I} = P_C x$. Then $(\forall i \in I)$ $\xi_i \pi_i \geqslant 0$.*

Proof. Let $i \in I$ and set $y = (\pi_1, \ldots, \pi_{i-1}, 0, \pi_{i+1}, \ldots, \pi_N)$. Then $y \in C$ and so $0 \leqslant \|x - y\|^2 - \|x - p\|^2 = (\xi_i - 0)^2 - (\xi_i - \pi_i)^2 = -\pi_i^2 + 2\pi_i\xi_i \leqslant 2\pi_i\xi_i$. ☐

29.2 Affine Subspaces

The following result complements Corollary 3.22.

Proposition 29.14 *Let C be a closed affine subspace of $\mathcal{H}$, let x and y be in $\mathcal{H}$, and let $z \in C$. Then the following hold:*

(i) *P_C is a weakly continuous affine operator.*
(ii) *$x - P_C x \perp C - C$.*
(iii) *$\|x - y\|^2 = \|P_C x - P_C y\|^2 + \|(\mathrm{Id} - P_C)x - (\mathrm{Id} - P_C)y\|^2$.*
(iv) *$\|(2P_C - \mathrm{Id})x - (2P_C - \mathrm{Id})y\| = \|x - y\|$.*
(v) *$\|P_C x - P_C y\|^2 = \langle x - y \mid P_C x - P_C y\rangle$.*
(vi) *$\|(\mathrm{Id} - P_C)x - (\mathrm{Id} - P_C)y\|^2 = \langle x - y \mid (\mathrm{Id} - P_C)x - (\mathrm{Id} - P_C)y\rangle$.*
(vii) *$\|x - P_C x\|^2 = \langle x - z \mid x - P_C x\rangle$.*

Proof. (i): This is a consequence of Proposition 4.16, Corollary 3.22(ii), and Lemma 2.41.

(ii): This follows from Corollary 3.22(i).

(iii): (ii) yields $\langle P_C x - P_C y \mid (\mathrm{Id} - P_C)x - (\mathrm{Id} - P_C)y\rangle = 0$. Hence, the conclusion follows from Lemma 2.17(ii) applied to $(x - y, P_C x - P_C y)$.

(iv)&(v): Combine (iii) with Lemma 2.17(ii) applied to $(x-y, P_C x - P_C y)$.

(vi): Combine (iii) with Lemma 2.17(ii) applied to $(x - y, (\mathrm{Id} - P_C)x - (\mathrm{Id} - P_C)y)$.

(vii): This follows from (vi). ☐

Proposition 29.15 *Suppose that $\{e_i\}_{i\in I}$ is a countable orthonormal subset of $\mathcal{H}$ and set $C = \overline{\mathrm{span}}\,\{e_i\}_{i\in I}$. Then*

$$(\forall x \in \mathcal{H}) \quad P_C x = \sum_{i\in I} \langle x \mid e_i\rangle\, e_i. \tag{29.12}$$

In particular, if $u \in \mathcal{H}$ and $\|u\| = 1$, then the projection of $x \in \mathcal{H}$ onto the line $U = \mathbb{R}\,u$ is given by $P_U x = \langle x \mid u\rangle\, u$.

Proof. If I is finite, then $C = \mathrm{span}\{e_i\}_{i\in I}$ and hence Example 3.10 yields the result. If I is countably infinite, we combine the finite case with Proposition 29.7 to obtain the conclusion. ☐

Proposition 29.16 *Let I be a nonempty finite set, let $(\omega_i)_{i\in I}$ be real numbers in $]0, 1]$ such that $\sum_{i\in I} \omega_i = 1$, and let $\boldsymbol{\mathcal{H}}$ be the real Hilbert space obtained by endowing the Cartesian product $\times_{i\in I} \mathcal{H}$ with the usual vector space structure and with the scalar product $(\boldsymbol{x}, \boldsymbol{y}) \mapsto \sum_{i\in I} \omega_i \langle x_i \mid y_i\rangle$, where $\boldsymbol{x} = (x_i)_{i\in I}$ and $\boldsymbol{y} = (y_i)_{i\in I}$. Set $\boldsymbol{D} = \big\{(x)_{i\in I} \;\big|\; x \in \mathcal{H}\big\}$, let $\boldsymbol{x} \in \boldsymbol{\mathcal{H}}$, and set $p = \sum_{i\in I} \omega_i x_i$. Then $P_{\boldsymbol{D}}\boldsymbol{x} = (p)_{i\in I}$.*

Proof. Set $\boldsymbol{p} = (p)_{i\in I}$, let $y \in \mathcal{H}$, and set $\boldsymbol{y} = (y)_{i\in I}$. It is clear that $\boldsymbol{p} \in \boldsymbol{D}$, that $\boldsymbol{y} \in \boldsymbol{D}$, and that $\boldsymbol{D}$ is a closed linear subspace of $\boldsymbol{\mathcal{H}}$. Furthermore, $\langle \boldsymbol{x} - \boldsymbol{p} \mid \boldsymbol{y}\rangle = \sum_{i\in I} \omega_i \langle x_i - p \mid y\rangle = \big\langle \sum_{i\in I} \omega_i x_i - p \mid y\big\rangle = 0$. Hence, we derive from Corollary 3.24(i) that $\boldsymbol{p} = P_{\boldsymbol{D}}\boldsymbol{x}$. ☐

Below, we provide examples of projectors onto affine subspaces.

Example 29.17 Let $\mathcal{K}$ be a real Hilbert space, let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$, let $y \in \operatorname{ran} L$, set $C = \big\{x \in \mathcal{H} \;\big|\; Lx = y\big\}$, and let $x \in \mathcal{H}$. Then the following hold:

(i) Suppose that $\operatorname{ran} L$ is closed. Then $(\forall z \in C)\ P_C x = x - L^\dagger L(x - z)$.
(ii) Suppose that $\operatorname{ran} L$ is closed. Then $P_C x = x - L^\dagger(Lx - y)$.
(iii) Suppose that LL^* is invertible. Then $P_C x = x - L^*(LL^*)^{-1}(Lx - y)$.
(iv) Suppose that L^*L is invertible. Then $C = \{(L^*L)^{-1}L^*y\}$.

Proof. In (iii)&(iv), $\operatorname{ran} LL^* = \mathcal{K}$ is closed and so is therefore $\operatorname{ran} L$ by Fact 2.26. Hence, in all cases, Proposition 3.30(ii) yields

$$LL^\dagger y = P_{\operatorname{ran} L}\, y = y. \tag{29.13}$$

(i): Take $z \in C$. Then $C = z + \ker L$ and it follows from Proposition 3.19 and Corollary 3.32(iv) that $P_C x = x - L^\dagger L(x - z)$.

(ii): The identity is obtained by setting $z = L^\dagger y$ in (i) and using (29.13).

(iii): It follows from (ii), (29.13), Corollary 3.32(iii), and Example 3.29 that $P_C x = x - L^\dagger L(x - L^\dagger y) = x - L^* L^{*\dagger}(x - L^\dagger y) = x - L^*(LL^*)^{-1}L(x - L^\dagger y) = x - L^*(LL^*)^{-1}(Lx - y)$.

(iv): Since $L^\dagger y \in C$ by (29.13), we derive from Example 3.29 and Fact 2.25(vi) that $C = L^\dagger y + \ker L = (L^*L)^{-1}L^*y + \{0\}$. ☐

Example 29.18 Suppose that u is a nonzero vector in $\mathcal{H}$, let $\eta \in \mathbb{R}$, and set $C = \big\{x \in \mathcal{H} \;\big|\; \langle x \mid u\rangle = \eta\big\}$. Then

$$(\forall x \in \mathcal{H}) \quad P_C x = x + \frac{\eta - \langle x \mid u\rangle}{\|u\|^2}\, u. \tag{29.14}$$

Proof. Set $\phi = \iota_{\{\eta\}}$ in Corollary 24.15, or set $L = \langle \cdot \mid u\rangle$ and $y = \eta$ in Example 29.17(iii), or see Example 3.23. ☐

Example 29.19 Let $\mathcal{K}$ be a real Hilbert space, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{K})$. Set $\boldsymbol{\mathcal{H}} = \mathcal{H} \oplus \mathcal{K}$ and $\boldsymbol{V} = \big\{(x, y) \in \boldsymbol{\mathcal{H}} \;\big|\; Lx = y\big\}$. Then, for every $(x, y) \in \boldsymbol{\mathcal{H}}$, the following hold:

(i) $P_{\boldsymbol{V}}(x, y) = \big(x - L^*(\mathrm{Id} + LL^*)^{-1}(Lx - y), y + (\mathrm{Id} + LL^*)^{-1}(Lx - y)\big)$.
(ii) $P_{\boldsymbol{V}}(x, y) = \big((\mathrm{Id} + L^*L)^{-1}(x + L^*y), L(\mathrm{Id} + L^*L)^{-1}(x + L^*y)\big)$.
(iii) $P_{\boldsymbol{V}^\perp}(x, y) = \big(L^*(\mathrm{Id} + LL^*)^{-1}(Lx - y), -(\mathrm{Id} + LL^*)^{-1}(Lx - y)\big)$.
(iv) $P_{\boldsymbol{V}^\perp}(x, y) = \big(x - (\mathrm{Id} + L^*L)^{-1}(x + L^*y), y - L(\mathrm{Id} + L^*L)^{-1}(x + L^*y)\big)$.

Proof. Set $\boldsymbol{T}\colon \boldsymbol{\mathcal{H}} \to \mathcal{K}\colon (x,y) \mapsto Lx - y$. Then $\boldsymbol{T} \in \mathcal{B}(\boldsymbol{\mathcal{H}},\mathcal{K})$, $\boldsymbol{T}^*\colon \mathcal{K} \to \boldsymbol{\mathcal{H}}\colon v \mapsto (L^*v, -v)$, $\boldsymbol{T}\boldsymbol{T}^* = \mathrm{Id} + LL^*$ and $\mathrm{Id} + L^*L$ are invertible by Example 21.3, and $\boldsymbol{V} = \ker \boldsymbol{T}$ is a closed linear subspace of $\boldsymbol{\mathcal{H}}$. Let us fix $(x,y) \in \boldsymbol{\mathcal{H}}$.

(i): Since $\boldsymbol{V} = \ker \boldsymbol{T}$, it follows from Example 29.17(iii) that

$$\begin{aligned} P_{\boldsymbol{V}}(x,y) &= (x,y) - \boldsymbol{T}^*(\boldsymbol{T}\boldsymbol{T}^*)^{-1}\boldsymbol{T}(x,y) \\ &= \big(x - L^*(\mathrm{Id} + LL^*)^{-1}(Lx-y), y + (\mathrm{Id}+LL^*)^{-1}(Lx-y)\big). \end{aligned} \tag{29.15}$$

(ii): On the one hand,

$$\begin{aligned} &(\mathrm{Id} + L^*L)\big(x - L^*(\mathrm{Id}+LL^*)^{-1}(Lx-y) - (\mathrm{Id}+L^*L)^{-1}(x+L^*y)\big) \\ &= x + L^*Lx - L^*(\mathrm{Id}+LL^*)^{-1}(Lx-y) \\ &\quad - L^*LL^*(\mathrm{Id}+LL^*)^{-1}(Lx-y) - x - L^*y \\ &= L^*(Lx-y) - L^*(\mathrm{Id}+LL^*)^{-1}(Lx-y) - L^*LL^*(\mathrm{Id}+LL^*)^{-1}(Lx-y) \\ &= L^*\big(\mathrm{Id}+LL^* - \mathrm{Id} - LL^*\big)(\mathrm{Id}+LL^*)^{-1}(Lx-y) \\ &= 0. \end{aligned} \tag{29.16}$$

Therefore, $x - L^*(\mathrm{Id}+LL^*)^{-1}(Lx-y) = (\mathrm{Id}+L^*L)^{-1}(x+L^*y)$. On the other hand,

$$\begin{aligned} &(\mathrm{Id}+LL^*)\big(y + (\mathrm{Id}+LL^*)^{-1}(Lx-y) - L(\mathrm{Id}+L^*L)^{-1}(x+L^*y)\big) \\ &= y + LL^*y + Lx - y - L(\mathrm{Id}+L^*L)^{-1}(x+L^*y) \\ &\quad - LL^*L(\mathrm{Id}+L^*L)^{-1}(x+L^*y) \\ &= L(x+L^*y) - L(\mathrm{Id}+L^*L)^{-1}(x+L^*y) - LL^*L(\mathrm{Id}+L^*L)^{-1}(x+L^*y) \\ &= \big(L(\mathrm{Id}+L^*L) - L - LL^*L\big)(\mathrm{Id}+L^*L)^{-1}(x+L^*y) \\ &= 0. \end{aligned} \tag{29.17}$$

and hence $y + (\mathrm{Id}+LL^*)^{-1}(Lx-y) = L(\mathrm{Id}+L^*L)^{-1}(x+L^*y)$. Altogether, (ii) follows from (i).

Finally, since $P_{\boldsymbol{V}^\perp} = \mathrm{Id} - P_{\boldsymbol{V}}$ (Corollary 3.24(v)), (iii) and (iv) follow from (i) and (ii), respectively. ☐

29.3 Projections onto Special Polyhedra

In the next two examples, we provide closed-form expressions for the projectors onto a half-space and onto a hyperslab, respectively.

Example 29.20 Let $u \in \mathcal{H}$, let $\eta \in \mathbb{R}$, and set $C = \big\{x \in \mathcal{H} \;\big|\; \langle x \mid u\rangle \leqslant \eta\big\}$. Then exactly one of the following holds:

(i) $u = 0$ and $\eta \geqslant 0$, in which case $C = \mathcal{H}$ and $P_C = \mathrm{Id}$.

(ii) $u = 0$ and $\eta < 0$, in which case $C = \varnothing$.
(iii) $u \neq 0$, in which case $C \neq \varnothing$ and

$$(\forall x \in \mathcal{H}) \quad P_C x = \begin{cases} x, & \text{if } \langle x \mid u \rangle \leqslant \eta; \\ x + \dfrac{\eta - \langle x \mid u \rangle}{\|u\|^2} u, & \text{if } \langle x \mid u \rangle > \eta. \end{cases} \tag{29.18}$$

Proof. (i)&(ii): Clear.
(iii): Set $\phi = \iota_{]-\infty,\eta]}$ in Corollary 24.15. □

Example 29.21 Let $u \in \mathcal{H}$, let η_1 and η_2 be in $\mathbb{R}$, and set

$$C = \big\{ x \in \mathcal{H} \bigm| \eta_1 \leqslant \langle x \mid u \rangle \leqslant \eta_2 \big\}. \tag{29.19}$$

Then exactly one of the following holds:

(i) $u = 0$ and $\eta_1 \leqslant 0 \leqslant \eta_2$, in which case $C = \mathcal{H}$ and $P_C = \mathrm{Id}$.
(ii) $u = 0$ and $[\ \eta_1 > 0$ or $\eta_2 < 0\ ]$, in which case $C = \varnothing$.
(iii) $u \neq 0$ and $\eta_1 > \eta_2$, in which case $C = \varnothing$.
(iv) $u \neq 0$ and $\eta_1 \leqslant \eta_2$, in which case $C \neq \varnothing$ and

$$(\forall x \in \mathcal{H}) \quad P_C x = \begin{cases} x + \dfrac{\eta_1 - \langle x \mid u \rangle}{\|u\|^2} u, & \text{if } \langle x \mid u \rangle < \eta_1; \\ x, & \text{if } \eta_1 \leqslant \langle x \mid u \rangle \leqslant \eta_2; \\ x + \dfrac{\eta_2 - \langle x \mid u \rangle}{\|u\|^2} u, & \text{if } \langle x \mid u \rangle > \eta_2. \end{cases} \tag{29.20}$$

Proof. (i)–(iii): Clear.
(iv): Set $\phi = \iota_{[\eta_1,\eta_2]}$ in Corollary 24.15. □

Next, we investigate the projector onto the intersection of two half-spaces with parallel boundaries.

Proposition 29.22 *Let u_1 and u_2 be in $\mathcal{H}$, and let η_1 and η_2 be in $\mathbb{R}$. Suppose that $\|u_1\|^2\|u_2\|^2 = |\langle u_1 \mid u_2 \rangle|^2$, i.e., $\{u_1, u_2\}$ is linearly dependent, and set*

$$C = \big\{ x \in \mathcal{H} \bigm| \langle x \mid u_1 \rangle \leqslant \eta_1 \big\} \cap \big\{ x \in \mathcal{H} \bigm| \langle x \mid u_2 \rangle \leqslant \eta_2 \big\}. \tag{29.21}$$

Then exactly one of the following cases occurs:

(i) *$u_1 = u_2 = 0$ and $0 \leqslant \min\{\eta_1, \eta_2\}$. Then $C = \mathcal{H}$ and $P_C = \mathrm{Id}$.*
(ii) *$u_1 = u_2 = 0$ and $\min\{\eta_1, \eta_2\} < 0$. Then $C = \varnothing$.*
(iii) *$u_1 \neq 0$, $u_2 = 0$, and $0 \leqslant \eta_2$. Then $C = \big\{ x \in \mathcal{H} \bigm| \langle x \mid u_1 \rangle \leqslant \eta_1 \big\}$ and*

$$(\forall x \in \mathcal{H}) \quad P_C x = \begin{cases} x, & \text{if } \langle x \mid u_1 \rangle \leqslant \eta_1; \\ x + \dfrac{\eta_1 - \langle x \mid u_1 \rangle}{\|u_1\|^2} u_1, & \text{if } \langle x \mid u_1 \rangle > \eta_1. \end{cases} \tag{29.22}$$

(iv) $u_1 \neq 0$, $u_2 = 0$, *and* $\eta_2 < 0$. *Then* $C = \varnothing$.

(v) $u_1 = 0$, $u_2 \neq 0$, *and* $0 \leqslant \eta_1$. *Then* $C = \big\{x \in \mathcal{H} \;\big|\; \langle x \mid u_2\rangle \leqslant \eta_2\big\}$ *and*

$$(\forall x \in \mathcal{H}) \quad P_C x = \begin{cases} x, & \text{if } \langle x \mid u_2\rangle \leqslant \eta_2; \\ x + \dfrac{\eta_2 - \langle x \mid u_2\rangle}{\|u_2\|^2}\, u_2, & \text{if } \langle x \mid u_2\rangle > \eta_2. \end{cases} \tag{29.23}$$

(vi) $u_1 = 0$, $u_2 \neq 0$, *and* $\eta_1 < 0$. *Then* $C = \varnothing$.

(vii) $u_1 \neq 0$, $u_2 \neq 0$, *and* $\langle u_1 \mid u_2\rangle > 0$. *Then* $C = \big\{x \in \mathcal{H} \;\big|\; \langle x \mid u\rangle \leqslant \eta\big\}$, *where* $u = \|u_2\|\, u_1$ *and* $\eta = \min\{\eta_1\|u_2\|, \eta_2\|u_1\|\}$, *and*

$$(\forall x \in \mathcal{H}) \quad P_C x = \begin{cases} x, & \text{if } \langle x \mid u\rangle \leqslant \eta; \\ x + \dfrac{\eta - \langle x \mid u\rangle}{\|u\|^2}\, u, & \text{if } \langle x \mid u\rangle > \eta. \end{cases} \tag{29.24}$$

(viii) $u_1 \neq 0$, $u_2 \neq 0$, $\langle u_1 \mid u_2\rangle < 0$, *and* $\eta_1\|u_2\| + \eta_2\|u_1\| < 0$. *Then* $C = \varnothing$.

(ix) $u_1 \neq 0$, $u_2 \neq 0$, $\langle u_1 \mid u_2\rangle < 0$, *and* $\eta_1\|u_2\| + \eta_2\|u_1\| \geqslant 0$. *Then* $C = \big\{x \in \mathcal{H} \;\big|\; \gamma_1 \leqslant \langle x \mid u\rangle \leqslant \gamma_2\big\} \neq \varnothing$, *where*

$$u = \|u_2\|\, u_1, \;\; \gamma_1 = -\eta_2\|u_1\|, \quad \text{and} \quad \gamma_2 = \eta_1\|u_2\|, \tag{29.25}$$

and

$$(\forall x \in \mathcal{H}) \quad P_C x = \begin{cases} x - \dfrac{\langle x \mid u\rangle - \gamma_1}{\|u\|^2}\, u, & \text{if } \langle x \mid u\rangle < \gamma_1; \\ x, & \text{if } \gamma_1 \leqslant \langle u \mid x\rangle \leqslant \gamma_2; \\ x - \dfrac{\langle x \mid u\rangle - \gamma_2}{\|u\|^2}\, u, & \text{if } \langle x \mid u\rangle > \gamma_2. \end{cases} \tag{29.26}$$

Proof. Let $x \in \mathcal{H}$.

(i), (ii), (iv), &(vi): Clear.

(iii)&(v): This follows from Example 29.20(iii).

(vii): Since u_1 and u_2 are linearly dependent and $\langle u_1 \mid u_2\rangle > 0$, we have $\|u_2\|u_1 = \|u_1\|u_2$. Now set $u = \|u_2\|u_1$. We have $\langle x \mid u_1\rangle \leqslant \eta_1 \Leftrightarrow \|u_2\|\,\langle x \mid u_1\rangle \leqslant \eta_1\|u_2\| \Leftrightarrow \langle x \mid u\rangle \leqslant \eta_1\|u_2\|$ and, similarly, $\langle x \mid u_2\rangle \leqslant \eta_2 \Leftrightarrow \langle x \mid u\rangle \leqslant \eta_2\|u_1\|$. Altogether, $x \in C \Leftrightarrow \langle x \mid u\rangle \leqslant \eta$ and the formula for P_C follows from Example 29.20(iii).

(viii)&(ix): Set $u = \|u_2\|u_1 = -\|u_1\|u_2$, $\gamma_1 = -\eta_2\|u_1\|$, and $\gamma_2 = \eta_1\|u_2\|$. Then $\langle x \mid u_1\rangle \leqslant \eta_1 \Leftrightarrow \|u_2\|\,\langle x \mid u_1\rangle \leqslant \eta_1\|u_2\| \Leftrightarrow \langle x \mid u\rangle \leqslant \gamma_2$, and $\langle x \mid u_2\rangle \leqslant \eta_2 \Leftrightarrow \|u_1\|\,\langle x \mid u_2\rangle \leqslant \eta_2\|u_1\| \Leftrightarrow -\eta_2\|u_1\| \leqslant \langle x \mid -\|u_1\|\, u_2\rangle \Leftrightarrow \gamma_1 \leqslant \langle x \mid u\rangle$. Altogether, $x \in C \Leftrightarrow \gamma_1 \leqslant \langle x \mid u\rangle \leqslant \gamma_2$, and the results follow from Example 29.21(iii)&(iv). ☐

Proposition 29.23 *Let u_1 and u_2 be in $\mathcal{H}$, and let η_1 and η_2 be in $\mathbb{R}$. Suppose that $\|u_1\|^2\|u_2\|^2 > |\langle u_1 \mid u_2\rangle|^2$, i.e., $\{u_1, u_2\}$ is linearly independent, set*

$$C = \big\{x \in \mathcal{H} \;\big|\; \langle x \mid u_1\rangle \leqslant \eta_1\big\} \cap \big\{x \in \mathcal{H} \;\big|\; \langle x \mid u_2\rangle \leqslant \eta_2\big\}, \tag{29.27}$$

and let $x \in \mathcal{H}$. Then $C \neq \varnothing$ and

$$P_C x = x - \nu_1 u_1 - \nu_2 u_2, \tag{29.28}$$

where exactly one of the following holds:

(i) $\langle x \mid u_1\rangle \leqslant \eta_1$ *and* $\langle x \mid u_2\rangle \leqslant \eta_2$. *Then* $\nu_1 = \nu_2 = 0$.
(ii) $\|u_2\|^2(\langle x \mid u_1\rangle - \eta_1) > \langle u_1 \mid u_2\rangle\,(\langle x \mid u_2\rangle - \eta_2)$ *and* $\|u_1\|^2(\langle x \mid u_2\rangle - \eta_2) > \langle u_1 \mid u_2\rangle\,(\langle x \mid u_1\rangle - \eta_1)$. *Then*

$$\nu_1 = \frac{\|u_2\|^2\big(\langle x \mid u_1\rangle - \eta_1\big) - \langle u_1 \mid u_2\rangle\,\big(\langle x \mid u_2\rangle - \eta_2\big)}{\|u_1\|^2\,\|u_2\|^2 - |\langle u_1 \mid u_2\rangle|^2} > 0 \tag{29.29}$$

and

$$\nu_2 = \frac{\|u_1\|^2\big(\langle x \mid u_2\rangle - \eta_2\big) - \langle u_1 \mid u_2\rangle\,\big(\langle x \mid u_1\rangle - \eta_1\big)}{\|u_1\|^2\,\|u_2\|^2 - |\langle u_1 \mid u_2\rangle|^2} > 0. \tag{29.30}$$

(iii) $\langle x \mid u_2\rangle > \eta_2$ *and* $\|u_2\|^2(\langle x \mid u_1\rangle - \eta_1) \leqslant \langle u_1 \mid u_2\rangle\,(\langle x \mid u_2\rangle - \eta_2)$. *Then*

$$\nu_1 = 0 \quad \textit{and} \quad \nu_2 = \frac{\langle x \mid u_2\rangle - \eta_2}{\|u_2\|^2} > 0. \tag{29.31}$$

(iv) $\langle x \mid u_1\rangle > \eta_1$ *and* $\|u_1\|^2(\langle x \mid u_2\rangle - \eta_2) \leqslant \langle u_1 \mid u_2\rangle\,(\langle x \mid u_1\rangle - \eta_1)$. *Then*

$$\nu_1 = \frac{\langle x \mid u_1\rangle - \eta_1}{\|u_1\|^2} > 0 \quad \textit{and} \quad \nu_2 = 0. \tag{29.32}$$

Proof. Set $L\colon \mathcal{H} \to \mathbb{R}^2 \colon y \mapsto (\langle y \mid u_1\rangle, \langle y \mid u_2\rangle)$ and

$$G = \begin{bmatrix} \|u_1\|^2 & \langle u_1 \mid u_2\rangle \\ \langle u_1 \mid u_2\rangle & \|u_2\|^2 \end{bmatrix}. \tag{29.33}$$

Since, by assumption, $\det G \neq 0$, G is invertible. Hence, $\operatorname{ran} L = \operatorname{ran} LL^* = \operatorname{ran} G = \mathbb{R}^2$ and, therefore, $C \neq \varnothing$. Proposition 27.21 applied to the objective function $(1/2)\|\cdot - x\|^2$ asserts the existence of ν_1 and ν_2 in $\mathbb{R}_+$ such that

$$P_C x = x - \nu_1 u_1 - \nu_2 u_2, \tag{29.34}$$

and such that the feasibility conditions

$$\begin{cases} \langle x \mid u_1\rangle - \nu_1\|u_1\|^2 - \nu_2\,\langle u_2 \mid u_1\rangle \leqslant \eta_1, \\ \langle x \mid u_2\rangle - \nu_1\,\langle u_1 \mid u_2\rangle - \nu_2\|u_2\|^2 \leqslant \eta_2, \end{cases} \tag{29.35}$$

hold, as well as the complementary slackness conditions

$$\nu_1\big(\langle x \mid u_1\rangle - \nu_1\|u_1\|^2 - \nu_2\langle u_2 \mid u_1\rangle - \eta_1\big) = 0 \tag{29.36}$$

and

$$\nu_2\big(\langle x \mid u_2\rangle - \nu_1\langle u_1 \mid u_2\rangle - \nu_2\|u_2\|^2 - \eta_2\big) = 0. \tag{29.37}$$

The linear independence of $\{u_1, u_2\}$ guarantees the uniqueness of (ν_1, ν_2). This leads to four conceivable cases.

(a) $\nu_1 = \nu_2 = 0$: Then (29.34) yields $x = P_C x \in C$. This verifies (i).

