MORGAN & CLAYPOOL PUBLISHERS

An Easy Path to Convex Analysis and Applications

Boris S. Mordukhovich
Nguyen Mau Nam

Synthesis Lectures on Mathematics and Statistics

Steven G. Krantz, *Series Editor*

An Easy Path to Convex Analysis and Applications

Synthesis Lectures on Mathematics and Statistics

Editor
Steven G. Krantz, ***Washington University, St. Louis***

An Easy Path to Convex Analysis and Applications
Boris S. Mordukhovich and Nguyen Mau Nam
2014

Applications of Affine and Weyl Geometry
Eduardo García-Río, Peter Gilkey, Stana Nikčević, and Ramón Vázquez-Lorenzo
2013

Essentials of Applied Mathematics for Engineers and Scientists, Second Edition
Robert G. Watts
2012

Chaotic Maps: Dynamics, Fractals, and Rapid Fluctuations
Goong Chen and Yu Huang
2011

Matrices in Engineering Problems
Marvin J. Tobias
2011

The Integral: A Crux for Analysis
Steven G. Krantz
2011

Statistics is Easy! Second Edition
Dennis Shasha and Manda Wilson
2010

Lectures on Financial Mathematics: Discrete Asset Pricing
Greg Anderson and Alec N. Kercheval
2010

An Easy Path to Convex Analysis and Applications

Boris S. Mordukhovich and Nguyen Mau Nam

www.morganclaypool.com

ISBN: 9781627052375 paperback
ISBN: 9781627052382 ebook

DOI 10.2200/S00554ED1V01Y201312MAS014

A Publication in the Morgan & Claypool Publishers series
SYNTHESIS LECTURES ON MATHEMATICS AND STATISTICS

Lecture #14
Series Editor: Steven G. Krantz, *Washington University, St. Louis*
Series ISSN
Synthesis Lectures on Mathematics and Statistics
Print 1938-1743 Electronic 1938-1751

An Easy Path to Convex Analysis and Applications

Boris S. Mordukhovich
Wayne State University

Nguyen Mau Nam
Portland State University

SYNTHESIS LECTURES ON MATHEMATICS AND STATISTICS #14

MORGAN & CLAYPOOL PUBLISHERS

ABSTRACT

Convex optimization has an increasing impact on many areas of mathematics, applied sciences, and practical applications. It is now being taught at many universities and being used by researchers of different fields. As convex analysis is the mathematical foundation for convex optimization, having deep knowledge of convex analysis helps students and researchers apply its tools more effectively. The main goal of this book is to provide an easy access to the most fundamental parts of convex analysis and its applications to optimization. Modern techniques of variational analysis are employed to clarify and simplify some basic proofs in convex analysis and build the theory of generalized differentiation for convex functions and sets in finite dimensions. We also present new applications of convex analysis to location problems in connection with many interesting geometric problems such as the Fermat-Torricelli problem, the Heron problem, the Sylvester problem, and their generalizations. Of course, we do not expect to touch every aspect of convex analysis, but the book consists of sufficient material for a first course on this subject. It can also serve as supplemental reading material for a course on convex optimization and applications.

KEYWORDS

Affine set, Carathéodory theorem, convex function, convex set, directional derivative, distance function, Fenchel conjugate, Fermat-Torricelli problem, generalized differentiation, Helly theorem, minimal time function, Nash equilibrium, normal cone, Radon theorem, optimal value function, optimization, smallest enclosing ball problem, set-valued mapping, subdifferential, subgradient, subgradient algorithm, support function, Weiszfeld algorithm

The first author dedicates this book to the loving memory of his father
Sholim Mordukhovich *(1924–1993),*
a kind man and a brave warrior.

The second author dedicates this book to the memory of his parents.

Contents

Preface

Some geometric properties of convex sets and, to a lesser extent, of convex functions had been studied before the 1960s by many outstanding mathematicians, first of all by Hermann Minkowski and Werner Fenchel. At the beginning of the 1960s convex analysis was greatly developed in the works of R. Tyrrell Rockafellar and Jean-Jacques Moreau who initiated a systematic study of this new field. As a fundamental part of variational analysis, convex analysis contains a generalized differentiation theory that can be used to study a large class of mathematical models with no differentiability assumptions imposed on their initial data. The importance of convex analysis for many applications in which convex optimization is the first to name has been well recognized. The presence of convexity makes it possible not only to comprehensively investigate qualitative properties of optimal solutions and derive necessary and sufficient conditions for optimality but also to develop effective numerical algorithms to solve convex optimization problems, even with nondifferentiable data. Convex analysis and optimization have an increasing impact on many areas of mathematics and applications including control systems, estimation and signal processing, communications and networks, electronic circuit design, data analysis and modeling, statistics, economics and finance, etc.

There are several fundamental books devoted to different aspects of convex analysis and optimization. Among them we mention "Convex Analysis" by Rockafellar [26], "Convex Analysis and Minimization Algorithms" (in two volumes) by Hiriart-Urruty and Lemaréchal [8] and its abridge version [9], "Convex Analysis and Nonlinear Optimization" by Borwein and Lewis [4], "Introductory Lectures on Convex Optimization" by Nesterov [21], and "Convex Optimization" by Boyd and Vandenberghe [3] as well as other books listed in the bibliography below.

In this big picture of convex analysis and optimization, our book serves as a bridge for students and researchers who have just started using convex analysis to reach deeper topics in the field. We give detailed proofs for most of the results presented in the book and also include many figures and exercises for better understanding the material. The powerful geometric approach developed in modern variational analysis is adopted and simplified in the convex case in order to provide the reader with an easy path to access generalized differentiation of convex objects in finite dimensions. In this way, the book also serves as a starting point for the interested reader to continue the study of nonconvex variational analysis and applications. It can be of interest from this viewpoint to experts in convex and variational analysis. Finally, the application part of this book not only concerns the classical topics of convex optimization related to optimality conditions and subgradient algorithms but also presents some recent while easily understandable qualitative and numerical results for important location problems.

The book consists of four chapters and is organized as follows. In Chapter 1 we study fundamental properties of convex sets and functions while paying particular attention to classes of convex functions important in optimization. Chapter 2 is mainly devoted to developing basic calculus rules for normals to convex sets and subgradients of convex functions that are in the mainstream of convex theory. Chapter 3 concerns some additional topics of convex analysis that are largely used in applications. Chapter 4 is fully devoted to applications of basic results of convex analysis to problems of convex optimization and selected location problems considered from both qualitative and numerical viewpoints. Finally, we present at the end of the book