(b) $\nu_1 > 0$ and $\nu_2 > 0$: In view of (29.33), (29.36) and (29.37) force $G\,\big[\nu_1\ \nu_2\big]^{\mathsf{T}} = Lx - \big[\eta_1\ \eta_2\big]^{\mathsf{T}}$, which yields (ii).

(c) $\nu_1 = 0$ and $\nu_2 > 0$: Condition (29.37) forces $\langle x \mid u_2\rangle - \nu_2\|u_2\|^2 = \eta_2$, which yields the formula for ν_2 as well as the equivalence $\nu_2 > 0 \Leftrightarrow \langle x \mid u_2\rangle > \eta_2$. In turn, (29.35) reduces to $\langle x \mid u_1\rangle - \eta_1 \leqslant \nu_2\langle u_1 \mid u_2\rangle$, which yields $\|u_2\|^2(\langle x \mid u_1\rangle - \eta_1) \leqslant \langle u_1 \mid u_2\rangle(\langle x \mid u_2\rangle - \eta_2)$, and $\langle x \mid u_2\rangle - \eta_2 = \nu_2\|u_2\|^2$. This verifies (iii).

(d) $\nu_1 > 0$ and $\nu_2 = 0$: This is analogous to (c) and yields (iv). ☐

The following notation will be convenient.

Definition 29.24 Set

$$H\colon \mathcal{H}\times\mathcal{H} \to 2^{\mathcal{H}}\colon (x,y) \mapsto \big\{z \in \mathcal{H} \;\big|\; \langle z - y \mid x - y\rangle \leqslant 0\big\} \tag{29.38}$$

and

$$\begin{aligned}
Q\colon\ &\mathcal{H}\times\mathcal{H}\times\mathcal{H} \to \mathcal{H}\\
&(x,y,z) \mapsto \begin{cases} z, & \text{if } \rho = 0 \text{ and } \chi \geqslant 0;\\[2ex] x + \Big(1 + \dfrac{\chi}{\nu}\Big)(z - y), & \text{if } \rho > 0 \text{ and } \chi\nu \geqslant \rho;\\[2ex] y + \dfrac{\nu}{\rho}\Big(\chi(x-y) + \mu(z-y)\Big), & \text{if } \rho > 0 \text{ and } \chi\nu < \rho,\end{cases}\\
&\qquad\qquad\qquad\text{where} \begin{cases} \chi = \langle x - y \mid y - z\rangle,\\ \mu = \|x-y\|^2,\\ \nu = \|y-z\|^2,\\ \rho = \mu\nu - \chi^2.\end{cases}
\end{aligned} \tag{29.39}$$

Corollary 29.25 (Haugazeau) *Let $(x,y,z) \in \mathcal{H}^3$ and set*

$$C = H(x,y) \cap H(y,z). \tag{29.40}$$

Moreover, set $\chi = \langle x - y \mid y - z\rangle$, $\mu = \|x-y\|^2$, $\nu = \|y-z\|^2$, and $\rho = \mu\nu - \chi^2$. Then exactly one of the following holds:

(i) *$\rho = 0$ and $\chi < 0$, in which case $C = \varnothing$.*

(ii) $[\,\rho = 0$ *and* $\chi \geqslant 0\,]$ *or* $\rho > 0$, *in which case* $C \neq \varnothing$ *and*

$$P_C x = Q(x, y, z). \tag{29.41}$$

Proof. We first observe that, by Cauchy–Schwarz, $\rho \geqslant 0$. Now set $u_1 = x - y$, $u_2 = y - z$, $\eta_1 = \langle y \mid u_1 \rangle$, and $\eta_2 = \langle z \mid u_2 \rangle$. Then $H(x, y) = \big\{ c \in \mathcal{H} \;\big|\; \langle c \mid u_1 \rangle \leqslant \eta_1 \big\}$, and $H(y, z) = \big\{ c \in \mathcal{H} \;\big|\; \langle c \mid u_2 \rangle \leqslant \eta_2 \big\}$.

(i): We have $\|u_1\|^2 \|u_2\|^2 = \langle u_1 \mid u_2 \rangle^2$ and $\langle u_1 \mid u_2 \rangle < 0$. Hence, $\|u_2\| u_1 = -\|u_1\| u_2$ and, in turn,

$$\begin{aligned}
\eta_1 \|u_2\| + \eta_2 \|u_1\| & \\
&= \langle y \mid x - y \rangle \, \|u_2\| + \langle z \mid y - z \rangle \, \|u_1\| \\
&= \langle y \mid u_1 \rangle \, \|u_2\| + \big(\langle z - y \mid y - z \rangle + \langle y \mid y - z \rangle \big) \|u_1\| \\
&= \langle y \mid u_1 \rangle \, \|u_2\| + \big(-\|u_2\|^2 + \langle y \mid u_2 \rangle \big) \|u_1\| \\
&= \langle y \mid \|u_2\| u_1 + \|u_1\| u_2 \rangle - \|u_1\| \, \|u_2\|^2 \\
&= \langle y \mid 0 \rangle - \|u_1\| \, \|u_2\|^2 \\
&< 0.
\end{aligned} \tag{29.42}$$

We therefore deduce from Proposition 29.22(viii) that $C = \varnothing$.

(ii): We verify the formula in each of the following three cases:

(a) $\rho = 0$ and $\chi \geqslant 0$: In view of (29.39), we must show that $P_C x = z$. We consider four subcases.

(a.1) $x = y = z$: Then $C = \mathcal{H}$ and therefore $P_C x = x = z$.

(a.2) $x \neq y = z$: Then $C = H(x, y)$ and it follows from (29.1) that $P_C x = y = z$.

(a.3) $x = y \neq z$: Then $C = H(y, z)$ and it follows from (29.1) that $P_C x = P_C y = z$.

(a.4) $x \neq y$ and $y \neq z$: Then $\{u_1, u_2\}$ is linearly dependent and $\chi = \langle u_1 \mid u_2 \rangle \geqslant 0$. Hence, $\chi > 0$. Now set $u = \|u_2\| u_1 = \|u_1\| u_2$ and $\eta = \min\{\eta_1 \|u_2\|, \eta_2 \|u_1\|\}$. We have

$$\langle x - z \mid u \rangle = \langle u_1 \mid u \rangle + \langle u_2 \mid u \rangle = \|u_1\|^2 \, \|u_2\| + \|u_1\| \, \|u_2\|^2 > 0. \tag{29.43}$$

On the other hand, $\eta_1 \|u_2\| - \eta_2 \|u_1\| = \langle y \mid u_1 \rangle \, \|u_2\| - \langle z \mid u_2 \rangle \, \|u_1\| = \langle y - z \mid u \rangle = \langle u_2 \mid u \rangle = \|u_1\| \, \|u_2\|^2 > 0$. Thus, $\eta = \eta_2 \|u_1\| = \langle z \mid u_2 \rangle \, \|u_1\| = \langle z \mid u \rangle$, and (29.43) yields $\langle x \mid u \rangle > \langle z \mid u \rangle = \eta$. Hence, by Proposition 29.22(vii) and (29.43),

$$\begin{aligned}
P_C x &= x + \frac{\eta - \langle x \mid u \rangle}{\|u\|^2} \, u \\
&= x - \frac{\langle x - z \mid u \rangle}{\|u\|^2} \, u
\end{aligned}$$

$$\begin{aligned}
&= x - \frac{\|u_1\|^2\,\|u_2\| + \|u_1\|\,\|u_2\|^2}{\|u_1\|^2\|u_2\|^2}\,\|u_1\|u_2 \\
&= x - \frac{\|u_1\|\,u_2 + \|u_2\|\,u_2}{\|u_2\|} \\
&= x - u_1 - u_2 \\
&= z. \qquad (29.44)
\end{aligned}$$

(b) $\rho > 0$ and $\chi\nu \geqslant \rho$: In view of (29.39), we must show that $P_C x = x + (1+\chi/\nu)(z-y)$. We have $\langle u_1 \mid u_2\rangle\,\|u_2\|^2 \geqslant \|u_1\|^2\|u_2\|^2 - \langle u_1 \mid u_2\rangle^2 > 0$, i.e.,

$$\langle u_1 \mid u_2\rangle \big(\|u_2\|^2 + \langle u_1 \mid u_2\rangle\big) \geqslant \|u_1\|^2\|u_2\|^2 > \langle u_1 \mid u_2\rangle^2 . \qquad (29.45)$$

Since $\chi = \langle u_1 \mid u_2\rangle > 0$, we have

$$\begin{aligned}
\langle x \mid u_2\rangle - \eta_2 &= \langle x - z \mid u_2\rangle \\
&= \langle u_1 + u_2 \mid u_2\rangle \\
&= \langle u_1 \mid u_2\rangle + \|u_2\|^2 \qquad (29.46) \\
&> 0. \qquad (29.47)
\end{aligned}$$

On the other hand, using (29.46) and (29.45), we obtain

$$\begin{aligned}
\langle u_1 \mid u_2\rangle \big(\langle x \mid u_2\rangle - \eta_2\big) &= \langle u_1 \mid u_2\rangle \big(\langle u_1 \mid u_2\rangle + \|u_2\|^2\big) \\
&\geqslant \|u_1\|^2\|u_2\|^2 \\
&= \|u_2\|^2\big(\langle x \mid u_1\rangle - \eta_1\big). \qquad (29.48)
\end{aligned}$$

Altogether, (29.47), (29.48), Proposition 29.23(iii), and (29.46) yield

$$\begin{aligned}
P_C x &= x - \frac{\langle x \mid u_2\rangle - \eta_2}{\|u_2\|^2}\,u_2 \\
&= x + \frac{\|u_2\|^2 + \langle u_1 \mid u_2\rangle}{\|u_2\|^2}\,(-u_2) \\
&= x + (1+\chi/\nu)(z-y). \qquad (29.49)
\end{aligned}$$

(c) $\rho > 0$ and $\chi\nu < \rho$: In view of (29.39), we must show that $P_C x = y + (\nu/\rho)(\chi(x-y) + \mu(z-y))$. We have

$$\begin{aligned}
\|u_2\|^2\big(\langle x \mid u_1\rangle - \eta_1\big) &= \|u_2\|^2\|u_1\|^2 \\
&= \rho + \chi^2 \qquad (29.50) \\
&> \chi(\nu+\chi) \\
&= \langle u_1 \mid u_2\rangle\,\langle u_1 + u_2 \mid u_2\rangle \\
&= \langle u_1 \mid u_2\rangle\,\langle x - z \mid u_2\rangle \\
&= \langle u_1 \mid u_2\rangle\,(\langle x \mid u_2\rangle - \eta_2). \qquad (29.51)
\end{aligned}$$

Next, by (29.46),

$$\begin{aligned}\|u_1\|^2\big(\langle x \mid u_2\rangle - \eta_2\big) &= \|u_1\|^2\big(\langle u_1 \mid u_2\rangle + \|u_2\|^2\big)\\ &> \|u_1\|^2 \langle u_1 \mid u_2\rangle\\ &= \big(\langle x \mid u_1\rangle - \eta_1\big)\langle u_1 \mid u_2\rangle. \end{aligned} \tag{29.52}$$

Altogether, it follows from (29.50), (29.51), (29.52), and Proposition 29.23(ii) that

$$\begin{aligned}P_C x &= x - \left(1 - \frac{\chi\nu}{\rho}\right)(x-y) - \frac{\nu\mu}{\rho}(y-z)\\ &= y + \frac{\nu}{\rho}\big(\chi(x-y) + \mu(z-y)\big),\end{aligned} \tag{29.53}$$

which concludes the proof. ☐

Proposition 29.26 *Let m be a strictly positive integer, set $I = \{1,\ldots,m\}$, let $(x_i)_{i\in I}$ be a family of vectors in $\mathcal{H}$, let $x \in \mathcal{H}$, and set $C = \operatorname{conv}\{x_i\}_{i\in I}$. Then $P_C x = \sum_{i\in I} \bar{\alpha}_i x_i$, where $(\bar{\alpha}_i)_{i\in I}$ is a solution to the problem*

$$\underset{\substack{(\alpha_i)_{i\in I}\in\mathbb{R}^m_+\\ \sum_{i\in I}\alpha_i = 1}}{\text{minimize}} \quad \sum_{i\in I}\sum_{j\in I}\alpha_i\alpha_j\langle x_i \mid x_j\rangle - 2\sum_{i\in I}\alpha_i\langle x_i \mid x\rangle. \tag{29.54}$$

Proof. Set $S = \big\{(\alpha_i)_{i\in I}\in\mathbb{R}^m_+ \;\big|\; \sum_{i\in I}\alpha_i = 1\big\}$ and $L\colon \mathbb{R}^m \to \mathcal{H}\colon (\alpha_i)_{i\in I} \mapsto \sum_{i\in I}\alpha_i x_i$. Then $L(S) = C$ and $d_C^2(x) = \min_{a\in S}\|x - La\|^2 = \min_{a\in S}(\|x\|^2 + \|La\|^2 - 2\langle La \mid x\rangle)$. ☐

We now obtain a formula for the projector of the probability simplex.

Example 29.27 (Probability simplex) Suppose that $\mathcal{H} = \mathbb{R}^N$, set $I = \{1,\ldots,N\}$ and $f\colon \mathcal{H} \to \mathbb{R}\colon (\xi_i)_{i\in I} \mapsto \max\{\xi_i\}_{i\in I}$, and let

$$\Delta = \left\{(\xi_i)_{i\in I}\in\mathbb{R}^N_+ \;\middle|\; \sum_{i\in I}\xi_i = 1\right\} \tag{29.55}$$

be the probability simplex. Then $P_\Delta = \operatorname{Id} - \operatorname{Prox}_f$ (see Example 24.25 for a formula for Prox_f and also Example 29.34).

Proof. After verifying that $\iota_\Delta^* = f$, we deduce from (14.6) that $P_\Delta = \operatorname{Prox}_{\iota_\Delta} = \operatorname{Id} - \operatorname{Prox}_{\iota_\Delta^*} = \operatorname{Id} - \operatorname{Prox}_f$. ☐

Example 29.28 (ℓ^1 unit ball) Suppose that $\mathcal{H} = \mathbb{R}^N$, and set $I = \{1,\ldots,N\}$ and

$$C = \big\{x\in\mathcal{H} \;\big|\; \|x\|_1 \leqslant 1\big\}. \tag{29.56}$$

Let $x = (\xi_i)_{i\in I} \in \mathcal{H}\smallsetminus C$, and set $y = (|\xi_i|)_{i\in I}$ and $P_\Delta y = (\pi_i)_{i\in I}$, where Δ is the probability simplex of (29.55). Then

$$P_C x = \big(\operatorname{sign}(\xi_i)\pi_i\big)_{i\in I}. \tag{29.57}$$

Proof. Since $x \notin C$, it follows from Proposition 11.5 that $P_C x \notin \operatorname{int} C$. Therefore $\|P_C x\|_1 = 1$. Now set $P_C x = (\chi_i)_{i\in I}$. First, suppose that $x \in \mathbb{R}^N_+$. By Proposition 29.13, $P_C x \in \mathbb{R}^N_+$ and therefore $P_C x = P_{C\cap\mathbb{R}^N_+} x$. Thus $\sum_{i\in I} \chi_i = \|P_C x\|_1 = 1$ and hence $P_C x \in \Delta$. Therefore, $P_C x = P_\Delta\, x$, as claimed. Finally, if $x \notin \mathbb{R}^N_+$, then there exists a diagonal matrix $L \in \mathbb{R}^{N\times N}$ with diagonal entries in $\{-1, 1\}$ such that $Lx = y$ and $L^{-1} = L^{\mathsf{T}} = L$. Since $L(C) = C$, the result follows from Proposition 29.2(ii). ☐

29.4 Projections Involving Convex Cones

Projections onto convex cones were discussed in Proposition 6.28 and in Theorem 6.30. Here are further properties.

Proposition 29.29 *Let K be a nonempty closed convex cone in $\mathcal{H}$, let $x \in \mathcal{H}$, and let $\rho \in \mathbb{R}_+$. Then $P_K(\rho x) = \rho P_K x$.*

Proof. See Exercise 29.9. ☐

Example 29.30 Let I be a nonempty set, suppose that $\mathcal{H} = \ell^2(I)$, and set $K = \ell^2_+(I)$. Then P_K is weakly sequentially continuous and, for every $x = (\xi_i)_{i\in I} \in \mathcal{H}$, $P_K x = (\max\{\xi_i, 0\})_{i\in I}$.

Proof. This is a consequence of Proposition 29.4 (see also Example 6.29). ☐

Example 29.31 Suppose that $u \in \mathcal{H} \smallsetminus \{0\}$, and let $K = \mathbb{R}_+ u$ be the associated ray. Then

$$(\forall x \in \mathcal{H}) \quad P_K x = \begin{cases} \dfrac{\langle x \mid u\rangle}{\|u\|^2} u, & \text{if } \ \langle x \mid u\rangle > 0;\\ 0, & \text{otherwise.} \end{cases} \tag{29.58}$$

Proof. Since $K^\ominus = \big\{x \in \mathcal{H} \ \big|\ \langle x \mid u\rangle \leqslant 0\big\}$, the result follows from Theorem 6.30(i) and Example 29.20(iii). ☐

Example 29.32 Suppose that $\mathcal{H} = \mathbb{S}^N$, let $K = \mathbb{S}^N_+$ be the closed convex cone of symmetric positive semidefinite matrices, and let $A \in \mathcal{H}$. Then there exist a diagonal matrix $\Lambda \in \mathbb{R}^{N\times N}$ and $U \in \mathbb{U}^{N\times N}$ such that $A = U\Lambda U^{\mathsf{T}}$ and $P_K A = U\Lambda_+ U^{\mathsf{T}}$, where Λ_+ is the diagonal matrix obtained from Λ by setting the negative entries to 0.

Proof. Set $A_+ = U\Lambda_+ U^{\mathsf{T}}$ and $\Lambda_- = \Lambda - \Lambda_+$. Then $A_+ \in K$ and $A - A_+ = U\Lambda_- U^{\mathsf{T}} \in -K = K^\ominus$ by Example 6.26. Furthermore, $\langle A_+ \mid A - A_+\rangle = \operatorname{tra}(A_+(A - A_+)) = \operatorname{tra}(U\Lambda_+ U^{\mathsf{T}} U\Lambda_- U^{\mathsf{T}}) = \operatorname{tra}(U\Lambda_+\Lambda_- U^{\mathsf{T}}) = \operatorname{tra}(\Lambda_+\Lambda_-) = 0$. Altogether, the result follows from Proposition 6.28. ☐

In the remainder of this section, we focus on the problem of projecting onto the intersection of a cone and a hyperplane.

Proposition 29.33 *Let K be a nonempty closed convex cone in $\mathcal{H}$, suppose that $u \in K$ satisfies $\|u\| = 1$, let $\eta \in \mathbb{R}_{++}$, set $C = K \cap \big\{x \in \mathcal{H} \;\big|\; \langle x \mid u\rangle = \eta\big\}$, and let $x \in \mathcal{H}$. Then $C \neq \varnothing$ and*

$$P_C x = P_K(\bar{\nu} u + x), \tag{29.59}$$

where $\bar{\nu} \in \mathbb{R}$ is a solution to the equation $\langle P_K(\bar{\nu}u + x) \mid u\rangle = \eta$.

Proof. The problem of finding $P_C x$ is a special case of the primal problem (28.71) in Example 28.19, in which $\mathcal{K} = \mathbb{R}$, $L = \langle \cdot \mid u\rangle$, and $r = \eta$. Since $\eta \in \operatorname{int} \langle K \mid u\rangle$, we derive from Example 28.19 that the dual problem (28.72), which becomes, via Theorem 6.30(iii),

$$\underset{\nu \in \mathbb{R}}{\text{minimize}} \;\; \phi(\nu), \quad \text{where} \quad \phi\colon \nu \mapsto \tfrac{1}{2} d^2_{K^\ominus}(\nu u + x) - \nu\eta, \tag{29.60}$$

has at least one solution $\bar{\nu}$. Furthermore, $\bar{\nu}$ is characterized by $\phi'(\bar{\nu}) = 0$, i.e., using Proposition 12.32, by $\langle P_K(\bar{\nu}u + x) \mid u\rangle = \eta$, and the primal solution is $P_C x = P_K(\bar{\nu}u + x)$. ☐

Here is an application of Proposition 29.33 to projecting onto a simplex; for an alternative approach, combine Example 29.27 and Proposition 29.1(ii).

Example 29.34 Let N be a strictly positive integer, suppose that $\mathcal{H} = \mathbb{R}^N$, let $\beta \in \mathbb{R}_{++}$, set $I = \{1, \ldots, N\}$, set $C = \big\{(\xi_i)_{i\in I} \in \mathbb{R}^N_+ \;\big|\; \sum_{i\in I} \xi_i = \beta\big\}$, let $x = (\xi_i)_{i\in I} \in \mathcal{H}$, and set

$$\phi\colon \mathbb{R} \to \mathbb{R}\colon t \mapsto \sum_{i\in I} \max\{\xi_i + t, 0\}. \tag{29.61}$$

Then ϕ is continuous and increasing, and there exists $s \in \mathbb{R}$ such that $\phi(s) = \beta$. Furthermore, s is unique, $P_C x = (\max\{\xi_i + s, 0\})_{i\in I}$, and $s \in \left]-\xi, \beta - \xi\right]$, where $\xi = \max_{i\in I} \xi_i$. (Thus, s may be found easily by the bisection method.)

Proof. Set $K = \mathbb{R}^N_+$, $u = (1, \ldots, 1)/\sqrt{N} \in \mathbb{R}^N$, and $\eta = \beta/\sqrt{N}$. Then $\|u\| = 1$ and $C = K \cap \big\{x \in \mathcal{H} \;\big|\; \langle x \mid u\rangle = \eta\big\}$. For every $i \in I$, the function $t \mapsto \max\{\xi_i + t, 0\}$ is increasing. Using Example 29.30, we deduce that $\phi\colon \mathbb{R} \to \mathbb{R}\colon t \mapsto \sum_{i\in I} \max\{\xi_i + t, 0\} = \sqrt{N}\langle P_K(x + \sqrt{N} t u) \mid u\rangle$ is increasing, and, by Proposition 29.33, that the equation $\phi(\bar{\nu}/\sqrt{N}) = \beta$ has a solution. Thus, there exists $s \in \mathbb{R}$ such that $\phi(s) = \beta$. Since $\beta > 0$, there exists $j \in I$ such that $\xi_j + s > 0$. Hence, $t \mapsto \max\{\xi_j + t, 0\}$ is strictly increasing on some open interval U containing s, which implies that ϕ is strictly increasing on U as well. This shows the uniqueness of s. Finally, $\phi(-\xi) = \sum_{i\in I} \max\{\xi_i - \xi, 0\} = \sum_{i\in I} 0 = 0 < \beta \leqslant \sum_{i\in I} \max\{\xi_i - \xi + \beta, 0\} = \phi(-\xi + \beta)$. ☐

29.5 Projections onto Epigraphs and Lower Level Sets

Proposition 29.35 *Let $f\colon \mathcal{H} \to \mathbb{R}$ be convex and continuous, set $C = \operatorname{epi} f$, and let $(z,\zeta) \in (\mathcal{H} \times \mathbb{R}) \smallsetminus C$. Then the inclusion $z \in x + (f(x) - \zeta)\partial f(x)$ has a unique solution, say $\bar{x}$, and $P_C(z,\zeta) = (\bar{x}, f(\bar{x}))$.*

Proof. Set $(\bar{x},\bar{\xi}) = P_C(z,\zeta)$. Then $(\bar{x},\bar{\xi})$ is the unique solution to the problem

$$\underset{\substack{(x,\xi)\in\mathcal{H}\oplus\mathbb{R}\\ f(x)+(-1)\xi\leqslant 0}}{\text{minimize}} \quad \tfrac{1}{2}\|x-z\|^2 + \tfrac{1}{2}|\xi-\zeta|^2. \tag{29.62}$$

In view of Proposition 27.21, $(\bar{x},\bar{\xi})$ is characterized by $f(\bar{x}) \leqslant \bar{\xi}$ and the existence of $\bar{\nu} \in \mathbb{R}_+$ such that $\bar{\nu}(f(\bar{x})-\bar{\xi}) = 0$, $z-\bar{x} \in \bar{\nu}\partial f(\bar{x})$, and $\zeta-\bar{\xi} = -\bar{\nu}$. If $\bar{\nu} = 0$, then $(\bar{x},\bar{\xi}) = (z,\zeta) \in C$, which is impossible. Hence $\bar{\nu} > 0$. We conclude that $\bar{\xi} = f(\bar{x})$ and therefore that $\bar{\nu} = f(\bar{x}) - \zeta$. ☐

Example 29.36 Let N be a strictly positive integer, suppose that $\mathcal{H} = \mathbb{R}^N$, set $I = \{1,\ldots,N\}$, set $f\colon \mathcal{H} \to \mathbb{R}\colon (\xi_i)_{i\in I} \mapsto (1/2)\sum_{i\in I}\alpha_i|\xi_i|^2$, where $(\alpha_i)_{i\in I} \in \mathbb{R}^N_{++}$. Suppose that $(z,\zeta) \in (\mathcal{H}\times\mathbb{R}) \smallsetminus (\operatorname{epi} f)$ and set $z = (\zeta_i)_{i\in I}$. By Proposition 29.35, the system of equations $\xi_i = \zeta_i - (f(x)-\zeta)\alpha_i\xi_i$, where $i \in I$, has a unique solution $\bar{x} = (\bar{\xi}_i)_{i\in I}$. Now set $\eta = f(\bar{x})$. Then

$$(\forall i \in I) \quad \bar{\xi}_i = \frac{\zeta_i}{(\eta-\zeta)\alpha_i + 1} \tag{29.63}$$

and hence

$$\eta = \tfrac{1}{2}\sum_{i\in I}\alpha_i\left(\frac{\zeta_i}{(\eta-\zeta)\alpha_i+1}\right)^2. \tag{29.64}$$

This one-dimensional equation can be solved numerically for η, and $\bar{x}$ can then be recovered from (29.63).

Proposition 29.37 *Let $f \in \Gamma_0(\mathcal{H})$ and let $z \in \mathcal{H}$. Suppose that $\zeta \in \left]\inf f(\mathcal{H}), f(z)\right[$, set $C = \operatorname{lev}_{\leqslant\zeta} f$, and suppose that $C \subset \operatorname{int}\operatorname{dom} f$. Then, in terms of the real variable ν, the equation $f(\operatorname{Prox}_{\nu f} z) = \zeta$ has at least one solution in $\mathbb{R}_{++}$ and, if $\bar{\nu}$ is such a solution, then $P_C z = \operatorname{Prox}_{\bar{\nu} f} z$.*

Proof. The vector $P_C z$ is the unique solution to the convex optimization problem

$$\underset{\substack{x\in\mathcal{H}\\ f(x)\leqslant\zeta}}{\text{minimize}} \ \tfrac{1}{2}\|x-z\|^2. \tag{29.65}$$

In view of Proposition 27.21, there exists $\bar{\nu} \in \mathbb{R}_+$ such that $f(P_C z) \leqslant \zeta$, $\bar{\nu}(f(P_C z) - \zeta) = 0$, and $z - P_C z \in \bar{\nu}\partial f(P_C z)$. Since $z \notin C$, we must have $\bar{\nu} > 0$. Hence, $f(P_C z) = \zeta$ and the result follows from (16.37). ☐

Example 29.38 Let C be a nonempty bounded closed convex subset of $\mathcal{H}$ and recall that $C^{\odot}$ denotes the polar set of C (see (7.7)). Suppose that $z \in \mathcal{H} \smallsetminus C^{\odot}$. Then, in terms of the real variable ν, the equation $\nu = \langle z - \nu P_C(z/\nu) \mid \nu P_C(z/\nu)\rangle$ has a unique solution $\bar{\nu} \in \mathbb{R}_{++}$. Moreover, $P_{C^{\odot}} z = z - \bar{\nu} P_C(z/\bar{\nu})$ and $\bar{\nu} = \langle P_{C^{\odot}} z \mid z - P_{C^{\odot}} z\rangle$.

Proof. Recall that $C^{\odot} = \operatorname{lev}_{\leqslant 1} \sigma_C = \operatorname{lev}_{\leqslant 1} \iota_C^*$ and that our assumptions imply that $\sigma_C(0) = 0 < 1 < \sigma_C(z)$. By Proposition 29.37, there exists $\bar{\nu} \in \mathbb{R}_{++}$ such that

$$P_{C^{\odot}} z = \operatorname{Prox}_{\bar{\nu}\iota_C^*} z \quad \text{and} \quad \iota_C^*(P_{C^{\odot}} z) = 1. \tag{29.66}$$

It follows from Theorem 14.3(ii) and Example 12.25 that $\operatorname{Prox}_{\bar{\nu}\iota_C^*} z = z - \bar{\nu} P_C(z/\bar{\nu})$, which provides the projection formula. In turn, we derive from (29.66) and Theorem 14.3(iii) that

$$\begin{aligned} 1 &= \iota_C^*(\operatorname{Prox}_{\bar{\nu}\iota_C^*} z) + \iota_C(P_C(z/\bar{\nu})) \\ &= \big\langle \operatorname{Prox}_{\bar{\nu}\iota_C^*} z \mid P_C(z/\bar{\nu})\big\rangle \\ &= \langle P_{C^{\odot}} z \mid z - P_{C^{\odot}} z\rangle / \bar{\nu}, \end{aligned} \tag{29.67}$$

which yields $\bar{\nu} = \langle P_{C^{\odot}} z \mid z - P_{C^{\odot}} z\rangle$. ☐

Example 29.39 Suppose that $\mathcal{H} = \mathbb{R}^N$, let $A \in \mathbb{R}^{N\times N}$ be symmetric and positive semidefinite, and let $u \in \mathbb{R}^N$. Set $f \colon \mathbb{R}^N \to \mathbb{R} \colon x \mapsto (1/2)\langle x \mid Ax\rangle + \langle x \mid u\rangle$, let $z \in \mathbb{R}^N$, suppose that $\zeta \in \left]\inf f(\mathcal{H}), f(z)\right[$, and set $C = \operatorname{lev}_{\leqslant \zeta} f$. Then $\bar{x} = P_C z$ is the unique solution to the system of equations $f(x) = \zeta$ and $x = (\mathrm{Id} + \bar{\nu} A)^{-1}(z - \bar{\nu} u)$, where $\bar{\nu} \in \mathbb{R}_{++}$.

Proof. This follows from Proposition 29.37. ☐

29.6 Subgradient Projectors

In this section we investigate a construct that provides an approximate projection onto a lower level set of a continuous convex function $f \colon \mathcal{H} \to \mathbb{R}$. Note that Proposition 16.17(ii) asserts that $\operatorname{dom} \partial f = \mathcal{H}$. Combining this with Fermat's rule (Theorem 16.3), we deduce that the following notion is well defined (see Figure 29.1 for an illustration).

Definition 29.40 Let $f \colon \mathcal{H} \to \mathbb{R}$ be a continuous convex function, let $\xi \in \mathbb{R}$ be such that $C = \operatorname{lev}_{\leqslant \xi} f \neq \varnothing$, and let s be a selection of ∂f. The *subgradient projector* onto C associated with (f, ξ, s) is

$$G \colon \mathcal{H} \to \mathcal{H} \colon x \mapsto \begin{cases} x + \dfrac{\xi - f(x)}{\|s(x)\|^2} s(x), & \text{if } f(x) > \xi; \\ x, & \text{if } f(x) \leqslant \xi. \end{cases} \tag{29.68}$$

The *subgradient projection* of $x \in \mathcal{H}$ onto C is Gx. If f is Gâteaux differentiable on $\mathcal{H} \smallsetminus C$, then the subgradient projector associated with (f, ξ) is

$$G\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \begin{cases} x + \dfrac{\xi - f(x)}{\|\nabla f(x)\|^2}\nabla f(x), & \text{if } f(x) > \xi;\\ x, & \text{if } f(x) \leqslant \xi. \end{cases} \tag{29.69}$$

We first collect some basic properties of subgradient projectors.

Proposition 29.41 *Let $f\colon \mathcal{H} \to \mathbb{R}$ be a continuous convex function, let $\xi \in \mathbb{R}$ be such that $C = \operatorname{lev}_{\leqslant \xi} f \neq \varnothing$, and let s be a selection of ∂f. Let G be the subgradient projector onto C associated with (f, ξ, s), let $x \in \mathcal{H}$, and set*

$$H = \big\{ y \in \mathcal{H} \;\big|\; \langle y - x \mid s(x)\rangle + f(x) \leqslant \xi \big\}. \tag{29.70}$$

Then the following hold:

(i) $\operatorname{Fix} G = C \subset H$.
(ii) $Gx = P_H x$.
(iii) *G is firmly quasinonexpansive.*
(iv) $\max\{f(x) - \xi, 0\} = \|s(x)\|\,\|Gx - x\|$.
(v) $\max\{f(x) - \xi, 0\}(x - Gx) = \|Gx - x\|^2\, s(x)$.
(vi) *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ and let $\bar{x}$ be a point in $\mathcal{H}$ such that $Gx_n - x_n \to 0$ and one of the following holds:*
   (a) $x_n \to \bar{x}$.
   (b) *$x_n \rightharpoonup \bar{x}$ and f is bounded on every bounded subset of $\mathcal{H}$.*

   Then $\bar{x} \in C$.
(vii) *Suppose that f is bounded on every bounded subset of $\mathcal{H}$. Then $\operatorname{Id} - G$ is demiclosed at 0.*
(viii) *G is continuous on C.*
(ix) *Suppose that f is Fréchet differentiable on $\mathcal{H} \smallsetminus C$. Then G is continuous.*

Proof. (i): The equality follows from (29.68). Now let $z \in C$. Since $s(x) \in \partial f(x)$, (16.1) yields $\langle z - x \mid s(x)\rangle + f(x) \leqslant f(z) \leqslant \xi$. Hence $z \in H$.

(ii): If $x \in C$, then (i) yields $Gx = x = P_H x$. If $x \notin C$, then $P_H x = Gx$ by Example 29.20(iii).

(iii): Let $z \in C$. We deduce from (i) that $z \in H$. In turn, (ii) and (3.10) imply that $\langle z - Gx \mid x - Gx\rangle = \langle z - P_H x \mid x - P_H x\rangle \leqslant 0$, and Proposition 4.2 yields the claim.

(iv): This follows from (29.68).

(v): This follows from (iv).

(vi): Using either Proposition 16.17(iii) in case (vi)(a) or Lemma 2.46 and Proposition 16.20 in case (vi)(b), we assume the existence of $\rho \in \mathbb{R}_{++}$ such that $(x_n)_{n\in\mathbb{N}}$ lies in $B(\bar{x}; \rho)$ and $\sigma = \sup \|\partial f(B(\bar{x}; \rho))\| < +\infty$. Hence, in both cases, we have

$$(\forall n \in \mathbb{N}) \quad \|s(x_n)\| \leqslant \sigma. \tag{29.71}$$

On the other hand, it follows from Theorem 9.1 that $\max\{f,\xi\}$ is weakly sequentially lower semicontinuous. Hence, by (iv),

$$\max\{f(\bar{x})-\xi,0\} \leqslant \underline{\lim}\max\{f(x_n)-\xi,0\} \leqslant \sigma\lim\|Gx_n-x_n\| = 0. \quad (29.72)$$

Thus, $f(\bar{x}) \leqslant \xi$.

(vii): See (vi)(b) and (i).

(viii): Let $\bar{x} \in C$ and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ that converges to $\bar{x}$. If $(x_{k_n})_{n\in\mathbb{N}}$ is a subsequence of $(x_n)_{n\in\mathbb{N}}$ lying in C, then $Gx_{k_n} = x_{k_n} \to \bar{x} = G\bar{x}$ by (i). Thus, we suppose that $(x_n)_{n\in\mathbb{N}}$ lies in $\mathcal{H}\smallsetminus C$. Then $(\forall n \in \mathbb{N})$ $f(x_n)-\xi \leqslant f(\bar{x})-\xi-\langle \bar{x}-x_n \mid s(x_n)\rangle \leqslant \|x_n-\bar{x}\|\,\|s(x_n)\|$. Hence $0 < (f(x_n)-\xi)/\|s(x_n)\| \leqslant \|x_n-\bar{x}\| \to 0$. In view of (29.68), $\|Gx_n-x_n\| \to 0$. Therefore, $Gx_n \to \bar{x} = G\bar{x}$ by (vi)(a) and (i).

(ix): By (viii), G is continuous on C. On the other hand, $\mathcal{H}\smallsetminus C$ is open, and G is continuous on this set by (29.69) and Corollary 17.43. $\square$

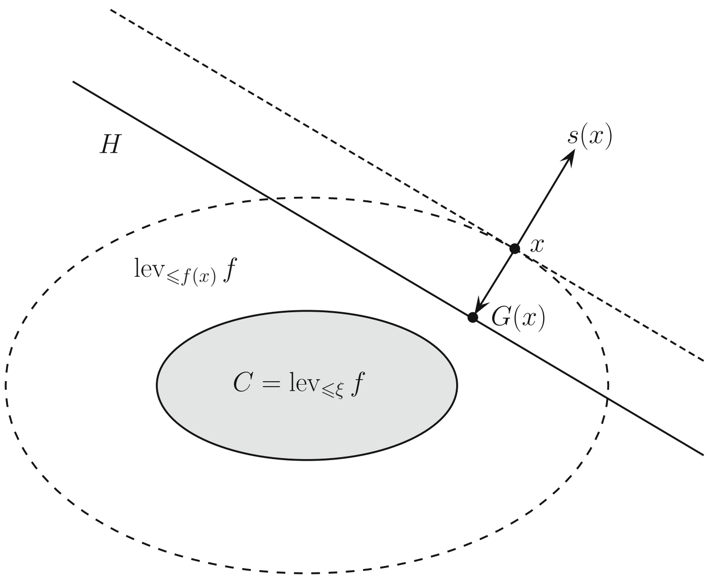

Fig. 29.1 Geometrical illustration of properties (i) and (ii) in Proposition 29.41. Let $x \in \mathcal{H}$ be such that $f(x) > \xi$. According to Lemma 27.20, $s(x)$ belongs to the normal cone to $\mathrm{lev}_{\leqslant f(x)} f$, and the subgradient projection $G(x)$ is the projection of x onto the half-space H of (29.70), which contains C.

Example 29.42 Let $h \in \Gamma_0(\mathcal{H})$ be such that $\min h(\mathcal{H}) = 0$, set $g = h \,\square\, (1/2)\|\cdot\|^2$, and let $\gamma \in \mathbb{R}_{++}$ be such that $f = g^\gamma$ is convex. Set $C = \operatorname{Argmin} h = \mathrm{lev}_{\leqslant 0} f$, and denote the subgradient projector onto C associated with $(f,0)$ by G. Then

$$G\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \begin{cases} x - \dfrac{g(x)}{\gamma}\dfrac{x - \operatorname{Prox}_h x}{\|x - \operatorname{Prox}_h x\|^2}, & \text{if } h(x) > 0; \\ x, & \text{if } h(x) = 0. \end{cases} \tag{29.73}$$

Proof. Combine Definition 12.23, (12.28), and Proposition 17.5. ☐

Example 29.43 Let C be a nonempty closed convex subset of $\mathcal{H}$, let $p \in [1, +\infty[$, and denote the subgradient projector onto C associated with $(d_C^p, 0)$ by G. Then

$$G = \Big(1 - \tfrac{1}{p}\Big)\operatorname{Id} + \tfrac{1}{p} P_C. \tag{29.74}$$

Proof. Apply Example 29.42 with $h = \iota_C$ and $\gamma = p/2$. ☐

The following example shows that subgradient projectors generalize classical projectors.

Example 29.44 Let C be a nonempty closed convex subset of $\mathcal{H}$, and denote the subgradient projector onto C associated with $(d_C, 0)$ by G. Then $G = P_C$.

Proof. Apply Example 29.43 with $p = 1$. Alternatively, observe that $C = \operatorname{lev}_{\leqslant 0} d_C$ and that d_C is Fréchet differentiable on $\mathcal{H} \smallsetminus C$ by Proposition 18.23(iii). Now let $x \in \mathcal{H}$. If $x \in C$, then $d_C(x) = 0$ and (29.69) yields $Gx = x = P_C x$. Next, suppose that $x \notin C$. Then $d_C(x) > 0$ and Proposition 18.23(iii) yields $\nabla d_C(x) = (x - P_C x)/d_C(x)$. Hence, it follows from (29.69) that $Gx = P_C x$. ☐

Example 29.45 Let $(C_i)_{i\in I}$ be a finite family of intersecting closed convex subsets of $\mathcal{H}$ with associated projectors $(P_i)_{i\in I}$, let $(\omega_i)_{i\in I}$ be a family in $]0, 1]$ such that $\sum_{i\in I} \omega_i = 1$, let $x \in \mathcal{H}$, set $f = (1/2)\sum_{i\in I} \omega_i d_{C_i}^2$, set $C = \operatorname{lev}_{\leqslant 0} f$, and denote the subgradient projector onto C associated with $(f, 0)$ by G. Then $C = \bigcap_{i\in I} C_i$ and

$$Gx = \begin{cases} x + \dfrac{\sum_{i\in I} \omega_i \|P_i x - x\|^2}{2\|\sum_{i\in I} \omega_i P_i x - x\|^2}\Big(\displaystyle\sum_{i\in I} \omega_i P_i x - x\Big), & \text{if } x \notin C; \\ x, & \text{if } x \in C. \end{cases} \tag{29.75}$$

Proof. This follows from (29.69) and Corollary 12.31. ☐

Example 29.46 Let $\mathcal{K}$ be a real Hilbert space, let $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, let $r \in \mathcal{K}$, let $\xi \in \mathbb{R}_{++}$, let $x \in \mathcal{H}$, suppose that $C = \big\{y \in \mathcal{H} \;\big|\; \|Ly - r\|^2 \leqslant \xi\big\} \neq \varnothing$, and denote the subgradient projector onto C associated with $y \mapsto (\|Ly - r\|^2, \xi)$ by G. Then

$$Gx = \begin{cases} x + \dfrac{\xi - \|Lx - r\|^2}{2\|L^*(Lx - r)\|^2} L^*(Lx - r), & \text{if } \|Lx - r\|^2 > \xi; \\ x, & \text{if } \|Lx - r\|^2 \leqslant \xi. \end{cases} \tag{29.76}$$

Proof. This follows from (29.69) and Example 2.60. ☐

In stark contrast to projection operators, which are firmly nonexpansive (see Proposition 4.16), subgradient projectors are not necessarily nonexpansive or even continuous, as demonstrated in the next example.

Example 29.47 Suppose that $\mathcal{H} = \mathbb{R}$, set $f\colon x \mapsto \max\{x+1, 2x+1\}$, let s be a selection of ∂f, and denote the subgradient projector onto $C = \operatorname{lev}_{\leqslant 0} f = \left]-\infty, -1\right]$ associated with $(f, 0, s)$ by G. Then $\partial f(0) = [1, 2]$ by Theorem 18.5, and (29.68) yields

$$(\forall x \in \mathcal{H}) \quad Gx = \begin{cases} x, & \text{if } \ x \leqslant -1; \\ -1, & \text{if } \ -1 < x < 0; \\ -1/s(0) \in [-1, -1/2], & \text{if } \ x = 0; \\ -1/2, & \text{if } \ x > 0. \end{cases} \tag{29.77}$$

Thus, G is discontinuous.

The next example illustrates the fact that a subgradient projector may fail to be weakly continuous.

Example 29.48 Suppose that $\mathcal{H} = \ell^2(I)$, where $I = \mathbb{N} \smallsetminus \{0\}$, and set

$$f\colon \mathcal{H} \to \mathbb{R}\colon (\xi_n)_{n\in I} \mapsto \sum_{n\in I} n\xi_n^{2n}. \tag{29.78}$$

By Example 16.22 and Exercise 17.20, f is convex, continuous, and Gâteaux differentiable with $\nabla f\colon \mathcal{H} \to \mathcal{H}\colon (\xi_i)_{i\in I} \mapsto (2n^2\xi_n^{2n-1})_{n\in I}$. Denote, for every $n \in I$, the nth standard unit vector in $\mathcal{H}$ by e_n, and set $(\forall n \in I)\ x_n = e_1 + e_n$. Then $x_n \rightharpoonup e_1$ and, for every $n \in I$ such that $n \geqslant 2$, we have $f(x_n) = 1 + n$ and $\nabla f(x_n) = 2e_1 + 2n^2 e_n$; thus, $f(x_n)/\|\nabla f(x_n)\| = (1+n)/\sqrt{4+4n^4} \to 0$. The subgradient projector G onto $C = \operatorname{lev}_{\leqslant 0} f = \{0\}$ associated with $(f, 0)$ therefore satisfies $\|x_n - Gx_n\| = f(x_n)/\|\nabla f(x_n)\| \to 0$. Furthermore, $G(e_1) = e_1 - (f(e_1)/\|\nabla f(e_1)\|^2)\nabla f(e_1) = \frac{1}{2}e_1$. To summarize, we have

$$x_n \rightharpoonup e_1,\ x_n - Gx_n \to 0,\ Gx_n \rightharpoonup e_1,\ Ge_1 = \tfrac{1}{2}e_1, \text{ and } \operatorname{Fix} G = \{0\}. \tag{29.79}$$

It follows that G is neither weakly continuous, nor nonexpansive in view of Corollary 4.28.

We conclude this section with subgradient projection algorithms. The first result provides a method for solving convex feasibility problems formulated in terms of convex inequalities.

Proposition 29.49 *Let $(f_i)_{i\in I}$ be a finite family of continuous convex functions from $\mathcal{H}$ to $\mathbb{R}$ such that $C = \bigcap_{i\in I} \operatorname{lev}_{\leqslant 0} f_i \neq \varnothing$ and suppose that one of the following holds:*

(i) *The functions* $(f_i)_{i\in I}$ *are bounded on every bounded subset of* $\mathcal{H}$.
(ii) *The functions* $(f_i^*)_{i\in I}$ *are supercoercive.*
(iii) $\mathcal{H}$ *is finite-dimensional.*

Let $\mathrm{i}\colon \mathbb{N} \to I$ *be such that there exists a strictly positive integer* m *for which*

$$(\forall i \in I)(\forall n \in \mathbb{N}) \quad i \in \{\mathrm{i}(n), \ldots, \mathrm{i}(n+m-1)\}, \tag{29.80}$$

and let $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $]0,2[$ *such that* $0 < \inf_{n\in\mathbb{N}} \lambda_n \leqslant \sup \lambda_n < 2$. *For every* $i \in I$, *fix a selection* s_i *of* ∂f_i *and denote by* G_i *the subgradient projector onto* $\operatorname{lev}_{\leqslant 0} f_i$ *associated with* $(f_i, 0, s_i)$. *Let* $x_0 \in \mathcal{H}$ *and set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n\big(G_{\mathrm{i}(n)}x_n - x_n\big). \tag{29.81}$$

Then $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* C.

Proof. It follows from Proposition 29.41(i)&(iii) and Proposition 5.13(ii)&(iv)–(v) that $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to C, that

$$\sum_{n\in\mathbb{N}} \|G_{\mathrm{i}(n)}x_n - x_n\|^2 < +\infty, \tag{29.82}$$

and that

$$\|x_{n+1} - x_n\| \to 0. \tag{29.83}$$

Now let x be a weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$. In view of Proposition 5.13(vi), it suffices to show that $(\forall i \in I)\ f_i(x) \leqslant 0$. Fix $i \in I$. It follows from (29.80) that there exist sequences $(k_n)_{n\in\mathbb{N}}$ and $(p_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$ such that $x_{k_n} \rightharpoonup x$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} k_n \leqslant p_n \leqslant k_n + m - 1 < k_{n+1} \leqslant p_{n+1}, \\ i = \mathrm{i}(p_n). \end{cases} \tag{29.84}$$

Hence, using (29.83),

$$\begin{aligned} \|x_{p_n} - x_{k_n}\| &\leqslant \sum_{l=k_n}^{k_n+m-2} \|x_{l+1} - x_l\| \\ &\leqslant (m-1) \max_{k_n \leqslant l \leqslant k_n+m-2} \|x_{l+1} - x_l\| \\ &\to 0, \end{aligned} \tag{29.85}$$

and therefore $x_{p_n} \rightharpoonup x$.

(i): It follows from (29.82) that $G_i x_{p_n} - x_{p_n} \to 0$. Hence, Proposition 29.41(vi)(b) implies that $f_i(x) \leqslant 0$.

(ii)&(iii): By Proposition 16.20, these follow from (i). ☐

As a corollary, we obtain Polyak's subgradient projection algorithm for constrained minimization.

Corollary 29.50 (Polyak's subgradient projection algorithm) *Let C be a closed convex subset of $\mathcal{H}$, let $f\colon \mathcal{H} \to \mathbb{R}$ be a continuous convex function such that* $\operatorname{Argmin}_C f \neq \varnothing$*, and suppose that one of the following holds:*

(i) *f is bounded on every bounded subset of $\mathcal{H}$.*
(ii) *f^* is supercoercive.*
(iii) *$\mathcal{H}$ is finite-dimensional.*

Let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence in $]0,2[$ such that $0 < \inf_{n\in\mathbb{N}} \alpha_n \leqslant \sup \alpha_n < 2$, let s be a selection ∂f, let $x_0 \in C$, set $\mu = \min f(C)$, and set

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \begin{cases} P_C\Bigg(x_n + \alpha_n \dfrac{\mu - f(x_n)}{\|s(x_n)\|^2} s(x_n)\Bigg), & \text{if } s(x_n) \neq 0;\\ x_n, & \text{if } s(x_n) = 0. \end{cases} \tag{29.86}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a minimizer of f over C.

Proof. In view of Fermat's rule (Theorem 16.3), we assume that $(\forall n \in \mathbb{N})$ $s(x_n) \neq 0$. Now, set $I = \{1,2\}$, $f_1 = d_C$, $f_2 = f - \mu$, $y_0 = x_0$, and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \lambda_n = 1 \ \text{ and } \ \mathrm{i}(n) = 1, & \text{if } n \text{ is odd};\\ \lambda_n = \alpha_n \ \text{ and } \ \mathrm{i}(n) = 2, & \text{if } n \text{ is even}. \end{cases} \tag{29.87}$$

Then $\operatorname{Argmin}_C f = \operatorname{lev}_{\leqslant 0} f_1 \cap \operatorname{lev}_{\leqslant 0} f_2$, f_1 is bounded on every bounded subset of $\mathcal{H}$, and (29.80) holds with $m = 2$. Hence, it follows from Proposition 29.49 that the sequence $(y_n)_{n\in\mathbb{N}}$ generated by (29.81), i.e.,

$$(\forall n \in \mathbb{N}) \quad y_{n+1} = y_n + \lambda_n\big(G_{\mathrm{i}(n)} y_n - y_n\big) \tag{29.88}$$

converges weakly to a point $x \in \operatorname{Argmin}_C f$. However, (29.86), Example 29.45, and (29.68) yield $(\forall n \in \mathbb{N})$ $x_n = y_{2n}$. We conclude that $x_n \rightharpoonup x$. ☐

Exercises

Exercise 29.1 Suppose that $\mathcal{H} = \mathsf{H} \times \mathbb{R}$, where H is a real Hilbert space, let $\rho \in \mathbb{R}_{++}$ and $\alpha \in \mathbb{R}_{++}$, and set $C_1 = B(0;\rho) \times \mathbb{R}$ and $C_2 = \mathsf{H} \times [-\alpha, +\alpha]$. Provide formulas for the projection operators P_{C_1} and P_{C_2}, and check that the projection onto the cylinder $C_1 \cap C_2$ is given by $P_{C_1\cap C_2} = P_{C_1} P_{C_2} = P_{C_2} P_{C_1}$.

Exercise 29.2 Show that Proposition 29.2(ii) fails if $L^* \neq L^{-1}$.

Exercise 29.3 Prove Proposition 29.4 using (29.1).

Exercise 29.4 Let C and D be nonempty closed convex subsets of $\mathcal{H}$ such that $C \perp D$. Show that $P_{C+D} = P_D + P_C \circ (\mathrm{Id} - P_D)$.

Exercise 29.5 Let C be a nonempty closed convex subset of $\mathcal{H}$, let $x \in \mathcal{H}$, and let $\lambda \in\,]0,1[$. Show that $\|P_{\lambda C} x\| \leqslant \|P_C x\|$.

Exercise 29.6 Given finite families $(u_i)_{i\in I}$ in $\mathcal{H} \smallsetminus \{0\}$ and $(\beta_i)_{i\in I}$ in $\mathbb{R}$, set $C = \bigcap_{i\in I} \big\{x \in \mathcal{H} \;\big|\; \langle x \mid u_i\rangle \leqslant \beta_i\big\}$. Let V be a closed linear subspace of $\bigcap_{i\in I}\{u_i\}^\perp$ and set $D = C \cap V^\perp$. Show that D is a polyhedron in $V^\perp$ and that $P_C = P_V + P_D = P_V + P_D \circ P_{V^\perp}$.

Exercise 29.7 Prove Proposition 29.10.

Exercise 29.8 Prove Example 29.19(ii) using Proposition 17.4.

Exercise 29.9 Prove Proposition 29.29.

Exercise 29.10 Prove Example 29.31 using (29.1).

Exercise 29.11 Let $\alpha \in \mathbb{R}_{++}$ and set $K_\alpha = \big\{(x,\rho) \in \mathcal{H}\times\mathbb{R} \;\big|\; \|x\| \leqslant \alpha\rho\big\}$. Use Exercise 6.4 to show that $K_\alpha^\ominus = -K_{1/\alpha}$ and that

$$\big(\forall (x,\rho)\in\mathcal{H}\times\mathbb{R}\big)\quad P_{K_\alpha}(x,\rho) = \begin{cases} (x,\rho), & \text{if } \|x\| \leqslant \alpha\rho;\\ (0,0), & \text{if } \alpha\|x\| \leqslant -\rho;\\ \dfrac{\alpha\|x\|+\rho}{\alpha^2+1}\Big(\alpha\dfrac{x}{\|x\|},1\Big), & \text{otherwise.}\end{cases} \tag{29.89}$$

Exercise 29.12 Let $z \in \mathcal{H}$ and let $\gamma \in [0,1]$. Suppose that $\|z\| = 1$ and set

$$K_{z,\gamma} = \big\{x \in \mathcal{H} \;\big|\; \gamma\|x\| - \langle x \mid z\rangle \leqslant 0\big\}. \tag{29.90}$$

Use Exercise 29.11 to prove that $K_{z,\gamma}^\ominus = K_{-z,\sqrt{1-\gamma^2}}$ and that

$$(\forall x\in\mathcal{H})\quad P_{K_{z,\gamma}}x = \begin{cases} x, & \text{if } x \in K_{z,\gamma};\\ 0, & \text{if } x\in K_{z,\gamma}^\ominus;\\ \delta y, & \text{otherwise,}\end{cases} \tag{29.91}$$

where $y = \gamma z + \sqrt{1-\gamma^2}(x - \langle x \mid z\rangle\, z)/\|x - \langle x \mid z\rangle\, z\|$ and $\delta = \langle x \mid y\rangle$.

Exercise 29.13 Consider Exercise 29.12 and its notation. Let $\varepsilon \in \mathbb{R}_{++}$, and let $w \in \mathcal{H}$ be such that $\|w\| = 1$ and $(\forall x \in K_{z,\varepsilon/(2+\varepsilon)})$ $\langle x \mid w\rangle \geqslant 0$. Prove that $\|w - z\| < \varepsilon/2$.

Exercise 29.14 Let N be a strictly positive integer, let $\beta \in \mathbb{R}_{++}$, and set $I = \{1,\ldots,N\}$. Using Example 29.34, show that the problem

$$\underset{\substack{y=(\eta_i)_{i\in I}\in\mathbb{R}^N_+\\ \sum_{i\in I}\eta_i=\beta}}{\text{minimize}} \;\; \|y\| \tag{29.92}$$

has a unique solution, namely $(\beta/N)_{i\in I}$.

Exercise 29.15 Suppose that $\mathcal{H} = \mathbb{R}$, set $f\colon \mathcal{H}\to\mathcal{H}\colon x\mapsto x^2$ and $C = \operatorname{epi} f$, and let $(z,\zeta) \in \mathbb{R}^2 \smallsetminus (\operatorname{epi} f)$. Prove that $P_C(z,\zeta) = (\bar{x},\bar{x}^2)$, where $\bar{x}\in\mathbb{R}$ is the unique solution to the cubic equation $2x^3 + (1-2\zeta)x - z = 0$.

Exercise 29.16 Let N be a strictly positive integer, suppose that $\mathcal{H} = \mathbb{R}^N$, set $I = \{1, \ldots, N\}$, let $\xi \in \mathbb{R}_{++}$, and let $(\alpha_i)_{i \in I} \in \mathbb{R}_+^N$. Set $f \colon \mathbb{R}^N \to \mathbb{R} \colon x \mapsto (1/2)\sum_{i \in I} \alpha_i |\xi_i|^2$ and let $z \in \mathbb{R}^N$ be such that $f(z) > \xi$. Set $C = \operatorname{lev}_{\leqslant \xi} f$ and $x = P_C z$. Show that, for every $i \in I$, $\xi_i = \zeta_i/(1 + \lambda\alpha_i)$, where

$$\frac{1}{2}\sum_{i \in I} \alpha_i \left(\frac{\zeta_i}{1 + \lambda\alpha_i}\right)^2 = \xi. \tag{29.93}$$

Exercise 29.17 Use Proposition 29.49 to provide a proof of Corollary 5.24 in the case when $\bigcap_{i \in I} C_i \neq \varnothing$.

Exercise 29.18 Let $f \colon \mathcal{H} \to \mathbb{R}$ be a continuous convex function, let $\xi \in \mathbb{R}$ be such that $C = \operatorname{lev}_{\leqslant \xi} f \neq \varnothing$, and let s be a selection of ∂f. Let G be the subgradient projector onto C associated with (f, ξ, s), and suppose that G is continuous. Show that f is Fréchet differentiable on $\mathcal{H} \smallsetminus C$.

Exercise 29.19 Let f be a continuous convex function from $\mathcal{H}$ to $\mathbb{R}$ such that $C = \operatorname{lev}_{\leqslant 0} f \neq \varnothing$ and which is bounded on every bounded subset of $\mathcal{H}$. Use Proposition 29.49 to devise an algorithm for finding a point in C.

Exercise 29.20 Let $(C_i)_{i \in I}$ be a finite family of closed convex subsets of $\mathcal{H}$ such that one of them is bounded and $C = \bigcap_{i \in I} C_i \neq \varnothing$. Use Exercise 29.19 and Example 29.45 to devise an algorithm to find a point in C.

Exercise 29.21 Construct a continuous Gâteaux differentiable convex function $h \colon \mathcal{H} \to \mathbb{R}$ such that the subgradient projector G associated with $(h, 0)$ onto $C = \operatorname{lev}_{\leqslant 0} h$ is not strong-to-weak continuous.

Hint: You may proceed along the following steps:

(i) Denote the function in Exercise 18.12 by f, and let $v \in (\operatorname{ran} \nabla f) \smallsetminus \{0\}$, say $v = \nabla f(y)$.
(ii) Set $g \colon \mathcal{H} \to \mathbb{R} \colon x \mapsto f(x) - \langle x \mid v\rangle$ and show that $\min g(\mathcal{H}) = g(y) < g(0) = 0$.
(iii) Set $h \colon \mathcal{H} \to \mathbb{R} \colon x \mapsto g(x) - g(y)/2$ and deduce that $y \in C$ but $0 \notin C$.
(iv) Since h is not Fréchet differentiable 0, obtain a sequence $(x_n)_{n \in \mathbb{N}}$ in $\mathcal{H}$ such that $x_n \to 0$, $\nabla h(x_n) \rightharpoonup \nabla h(0) = -v$, and $\nabla h(x_n) \not\to -v$.
(v) Prove that $(Gx_n)_{n \in \mathbb{N}}$ does not converge weakly to $G0$.

Exercise 29.22 Let $f \colon \mathcal{H} \to \mathbb{R}$ be convex and continuous, suppose that $C = \operatorname{lev}_{\leqslant 0} f \neq \varnothing$, let s be a selection of ∂f, and denote the subgradient projector onto C associated with $(f, 0, s)$ by G. Let $x_0 \in \mathcal{H}$, set

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = Gx_n, \tag{29.94}$$

and assume that $(x_n)_{n \in \mathbb{N}}$ lies in $\mathcal{H} \smallsetminus C$. Show that $\sum_{n \in \mathbb{N}} f^2(x_n)/\|s(x_n)\|^2 < +\infty$ and that, if $\operatorname{int} C \neq \varnothing$, then $\sum_{n \in \mathbb{N}} f(x_n)/\|s(x_n)\| < +\infty$.

Exercise 29.23 Let $f \in \Gamma_0(\mathcal{H})$, let $\xi \in \left]\inf f(\mathcal{H}), +\infty\right[$, let $x_0 \in \operatorname{lev}_{<\xi} f$, and set $C = \operatorname{lev}_{\leqslant \xi} f$. Show that $(\forall x \in \operatorname{dom} f \smallsetminus C)\ d_C(x) \leqslant \|x - x_0\|(f(x) - \xi)/(f(x) - f(x_0))$.

Chapter 30
Best Approximation Algorithms

Best approximation algorithms were already discussed in Corollary 5.30, in Example 28.18, and in Example 28.19. In this chapter, we provide further approaches for computing the projection onto the intersection of finitely many closed convex sets. The methods we present, all of which employ the individual projectors onto the given sets, are Halpern's algorithm, Dykstra's algorithm, and Haugazeau's algorithm. Applications to solving monotone inclusion and minimization problems with strongly convergent algorithms are also given.

30.1 Halpern's Algorithm

Let T be a nonexpansive operator such that $\operatorname{Fix} T \neq \varnothing$. As discussed in Section 5.2, the Banach–Picard iteration process of (1.69) may fail to produce a point in $\operatorname{Fix} T$. The Krasnosel'skiĭ–Mann algorithm of Theorem 5.15 provides a relaxation method that converges weakly to a point in $\operatorname{Fix} T$. The next theorem describes a so-called vanishing viscosity method, originally proposed by Halpern. This method converges strongly to the best approximation to a point $x \in \mathcal{H}$ from $\operatorname{Fix} T$.

Theorem 30.1 *Let D be a nonempty closed convex subset of $\mathcal{H}$, let $T\colon D \to D$ be a nonexpansive operator such that $\operatorname{Fix} T \neq \varnothing$, let $x \in D$, and let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,1[$ such that the following hold:*

(i) $\lambda_n \to 0$.
(ii) $\sum_{n\in\mathbb{N}} \lambda_n = +\infty$.
(iii) $\sum_{n\in\mathbb{N}} |\lambda_{n+1} - \lambda_n| < +\infty$.

The original version of this chapter was revised. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-48311-5_31

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_30

Let $x_0 \in D$ and set

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \lambda_n x + (1 - \lambda_n) T x_n. \tag{30.1}$$

Then $x_n \to P_{\operatorname{Fix} T} x$.

Proof. Since $-\ln$ is convex, Proposition 17.7 yields

$$\big(\forall \xi \in \left]0,1\right[\big) \quad \ln(1-\xi) \leqslant -\xi. \tag{30.2}$$

Hence, for every n and every m in $\mathbb{N}$ such that $n \geqslant m$, we have

$$\prod_{k=m}^{n} (1-\lambda_k) = \exp\left(\ln \prod_{k=m}^{n} (1-\lambda_k) \right) \leqslant \exp\left(\sum_{k=m}^{n} -\lambda_k \right) \tag{30.3}$$

and, letting n tend to $+\infty$ and recalling (ii), we obtain

$$\prod_{k=m}^{+\infty} (1-\lambda_k) = 0. \tag{30.4}$$

Now set $C = \operatorname{Fix} T$, which is closed and convex by Proposition 4.23(ii). We consider two cases.

(a) $x_0 = x$: Let $y \in C$. If $\|x_n - y\| \leqslant \|x - y\|$ for some $n \in \mathbb{N}$, then $\|x_{n+1} - y\| \leqslant \lambda_n \|x-y\| + (1-\lambda_n)\|Tx_n - y\| \leqslant \lambda_n \|x-y\| + (1-\lambda_n)\|x_n - y\| \leqslant \lambda_n \|x-y\| + (1-\lambda_n)\|x-y\| = \|x-y\|$. It follows by induction that

$$(\forall n \in \mathbb{N}) \quad \|Tx_n - y\| = \|Tx_n - Ty\| \leqslant \|x_n - y\| \leqslant \|x - y\|. \tag{30.5}$$

In view of (30.1), we obtain that

$$(x_n)_{n \in \mathbb{N}} \text{ is bounded,} \tag{30.6}$$

that

$$(Tx_n)_{n \in \mathbb{N}} \text{ is bounded,} \tag{30.7}$$

and therefore, in view of (i), that

$$||x_{n+1} - Tx_n|| = \lambda_n ||x - Tx_n|| \to 0. \tag{30.8}$$

By (30.6) and (30.7),

$$\mu = \sup_{n \in \mathbb{N}} \max\big\{\|x_{n+1} - x_n\|, \|x - Tx_n\|\big\} < +\infty. \tag{30.9}$$

Hence since, for every $n \in \mathbb{N}$,

$$x_{n+2} - x_{n+1} = (\lambda_{n+1} - \lambda_n)(x - Tx_n) + (1 - \lambda_{n+1})(Tx_{n+1} - Tx_n), \tag{30.10}$$

it follows from (30.9) and the nonexpansiveness of T that

$$\|x_{n+2} - x_{n+1}\| \leqslant \mu|\lambda_{n+1} - \lambda_n| + (1 - \lambda_{n+1})\|x_{n+1} - x_n\|. \tag{30.11}$$

Thus, for every n and every m in $\mathbb{N}$ such that $n \geqslant m$, we deduce from (30.2) using induction that

$$\begin{aligned}\|x_{n+2} - x_{n+1}\| &\leqslant \mu \sum_{k=m}^{n} |\lambda_{k+1} - \lambda_k| + \|x_{m+1} - x_m\| \prod_{k=m}^{n} (1 - \lambda_{k+1}) \\ &\leqslant \mu \sum_{k=m}^{n} |\lambda_{k+1} - \lambda_k| + \mu \prod_{k=m}^{n} (1 - \lambda_{k+1}). \end{aligned} \tag{30.12}$$

Letting n and then m tend to $+\infty$ in (30.12), and recalling (iii) and (30.4), we obtain $\overline{\lim}\, \|x_{n+2}-x_{n+1}\| \leqslant 0$. Hence $x_{n+1}-x_n \to 0$ and, by nonexpansiveness of T,

$$Tx_{n+1} - Tx_n \to 0. \tag{30.13}$$

Combining (30.8) and (30.13) yields

$$x_n - Tx_n \to 0. \tag{30.14}$$

By (30.6) and Lemma 2.45, there exists a subsequence $(x_{k_n})_{n\in\mathbb{N}}$ of $(x_n)_{n\in\mathbb{N}}$ such that

$$\lim \langle Tx_{k_n} - P_C x \mid x - P_C x\rangle = \overline{\lim}\, \langle Tx_n - P_C x \mid x - P_C x\rangle \tag{30.15}$$

and $(x_{k_n})_{n\in\mathbb{N}}$ converges weakly to some point $z \in D$. Corollary 4.28 and (30.14) yield $z \in C$. It therefore follows from (30.14) and (3.10) that

$$\begin{aligned}\overline{\lim}\, \langle Tx_n - P_C x \mid x - P_C x\rangle &= \lim \langle Tx_{k_n} - P_C x \mid x - P_C x\rangle \\ &= \lim \langle Tx_{k_n} - x_{k_n} \mid x - P_C x\rangle \\ &\quad + \lim \langle x_{k_n} - P_C x \mid x - P_C x\rangle \\ &= \langle z - P_C x \mid x - P_C x\rangle \\ &\leqslant 0. \end{aligned} \tag{30.16}$$

Now let $\varepsilon \in \mathbb{R}_{++}$. By (30.16) and (i), there exists $m \in \mathbb{N}$ such that, if $n \in \mathbb{N}$ and $n \geqslant m$, then

$$\langle Tx_n - P_C x \mid x - P_C x\rangle \leqslant \varepsilon \quad \text{and} \quad \lambda_n d_C^2(x) \leqslant \varepsilon. \tag{30.17}$$

Thus, for every $n \in \mathbb{N}$ such that $n \geqslant m$,

$$\begin{aligned}\|x_{n+1} - P_C x\|^2 &= \|\lambda_n(x - P_C x) + (1 - \lambda_n)(Tx_n - P_C x)\|^2 \\ &= \lambda_n^2 d_C^2(x) + (1 - \lambda_n)^2\|Tx_n - P_C x\|^2 \\ &\quad + 2\lambda_n(1 - \lambda_n) \langle x - P_C x \mid Tx_n - P_C x\rangle \\ &\leqslant \lambda_n\varepsilon + (1 - \lambda_n)\|x_n - P_C x\|^2 + 2\lambda_n\varepsilon. \end{aligned} \tag{30.18}$$

It follows by induction that, for every $n \in \mathbb{N}$ such that $n \geqslant m$,

$$\|x_{n+1} - P_C x\|^2 \leqslant 3\varepsilon + \|x_m - P_C x\|^2 \prod_{k=m}^{n} (1 - \lambda_k). \tag{30.19}$$

In view of (30.4), taking the limit superior over n yields $\overline{\lim} \|x_{n+1} - P_C x\|^2 \leqslant 3\varepsilon$. Since ε was chosen arbitrarily, we conclude that $x_n \to P_C x$.

(b) $x_0 \neq x$: Let $(y_n)_{n\in\mathbb{N}}$ be the sequence generated by (30.1) with starting point $y_0 = x$. By (a), $y_n \to P_C x$. On the other hand, it follows from (30.1) and (30.4) that

$$(\forall n \in \mathbb{N}) \quad \|x_{n+1} - y_{n+1}\| \leqslant \|x_0 - y_0\| \prod_{k=0}^{n} (1 - \lambda_k) \to 0. \tag{30.20}$$

Altogether, $x_n \to P_C x$. ☐

Next, we describe applications of Theorem 30.1 to the problem of finding the best approximation from a set of common fixed points.

Corollary 30.2 *Let D be a nonempty closed convex subset of $\mathcal{H}$, let $(T_i)_{i\in I}$ be a finite family of nonexpansive operators from D to D such that $C = \bigcap_{i\in I} \operatorname{Fix} T_i \neq \varnothing$, let $(\omega_i)_{i\in I}$ be real numbers in $]0,1]$ such that $\sum_{i\in I} \omega_i = 1$, and let $x \in D$. Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,1[$ such that $\lambda_n \to 0$, $\sum_{n\in\mathbb{N}} \lambda_n = +\infty$, and $\sum_{n\in\mathbb{N}} |\lambda_{n+1} - \lambda_n| < +\infty$, let $x_0 \in D$, and set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \lambda_n x + (1 - \lambda_n) \sum_{i\in I} \omega_i T_i x_n. \tag{30.21}$$

Then $x_n \to P_C x$.

Proof. Apply Theorem 30.1 to $T = \sum_{i\in I} \omega_i T_i$ using Proposition 4.47. ☐

Corollary 30.3 *Let D be a nonempty closed convex subset of $\mathcal{H}$, let m be a strictly positive integer, set $I = \{1, \dots, m\}$, let $(T_i)_{i\in I}$ be a family of averaged nonexpansive operators from D to D such that $C = \bigcap_{i\in I} \operatorname{Fix} T_i \neq \varnothing$, and let $x \in D$. Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,1[$ such that $\lambda_n \to 0$, $\sum_{n\in\mathbb{N}} \lambda_n = +\infty$, and $\sum_{n\in\mathbb{N}} |\lambda_{n+1} - \lambda_n| < +\infty$, let $x_0 \in D$, and set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \lambda_n x + (1 - \lambda_n) T_1 \cdots T_m x_n. \tag{30.22}$$

Then $x_n \to P_C x$.

Proof. Apply Theorem 30.1 to $T = T_1 \cdots T_m$ using Corollary 4.51. ☐

Example 30.4 Let m be a strictly positive integer, set $I = \{1, \dots, m\}$, let $(C_i)_{i\in I}$ be a family of closed convex subsets of $\mathcal{H}$ such that $C = \bigcap_{i\in I} C_i \neq \varnothing$, let $(\omega_i)_{i\in I}$ be real numbers in $]0,1]$ such that $\sum_{i\in I} \omega_i = 1$, and let $x \in \mathcal{H}$. Let $x_0 \in \mathcal{H}$, let $y_0 \in \mathcal{H}$, and set

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_{n+1} = \dfrac{1}{n+2}x + \dfrac{n+1}{n+2}\displaystyle\sum_{i\in I} \omega_i P_{C_i} x_n, \\ y_{n+1} = \dfrac{1}{n+2}x + \dfrac{n+1}{n+2} P_{C_1} \cdots P_{C_m} y_n. \end{cases} \tag{30.23}$$

Then $x_n \to P_C x$ and $y_n \to P_C x$.

Corollary 30.5 (Linear mean ergodic theorem) *Let $T \in \mathcal{B}(\mathcal{H})$ be nonexpansive and let $x \in \mathcal{H}$. Then*

$$\frac{1}{n+1}\sum_{k=0}^{n} T^k x \to P_{\ker(\mathrm{Id}-T)}x. \tag{30.24}$$

Proof. Apply Theorem 30.1 with $(\lambda_n)_{n\in\mathbb{N}} = (1/(n+2))_{n\in\mathbb{N}}$ and $x_0 = x$. $\square$

30.2 Dykstra's Algorithm

Corollary 5.24 features a weakly convergent periodic projection algorithm for finding a common point of closed convex sets. We saw in Corollary 5.30 that, when all the sets are closed affine subspaces, the iterates produced by this algorithm actually converge to the projection of the initial point x_0 onto the intersection of the sets. As shown in Figure 30.1, this is however not true in general. In this section, we describe a strongly convergent projection algorithm for finding the projection of a point onto the intersection of closed convex sets. The following technical result will be required.

Lemma 30.6 *Let $(\rho_n)_{n\in\mathbb{N}} \in \ell^2_+(\mathbb{N})$, let $m \in \mathbb{N}$, and set $(\forall n \in \mathbb{N})$ $\sigma_n = \sum_{k=0}^{n} \rho_k$. Then $\underline{\lim}\, \sigma_n(\sigma_n - \sigma_{n-m-1}) = 0$.*

Proof. Since, by Cauchy–Schwarz,

$$\sigma_n\big(\sigma_n - \sigma_{n-m-1}\big) \leqslant \sqrt{n+1}\sqrt{\sum_{k=0}^{n} \rho_k^2}\,\big(\sigma_n - \sigma_{n-m-1}\big), \tag{30.25}$$

for every integer $n > m$, it suffices to show that $\underline{\lim}\, \sqrt{n+1}(\sigma_n - \sigma_{n-m-1}) = 0$. We shall prove this by contradiction: assume that there exist $\varepsilon \in \mathbb{R}_{++}$ and an integer $n_0 > m$ such that, for every integer $n \geqslant n_0$, $\sigma_n - \sigma_{n-m-1} \geqslant \varepsilon/\sqrt{n+1}$ which, by Cauchy–Schwarz, implies that

$$\frac{\varepsilon^2}{n+1} \leqslant \big(\rho_{n-m} + \cdots + \rho_n\big)^2 \leqslant (m+1)\big(\rho_{n-m}^2 + \cdots + \rho_n^2\big). \tag{30.26}$$

Summing (30.26) over all integers $n \geqslant n_0$ yields a contradiction. $\square$

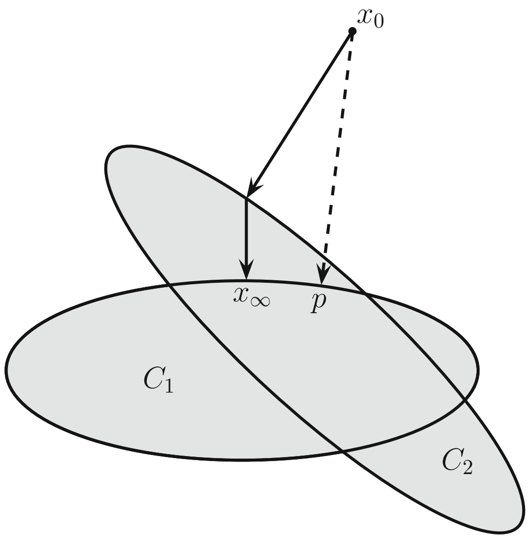

Fig. 30.1 The method of alternating projections $x_{n+1} = P_{C_1}P_{C_2}x_n$ converges (in one iteration) to x_∞ and fails to produce the best approximation p to x_0 from $C_1 \cap C_2$.

The periodic projection algorithm described next was first studied by Dykstra in the case of closed convex cones.

Theorem 30.7 (Dykstra's algorithm) *Let m be a strictly positive integer, set $I = \{1, \ldots, m\}$, let $(C_i)_{i\in I}$ be a family of closed convex subsets of $\mathcal{H}$ such that $C = \bigcap_{i\in I} C_i \neq \varnothing$, and let $x_0 \in \mathcal{H}$. Set*

$$\mathrm{i}\colon \mathbb{N} \to I\colon n \mapsto 1 + \mathrm{rem}(n-1, m), \tag{30.27}$$

where $\mathrm{rem}(\cdot, m)$ is the remainder function of the division by m. For every strictly positive integer n, set $P_n = P_{C_n}$, where $C_n = C_{\mathrm{i}(n)}$ if $n > m$. Moreover, set $q_{-(m-1)} = \cdots = q_{-1} = q_0 = 0$ and

$$\begin{array}{l} \text{for } n = 1, 2, \ldots \\ \left\lfloor \begin{array}{l} x_n = P_n(x_{n-1} + q_{n-m}), \\ q_n = x_{n-1} + q_{n-m} - x_n. \end{array} \right. \end{array} \tag{30.28}$$

Then $x_n \to P_C x_0$.

Proof. It follows from Theorem 3.16 that

$$(\forall n \in \mathbb{N} \smallsetminus \{0\}) \quad x_n \in C_n \quad \text{and} \quad (\forall y \in C_n) \ \ \langle x_n - y \mid q_n\rangle \geqslant 0. \tag{30.29}$$

Moreover, for every integer $n \geqslant 1$, (30.28) yields

$$x_{n-1} - x_n = q_n - q_{n-m} \tag{30.30}$$

and, therefore,

$$x_0 - x_n = \sum_{k=n-m+1}^{n} q_k. \tag{30.31}$$

Let $z \in C$, let $n \in \mathbb{N}$, and let $x_{-(m-1)}, \dots, x_{-1}$ be arbitrary vectors in $\mathcal{H}$. We derive from (30.30) that

$$\begin{aligned}
&\|x_n - z\|^2 \\
&\quad = \|x_{n+1} - z\|^2 + \|x_n - x_{n+1}\|^2 + 2\left\langle x_{n+1} - z \mid x_n - x_{n+1}\right\rangle \\
&\quad = \|x_{n+1} - z\|^2 + \|x_n - x_{n+1}\|^2 + 2\left\langle x_{n+1} - z \mid q_{n+1} - q_{n+1-m}\right\rangle \\
&\quad = \|x_{n+1} - z\|^2 + \|x_n - x_{n+1}\|^2 + 2\left\langle x_{n+1} - z \mid q_{n+1}\right\rangle \\
&\qquad + 2\left\langle x_{n+1-m} - x_{n+1} \mid q_{n+1-m}\right\rangle - 2\left\langle x_{n-m+1} - z \mid q_{n-m+1}\right\rangle.
\end{aligned} \tag{30.32}$$

Now let $l \in \mathbb{N}$ be such that $n \geqslant l$. Using (30.32) and induction, we obtain

$$\begin{aligned}
\|x_l - z\|^2 = \|x_n - z\|^2 &+ \sum_{k=l+1}^{n} \Big(\|x_k - x_{k-1}\|^2 + 2\left\langle x_{k-m} - x_k \mid q_{k-m}\right\rangle \Big) \\
&+ 2 \sum_{k=n-m+1}^{n} \left\langle x_k - z \mid q_k\right\rangle - 2 \sum_{k=l-(m-1)}^{l} \left\langle x_k - z \mid q_k\right\rangle.
\end{aligned} \tag{30.33}$$

In particular, when $l = 0$, we see that

$$\begin{aligned}
\|x_0 - z\|^2 = \|x_n - z\|^2 &+ \sum_{k=1}^{n} \Big(\|x_k - x_{k-1}\|^2 + 2\left\langle x_{k-m} - x_k \mid q_{k-m}\right\rangle \Big) \\
&+ 2 \sum_{k=n-m+1}^{n} \left\langle x_k - z \mid q_k\right\rangle.
\end{aligned} \tag{30.34}$$

Since all the summands on the right-hand side of (30.34) lie in $\mathbb{R}_+$, it follows that $(x_k)_{k \in \mathbb{N}}$ is bounded and

$$\sum_{k \in \mathbb{N}} \|x_{k-1} - x_k\|^2 < +\infty. \tag{30.35}$$

Using (30.31) and (30.29), we get

$$\begin{aligned}
\left\langle z - x_n \mid x_0 - x_n\right\rangle &= \sum_{k=n-m+1}^{n} \left\langle z - x_k \mid q_k\right\rangle + \sum_{k=n-m+1}^{n} \left\langle x_k - x_n \mid q_k\right\rangle \\
&\leqslant \sum_{k=n-m+1}^{n} \left\langle x_k - x_n \mid q_k\right\rangle.
\end{aligned} \tag{30.36}$$

On the other hand, for every $k \in \{n-m+1, \ldots, n-1\}$, the identities

$$\begin{aligned} q_k &= q_k - 0 \\ &= q_k - q_{\mathrm{i}(k)-m} \\ &= (q_k - q_{k-m}) + (q_{k-m} - q_{k-2m}) + \cdots + (q_{\mathrm{i}(k)} - q_{\mathrm{i}(k)-m}) \end{aligned} \tag{30.37}$$

and (30.30) result in

$$\sum_{k=n-m+1}^{n-1} \|q_k\| \leqslant \sum_{k=1}^{n-1} \|q_k - q_{k-m}\| = \sum_{k=1}^{n-1} \|x_{k-1} - x_k\|. \tag{30.38}$$

Hence, using the triangle inequality, we obtain

$$\begin{aligned} \sum_{k=n-m+1}^{n-1} |\langle x_k - x_n \mid q_k\rangle| &\leqslant \sum_{k=n-m+1}^{n-1} \|x_k - x_n\| \, \|q_k\| \\ &\leqslant \sum_{l=n-m+2}^{n} \|x_{l-1} - x_l\| \sum_{k=n-m+1}^{n-1} \|q_k\| \\ &\leqslant \sum_{l=n-m+2}^{n} \|x_{l-1} - x_l\| \sum_{k=1}^{n-1} \|x_{k-1} - x_k\|. \end{aligned} \tag{30.39}$$

In view of (30.35) and Lemma 30.6, we deduce that

$$\underline{\lim} \sum_{k=n-m+1}^{n} |\langle x_k - x_n \mid q_k\rangle| = 0. \tag{30.40}$$

Let $(x_{p_n})_{n\in\mathbb{N}}$ be a subsequence of $(x_n)_{n\in\mathbb{N}}$ such that

$$\lim \sum_{k=p_n-m+1}^{p_n} |\langle x_k - x_{p_n} \mid q_k\rangle| = 0, \tag{30.41}$$

such that $(x_{p_n})_{n\in\mathbb{N}}$ converges weakly, say $x_{p_n} \rightharpoonup x \in \mathcal{H}$, and such that $\lim \|x_{p_n}\|$ exists. In view of the definition of i, we also assume that there exists $j \in I$ such that $(\forall n \in \mathbb{N})$ $\mathrm{i}(p_n) = j$. By construction, $(x_{p_n})_{n\in\mathbb{N}}$ lies in C_j, and Corollary 3.35 therefore yields $x \in C_j$. Furthermore, since $x_n - x_{n+1} \to 0$ by (30.35) and since, for every $n \in \mathbb{N} \smallsetminus \{0\}$, $x_n \in C_n$,

$$x \in C. \tag{30.42}$$

Combining (30.36) and (30.41) yields

$$\overline{\lim} \, \langle z - x_{p_n} \mid x_0 - x_{p_n}\rangle \leqslant 0. \tag{30.43}$$

Hence, for every $z \in C$, it follows from Lemma 2.42 that

$$\begin{aligned}\langle z - x \mid x_0 - x \rangle &= \|x\|^2 - \langle z \mid x \rangle - \langle x \mid x_0 \rangle + \langle z \mid x_0 \rangle \\ &\leqslant \underline{\lim} \big(\|x_{p_n}\|^2 - \langle z \mid x_{p_n} \rangle - \langle x_{p_n} \mid x_0 \rangle + \langle z \mid x_0 \rangle \big) \\ &\leqslant \overline{\lim} \big(\|x_{p_n}\|^2 - \langle z \mid x_{p_n} \rangle - \langle x_{p_n} \mid x_0 \rangle + \langle z \mid x_0 \rangle \big) \\ &= \overline{\lim} \, \langle z - x_{p_n} \mid x_0 - x_{p_n} \rangle \\ &\leqslant 0. \end{aligned} \tag{30.44}$$

In view of (30.42) and (30.44), we derive from (3.10) that

$$x = P_C x_0. \tag{30.45}$$

Using (30.44), with $z = x$, yields $\|x_{p_n}\| \to \|x\|$. Since $x_{p_n} \rightharpoonup x$, we deduce from Corollary 2.52 that

$$x_{p_n} \to x. \tag{30.46}$$

In turn, by (30.36) and (30.41),

$$\begin{aligned} 0 \leftarrow & \langle x - x_{p_n} \mid x_0 - x_{p_n} \rangle \\ &= \sum_{k=p_n-m+1}^{p_n} \langle x - x_k \mid q_k \rangle + \sum_{k=p_n-m+1}^{p_n} \langle x_k - x_{p_n} \mid q_k \rangle \\ &\leqslant \sum_{k=p_n-m+1}^{p_n} \langle x_k - x_{p_n} \mid q_k \rangle \\ &\to 0. \end{aligned} \tag{30.47}$$

This implies that

$$\lim \sum_{k=p_n-m+1}^{p_n} \langle x - x_k \mid q_k \rangle = 0. \tag{30.48}$$

We now show by contradiction that the entire sequence $(x_n)_{n\in\mathbb{N}}$ converges to x. To this end, let us assume that there exist $\varepsilon \in \mathbb{R}_{++}$ and a subsequence $(x_{l_n})_{n\in\mathbb{N}}$ of $(x_n)_{n\in\mathbb{N}}$ such that $(\forall n \in \mathbb{N})$ $\|x_{l_n} - x\| \geqslant \varepsilon$. After passing to a subsequence of $(x_{l_n})_{n\in\mathbb{N}}$ and relabeling if necessary, we assume furthermore that $(\forall n \in \mathbb{N})$ $l_n > p_n$. Then (30.33), (30.29), and (30.48) yield

$$\begin{aligned} 0 \leftarrow & \|x_{p_n} - x\|^2 \\ &= \|x_{l_n} - x\|^2 + \sum_{k=p_n+1}^{l_n} \big(\|x_k - x_{k-1}\|^2 + 2 \langle x_{k-m} - x_k \mid q_{k-m} \rangle \big) \\ &\quad + 2 \sum_{k=l_n-m+1}^{l_n} \langle x_k - x \mid q_k \rangle - 2 \sum_{k=p_n-m+1}^{p_n} \langle x_k - x \mid q_k \rangle \end{aligned}$$

$$\begin{aligned}
&\geqslant \|x_{l_n} - x\|^2 + 2 \sum_{k=p_n-m+1}^{p_n} \langle x - x_k \mid q_k \rangle \\
&\geqslant 2 \sum_{k=p_n-m+1}^{p_n} \langle x - x_k \mid q_k \rangle \\
&\to 0. \qquad (30.49)
\end{aligned}$$

It follows that $x_{l_n} \to x$, which is impossible. □

Theorem 30.7 allows for a different proof of Corollary 5.30; we leave the verification of the details as Exercise 30.3.

30.3 Haugazeau's Algorithm

In this section we consider the problem of finding the best approximation to a point z from the set of common fixed points of firmly quasinonexpansive operators.

Theorem 30.8 *Let $(T_i)_{i\in I}$ be a finite family of firmly quasinonexpansive operators from $\mathcal{H}$ to $\mathcal{H}$ such that the operators $(\mathrm{Id} - T_i)_{i\in I}$ are demiclosed at 0 and $C = \bigcap_{i\in I} \operatorname{Fix} T_i \neq \varnothing$. Let $x_0 \in \mathcal{H}$, let $\mathrm{i}\colon \mathbb{N} \to I$ be such that there exists a strictly positive integer m for which*

$$(\forall i \in I)(\forall n \in \mathbb{N}) \quad i \in \{\mathrm{i}(n), \ldots, \mathrm{i}(n+m-1)\}, \tag{30.50}$$

and set

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = Q\big(x_0, x_n, T_{\mathrm{i}(n)} x_n\big), \tag{30.51}$$

where Q is defined in (29.39). Then $x_n \to P_C x_0$.

Proof. Throughout, the notation (29.38) will be used. We first observe that Proposition 4.23(ii) implies that C is a nonempty closed convex set. Also, for every nonempty closed convex subset D of $\mathcal{H}$, (29.1) implies that

$$(\forall x \in \mathcal{H}) \quad P_D x \in D \quad \text{and} \quad D \subset H(x, P_D x). \tag{30.52}$$

In view of Corollary 29.25, to check that the sequence $(x_n)_{n\in\mathbb{N}}$ is well defined, we must show that $(\forall n \in \mathbb{N})\ H(x_0, x_n) \cap H(x_n, T_{\mathrm{i}(n)} x_n) \neq \varnothing$. To this end, it is sufficient to show that $C \subset \bigcap_{n\in\mathbb{N}} H(x_0, x_n) \cap H(x_n, T_{\mathrm{i}(n)} x_n)$, i.e., since (30.50) and Corollary 4.25 yield

$$C = \bigcap_{n\in\mathbb{N}} \operatorname{Fix} T_{\mathrm{i}(n)} \subset \bigcap_{n\in\mathbb{N}} H\big(x_n, T_{\mathrm{i}(n)} x_n\big), \tag{30.53}$$

that $C \subset \bigcap_{n\in\mathbb{N}} H(x_0, x_n)$. For $n = 0$, it is clear that $C \subset H(x_0, x_n) = \mathcal{H}$. Furthermore, for every $n \in \mathbb{N}$, it follows from (30.53), (30.52), and Corollary 29.25 that

$$
\begin{aligned}
C \subset H(x_0, x_n) &\Rightarrow C \subset H(x_0, x_n) \cap H(x_n, T_{\mathrm{i}(n)} x_n) \\
&\Rightarrow C \subset H\big(x_0, Q(x_0, x_n, T_{\mathrm{i}(n)} x_n)\big) \\
&\Leftrightarrow C \subset H(x_0, x_{n+1}),
\end{aligned}
\tag{30.54}
$$

which establishes the assertion by induction. Now let $n \in \mathbb{N}$. We observe that Corollary 29.25 yields

$$
x_{n+1} \in H(x_0, x_n) \cap H\big(x_n, T_{\mathrm{i}(n)} x_n\big). \tag{30.55}
$$

Hence, since x_n is the projection of x_0 onto $H(x_0, x_n)$ and since $x_{n+1} = Q(x_0, x_n, T_{\mathrm{i}(n)} x_n) \in H(x_0, x_n)$, we have $\|x_0 - x_n\| \leqslant \|x_0 - x_{n+1}\|$. On the other hand, since $P_C x_0 \in C \subset H(x_0, x_n)$, we have $\|x_0 - x_n\| \leqslant \|x_0 - P_C x_0\|$. Altogether, $(\|x_0 - x_n\|)_{n \in \mathbb{N}}$ converges and

$$
\lim \|x_0 - x_n\| \leqslant \|x_0 - P_C x_0\|. \tag{30.56}
$$

For every $n \in \mathbb{N}$, the inclusion $x_{n+1} \in H(x_0, x_n)$ implies that

$$
\begin{aligned}
\|x_0 - x_{n+1}\|^2 - \|x_0 - x_n\|^2 &= \|x_{n+1} - x_n\|^2 + 2 \langle x_{n+1} - x_n \mid x_n - x_0 \rangle \\
&\geqslant \|x_{n+1} - x_n\|^2,
\end{aligned}
\tag{30.57}
$$

and it follows from the convergence of $(\|x_0 - x_n\|)_{n \in \mathbb{N}}$ that

$$
x_{n+1} - x_n \to 0. \tag{30.58}
$$

In turn since, for every $n \in \mathbb{N}$, the inclusion $x_{n+1} \in H(x_n, T_{\mathrm{i}(n)} x_n)$ implies that

$$
\begin{aligned}
\|x_{n+1} - x_n\|^2 &= \|x_{n+1} - T_{\mathrm{i}(n)} x_n\|^2 + 2 \big\langle x_{n+1} - T_{\mathrm{i}(n)} x_n \mid T_{\mathrm{i}(n)} x_n - x_n \big\rangle \\
&\quad + \|x_n - T_{\mathrm{i}(n)} x_n\|^2 \\
&\geqslant \|x_{n+1} - T_{\mathrm{i}(n)} x_n\|^2 + \|x_n - T_{\mathrm{i}(n)} x_n\|^2,
\end{aligned}
\tag{30.59}
$$

we obtain

$$
x_n - T_{\mathrm{i}(n)} x_n \to 0. \tag{30.60}
$$

Now let $i \in I$, and let x be a weak sequential cluster point of $(x_n)_{n \in \mathbb{N}}$. In view of (30.50), there exist sequences $(k_n)_{n \in \mathbb{N}}$ and $(p_n)_{n \in \mathbb{N}}$ in $\mathbb{N}$ such that $x_{k_n} \rightharpoonup x$ and

$$
(\forall n \in \mathbb{N}) \quad
\begin{cases}
k_n \leqslant p_n \leqslant k_n + m - 1 < k_{n+1} \leqslant p_{n+1}, \\
i = \mathrm{i}(p_n).
\end{cases}
\tag{30.61}
$$

Moreover, it follows from (30.58) that

$$\begin{aligned}\|x_{p_n} - x_{k_n}\| &\leqslant \sum_{l=k_n}^{k_n+m-2} \|x_{l+1} - x_l\| \\ &\leqslant (m-1) \max_{k_n \leqslant l \leqslant k_n+m-2} \|x_{l+1} - x_l\| \\ &\to 0, \end{aligned} \tag{30.62}$$

and, in turn, that $x_{p_n} \rightharpoonup x$. Hence, since $x_{p_n} - T_i x_{p_n} = x_{p_n} - T_{\mathrm{i}(p_n)} x_{p_n} \to 0$ by (30.60), we deduce from the demiclosedness of $\mathrm{Id} - T_i$ at 0 that $x \in \operatorname{Fix} T_i$ and, since i was arbitrarily chosen in I, that $x \in C$. Thus, in view of (30.56) and Proposition 4.21, $x_n \to P_C x_0$. ☐

Our first application concerns the problem of finding the projection of a point onto the set of solutions to systems of convex inequalities.

Corollary 30.9 *Let $(f_i)_{i\in I}$ be a finite family of continuous convex functions from $\mathcal{H}$ to $\mathbb{R}$ such that $C = \bigcap_{i\in I} \operatorname{lev}_{\leqslant 0} f_i \neq \varnothing$ and suppose that one of the following holds:*

(i) *The functions $(f_i)_{i\in I}$ are bounded on every bounded subset of $\mathcal{H}$.*
(ii) *The functions $(f_i^*)_{i\in I}$ are supercoercive.*
(iii) *$\mathcal{H}$ is finite-dimensional.*

Let $\mathrm{i}\colon \mathbb{N} \to I$ be such that there exists a strictly positive integer m for which

$$(\forall i \in I)(\forall n \in \mathbb{N}) \quad i \in \{\mathrm{i}(n), \ldots, \mathrm{i}(n+m-1)\}. \tag{30.63}$$

For every $i \in I$, fix a selection s_i of ∂f_i and denote by G_i the subgradient projector onto $\operatorname{lev}_{\leqslant 0} f_i$ associated with $(f_i, 0, s_i)$. Let $x_0 \in \mathcal{H}$ and set

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = Q\big(x_0, x_n, G_{\mathrm{i}(n)} x_n\big), \tag{30.64}$$

where Q is defined in (29.39). Then $x_n \to P_C x_0$.

Proof. (i): In view of Proposition 29.41(i)&(iii)&(vii), (30.64) is a special case of (30.51). Hence the claim follows from Theorem 30.8.

(ii)&(iii): Combine Proposition 16.20 and (i). ☐

Corollary 30.10 *Let $(T_i)_{i\in I}$ be a finite family of firmly nonexpansive operators from $\mathcal{H}$ to $\mathcal{H}$ such that $C = \bigcap_{i\in I} \operatorname{Fix} T_i \neq \varnothing$. Let $x_0 \in \mathcal{H}$, let $\mathrm{i}\colon \mathbb{N} \to I$ be such that there exists a strictly positive integer m for which*

$$(\forall i \in I)(\forall n \in \mathbb{N}) \quad i \in \{\mathrm{i}(n), \ldots, \mathrm{i}(n+m-1)\}, \tag{30.65}$$

and set

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = Q\big(x_0, x_n, T_{\mathrm{i}(n)} x_n\big), \tag{30.66}$$

where Q is defined in (29.39). Then $x_n \to P_C x_0$.

Proof. Combine Theorem 30.8 and Theorem 4.27. □

The next application yields a strongly convergent proximal-point algorithm that finds the zero of a maximally monotone operator at minimal distance from the starting point.

Corollary 30.11 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and such that $0 \in \operatorname{ran} A$, and let $x_0 \in \mathcal{H}$. Set $(\forall n \in \mathbb{N})$ $x_{n+1} = Q(x_0, x_n, J_A x_n)$, where Q is defined in (29.39). Then $x_n \to P_{\operatorname{zer} A}\, x_0$.*

Proof. Set $T_1 = J_A$. Then T_1 is firmly nonexpansive by Corollary 23.9, and Proposition 23.38 asserts that $\operatorname{Fix} T_1 = \operatorname{zer} A \neq \varnothing$. Thus, the result is an application of Theorem 30.8 with $I = \{1\}$, $m = 1$, and $\mathrm{i}\colon n \mapsto 1$. □

The next result provides a strongly convergent forward-backward algorithm.

Corollary 30.12 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $\beta \in \mathbb{R}_{++}$, let $B\colon \mathcal{H} \to \mathcal{H}$ be β-cocoercive, and let $\gamma \in \left]0, 2\beta\right[$. Suppose that $\operatorname{zer}(A+B) \neq \varnothing$, let $x_0 \in \mathcal{H}$, and set*

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
y_n = x_n - \gamma B x_n, \\
z_n = (1/2)\big(x_n + J_{\gamma A} y_n\big), \\
x_{n+1} = Q\big(x_0, x_n, z_n\big),
\end{array}
\right.
\end{array}
\tag{30.67}
$$

where Q is defined in (29.39). Then $x_n \to P_{\operatorname{zer}(A+B)}\, x_0$.

Proof. Set $T = (1/2)(\mathrm{Id} + J_{\gamma A} \circ (\mathrm{Id} - \gamma B))$ and note that (30.67) yields

$$
(\forall n \in \mathbb{N}) \quad x_{n+1} = Q\big(x_0, x_n, T x_n\big). \tag{30.68}
$$

By Corollary 23.11(i), $J_{\gamma A}$ is nonexpansive. Moreover, by Proposition 4.39 and Remark 4.34(i), $\mathrm{Id} - \gamma B$ is nonexpansive. Thus, $J_{\gamma A} \circ (\mathrm{Id} - \gamma B)$ is nonexpansive and, appealing to Proposition 4.4, we deduce that T is firmly nonexpansive. Altogether, since $\operatorname{Fix} T = \operatorname{Fix}(2T - \mathrm{Id}) = \operatorname{zer}(A + B)$ by Proposition 26.1(iv)(a), the result follows from Theorem 30.8 with $I = \{1\}$, $m = 1$, and $\mathrm{i}\colon n \mapsto 1$. □

Example 30.13 Let $f \in \Gamma_0(\mathcal{H})$, let $\beta \in \mathbb{R}_{++}$, let $g\colon \mathcal{H} \to \mathbb{R}$ be convex and differentiable with a $1/\beta$-Lipschitz continuous gradient, and let $\gamma \in \left]0, 2\beta\right[$. Furthermore, suppose that $\operatorname{Argmin}(f + g) \neq \varnothing$, let $x_0 \in \mathcal{H}$, and set

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
y_n = x_n - \gamma \nabla g(x_n), \\
z_n = (1/2)\big(x_n + \operatorname{Prox}_{\gamma f} y_n\big), \\
x_{n+1} = Q\big(x_0, x_n, z_n\big),
\end{array}
\right.
\end{array}
\tag{30.69}
$$

where Q is defined in (29.39). Then $x_n \to P_{\operatorname{Argmin}(f+g)}\, x_0$.

Proof. This is an application of Corollary 30.12 to $A = \partial f$ and $B = \nabla g$, using Corollary 18.17. Indeed, A and B are maximally monotone by Theorem 20.25 and, since $\operatorname{dom} g = \mathcal{H}$, Corollary 27.3 yields $\operatorname{Argmin}(f+g) = \operatorname{zer}(A+B)$. ☐

Remark 30.14 The principle employed in the proof of Corollary 30.12 extends as follows. Let $(R_i)_{i\in I}$ be a finite family of nonexpansive operators from $\mathcal{H}$ to $\mathcal{H}$ such that $\bigcap_{i\in I} \operatorname{Fix} R_i \neq \varnothing$ and set $(\forall i \in I)\ T_i = (1/2)(R_i + \mathrm{Id})$. Then $\bigcap_{i\in I} \operatorname{Fix} T_i = \bigcap_{i\in I} \operatorname{Fix} R_i$, and it follows from Proposition 4.4 that the operators $(T_i)_{i\in I}$ are firmly nonexpansive. Thus, Theorem 30.8 can be used to find the best approximation to a point $x_0 \in \mathcal{H}$ from $\bigcap_{i\in I} \operatorname{Fix} R_i$.

Our last result describes an alternative to Dykstra's algorithm of Theorem 30.7 for finding the projection onto the intersection of closed convex sets using the projectors onto the individual sets periodically.

Corollary 30.15 (Haugazeau's algorithm) *Let m be a strictly positive integer, set $I = \{1,\dots,m\}$, let $(C_i)_{i\in I}$ be a family of closed convex subsets of $\mathcal{H}$ such that $C = \bigcap_{i\in I} C_i \neq \varnothing$, and let $x_0 \in \mathcal{H}$. For every $i \in I$, denote by P_i the projector onto C_i. Let $\operatorname{rem}(\cdot, m)$ be the remainder function of division by m, and set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = Q\big(x_0, x_n, P_{1+\operatorname{rem}(n,m)} x_n\big), \tag{30.70}$$

where Q is defined in (29.39). Then $x_n \to P_C x_0$.

Proof. In view of Proposition 4.16, this is an application of Theorem 30.8 with $(T_i)_{i\in I} = (P_i)_{i\in I}$ and $\mathrm{i}\colon n \mapsto 1 + \operatorname{rem}(n,m)$. ☐

Exercises

Exercise 30.1 Suppose that $\mathcal{H} = \mathbb{R}$, and set $D = [-1,1]$, $x = 0$, and $x_0 = 1$. By considering $T_1\colon D \to D\colon x \mapsto 1$ and $T_2\colon D \to D\colon x \mapsto -x$ show that Theorem 30.1 fails when (i) or (ii) is not satisfied.

Exercise 30.2 Give a detailed proof of Corollary 30.5.

Exercise 30.3 Use Theorem 30.7 to derive Corollary 5.30.

Exercise 30.4 Consider the application of Dykstra's algorithm (30.28) to $m = 2$ intersecting closed convex subsets C and D of $\mathcal{H}$.

(i) Show that the algorithm can be cast in the following form:

$$p_0 = q_0 = 0 \quad \text{and} \quad \begin{array}{l} \text{for } n = 0,1,\dots \\ \left\lfloor \begin{array}{l} y_n = P_C(x_n + p_n), \\ p_{n+1} = x_n + p_n - y_n, \\ x_{n+1} = P_D(y_n + q_n), \\ q_{n+1} = y_n + q_n - x_{n+1}. \end{array} \right. \end{array} \tag{30.71}$$

(ii) Suppose that $\mathcal{H} = \mathbb{R}^2$, $x_0 = (2, 2)$, $C = \big\{(\xi_1, \xi_2) \in \mathcal{H} \;\big|\; \xi_2 \leqslant 0\big\}$, and $D = \big\{(\xi_1, \xi_2) \in \mathcal{H} \;\big|\; \xi_1 + \xi_2 \leqslant 0\big\}$. Compute the values of x_n for $n \in \{0, \ldots, 5\}$ and draw a picture showing the progression of the iterates toward $P_{C\cap D}x_0$.

Exercise 30.5 Let $T\colon \mathcal{H} \to \mathcal{H}$ be nonexpansive and such that $\operatorname{Fix} T \neq \varnothing$. Devise a strongly convergent algorithm based on (30.51) to find the minimal norm fixed point of T.

Exercise 30.6 Let $(T_i)_{i\in I}$ be a finite family of nonexpansive operators from $\mathcal{H}$ to $\mathcal{H}$ such that $C = \bigcap_{i\in I} \operatorname{Fix} T_i \neq \varnothing$, let $x_0 \in \mathcal{H}$, let $(\omega_i)_{i\in I}$ be real numbers in $]0, 1]$ such that $\sum_{i\in I} \omega_i = 1$, and suppose that $(\alpha_i)_{i\in I}$ are real numbers in $]0, 1[$ such that, for every $i \in I$, T_i is α_i-averaged. Use Theorem 30.8 to devise a strongly convergent algorithm involving $(\alpha_i)_{i\in I}$ to find $P_C x_0$.

Correction to: Convex Analysis and Monotone Operator Theory in Hilbert Spaces

Correction to:
H.H. Bauschke, P.L. Combettes,
***Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,**
https://doi.org/10.1007/978-3-319-48311-5

The original version of this book was inadvertently published without updating the following corrections in Chapters 1–3, 5–13, 16–20, 23, 24, 26, 29, 30, and back matter. These are corrected now.

The updated version of this book can be found at
https://doi.org/10.1007/978-3-319-48311-5_1
https://doi.org/10.1007/978-3-319-48311-5_2
https://doi.org/10.1007/978-3-319-48311-5_3
https://doi.org/10.1007/978-3-319-48311-5_5
https://doi.org/10.1007/978-3-319-48311-5_6
https://doi.org/10.1007/978-3-319-48311-5_7
https://doi.org/10.1007/978-3-319-48311-5_8
https://doi.org/10.1007/978-3-319-48311-5_9
https://doi.org/10.1007/978-3-319-48311-5_10
https://doi.org/10.1007/978-3-319-48311-5_11
https://doi.org/10.1007/978-3-319-48311-5_12
https://doi.org/10.1007/978-3-319-48311-5_13
https://doi.org/10.1007/978-3-319-48311-5_16
https://doi.org/10.1007/978-3-319-48311-5_17
https://doi.org/10.1007/978-3-319-48311-5_18
https://doi.org/10.1007/978-3-319-48311-5_19
https://doi.org/10.1007/978-3-319-48311-5_20
https://doi.org/10.1007/978-3-319-48311-5_23
https://doi.org/10.1007/978-3-319-48311-5_24
https://doi.org/10.1007/978-3-319-48311-5_26
https://doi.org/10.1007/978-3-319-48311-5_29
https://doi.org/10.1007/978-3-319-48311-5_30
https://doi.org/10.1007/978-3-319-48311-5

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5_31

For **Figures 8.1, 13.1, 16.1, 18.1, 20.1, and 26.1**—The usage of font/Symbol $\mathcal{H}$ invariably differs from universally adhered display standards throughout the text content and the same has been incorporated to resolve the issue.

$\boxed{1}$ **Second line after Eq. (1.55):** The identity $\operatorname{diam} B(x_n; \varepsilon_n) = 2\varepsilon_n$ has been replaced by $\operatorname{diam} B(x_n; \varepsilon_n) \leqslant 2\varepsilon_n$.

$\boxed{2}$ **Second line after Eq. (2.17):** $n \times n$ has been replaced by $N \times N$.

$\boxed{3}$ **Fact 2.63:** The assumption:
let V be a neighborhood of Tx, and let $R\colon V \to \mathcal{K}$. Suppose that R is Fréchet differentiable at x and that T is Gâteaux differentiable at Tx

has been replaced by:

let V be a neighborhood of Tx such that $T(U) \subset V$, and let $R\colon V \to \mathcal{K}$. Suppose that T is Gâteaux differentiable at x and that R is Fréchet differentiable at Tx

$\boxed{4}$ **Exercise 3.2:** The phase "a nonempty set" has been replaced by "a nonempty finite set".

$\boxed{5}$ **First line of Exercise 5.9:** The phrase "Let m be a strictly positive integer" has been replaced by "Let $m \geqslant 2$ be an integer"

$\boxed{6}$ **Second line after Eq. (6.3):** The expression $\sum_{i \in I} \alpha_i = 1$ has been replaced by $\sum_{i \in I} \alpha_i \leqslant 1$.

$\boxed{7}$ **Example 7.10:** The assumption "Suppose that $\mathcal{H}$ is finite-dimensional" has been added.

$\boxed{8}$ **Proof of Proposition 8.14(i):** The phrase "the convexity of ϕ on C" has been replaced by "the convexity of ϕ on I".

$\boxed{9}$ **Example 8.33, last line of the proof:** The phrase "so is $\phi = f(\cdot, 1)$" has been replaced by "so is $\phi^{\diamond} = f(\cdot, 1)$"

$\boxed{10}$ **Exercise 8.12(ii):** The expression $|x^{1/p} + 1|^p$ has been replaced by $-|x^{1/p} + 1|^p$

$\boxed{11}$ **Exercise 8.19:** The expression $0 \in C$ has been replaced by $0 \in \operatorname{ri} C$.

$\boxed{12}$ **Equation (9.39):** $\mu(\omega)$ has been replaced by $\mu(\Omega)$

$\boxed{13}$ **Right-hand side of Equation (9.40):** $\mu(\omega)$ has been replaced by $\mu(\Omega)$

$\boxed{14}$ **Proposition 11.8(ii):** The expression $C \cap \operatorname{Argmin} f$ has been replaced by $(\operatorname{int} C) \cap \operatorname{Argmin} f$.

$\boxed{15}$ **Definition 12.34:** The phrase "it is exact at a point $y \in \mathcal{H}$" has been replaced by "it is exact at a point $y \in \mathcal{K}$"

$\boxed{16}$ **Exercise 13.4, Equation (13.39):** The expression $\frac{1}{2}u^2 - |u| - \frac{1}{2}$ has been replaced by $\frac{1}{2}u^2 + |u| - \frac{1}{2}$.

$\boxed{17}$ **Proposition 13.24 and Corollary 13.25:** The assumption "Let $\mathcal{K}$ be a real Hilbert space" has been added.

$\boxed{18}$ **Proof of Proposition 13.45:** The phrase "in view of Proposition 13.13" has been removed.

$\boxed{19}$ **In Eq. (13.26):** The expression $\sup_{x_i \in \mathcal{H}}$ has been replaced by $\sup_{x_i \in \mathcal{H}_i}$.

$\boxed{20}$ **Equation (17.45):** The expression $N_C x \cap B\big(0; \phi'(d_C(x))\big)$ has been replaced by $N_C x \cap B\big(0; \phi'(0)\big)$.

$\boxed{21}$ **Corollary 18.20, first sum in Eq. (18.40):** The expression $\sum_{i=1}^{m} \alpha_i \operatorname{Prox}_{f_i}$ has been replaced by $\sum_{i \in I} \alpha_i \operatorname{Prox}_{f_i}$

$\boxed{22}$ **Fourth line after Eq. (18.40):** The expression $\nabla f = \sum_{i \in \mathcal{H}} \alpha_i \operatorname{Prox}_{f_i}$ has been replaced by $\nabla f = \sum_{i \in I} \alpha_i \operatorname{Prox}_{f_i}$.

$\boxed{23}$ **Exercise 20.1(iii):** The phrase "is differentiable" has been replaced by "is convex and differentiable".

$\boxed{24}$ **Equation (24.81):** The expression

$$\xi \mapsto \begin{cases} \xi \operatorname{arctanh}^{-1}(\xi) + \dfrac{1}{2}\Big(\ln\big(1-\xi^2\big) - \xi^2\Big), & \text{if } |\xi| < 1; \\ +\infty, & \text{if } |\xi| \geqslant 1. \end{cases}$$

has been replaced by

$$\xi \mapsto \begin{cases} \dfrac{(1+\xi)\ln(1+\xi) + (1-\xi)\ln(1-\xi) - \xi^2}{2}, & \text{if } |\xi| < 1; \\ \ln(2) - 1/2, & \text{if } |\xi| = 1; \\ +\infty, & \text{if } |\xi| > 1. \end{cases}$$

$\boxed{25}$ **Example 24.51:** The expression "is a proximal thresholder" has been replaced by "If $\mathcal{H} = \mathbb{R}$, then Prox_f is a proximal thresholder".

$\boxed{26}$ **Exercise 24.6:** The phrase "Show that $(\forall x \in \mathcal{H})$" has been replaced by "Show that $(\forall x \in \mathbb{R})$"

$\boxed{27}$ **Corollary 26.6: The statement:**
Then $x \in \operatorname{zer}(A+B)$, $u \in \operatorname{zer}(-A^{-1} \circ (-\mathrm{Id}) + B^{-1})$, $(x, -u) \in \operatorname{gra} A$, and $(x, u) \in \operatorname{gra} B$

has been replaced by:

Then $x \in \operatorname{zer}(A+B)$, $-u \in \operatorname{zer}(-A^{-1} \circ (-\mathrm{Id}) + B^{-1})$, $(x, u) \in \operatorname{gra} A$, and $(x, -u) \in \operatorname{gra} B$

$\boxed{28}$ **Theorem 26.34, line after Eq. (26.97):** The inclusion $\gamma \in \left]0, 1/\|L\|^2\right[$ has been replaced by $\gamma \in \left]0, 1/\|L\|\right[$

$\boxed{29}$ **Statement of Proposition 29.4:** The identity $\boldsymbol{C} = \bigtimes_{i\in I} C_i$ has been replaced by $\boldsymbol{C} = \mathcal{H} \cap \bigtimes_{i\in I} C_i$

$\boxed{30}$ **Exercise 29.23:** The statement "Show that $(\forall x \in \operatorname{dom} f \smallsetminus C)$" has been replaced by "Show that $(\forall x \in \operatorname{dom} f \smallsetminus C)$ $d_C(x) \leqslant \|x - x_0\|(f(x) - \xi)/(f(x) - f(x_0))$."

$\boxed{31}$ **Example 29.47:** The expression $C = \operatorname{lev}_{\leqslant 0} f = \left]-\infty, -1/2\right]$ has been replaced by $C = \operatorname{lev}_{\leqslant 0} f = \left]-\infty, -1\right]$.

$\boxed{32}$ **Equation (30.8) and the sentence before:**

and therefore that

$$(x_{n+1} - Tx_n)_{n\in\mathbb{N}} \text{ is bounded.}$$

has been replaced by

and therefore, in view of (i), that

$$\|x_{n+1} - Tx_n\| = \lambda_n \|x - Tx_n\| \to 0.$$

Bibliographical Pointers

Seminal references for convex analysis, monotone operator theory, and nonexpansive operators are [65], [84], [96], [174], [213], [233], [256], [258], [271], [273], [313], [364], [365]. Historical comments can be found in [175], [201], [220], [261], [262], [295], [313], [320], [366], [367], [368], [369].

Further accounts and applications can be found in [11], [15], [21], [22], [23], [41], [55], [60], [61],[62], [71], [73], [75], [76], [81], [98], [103], [109], [110], [114], [117], [120], [145], [146], [147], [155], [162], [164], [169], [179], [180], [181], [182], [184], [195], [202], [203], [204], [208], [221], [224], [227], [232], [242], [243], [253], [261], [262], [263], [279], [288], [294], [298], [303], [318], [320], [323], [326], [329], [331], [343], [351], [353], [358], [360], [363], [366], [367], [368], [369].

It is virtually impossible to compile a complete bibliography on the topics covered by this book, most of which are quite mature. What follows is an attempt to list work that is relevant to the individual chapters, without making any claim to exhaustiveness.

Chapter 1: [6], [20], [23], [60], [75], [85], [116], [149], [160], [161], [164], [200], [214], [227], [273], [320], [325], [340], [358], [363].

Chapter 2: [2], [6], [79], [80], [84], [142], [146], [147], [150], [156], [169], [192], [199], [204], [224], [238], [244], [253], [280], [282], [291], [322], [325], [336], [340].

Chapter 3: [5], [101], [157], [147], [187], [192], [204], [219], [304], [313], [321], [334], [342], [351], [363].

Chapter 4: [26], [31], [35], [36], [50], [92], [93], [99], [100], [125], [138], [140], [181], [182], [281], [301], [365].

Chapter 5: [24], [25], [26], [27], [33], [36], [83], [86], [88], [96], [99], [100], [111], [113], [118], [119], [122], [124], [125], [141], [165], [166], [167], [172], [186], [188], [193], [223], [247], [249], [276], [285], [299], [300], [324], [355].

Chapter 6: [66], [68], [69], [77], [80], [204], [205], [215], [222], [264], [329], [363], [365].

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5

Chapter 7: [63], [202], [204], [218], [288].

Chapter 8: [64], [66], [73], [78], [80], [202], [203], [210], [245], [273], [296], [311], [313], [363].

Chapter 9: [6], [71], [84], [127], [273], [313], [329], [363].

Chapter 10: [81], [107], [190], [232], [274], [292], [313], [354], [362], [363].

Chapter 11: [75], [151], [181], [233], [292], [356], [363].

Chapter 12: [9], [11], [18], [198], [202], [264], [265], [267], [268], [269], [271], [273], [274], [313], [320], [363].

Chapter 13: [23], [71], [108], [173], [202], [246], [271], [273], [311], [313], [318], [320], [363].

Chapter 14: [36], [45], [73], [139], [251], [270], [271], [273], [313], [320], [363].

Chapter 15: [13], [69], [70], [71], [104], [185], [204], [212], [273], [308], [310], [313], [333], [363].

Chapter 16: [19], [36], [73], [91], [168], [259], [263], [266], [273], [307], [313], [318], [363].

Chapter 17: [37], [66], [71], [73], [180], [207], [241], [273], [288], [313], [335], [363].

Chapter 18: [8], [28], [37], [38], [44], [66], [72], [73], [74], [153], [163], [180], [288], [313], [349], [363].

Chapter 19: [22], [128], [154], [164], [208], [310], [313], [318], [363].

Chapter 20: [30], [39], [47], [53], [84], [94], [99], [103], [105], [106], [176], [192], [226], [252], [271], [273], [289], [315], [329], [331], [333], [338], [367], [368].

Chapter 21: [34], [59], [67], [84], [97], [102], [103], [105], [144], [147], [211], [216], [225], [257], [258], [271], [277], [288], [302], [312], [313], [316], [328], [329], [330], [331], [332], [352], [357], [367], [368].

Chapter 22: [30], [112], [159], [191], [209], [309], [314], [368].

Chapter 23: [11], [31], [34], [46], [48], [84], [129], [138], [158], [183], [217], [258], [319], [350].

Chapter 24: [11], [71], [89], [115], [131], [133], [134], [135], [138], [139], [143], [189], [230], [234], [260], [265], [268], [271], [272], [275], [344], [361].

Chapter 25: [4], [10], [16], [39], [47], [49], [54], [56], [57], [76], [77], [87], [130], [176], [177], [231], [278], [283], [315], [328], [330], [331], [333], [367], [368].

Chapter 26: [1], [3], [14], [17], [29], [40], [90], [95], [125], [126], [132], [137], [152], [158], [170], [171], [230], [236], [237] [239], [240], [254], [255], [286], [287], [306], [327], [338], [341], [347], [348], [366].

Chapter 27: [12], [85], [201], [202], [203], [208], [229], [313], [339], [345], [363], [366].

Chapter 28: [14], [125], [128], [136], [139], [152], [189], [233], [248], [250], [254], [297], [313], [346], [347].

Chapter 29: [33], [42], [51], [52], [58], [120], [121], [147], [178], [197], [284], [293], [365].

Chapter 30: [32], [43], [82], [119], [123], [194], [197], [228], [290], [337], [359].

Symbols and Notation

Real Line

$\mathbb{R}$		The set of real numbers
$\xi \geqslant 0$	p. 5	The real number ξ is positive
$\xi > 0$	p. 5	The real number ξ is strictly positive
$\xi \leqslant 0$	p. 5	The real number ξ is negative
$\xi < 0$	p. 5	The real number ξ is strictly negative
$\mathbb{R}_+$	p. 5	The set of positive real numbers $[0, +\infty[$
$\mathbb{R}_{++}$	p. 5	The set of strictly positive real numbers $]0, +\infty[$
$\mathbb{R}_-$	p. 5	The set of negative real numbers $]-\infty, 0]$
$\mathbb{R}_{--}$	p. 5	The set of strictly negative real numbers $]-\infty, 0[$
$\mathbb{Q}$		The set of rational numbers
$\mathbb{Z}$		The set of integers
$\mathbb{N}$	p. 4	The set of positive integers $\{0, 1, \dots\}$
inf, min	p. 5	Infimum and minimum
sup, max	p. 5	Supremum and maximum
$\overline{\lim}\, \xi_a$	p. 5	Limit superior of a net $(\xi_a)_{a \in A}$
$\underline{\lim}\, \xi_a$	p. 5	Limit inferior of a net $(\xi_a)_{a \in A}$
$\alpha \downarrow \mu$		α is in $]\mu, +\infty[$ and converges to μ.
$\operatorname{rem}(n, m)$		Remainder of the division of n by m
W	p. 236	Lambert W-function

Sets

$2^{\mathcal{X}}$	p. 2	Power set of a set $\mathcal{X}$
$\complement C$		The complement of a set C
$C \times D$		Cartesian product of the sets C and D
$C + D$	p. 1	

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5

$C - D$	p. 1	
λC	p. 1	Scaling of a set C by a real number λ
ΛC	p. 1	$\Lambda C = \bigcup_{\lambda \in \Lambda} \lambda C$, where $\Lambda \subset \mathbb{R}$
Λz	p. 1	$\Lambda z = \{\lambda z \mid \lambda \in \Lambda\}$, where $\Lambda \subset \mathbb{R}$
$z + C$	p. 1	Translation of a set C by a vector z
$C - z$	p. 1	Translation of a set C by a vector $-z$
$\operatorname{span} C$	p. 1	Span of a set C
$\overline{\operatorname{span}}\, C$	p. 1	Closed span of a set C
$\operatorname{aff} C$	p. 1	Affine hull of a set C
$A(C)$	p. 2	Image of a set C by an operator A
$\operatorname{diam} C$	p. 16	Diameter of a set C
$\overline{C}$	p. 7	Closure of a set C
$\operatorname{int} C$	p. 7	Interior of a set C
$\operatorname{bdry} C$	p. 7	Boundary of a set C
$\operatorname{core} C$	p. 114	Core of a set C
$\operatorname{sri} C$	p. 114	Strong relative interior of a set C
$\operatorname{qri} C$	p. 115	Quasirelative interior of a set C
$\operatorname{ri} C$	p. 115	Relative interior of a set C
$\operatorname{conv} C$	p. 49	Convex hull of a set C
$\overline{\operatorname{conv}}\, C$	p. 49	Closed convex hull of a set C
P_C	p. 50	Projector onto a nonempty closed convex set C
$Q(x, y, z)$	p. 545	Projector arising in Haugazeau's algorithm
$\operatorname{cone} C$	p. 111	Conical hull of a set C
$\overline{\operatorname{cone}}\, C$	p. 111	Closed conical hull of a set C
$C^{\perp}$	p. 27	Orthogonal complement of a set C
$C^{\ominus}$	p. 122	Polar cone of a set C
$C^{\oplus}$	p. 122	Dual cone of a set C
$C^{\odot}$	p. 136	Polar set of a set C
N_C	p. 126	Normal cone operator of a set C
T_C	p. 125	Tangent cone operator of a set C
$\operatorname{rec} C$	p. 128	Recession cone of a set C
$\operatorname{bar} C$	p. 128	Barrier cone of a set C
$\operatorname{spts} C$	p. 133	Set of support points of a set C
$\overline{\operatorname{spts}}\, C$	p. 133	Closure of the set of support points of a set C
σ_C	p. 135	Support function of a set C
1_C		Characteristic function of a set C
ι_C	p. 12	Indicator function of a set C
d_C	p. 16	Distance function to a set C
m_C	p. 149	Minkowski gauge of a set C
soft_C	p. 428	Soft-thresholder on a set C
$C_{\downarrow}$	p. 439	Set of vectors of $C \subset \mathbb{R}^N$ with decreasingly ordered entries
$H(x, y)$	p. 545	Half-space arising in Haugazeau's algorithm
$[x, y]$, $]x, y]$, $[x, y[$, $]x, y[$	p. 2	Line segments between x and y

Topology

$\mathcal{V}(x)$	p. 7	Family of all neighborhoods of x
$x_a \to x$	p. 7	The net $(x_a)_{a\in A}$ converges to x
$\overline{C}$	p. 7	Closure of a set C
$\operatorname{int} C$	p. 7	Interior of a set C
$\operatorname{bdry} C$	p. 7	Boundary of a set C
$B(x;\rho)$	p. 16	Closed ball with center x and radius ρ
F_σ	p. 24	
G_δ	p. 19	

Functions

$\Gamma(\mathcal{H})$	p. 157	Set of lower semicontinuous convex functions from $\mathcal{H}$ to $[-\infty,+\infty]$
$\Gamma_0(\mathcal{H})$	p. 160	Set of proper lower semicontinuous convex functions from $\mathcal{H}$ to $]-\infty,+\infty]$
$\bigoplus_{i\in I} f_i$	p. 28	Separable sum of functions
$\tau_y f$	p. 7	Translation of a function f by a vector y
$f^{\vee}$	p. 7	Reversal of a function f
$f\vert_C$		Restriction of a function f to a set C
$\operatorname{dom} f$	p. 5	Domain of a function f
$\overline{\operatorname{dom}}\, f$	p. 6	Closure of the domain of a function f
$\operatorname{gra} f$	p. 6	Graph of a function f
$\operatorname{epi} f$	p. 6	Epigraph of a function f
$\overline{\operatorname{epi}}\, f$	p. 6	Closure of the epigraph of a function f
$\operatorname{lev}_{\leqslant \xi} f$	p. 6	Lower level set of a function f
$\operatorname{lev}_{< \xi} f$	p. 6	Strict lower level set of a function f
$\operatorname{cont} f$	p. 10	Domain of continuity of a function f
$\bar{f}$	p. 14	Lower semicontinuous envelope of a function f
$\breve{f}$	p. 158	Lower semicontinuous convex envelope of a function f
$\operatorname{Argmin} f$	p. 190	Set of global minimizers of a function f
$\operatorname{Argmin}_C f$	p. 190	$\operatorname{Argmin}(f+\iota_C)$
$\operatorname{argmin} f$	p. 190	The unique minimizer of a function f
$f \,\square\, g$	p. 203	Infimal convolution of the functions f and g
$f \boxdot g$	p. 203	Exact infimal convolution of the functions f and g
$L \triangleright f$	p. 214	Infimal postcomposition of an operator L and a function f
$L \mathbin{\cdot\!\!\!\triangleright} f$	p. 214	Exact infimal postcomposition of an operator L and a function f
${}^{\gamma}f$	p. 210	Moreau envelope of index γ of a function f
Prox_f	p. 211	Proximity operator of a function f
$f^{\diamond}$	p. 148	Adjoint of a function f
f^{*}	p. 219	Conjugate of a function f

f^{**}	p. 219	Biconjugate of a function f
$\operatorname{rec} f$	p. 164	Recession function of a function f
$\operatorname{pav}(f, g)$	p. 239	Proximal average of the functions f and g
∂f	p. 263	Subdifferential of a function f
∇f	p. 41	Gradient operator of a function f
$\mathsf{D} f$	p. 40	Gâteaux derivative of a function f
$\mathsf{D}^2 f$	p. 40	Second Gâteaux derivative of a function f
$f'(x; y)$	p. 289	Directional derivative of f at x in the direction y
f'_+	p. 289	Right derivative of a function $f\colon \mathbb{R} \to \left]-\infty, +\infty\right]$
f'_-	p. 289	Left derivative of a function $f\colon \mathbb{R} \to \left]-\infty, +\infty\right]$
1_C		Characteristic function of a set C
ι_C	p. 12	Indicator function of a set C
d_C	p. 16	Distance function to a set C
σ_C	p. 135	Support function of a set C
m_C	p. 149	Minkowski gauge of a set C
F^{T}	p. 230	Transposition of the bivariate function F
$\mathsf{D}^* F$	p. 283	Coderivative of the bivariate function F

Set-valued Operators

$A\colon \mathcal{X} \to 2^{\mathcal{Y}}$	p. 2	A is a set-valued operator from $\mathcal{X}$ to $\mathcal{Y}$
$\operatorname{gra} A$	p. 2	Graph of an operator A
$\operatorname{dom} A$	p. 2	Domain of an operator A
$\overline{\operatorname{dom}}\, A$	p. 2	Closure of the domain of an operator A
$\operatorname{ran} A$	p. 2	Range of an operator A
$\overline{\operatorname{ran}}\, A$	p. 2	Closure of the range of an operator A
$\operatorname{zer} A$	p. 2	Set of zeros of an operator A
A^{-1}	p. 2	Inverse of an operator A
A_V	p. 359	Partial inverse of the operator A with respect to the closed linear subspace V
$\tau_y A$	p. 3	Translation of an operator A by y
$A^{\vee}$	p. 3	Reversal of an operator A
λA	p. 3	Scaling of the operator A by $\lambda \in \mathbb{R}$
$A + B$	p. 3	Sum of the operators A and B
$A \circ B$	p. 2	Composition of the operators A and B
$A \,\Box\, B$	p. 457	Parallel sum of the operators A and B
$L \triangleright A$	p. 459	Parallel composition of the operator A by the bounded linear operator L
F_A	p. 363	Fitzpatrick function of an operator A
J_A	p. 393	Resolvent of an operator A
R_A	p. 396	Reflected resolvent of an operator A
${}^{\gamma}A$	p. 393	Yosida approximation of an operator A of index $\gamma \in \mathbb{R}_{++}$
${}^{0}Ax$	p. 407	The element of minimal norm in Ax

Single-valued Operators

$T\colon \mathcal{X} \to \mathcal{Y}$	p. 2	T is an operator from $\mathcal{X}$ to $\mathcal{Y}$ defined everywhere on $\mathcal{X}$
$T^{\vee}$	p. 3	Reversal of an operator T
$\operatorname{Fix} T$	p. 21	Set of fixed points of an operator $T\colon \mathcal{X} \to \mathcal{Y}$
$\mathsf{D}T$	p. 40	Gâteaux derivative of an operator T
D^2T	p. 40	Second Gâteaux derivative of an operator T
$T\|_C$		Restriction of an operator T to a set C
$T^{-1}(C)$	p. 3	Inverse image of a set C by an operator T
$\ker L$	p. 33	Kernel of a linear operator L
$\\|L\\|$	p. 32	Norm of a linear operator L
L^*	p. 32	Adjoint of a bounded linear operator L
$L^\dagger$	p. 57	Generalized inverse of a bounded linear operator L
$A \succcurlyeq B$	p. 224	Löwner partial ordering of the operators A and B
$A \succ B$	p. 225	Strict Löwner partial ordering of the operators A and B
A^{T}		Transpose of a matrix A
$\operatorname{tra}(A)$	p. 28	Trace of a matrix A
Diag	p. 444	Creates a diagonal matrix
$\\|A\\|_2$	p. 32	Spectral norm of a matrix A
$\\|A\\|_{\mathsf{F}}$	p. 28	Frobenius norm of a matrix A
$\\|A\\|_{\mathsf{nuc}}$	p. 445	Nuclear norm of a matrix A
$\\|A\\|_{\mathsf{Sch}}$	p. 445	Schatten p-norm of a matrix A
$A \succ 0$	p. 443	The matrix A is positive definite

Banach Spaces

$\mathcal{H}, \mathcal{H}_i, \mathcal{K}, \mathcal{K}_i$	p. 27	Real Hilbert spaces
$\langle \cdot \mid \cdot \rangle$	p. 27	Scalar product
$\\| \cdot \\|$	p. 27	Norm
d	p. 27	Distance
$\to$	p. 34	Strong convergence in a Hilbert space
$\rightharpoonup$	p. 34	Weak convergence in a Hilbert space
Id	p. 27	Identity operator
$\mathcal{B}(\mathcal{X}, \mathcal{Y})$	p. 31	Space of bounded linear operators from $\mathcal{X}$ to $\mathcal{Y}$ with domain $\mathcal{H}$
$\mathcal{B}(\mathcal{X})$	p. 31	Space of bounded linear operators from $\mathcal{X}$ to $\mathcal{X}$ with domain $\mathcal{X}$
$\bigoplus_{i\in I} \mathcal{H}_i$	p. 28	Hilbert direct sum
$\mathcal{H}^{\text{weak}}$	p. 34	A real Hilbert space endowed with the weak topology
$\mathbb{R}^N$	p. 28	The standard N-dimensional Euclidean space
$\mathbb{R}^N_+$	p. 5	The positive orthant in $\mathbb{R}^N$

$\mathbb{R}^N_-$	p. 5	The negative orthant in $\mathbb{R}^N$
$\|\cdot\|_p$	p. 169	ℓ^p norm on $\mathbb{R}^N$
$\|\cdot\|_{\mathsf{lgl}}$	p. 215	Latent group lasso penalty
$(\Omega, \mathcal{F}, \mu)$	p. 28	Measure space
μ-a.e.	p. 28	μ-almost everywhere
$(\Omega, \mathcal{F}, \mathsf{P})$	p. 29	Probability space
$\mathsf{E}X$	p. 29	Expected value of a random variable X
x'	p. 29	Time derivative of a function $x\colon [0,T] \to \mathsf{H}$
$L^p((\Omega, \mathcal{F}, \mu); \mathsf{H})$	p. 28	Measurable functions $x\colon \Omega \to \mathsf{H}$ such that $\|x\|^p_{\mathsf{H}}$ is μ-integrable
$L^2([0,T])$	p. 29	Measurable functions $x\colon [0,T] \to \mathbb{R}$ such that $|x|^2$ is Lebesgue integrable
$L^2([0,T]; \mathsf{H})$	p. 29	Measurable functions $x\colon [0,T] \to \mathsf{H}$ such that $\|x\|^2_{\mathsf{H}}$ is Lebesgue integrable
$W^{1,2}([0,T]; \mathsf{H})$	p. 29	Functions $x \in L^2([0,T]; \mathsf{H})$ such that $x' \in L^2([0,T]; \mathsf{H})$
$L^2(\Omega)$	p. 29	Measurable functions $x\colon \Omega \to \mathbb{R}$ such that $|x|^p$ is Lebesgue integrable
$\ell^2(I)$	p. 28	Hilbert space of square-summable functions from I to $\mathbb{R}$
$\ell^2_+(I)$	p. 113	Set of square-summable functions from I to $\mathbb{R}_+$
$\ell^2_-(I)$	p. 113	Set of square-summable functions from I to $\mathbb{R}_-$
$\ell^1_+(\mathbb{N})$	p. 103	Set of sequences $(\xi_n)_{n\in\mathbb{N}}$ in $\mathbb{R}_+$ such that $\sum_{n\in\mathbb{N}} |\xi_n| < +\infty$.
$\mathbb{R}^{M\times N}$	p. 28	Space of real $M \times N$ matrices
$\mathbb{S}^N$	p. 28	Space of real $N \times N$ symmetric matrices
$\mathbb{S}^N_+$	p. 549	Set of real $N \times N$ symmetric positive semidefinite matrices
$\mathbb{U}^N$	p. 440	Set of $N \times N$ orthogonal matrices

References

[1] M. A. Alghamdi, A. Alotaibi, P. L. Combettes, and N. Shahzad, *A primal-dual method of partial inverses for composite inclusions*, Optim. Lett., 8 (2014), pp. 2271–2284.

[2] C. D. Aliprantis and K. C. Border, *Infinite Dimensional Analysis*, Springer-Verlag, Berlin, third ed., 2006.

[3] A. Alotaibi, P. L. Combettes, and N. Shahzad, *Solving coupled composite monotone inclusions by successive Fejér approximations of their Kuhn-Tucker set*, SIAM J. Optim., 24 (2014), pp. 2076–2095.

[4] W. N. Anderson, Jr. and R. J. Duffin, *Series and parallel addition of matrices*, J. Math. Anal. Appl., 26 (1969), pp. 576–594.

[5] N. Aronszajn, *Introduction to the Theory of Hilbert Spaces*, Research Foundation, Stillwater, OK, 1950.

[6] R. B. Ash, *Real Analysis and Probability*, Academic Press, New York, 1972.

[7] E. Asplund, *A monotone convergence theorem for sequences of nonlinear mappings*, in Nonlinear Functional Analysis (Proc. Sympos. Pure Math., Vol. XVIII, Part 1, Chicago, Ill., 1968), Amer. Math. Soc., Providence, R.I., 1970, pp. 1–9.

[8] E. Asplund and R. T. Rockafellar, *Gradients of convex functions*, Trans. Amer. Math. Soc., 139 (1969), pp. 443–467.

[9] H. Attouch, *Convergence de fonctions convexes, des sous-différentiels et semi-groupes associés*, C. R. Acad. Sci. Paris Sér. A, 284 (1977), pp. 539–542.

[10] ——, *On the maximality of the sum of two maximal monotone operators*, Nonlinear Anal., 5 (1981), pp. 143–147.

[11] ——, *Variational Convergence for Functions and Operators*, Pitman, Boston, MA, 1984.

[12] ——, *Viscosity solutions of minimization problems*, SIAM J. Optim., 6 (1996), pp. 769–806.

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5

[13] H. Attouch and H. Brézis, *Duality for the sum of convex functions in general Banach spaces*, in Aspects of Mathematics and Its Applications, vol. 34, North-Holland, Amsterdam, 1986, pp. 125–133.

[14] H. Attouch, L. M. Briceño-Arias, and P. L. Combettes, *A parallel splitting method for coupled monotone inclusions*, SIAM J. Control Optim., 48 (2010), pp. 3246–3270.

[15] H. Attouch, G. Buttazzo, and G. Michaille, *Variational Analysis in Sobolev and BV Spaces*, SIAM, Philadelphia, PA, second ed., 2014.

[16] H. Attouch, H. Riahi, and M. Théra, *Somme ponctuelle d'opérateurs maximaux monotones*, Serdica Math. J., 22 (1996), pp. 267–292.

[17] H. Attouch and M. Théra, *A general duality principle for the sum of two operators*, J. Convex Anal., 3 (1996), pp. 1–24.

[18] H. Attouch and R. J.-B. Wets, *A convergence theory for saddle functions*, Trans. Amer. Math. Soc., 280 (1983), pp. 1–41.

[19] J.-P. Aubin and I. Ekeland, *Applied Nonlinear Analysis*, Wiley, New York, 1984.

[20] J.-P. Aubin and H. Frankowska, *Set-Valued Analysis*, Birkhäuser, Boston, MA, 1990.

[21] A. Auslender, *Optimisation – Méthodes Numériques*, Masson, Paris, 1976.

[22] A. Auslender and M. Teboulle, *Asymptotic Cones and Functions in Optimization and Variational Inequalities*, Springer-Verlag, New York, 2003.

[23] D. Azé, *Éléments d'Analyse Convexe et Variationnelle*, Ellipses, Paris, 1997.

[24] J.-B. Baillon, *Un théorème de type ergodique pour les contractions non linéaires dans un espace de Hilbert*, C. R. Acad. Sci. Paris Sér. A, 280 (1975), pp. 1511–1514.

[25] ———, *Quelques propriétés de convergence asymptotique pour les contractions impaires*, C. R. Acad. Sci. Paris Sér. A, 283 (1976), pp. 587–590.

[26] J.-B. Baillon, R. E. Bruck, and S. Reich, *On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces*, Houston J. Math., 4 (1978), pp. 1–9.

[27] J.-B. Baillon, P. L. Combettes, and R. Cominetti, *There is no variational characterization of the cycles in the method of periodic projections*, J. Funct. Anal., 262 (2012), pp. 400–408.

[28] J.-B. Baillon and G. Haddad, *Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones*, Israel J. Math., 26 (1977), pp. 137–150.

[29] A. B. Bakušinskiĭ and B. T. Polyak, *The solution of variational inequalities*, Soviet Math. Dokl., 15 (1974), pp. 1705–1710.

[30] S. Bartz, H. H. Bauschke, J. M. Borwein, S. Reich, and X. Wang, *Fitzpatrick functions, cyclic monotonicity and Rockafellar's antiderivative*, Nonlinear Anal., 66 (2007), pp. 1198–1223.

[31] S. Bartz, H. H. Bauschke, S. M. Moffat, and X. Wang, *The resolvent average of monotone operators: dominant and recessive properties*, SIAM J. Optim., 26 (2016), pp. 602–634.

[32] H. H. Bauschke, *The approximation of fixed points of compositions of nonexpansive mappings in Hilbert space*, J. Math. Anal. Appl., 202 (1996), pp. 150–159.

[33] ———, *Projection Algorithms and Monotone Operators*, PhD thesis, Simon Fraser University, Burnaby, B.C., Canada, August 1996.

[34] ———, *Fenchel duality, Fitzpatrick functions and the extension of firmly nonexpansive mappings*, Proc. Amer. Math. Soc., 135 (2007), pp. 135–139.

[35] ———, *New demiclosedness principles for (firmly) nonexpansive operators*, in Computational and Analytical Mathematics, D. H. Bailey, H. H. Bauschke, P. Borwein, F. Garvan, M. Théra, J. D. Vanderwerff, and H. Wolkowicz, eds., Berlin, 2013, Springer, pp. 19–28.

[36] H. H. Bauschke and J. M. Borwein, *On projection algorithms for solving convex feasibility problems*, SIAM Rev., 38 (1996), pp. 367–426.

[37] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, *Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces*, Commun. Contemp. Math., 3 (2001), pp. 615–647.

[38] ———, *Bregman monotone optimization algorithms*, SIAM J. Control Optim., 42 (2003), pp. 596–636.

[39] H. H. Bauschke, J. M. Borwein, and X. Wang, *Fitzpatrick functions and continuous linear monotone operators*, SIAM J. Optim., 18 (2007), pp. 789–809.

[40] H. H. Bauschke, R. I. Boţ, W. L. Hare, and W. M. Moursi, *Attouch-Théra duality revisited: paramonotonicity and operator splitting*, J. Approx. Theory, 164 (2012), pp. 1065–1084.

[41] H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, eds., *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, Springer-Verlag, New York, 2011.

[42] H. H. Bauschke, J. Chen, and X. Wang, *A projection method for approximating fixed points of quasi nonexpansive mappings without the usual demiclosedness condition*, J. Nonlinear Convex Anal., 15 (2014), pp. 129–135.

[43] H. H. Bauschke and P. L. Combettes, *A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces*, Math. Oper. Res., 26 (2001), pp. 248–264.

[44] ———, *The Baillon-Haddad theorem revisited*, J. Convex Anal., 17 (2010), pp. 781–787.

[45] H. H. Bauschke, R. Goebel, Y. Lucet, and X. Wang, *The proximal average: basic theory*, SIAM J. Optim., 19 (2008), pp. 766–785.
[46] H. H. Bauschke, W. L. Hare, and W. M. Moursi, *Generalized solutions for the sum of two maximally monotone operators*, SIAM J. Control Optim., 52 (2014), pp. 1034–1047.
[47] H. H. Bauschke, D. A. McLaren, and H. S. Sendov, *Fitzpatrick functions: inequalities, examples, and remarks on a problem by S. Fitzpatrick*, J. Convex Anal., 13 (2006), pp. 499–523.
[48] H. H. Bauschke, S. M. Moffat, and X. Wang, *Firmly nonexpansive mappings and maximally monotone operators: correspondence and duality*, Set-Valued Var. Anal., 20 (2012), pp. 131–153.
[49] ———, *Near equality, near convexity, sums of maximally monotone operators, and averages of firmly nonexpansive mappings*, Math. Program., 139 (2013), pp. 55–70.
[50] H. H. Bauschke, D. Noll, and H. M. Phan, *Linear and strong convergence of algorithms involving averaged nonexpansive operators*, J. Math. Anal. Appl., 421 (2015), pp. 1–20.
[51] H. H. Bauschke, C. Wang, X. Wang, and J. Xu, *On subgradient projectors*, SIAM J. Optim., 25 (2015), pp. 1064–1082.
[52] ———, *On the finite convergence of a projected cutter method*, J. Optim. Theory Appl., 165 (2015), pp. 901–916.
[53] H. H. Bauschke, X. Wang, and L. Yao, *Monotone linear relations: maximality and Fitzpatrick functions*, J. Convex Anal., 16 (2009), pp. 673–686.
[54] ———, *Rectangularity and paramonotonicity of maximally monotone operators*, Optimization, 63 (2014), pp. 487–504.
[55] A. Beck, *Introduction to Nonlinear Optimization*, SIAM, Philadelphia, PA, 2014.
[56] S. R. Becker and P. L. Combettes, *An algorithm for splitting parallel sums of linearly composed monotone operators, with applications to signal recovery*, J. Nonlinear Convex Anal., 15 (2014), pp. 137–159.
[57] A. Ben-Israel and T. N. E. Greville, *Generalized Inverses*, Springer-Verlag, New York, second ed., 2003.
[58] A. Ben-Tal and A. Nemirovski, *Non-Euclidean restricted memory level method for large-scale convex optimization*, Math. Program. Ser. A, 102 (2005), pp. 407–456.
[59] H. Berens and U. Westphal, *Kodissipative metrische Projektionen in normierten linearen Räumen*, in Linear Spaces and Approximation, P. L. Butzer and B. Szőkefalvi-Nagy, eds., Birkhäuser, Basel, 1978, pp. 119–130.
[60] C. Berge, *Espaces Topologiques: Fonctions Multivoques*, Dunod, Paris, deuxième éd., 1966. English translation of the first edition: *Topological Spaces – Including a Treatment of Multi-Valued Functions, Vector Spaces and Convexity*, Dover, Mineola, NY, 2010.

[61] C. Berge and A. Ghouila-Houri, *Programmes, Jeux et Réseaux de Transport*, Dunod, Paris, 1962. English translation: *Programming, Games and Transportation Networks*, Wiley, New York, 1965.

[62] D. P. Bertsekas, *Convex Analysis and Optimization*, Athena Scientific, Nashua, NH, 2003.

[63] E. Bishop and R. R. Phelps, *The support functionals of a convex set*, in Proc. Sympos. Pure Math., Vol. VII, Amer. Math. Soc., Providence, R.I., 1963, pp. 27–35.

[64] H. Blumberg, *On convex functions*, Trans. Amer. Math. Soc., 20 (1919), pp. 40–44.

[65] T. Bonnesen and W. Fenchel, *Theorie der konvexen Körper*, Springer-Verlag, Berlin, 1934. English translation: *Theory of Convex Bodies*, BCS Associates, Moscow, ID, 1987.

[66] J. M. Borwein and M. Fabián, *On convex functions having points of Gâteaux differentiability which are not points of Fréchet differentiability*, Canad. J. Math., 45 (1993), pp. 1121–1134.

[67] J. M. Borwein, S. Fitzpatrick, and R. Girgensohn, *Subdifferentials whose graphs are not norm×weak* closed*, Canad. Math. Bull., 46 (2003), pp. 538–545.

[68] J. M. Borwein and R. Goebel, *Notions of relative interior in Banach spaces*, J. Math. Sci. (N. Y.), 115 (2003), pp. 2542–2553.

[69] J. M. Borwein and A. S. Lewis, *Partially finite convex programming I – Quasi relative interiors and duality theory*, Math. Programming Ser. B, 57 (1992), pp. 15–48.

[70] ——, *Partially finite convex programming II – Explicit lattice models*, Math. Programming Ser. B, 57 (1992), pp. 49–83.

[71] ——, *Convex Analysis and Nonlinear Optimization*, Springer-Verlag, New York, second ed., 2006.

[72] J. M. Borwein and J. D. Vanderwerff, *Convex functions of Legendre type in general Banach spaces*, J. Convex Anal., 8 (2001), pp. 569–581.

[73] ——, *Convex Functions*, Cambridge University Press, 2010.

[74] J. M. Borwein and H. Wiersma, *Asplund decomposition of monotone operators*, SIAM J. Optim., 18 (2007), pp. 946–960.

[75] J. M. Borwein and Q. J. Zhu, *Techniques of Variational Analysis*, Springer-Verlag, New York, 2005.

[76] R. I. Boţ, *Conjugate Duality in Convex Optimization*, vol. 637 of Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin, 2010.

[77] R. I. Boţ and S. László, *On the generalized parallel sum of two maximal monotone operators of Gossez type (D)*, J. Math. Anal. Appl., 391 (2012), pp. 82–98.

[78] N. Bourbaki, *Fonctions d'une Variable Réelle (Théorie Élémentaire) – Chapitres I à III*, Hermann, Paris, 1949.

[79] ———, *Intégration, Chapitres 1 à 4*, Hermann, Paris, 1965. English translation: *Integration I*, Springer-Verlag, New York, 2004.

[80] ———, *Espaces Vectoriels Topologiques, Chapitres 1 à 5*, Masson, Paris, 1981. English translation: *Topological Vector Spaces – Chapters 1–5*, Springer-Verlag, New York, 1987.

[81] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.

[82] J. P. Boyle and R. L. Dykstra, *A method for finding projections onto the intersection of convex sets in Hilbert spaces*, in Advances in Order Restricted Statistical Inference, vol. 37 of Lecture Notes in Statistics, Springer-Verlag, Berlin, 1986, pp. 28–47.

[83] L. M. Brègman, *The method of successive projection for finding a common point of convex sets*, Soviet Math. Dokl., 6 (1965), pp. 688–692.

[84] H. Brézis, *Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert*, North-Holland/Elsevier, New York, 1973.

[85] ———, *Analyse Fonctionnelle*, Dunod, Paris, 1983.

[86] H. Brézis and F. E. Browder, *Nonlinear ergodic theorems*, Bull. Amer. Math. Soc., 82 (1976), pp. 959–961.

[87] H. Brézis and A. Haraux, *Image d'une somme d'opérateurs monotones et applications*, Israel J. Math., 23 (1976), pp. 165–186.

[88] H. Brézis and P.-L. Lions, *Produits infinis de résolvantes*, Israel J. Math., 29 (1978), pp. 329–345.

[89] L. M. Briceño-Arias and P. L. Combettes, *Convex variational formulation with smooth coupling for multicomponent signal decomposition and recovery*, Numer. Math. Theory Methods Appl., 2 (2009), pp. 485–508.

[90] ———, *A monotone + skew splitting model for composite monotone inclusions in duality*, SIAM J. Optim., 21 (2011), pp. 1230–1250.

[91] A. Brøndsted and R. T. Rockafellar, *On the subdifferentiability of convex functions*, Proc. Amer. Math. Soc., 16 (1965), pp. 605–611.

[92] F. E. Browder, *Existence of periodic solutions for nonlinear equations of evolution*, Proc. Nat. Acad. Sci. U.S.A., 53 (1965), pp. 1100–1103.

[93] ———, *Fixed-point theorems for noncompact mappings in Hilbert space*, Proc. Nat. Acad. Sci. U.S.A., 53 (1965), pp. 1272–1276.

[94] ———, *Multi-valued monotone nonlinear mappings and duality mappings in Banach spaces*, Trans. Amer. Math. Soc., 118 (1965), pp. 338–351.

[95] ———, *Existence and approximation of solutions of nonlinear variational inequalities*, Proc. Nat. Acad. Sci. U.S.A., 56 (1966), pp. 1080–1086.

[96] ———, *Convergence theorems for sequences of nonlinear operators in Banach spaces*, Math. Z., 100 (1967), pp. 201–225.

[97] ——, *Nonlinear maximal monotone operators in Banach space*, Math. Ann., 175 (1968), pp. 89–113.

[98] ——, *Nonlinear Operators and Nonlinear Equations of Evolution in Banach Spaces*, in Nonlinear Functional Analysis (Proc. Sympos. Pure Math., Vol. XVIII, Part 2, Chicago, IL, 1968), Amer. Math. Soc., Providence, RI, 1976, pp. 1–308. Preprint version 1968.

[99] F. E. Browder and W. V. Petryshyn, *Construction of fixed points of nonlinear mappings in Hilbert space*, J. Math. Anal. Appl., 20 (1967), pp. 197–228.

[100] R. E. Bruck and S. Reich, *Nonexpansive projections and resolvents of accretive operators in Banach spaces*, Houston J. Math., 3 (1977), pp. 459–470.

[101] H. Brunn, *Über Kerneigebiete*, Math. Ann., 73 (1913), pp. 436–440.

[102] L. N. H. Bunt, *Bijdrage tot de Theorie der Convexe Puntverzamelingen*, PhD thesis, Rijksuniversiteit Groningen, 1934.

[103] R. S. Burachik and A. N. Iusem, *Set-Valued Mappings and Enlargements of Monotone Operators*, Springer-Verlag, New York, 2008.

[104] R. S. Burachik and V. Jeyakumar, *A new geometric condition for Fenchel's duality in infinite dimensional spaces*, Math. Program. Ser. B, 104 (2005), pp. 229–233.

[105] R. S. Burachik and B. F. Svaiter, *ϵ-enlargements of maximal monotone operators in Banach spaces*, Set-Valued Anal., 7 (1999), pp. 117–132.

[106] ——, *Maximal monotonicity, conjugation and the duality product*, Proc. Amer. Math. Soc., 131 (2003), pp. 2379–2383.

[107] D. Butnariu and A. N. Iusem, *Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization*, Kluwer, Dordrecht, 2000.

[108] C. Castaing and M. Valadier, *Convex Analysis and Measurable Multifunctions*, vol. 580 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1977.

[109] J. Céa, *Optimisation – Théorie et Algorithmes*, Dunod, Paris, 1971.

[110] A. Cegielski, *Iterative Methods for Fixed Point Problems in Hilbert Spaces*, Springer, New York, 2012.

[111] Y. Censor, T. Elfving, and G. T. Herman, *Averaging strings of sequential iterations for convex feasibility problems*, in Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications, D. Butnariu, Y. Censor, and S. Reich, eds., Amsterdam, 2001, North-Holland, pp. 101–113.

[112] Y. Censor, A. N. Iusem, and S. A. Zenios, *An interior point method with Bregman functions for the variational inequality problem with paramonotone operators*, Math. Programming Ser. A, 81 (1998), pp. 373–400.

[113] Y. Censor and E. Tom, *Convergence of string-averaging projection schemes for inconsistent convex feasibility problems*, Optim. Methods Softw., 18 (2003), pp. 543–554.

[114] Y. Censor and S. A. Zenios, *Parallel Optimization*, Oxford University Press, New York, 1997.

[115] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, *A variational formulation for frame-based inverse problems*, Inverse Problems, 23 (2007), pp. 1495–1518.

[116] G. Choquet, *Cours de Topologie*, Masson, Paris, 1964. English translation: *Topology*, Wiley, New York, 1966.

[117] J. Coleman and G. Walker, *The mathematics of chaos*, in Pandemonium, Killing Joke, ed., London, 1994, Butterfly Rec.

[118] P. L. Combettes, *The foundations of set theoretic estimation*, Proc. IEEE, 81 (1993), pp. 182–208.

[119] ——, *Construction d'un point fixe commun à une famille de contractions fermes*, C. R. Acad. Sci. Paris Sér. I Math., 320 (1995), pp. 1385–1390.

[120] ——, *The convex feasibility problem in image recovery*, in Advances in Imaging and Electron Physics, P. Hawkes, ed., vol. 95, Academic Press, San Diego, CA, 1996, pp. 155–270.

[121] ——, *Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections*, IEEE Trans. Image Proc., 6 (1997), pp. 493–506.

[122] ——, *Hilbertian convex feasibility problem: Convergence of projection methods*, Appl. Math. Optim., 35 (1997), pp. 311–330.

[123] ——, *Strong convergence of block-iterative outer approximation methods for convex optimization*, SIAM J. Control Optim., 38 (2000), pp. 538–565.

[124] ——, *Quasi-Fejérian analysis of some optimization algorithms*, in Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications, D. Butnariu, Y. Censor, and S. Reich, eds., Amsterdam, 2001, North-Holland, pp. 115–152.

[125] ——, *Solving monotone inclusions via compositions of nonexpansive averaged operators*, Optimization, 53 (2004), pp. 475–504.

[126] ——, *Iterative construction of the resolvent of a sum of maximal monotone operators*, J. Convex Anal., 16 (2009), pp. 727–748.

[127] ——, *Perspective functions: Properties, constructions, and examples*, Set-Valued Var. Anal., 26 (2018), pp. 247–264.

[128] P. L. Combettes, Đinh Dũng, and B. C. Vũ, *Dualization of signal recovery problems*, Set-Valued Var. Anal., 18 (2010), pp. 373–404.

[129] P. L. Combettes and S. A. Hirstoaga, *Equilibrium programming in Hilbert spaces*, J. Nonlinear Convex Anal., 6 (2005), pp. 117–136.

[130] ——, *Approximating curves for nonexpansive and monotone operators*, J. Convex Anal., 13 (2006), pp. 633–646.

[131] P. L. Combettes and C. L. Müller, *Perspective functions: Proximal calculus and applications in high-dimensional statistics*, J. Math. Anal. Appl., 457 (2018), pp. 1283–1306.

[132] P. L. Combettes and Q. V. Nguyen, *Solving composite monotone inclusions in reflexive Banach spaces by constructing best Bregman approximations from their Kuhn-Tucker set*, J. Convex Anal., 23 (2016), pp. 481–510.

[133] P. L. Combettes and J.-C. Pesquet, *A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery*, IEEE J. Select. Topics Signal Process., 1 (2007), pp. 564–574.

[134] ——, *Proximal thresholding algorithm for minimization over orthonormal bases*, SIAM J. Optim., 18 (2007), pp. 1351–1376.

[135] ——, *A proximal decomposition method for solving convex variational inverse problems*, Inverse Problems, 24 (2008), art. 065014, 27 pp.

[136] ——, *Proximal splitting methods in signal processing*, in Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Springer, New York, 2011, pp. 185–212.

[137] ——, *Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators*, Set-Valued Var. Anal., 20 (2012), pp. 307–330.

[138] P. L. Combettes and B. C. Vũ, *Variable metric forward-backward splitting with applications to monotone inclusions in duality*, Optimization, 63 (2014), pp. 1289–1318.

[139] P. L. Combettes and V. R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul., 4 (2005), pp. 1168–1200.

[140] P. L. Combettes and I. Yamada, *Compositions and convex combinations of averaged nonexpansive operators*, J. Math. Anal. Appl., 425 (2015), pp. 55–70.

[141] R. Cominetti, J. A. Soto, and J. Vaisman, *On the rate of convergence of Krasnosel'skiĭ-Mann iterations and their connection with sums of Bernoullis*, Israel J. Math., 199 (2014), pp. 757–772.

[142] J. B. Conway, *A Course in Functional Analysis*, Springer-Verlag, New York, second ed., 1990.

[143] C. Davis, *All convex invariant functions of hermitian matrices*, Arch. Math., 8 (1957), pp. 276–278.

[144] H. Debrunner and P. Flor, *Ein Erweiterungssatz für monotone Mengen*, Arch. Math., 15 (1964), pp. 445–447.

[145] V. F. Demyanov and A. M. Rubinov, *Priblizhennye Metody Resheniya Ékstremal'nykh Zadach*, Leningrad University Press, Saint Petersburg, 1968. English translation: *Approximate Methods in Optimization Problems*, Elsevier, New York, 1970.

[146] Z. Denkowski, S. Migórski, and N. S. Papageorgiou, *An Introduction to Nonlinear Analysis – Theory*, Kluwer, Dordrecht, 2003.

[147] F. DEUTSCH, *Best Approximation in Inner Product Spaces*, Springer-Verlag, New York, 2001.

[148] J. DIESTEL, *Sequences and Series in Banach Spaces*, Springer-Verlag, New York, NY, 1984.

[149] J. A. DIEUDONNÉ, *Foundations of Modern Analysis*, Academic Press, New York, second ed., 1969.

[150] ———, *History of Functional Analysis*, North-Holland Publishing Co., Amsterdam, 1981.

[151] A. L. DONTCHEV AND T. ZOLEZZI, *Well-Posed Optimization Problems*, vol. 1543 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1993.

[152] J. DOUGLAS AND H. H. RACHFORD, *On the numerical solution of heat conduction problems in two and three space variables*, Trans. Amer. Math. Soc., 82 (1956), pp. 421–439.

[153] A. YA. DUBOVITSKII AND A. A. MILYUTIN, *Extremum problems in the presence of restrictions*, USSR Comp. Math. Math. Phys., 5 (1965), pp. 1–80.

[154] R. J. DUFFIN, *Convex analysis treated by linear programming*, Math. Programming, 4 (1973), pp. 125–143.

[155] N. DUNFORD AND J. T. SCHWARTZ, *Linear Operators I – General Theory*, Wiley, New York, 1958.

[156] W. F. EBERLEIN, *Weak compactness in Banach spaces. I*, Proc. Nat. Acad. Sci. U.S.A., 33 (1947), pp. 51–53.

[157] C. ECKART AND G. YOUNG, *The approximation of one matrix by another of lower rank*, Psychometrika, 1 (1936), pp. 211–218.

[158] J. ECKSTEIN AND D. P. BERTSEKAS, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program., Ser. A, 55 (1992), pp. 293–318.

[159] M. R. EDWARDS, *Five classes of monotone linear relations and operators*, in Computational and Analytical Mathematics, D. H. Bailey, H. H. Bauschke, P. Borwein, F. Garvan, M. Théra, J. D. Vanderwerff, and H. Wolkowicz, eds., Berlin, 2013, Springer.

[160] I. EKELAND, *On the variational principle*, J. Math. Anal. Appl., 47 (1974), pp. 324–353.

[161] ———, *Nonconvex minimization problems*, Bull. Amer. Math. Soc., 1 (1979), pp. 443–474.

[162] ———, *Convexity Methods in Hamiltonian Mechanics*, Springer-Verlag, Berlin, 1990.

[163] I. EKELAND AND G. LEBOURG, *Generic Fréchet-differentiability and perturbed optimization problems in Banach spaces*, Trans. Amer. Math. Soc., 224 (1976), pp. 193–216.

[164] I. EKELAND AND R. TÉMAM, *Analyse Convexe et Problèmes Variationnels*, Dunod, Paris, 1974. English translation: *Convex Analysis and Variational Problems*, SIAM, Philadelphia, PA, 1999.

[165] I. I. Eremin, *Generalization of the relaxation method of Motzkin and Agmon*, Uspehi Mat. Nauk, 20 (1965), pp. 183–187.

[166] ——, *Methods of Fejér approximations in convex programming*, Math. Notes, 3 (1968), pp. 139–149.

[167] ——, *Fejér mappings and convex programming*, Siberian Math. J., 10 (1969), pp. 762–772.

[168] E. Ernst and M. Théra, *Boundary half-strips and the strong chip*, SIAM J. Optim., 18 (2007), pp. 834–852.

[169] M. Fabian, P. Habala, P. Hájek, V. Montesinos Santalucía, J. Pelant, and V. Zizler, *Functional Analysis and Infinite-Dimensional Geometry*, Springer-Verlag, New York, NY, 2001.

[170] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems I*, Springer-Verlag, New York, 2003.

[171] ——, *Finite-Dimensional Variational Inequalities and Complementarity Problems II*, Springer-Verlag, New York, 2003.

[172] L. Fejér, *Über die Lage der Nullstellen von Polynomen, die aus Minimumforderungen gewisser Art entspringen*, Math. Ann., 85 (1922), pp. 41–48.

[173] W. Fenchel, *On conjugate convex functions*, Canadian J. Math., 1 (1949), pp. 73–77.

[174] ——, *Convex Cones, Sets and Functions*, tech. report, Princeton University, Princeton, NJ, 1953. From 1951 lecture notes.

[175] ——, *Convexity through the ages*, in Convexity and Its Applications, Birkhäuser, Basel, 1983, pp. 120–130.

[176] S. Fitzpatrick, *Representing monotone operators by convex functions*, in Workshop/Miniconference on Functional Analysis and Optimization, vol. 20 of Proc. Centre Math. Anal. Austral. Nat. Univ., Austral. Nat. Univ., 1988, pp. 59–65.

[177] S. Fitzpatrick and R. R. Phelps, *Bounded approximants to monotone operators on Banach spaces*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 9 (1992), pp. 573–595.

[178] E. M. Gafni and D. P. Bertsekas, *Two-metric projection methods for constrained optimization*, SIAM J. Control Optim., 22 (1984), pp. 936–964.

[179] N. Ghoussoub, *Self-Dual Partial Differential Systems and Their Variational Principles*, Springer, New York, 2009.

[180] J. R. Giles, *Convex Analysis with Application in the Differentiation of Convex Functions*, Pitman, Boston, MA, 1982.

[181] K. Goebel and W. A. Kirk, *Topics in Metric Fixed Point Theory*, Cambridge University Press, Cambridge, 1990.

[182] K. Goebel and S. Reich, *Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings*, Marcel Dekker, New York, 1984.

[183] R. Goebel, *Self-dual smoothing of convex and saddle functions*, J. Convex Anal., 15 (2008), pp. 179–190.

[184] E. G. Golshtein and N. V. Tretyakov, *Modifitsirovannye Funktsii Lagranzha – Teoriya i Metody Optimizatsii*, Nauka, Moscow, 1989. English translation: *Modified Lagrangians and Monotone Maps in Optimization*, Wiley, New York, 1996.

[185] M. S. Gowda and M. Teboulle, *A comparison of constraint qualifications in infinite-dimensional convex programming*, SIAM J. Control Optim., 28 (1990), pp. 925–935.

[186] C. W. Groetsch, *A note on segmenting Mann iterates*, J. Math. Anal. Appl., 40 (1972), pp. 369–372.

[187] ——, *Generalized Inverses of Linear Operators: Representation and Approximation*, Marcel Dekker, New York, 1977.

[188] L. G. Gubin, B. T. Polyak, and E. V. Raik, *The method of projections for finding the common point of convex sets*, USSR Comp. Math. Math. Phys., 7 (1967), pp. 1–24.

[189] O. Güler, *On the convergence of the proximal point algorithm for convex minimization*, SIAM J. Control Optim., 29 (1991), pp. 403–419.

[190] N. Hadjisavvas, S. Komlósi, and S. Schaible, eds., *Handbook of Generalized Convexity and Generalized Monotonicity*, Springer-Verlag, New York, 2005.

[191] N. Hadjisavvas and S. Schaible, *On a generalization of paramonotone maps and its application to solving the Stampacchia variational inequality*, Optimization, 55 (2006), pp. 593–604.

[192] P. R. Halmos, *A Hilbert Space Problem Book*, Springer-Verlag, New York, second ed., 1982.

[193] I. Halperin, *The product of projection operators*, Acta Sci. Math. (Szeged), 23 (1962), pp. 96–99.

[194] B. Halpern, *Fixed points of nonexpanding maps*, Bull. Amer. Math. Soc., 73 (1967), pp. 957–961.

[195] A. Haraux, *Nonlinear Evolution Equations: Global Behavior of Solutions*, vol. 841 of Lecture Notes in Mathematics, Springer-Verlag, New York, 1981.

[196] G. H. Hardy, J. E. Littlewood, and Pólya, *Inequalities*, Cambridge University Press, Cambridge, 1934.

[197] Y. Haugazeau, *Sur les Inéquations Variationnelles et la Minimisation de Fonctionnelles Convexes*, thèse de doctorat, Université de Paris, 1968.

[198] F. Hausdorff, *Über halbstetige Funktionen und deren Verallgemeinerung*, Math. Z., 5 (1919), pp. 292–309.

[199] G. Helmberg, *Curiosities concerning weak topology in Hilbert space*, Amer. Math. Monthly, 113 (2006), pp. 447–452.

[200] H. Heuser, *Funktionalanalysis*, B. G. Teubner, Stuttgart, vierte Aufl., 2006.

[201] J.-B. Hiriart-Urruty, *Du calcul différentiel au calcul variationnel : un aperçu de l'évolution de P. Fermat à nos jours*, Quadrature, no. 70, (2008), pp. 8–18.

[202] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms I*, Springer-Verlag, Berlin, 1993.

[203] ———, *Convex Analysis and Minimization Algorithms II*, Springer-Verlag, Berlin, 1993.

[204] R. B. Holmes, *Geometric Functional Analysis and Its Applications*, Springer-Verlag, New York, 1975.

[205] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1990.

[206] H. S. Hundal, *An alternating projection that does not converge in norm*, Nonlinear Anal., 57 (2004), pp. 35–61.

[207] A. D. Ioffe and V. L. Levin, *Subdifferentials of convex functions*, Trans. Moscow Math. Soc., 26(1972) (1974), pp. 1–72.

[208] A. D. Ioffe and V. M. Tihomirov, *Teoriya Ékstremal'nykh Zadach*, Nauka, Moscow, 1974. English translation: *Theory of Extremal Problems*, North-Holland, Amsterdam, 1979.

[209] A. N. Iusem, *On some properties of paramonotone operators*, J. Convex Anal., 5 (1998), pp. 269–278.

[210] J. L. W. V. Jensen, *Sur les fonctions convexes et les inégalités entre les valeurs moyennes*, Acta Math., 30 (1906), pp. 175–193.

[211] B. Jessen, *To saetninger om konvekse punktmaengder*, Mat. Tiddskrift B, 1940 (1940), pp. 66–70.

[212] V. Jeyakumar and H. Wolkowicz, *Generalizations of Slater's constraint qualification for infinite convex programs*, Math. Programming Ser. B, 57 (1992), pp. 85–101.

[213] R. I. Kačurovskiĭ, *On monotone operators and convex functionals*, Uspehi Mat. Nauk, 15 (1960), pp. 213–215.

[214] J. L. Kelley, *General Topology*, Van Nostrand, New York, 1955. Reprinted by Springer-Verlag, 1975.

[215] J. L. Kelley and I. Namioka, *Linear Topological Spaces*, Springer-Verlag, New York, 1976.

[216] P. S. Kenderov, *Monotone operators in Asplund spaces*, C. R. Acad. Bulgare Sci., 30 (1977), pp. 963–964.

[217] M. D. Kirszbraun, *Über die zusammenziehenden und Lipschitzschen Transformationen*, Fundam. Math., 22 (1934), pp. 77–108.

[218] V. Klee, *Convex sets in linear spaces*, Duke Math. J., 18 (1951), pp. 443–466.

[219] ———, *Convexity of Chebyshev sets*, Math. Ann., 142 (1961), pp. 292–304.

[220] ———, *What is a convex set?*, Amer. Math. Monthly, 78 (1971), pp. 616–631.

[221] D. E. Knuth, *Things a Computer Scientist Rarely Talks About*, Center for the Study of Language and Information, Stanford, CA, 2001.

[222] V. Komornik, *A simple proof of Farkas' lemma*, Amer. Math. Monthly, 105 (1998), pp. 949–950.

[223] M. A. Krasnosel'skiĭ, *Two remarks on the method of successive approximations*, Uspekhi Mat. Nauk, 10 (1955), pp. 123–127.

[224] E. Kreyszig, *Introductory Functional Analysis with Applications*, Wiley, New York, 1989.

[225] N. Kritikos, *Sur quelques propriétés des ensembles convexes*, Bull. Math. Soc. Roumaine Sci., 40 (1938), pp. 87–92.

[226] N. Krylov, *Properties of monotone mappings*, Lithuanian Math. J., 22 (1982), pp. 140–145.

[227] P.-J. Laurent, *Approximation et Optimisation*, Hermann, Paris, 1972.

[228] P.-J. Laurent and B. Martinet, *Méthodes duales pour le calcul du minimum d'une fonction convexe sur une intersection de convexes*, in Symposium on Optimization, vol. 132 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1970, pp. 159–180.

[229] P. D. Lax and A. N. Milgram, *Parabolic equations*, in Contributions to the Theory of Partial Differential Equations, Annals of Mathematics Studies, no. 33, Princeton University Press, Princeton, NJ, 1954, pp. 167–190.

[230] B. Lemaire, *The proximal algorithm*, in New Methods in Optimization and Their Industrial Uses, J.-P. Penot, ed., Birkhäuser, Basel, 1989, pp. 73–87.

[231] C. Lescarret, *Cas d'addition des applications monotones maximales dans un espace de Hilbert*, C. R. Acad. Sci. Paris Sér. A, 261 (1965), pp. 1160–1163.

[232] E. S. Levitin, *Teoriya Vozmushcheniĭ v Matematicheskom Programmirovanii i ee Prilozheniya*, Nauka, Moscow, 1992. English translation: *Perturbation Theory in Mathematical Programming and Its Applications*, Wiley, New York, 1994.

[233] E. S. Levitin and B. T. Polyak, *Constrained minimization methods*, USSR Comp. Math. Math. Phys., 6 (1966), pp. 1–50.

[234] A. S. Lewis, *Convex analysis on the hermitian matrices*, SIAM J. Optim., 6 (1996), pp. 164–177.

[235] A. S. Lewis and H. S. Sendov, *Nonsmooth analysis of singular values. I. Theory*, Set-Valued Anal., 13 (2005), pp. 213–241.

[236] J. Lieutaud, *Approximation d'Opérateurs par des Méthodes de Décomposition*, thèse de doctorat, Université de Paris, 1969.

[237] J.-L. Lions, *Quelques Méthodes de Résolution des Problèmes aux Limites Non Linéaires*, Dunod, Paris, 1969.

[238] J.-L. Lions and E. Magenes, *Problèmes aux Limites Non Homogènes et Applications – Vol. 1*, Dunod, Paris, 1968. English translation: *Non-Homogeneous Boundary Value Problems and Applications*, Springer-Verlag, New York, 1972.

[239] J.-L. Lions and G. Stampacchia, *Inéquations variationnelles non coercives*, C. R. Acad. Sci. Paris Sér. A, 261 (1965), pp. 25–27.

[240] P.-L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16 (1979), pp. 964–979.

[241] D. T. Luc and S. Swaminathan, *A characterization of convex functions*, Nonlinear Anal., 20 (1993), pp. 697–701.

[242] R. Lucchetti, *Convexity and Well-Posed Problems*, Springer-Verlag, New York, 2006.

[243] D. G. Luenberger, *Optimization by Vector Space Methods*, Wiley, New York, 1969.

[244] L. A. Lusternik and V. I. Sobolev, *Èlementy Funktsionalnogo Analiza*, Nauka, Moscow, second ed., 1965. English translation: *Elements of Functional Analysis*, Wiley, New York, 1974.

[245] G. G. Magaril-Il'yaev and V. M. Tikhomirov, *Vypuklyi Analiz Teoriya i Prilozheniya*, YRCC, Moscow, 2000. English translation: *Convex Analysis: Theory and Applications*, AMS, Providence, RI, 2003.

[246] S. Mandelbrojt, *Sur les fonctions convexes*, C. R. Acad. Sci. Paris Sér. A, 209 (1939), pp. 977–978.

[247] W. R. Mann, *Mean value methods in iteration*, Proc. Amer. Math. Soc., 4 (1953), pp. 506–510.

[248] B. Martinet, *Régularisation d'inéquations variationnelles par approximations successives*, Rev. Française Informat. Recherche Opérationnelle, 4 (1970), pp. 154–158.

[249] ——, *Algorithmes pour la Résolution de Problèmes d'Optimisation et de Minimax*, thèse de doctorat, Université de Grenoble, 1972. http://hal.archives-ouvertes.fr/docs/00/28/42/55/PDF/These_Martinet.Bernard_1972.pdf.

[250] ——, *Détermination approchée d'un point fixe d'une application pseudo-contractante. Cas de l'application prox*, C. R. Acad. Sci. Paris Sér. A, 274 (1972), pp. 163–165.

[251] J.-E. Martínez-Legaz, *Generalized conjugation and related topics*, in Generalized Convexity and Fractional Programming with Economic Applications (Pisa, 1988), vol. 345 of Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin, 1990, pp. 168–197.

[252] J.-E. Martínez-Legaz and M. Théra, *A convex representation of maximal monotone operators*, J. Nonlinear Convex Anal., 2 (2001), pp. 243–247.

[253] R. E. Megginson, *An Introduction to Banach Space Theory*, Springer-Verlag, New York, 1998.

[254] B. Mercier, *Lectures on Topics in Finite Element Solution of Elliptic Problems*, Tata Institute of Fundamental Research, Bombay, 1979.

[255] ——, *Inéquations Variationnelles de la Mécanique*, Publications Mathématiques d'Orsay no. 80, Université de Paris-Sud, Département de Mathématique, Orsay, 1980.

[256] H. Minkowski, *Gesammelte Abhandlungen*, Teubner, Berlin, 1911.

[257] G. J. Minty, *On the maximal domain of a "monotone" function*, Michigan Math. J., 8 (1961), pp. 135–137.

[258] ———, *Monotone (nonlinear) operators in Hilbert space*, Duke Math. J., 29 (1962), pp. 341–346.

[259] ———, *On the monotonicity of the gradient of a convex function*, Pacific J. Math., 14 (1964), pp. 243–247.

[260] L. Mirsky, *A trace inequality of John von Neumann*, Monatsh. Math., 79 (1975), pp. 303–306.

[261] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation I – Basic Theory*, Springer-Verlag, Berlin, 2006.

[262] ———, *Variational Analysis and Generalized Differentiation II – Applications*, Springer-Verlag, Berlin, 2006.

[263] B. S. Mordukhovich and N. M. Nam, *An Easy Path to Convex Analysis and Applications*, Morgan and Claypool Publishers, Williston, VT, 2014.

[264] J. J. Moreau, *Décomposition orthogonale d'un espace hilbertien selon deux cônes mutuellement polaires*, C. R. Acad. Sci. Paris Sér. A, 255 (1962), pp. 238–240.

[265] ———, *Fonctions convexes duales et points proximaux dans un espace hilbertien*, C. R. Acad. Sci. Paris Sér. A, 255 (1962), pp. 2897–2899.

[266] ———, *Fonctionnelles sous-différentiables*, C. R. Acad. Sci. Paris Sér. A, 257 (1963), pp. 4117–4119.

[267] ———, *Inf-convolution des fonctions numériques sur un espace vectoriel*, C. R. Acad. Sci. Paris Sér. A, 256 (1963), pp. 5047–5049.

[268] ———, *Propriétés des applications prox*, C. R. Acad. Sci. Paris Sér. A, 256 (1963), pp. 1069–1071.

[269] ———, *Remarques sur les fonctions à valeurs dans* $[-\infty, +\infty]$ *définies sur un demi-groupe*, C. R. Acad. Sci. Paris Sér. A, 257 (1963), pp. 3107–3109.

[270] ———, *Sur la fonction polaire d'une fonction semi-continue supérieurement*, C. R. Acad. Sci. Paris Sér. A, 258 (1964), pp. 1128–1130.

[271] ———, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.

[272] ———, *Convexity and duality*, in Functional Analysis and Optimization, Academic Press, New York, 1966, pp. 145–169.

[273] ———, *Fonctionnelles Convexes*, Séminaire Jean Leray sur les Équations aux Dérivées Partielles, no. 2, Collège de France, Paris, 1966–1967. http://www.numdam.org/item?id=SJL_1966-1967___2_1_0. Corrected reprint: Edizioni del Dipartimento di Ingegneria Civile, Università di Roma Tor Vergata, Rome, 2003.

[274] ———, *Inf-convolution, sous-additivité, convexité des fonctions numériques*, J. Math. Pures Appl., 49 (1970), pp. 109–154.

[275] ———, *Weak and strong solutions of dual problems*, in Contributions to Nonlinear Functional Analysis, Academic Press, New York, 1971, pp. 181–214.

[276] ———, *Un cas de convergence des itérées d'une contraction d'un espace hilbertien*, C. R. Acad. Sci. Paris Sér. A, 286 (1978), pp. 143–144.

[277] T. S. Motzkin, *Sur quelques propriétés caractéristiques des ensembles convexes*, Atti Accad. Naz. Lincei, Rend., VI, 21 (1935), pp. 562–567.

[278] A. Moudafi, *On the stability of the parallel sum of maximal monotone operators*, J. Math. Anal. Appl., 199 (1996), pp. 478–488.

[279] Yu. Nesterov, *Introductory Lectures on Convex Optimization*, Kluwer, Boston, MA, 2004.

[280] J. Neveu, *Bases Mathématiques du Calcul des Probabilités*, Masson, Paris, deuxième éd., 1970. English translation of the first edition: *Mathematical Foundations of the Calculus of Probability*, Holden-Day, San Francisco, 1965.

[281] N. Ogura and I. Yamada, *Non-strictly convex minimization over the fixed point set of an asymptotically shrinking nonexpansive mapping*, Numer. Funct. Anal. Optim., 23 (2002), pp. 113–137.

[282] Z. Opial, *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bull. Amer. Math. Soc., 73 (1967), pp. 591–597.

[283] G. B. Passty, *The parallel sum of nonlinear monotone operators*, Nonlinear Anal., 10 (1986), pp. 215–227.

[284] B. Pauwels, *Opérateurs de Projection Sous-Différentielle*, master's thesis, Université Pierre et Marie Curie, Paris, France, 2012. http://arxiv.org/abs/1403.7237.

[285] A. Pazy, *On the asymptotic behavior of iterates of nonexpansive mappings in Hilbert space*, Israel J. Math., 26 (1977), pp. 197–204.

[286] D. W. Peaceman and H. H. Rachford, *The numerical solution of parabolic and elliptic differential equations*, J. Soc. Ind. Appl. Math., 3 (1955), pp. 28–41.

[287] T. Pennanen, *Dualization of generalized equations of maximal monotone type*, SIAM J. Optim., 10 (2000), pp. 809–835.

[288] R. R. Phelps, *Convex Functions, Monotone Operators and Differentiability*, vol. 1364 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, second ed., 1993.

[289] R. R. Phelps and S. Simons, *Unbounded linear monotone operators on nonreflexive Banach spaces*, J. Convex Anal., 5 (1998), pp. 303–328.

[290] G. Pierra, *Éclatement de contraintes en parallèle pour la minimisation d'une forme quadratique*, in Optimization Techniques Modeling and Optimization in the Service of Man Part 2, vol. 41 of Lecture Notes in Computer Science, Berlin/Heidelberg, 1976, Springer-Verlag, pp. 200–218.

[291] A. Pietsch, *History of Banach Spaces and Linear Operators*, Birkhäuser, Boston, MA, 2007.

[292] B. T. Polyak, *Existence theorems and convergence of minimizing sequences in extremum problems with restrictions*, Soviet Math. Dokl., 7 (1966), pp. 72–75.

[293] ——, *Minimization of unsmooth functionals*, USSR Comp. Math. Math. Phys., 9 (1969), pp. 14–29.

[294] ——, *Vvedenie v Optimizatsiyu*, Nauka, Moscow, 1983. English translation: *Introduction to Optimization*, Optimization Software, New York, 1987.

[295] ——, *History of mathematical programming in the USSR: analyzing the phenomenon*, Math. Programming Ser. B, 91 (2002), pp. 401–416.

[296] T. Popoviciu, *Les Fonctions Convexes*, Hermann, Paris, 1944.

[297] L. C. Potter and K. S. Arun, *A dual approach to linear inverse problems with convex constraints*, SIAM J. Control Optim., 31 (1993), pp. 1080–1092.

[298] B. N. Pshenichnyĭ and Yu. M. Danilin, *Chislennye Metody v Ékstremal'nykh Zadachakh*, Nauka, Moscow, 1975. English translation: *Numerical Methods in Extremal Problems*, Mir, Moscow, 1978.

[299] E. Raik, *A class of iterative methods with Fejér-monotone sequences*, Eesti NSV Tead. Akad. Toimetised Füüs.-Mat., 18 (1969), pp. 22–26.

[300] S. Reich, *Weak convergence theorems for nonexpansive mappings in Banach spaces*, J. Math. Anal. Appl., 67 (1979), pp. 274–276.

[301] ——, *A limit theorem for projections*, Linear and Multilinear Algebra, 13 (1983), pp. 281–290.

[302] S. Reich and S. Simons, *Fenchel duality, Fitzpatrick functions and the Kirszbraun-Valentine extension theorem*, Proc. Amer. Math. Soc., 133 (2005), pp. 2657–2660.

[303] M. Reichert-Hahn, *Einführung in die lineare Funktionalanalysis*, tech. report, Goethe Universität, Frankfurt am Main, Germany, 2010. Lecture notes.

[304] F. Riesz, *Zur Theorie des Hilbertschen Raumes*, Acta Sci. Math. (Szeged), 7 (1934), pp. 34–38.

[305] F. Riesz and B. Sz.-Nagy, *Leçons d'Analyse Fonctionnelle*, Gauthier-Villars, Paris, cinquième éd., 1968. English translation of the second edition: *Functional Analysis*, Dover, Mineola, NY, 1990.

[306] S. M. Robinson, *Composition duality and maximal monotonicity*, Math. Program., 85 (1999), pp. 1–13.

[307] R. T. Rockafellar, *Convex Functions and Dual Extremum Problems*, PhD thesis, Harvard University, 1963.

[308] ——, *Duality theorems for convex functions*, Bull. Amer. Math. Soc., 70 (1964), pp. 189–192.

[309] ——, *Characterization of the subdifferentials of convex functions*, Pacific J. Math., 17 (1966), pp. 497–510.

[310] ——, *Duality and stability in extremum problems involving convex functions*, Pacific J. Math., 21 (1967), pp. 167–187.

[311] ——, *Integrals which are convex functionals*, Pacific J. Math., 24 (1968), pp. 525–539.

[312] ——, *Local boundedness of nonlinear, monotone operators*, Michigan Math. J., 16 (1969), pp. 397–407.

[313] ——, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
[314] ——, *On the maximal monotonicity of subdifferential mappings*, Pacific J. Math., 33 (1970), pp. 209–216.
[315] ——, *On the maximality of sums of nonlinear monotone operators*, Trans. Amer. Math. Soc., 149 (1970), pp. 75–88.
[316] ——, *On the virtual convexity of the domain and range of a nonlinear maximal monotone operator*, Math. Ann., 185 (1970), pp. 81–90.
[317] ——, *Convex integral functionals and duality*, in Contributions to Nonlinear Functional Analysis, E. H. Zarantonello, ed., Academic Press, New York, 1971, pp. 215–236.
[318] ——, *Conjugate Duality and Optimization*, SIAM, Philadelphia, PA, 1974.
[319] ——, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 14 (1976), pp. 877–898.
[320] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.
[321] H. Rubin and O. Wesler, *A note on convexity in Euclidean n-space*, Proc. Amer. Math. Soc., 9 (1958), pp. 522–523.
[322] W. Rudin, *Functional Analysis*, McGraw-Hill Inc., New York, second ed., 1991.
[323] A. Ruszczyński, *Nonlinear Optimization*, Princeton University Press, Princeton, NJ, 2006.
[324] H. Schaefer, *Über die Methode sukzessiver Approximationen*, Jber. Deutsch. Math. Verein., 59 (1957), pp. 131–140.
[325] L. Schwartz, *Analyse – Topologie Générale et Analyse Fonctionnelle*, Hermann, Paris, deuxième éd., 1980.
[326] R. E. Showalter, *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations*, American Mathematical Society, Providence, RI, 1997.
[327] M. Sibony, *Méthodes itératives pour les équations et inéquations aux dérivées partielles non linéaires de type monotone*, Calcolo, 7 (1970), pp. 65–183.
[328] S. Simons, *The range of a monotone operator*, J. Math. Anal. Appl., 199 (1996), pp. 176–201.
[329] ——, *Minimax and Monotonicity*, vol. 1693 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1998.
[330] ——, *LC-functions and maximal monotonicity*, J. Nonlinear Convex Anal., 7 (2006), pp. 123–138.
[331] ——, *From Hahn-Banach to Monotonicity*, vol. 1693 of Lecture Notes in Mathematics, Springer-Verlag, New York, second ed., 2008.
[332] S. Simons and C. Zălinescu, *A new proof for Rockafellar's characterization of maximal monotone operators*, Proc. Amer. Math. Soc., 132 (2004), pp. 2969–2972.

[333] ———, *Fenchel duality, Fitzpatrick functions and maximal monotonicity*, J. Nonlinear Convex Anal., 6 (2005), pp. 1–22.

[334] I. Singer, *The Theory of Best Approximation and Functional Analysis*, SIAM, Philadelphia, PA, 1974.

[335] V. Šmulian, *Sur la dérivabilité de la norme dans l'espace de Banach*, C. R. (Doklady) Acad. Sci. URSS (N. S.), 27 (1940), pp. 643–648.

[336] ———, *Über lineare topologische Räume*, Rec. Math. [Mat. Sbornik] N. S., 7 (1940), pp. 425–448.

[337] M. V. Solodov and B. F. Svaiter, *Forcing strong convergence of proximal point iterations in a Hilbert space*, Math. Program., 87 (2000), pp. 189–202.

[338] J. E. Spingarn, *Partial inverse of a monotone operator*, Appl. Math. Optim., 10 (1983), pp. 247–265.

[339] G. Stampacchia, *Formes bilinéaires coercitives sur les ensembles convexes*, C. R. Acad. Sci. Paris Sér. A, 258 (1964), pp. 4413–4416.

[340] L. A. Steen and J. A. Seebach, Jr., *Counterexamples in Topology*, Dover, Mineola, NY, 1995.

[341] B. F. Svaiter, *On weak convergence of the Douglas-Rachford method*, SIAM J. Control Optim., 49 (2011), pp. 280–287.

[342] B. Sz.-Nagy, *On the set of positive functions in L_2*, Ann. of Math., 39 (1938), pp. 1–13.

[343] W. Takahashi, *Nonlinear Functional Analysis*, Yokohama Publishers, Yokohama, 2000.

[344] C. M. Theobald, *An inequality for the trace of the product of two symmetric matrices*, Math. Proc. Cambridge Philos. Soc., 77 (1975), pp. 265–267.

[345] A. N. Tikhonov and V. Y. Arsenin, *Metody Resheniya Nekorrektnykh Zadach*, Nauka, Moscow, 1974. English translation: *Solutions of Ill-Posed Problems*, Winston, New York, 1977.

[346] P. Tseng, *Further applications of a splitting algorithm to decomposition in variational inequalities and convex programming*, Math. Programming Ser. B, 48 (1990), pp. 249–263.

[347] ———, *Applications of a splitting algorithm to decomposition in convex programming and variational inequalities*, SIAM J. Control Optim., 29 (1991), pp. 119–138.

[348] ———, *A modified forward-backward splitting method for maximal monotone mappings*, SIAM J. Control Optim., 38 (2000), pp. 431–446.

[349] M. Valadier, *Sous-différentiels d'une borne supérieure et d'une somme continue de fonctions convexes*, C. R. Acad. Sci. Paris Sér. A, 268 (1969), pp. 39–42.

[350] F. A. Valentine, *A Lipschitz condition preserving extension for a vector function*, Amer. J. Math., 67 (1945), pp. 83–93.

[351] ———, *Convex Sets*, McGraw-Hill Inc., New York, 1964.

[352] L. Veselý, *Local uniform boundedness principle for families of ϵ-monotone operators*, Nonlinear Anal., 24 (1995), pp. 1299–1304.

[353] C. Villani, *Topics in Optimal Transportation*, Amer. Math. Soc., Providence, RI, 2003.

[354] A. A. Vladimirov, J. E. Nesterov, and J. N. Čekanov, *Uniformly convex functionals*, Vestnik Moskov. Univ. Ser. XV Vychisl. Mat. Kibernet., 3 (1978), pp. 12–23.

[355] J. von Neumann, *On rings of operators. Reduction theory*, Ann. of Math., 50 (1949), pp. 401–485.

[356] J. Warga, *Minimizing certain convex functions*, J. Soc. Indust. Appl. Math., 11 (1963), pp. 588–593.

[357] U. Westphal and T. Schwartz, *Farthest points and monotone operators*, Bull. Austral. Math. Soc., 58 (1998), pp. 75–92.

[358] A. Wilansky, *Functional Analysis*, Blaisdell (Ginn), New York, 1964.

[359] R. Wittmann, *Approximation of fixed points of nonexpansive mappings*, Arch. Math., 58 (1992), pp. 486–491.

[360] I. Yamada, *Kougaku no Tameno Kansu Kaiseki (Functional Analysis for Engineering)*, Suurikougaku-Sha/Saiensu-Sha, Tokyo, 2009.

[361] Y.-L. Yu, *On decomposing the proximal map*, in Advances in Neural Information Processing Systems 26, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, eds., Curran Associates, Inc., 2013, pp. 91–99.

[362] C. Zălinescu, *On uniformly convex functions*, J. Math. Anal. Appl., 95 (1983), pp. 344–374.

[363] ———, *Convex Analysis in General Vector Spaces*, World Scientific Publishing Co. Inc., River Edge, NJ, 2002.

[364] E. H. Zarantonello, *Solving functional equations by contractive averaging*, Tech. Report 160, Math. Res. Center U.S. Army, Madison, University of Wisconsin, June 1960.

[365] ———, *Projections on convex sets in Hilbert space and spectral theory. I. Projections on convex sets*, in Contributions to Nonlinear Functional Analysis, E. H. Zarantonello, ed., Academic Press, New York, 1971, pp. 237–341.

[366] E. Zeidler, *Nonlinear Functional Analysis and Its Applications III: Variational Methods and Optimization*, Springer-Verlag, New York, 1985.

[367] ———, *Nonlinear Functional Analysis and Its Applications II/A: Linear Monotone Operators*, Springer-Verlag, New York, 1990.

[368] ———, *Nonlinear Functional Analysis and Its Applications II/B: Nonlinear Monotone Operators*, Springer-Verlag, New York, 1990.

[369] ———, *Nonlinear Functional Analysis and Its Applications I: Fixed Point Theorems*, Springer-Verlag, New York, 1993.

Index

H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics,
DOI 10.1007/978-3-319-48311-5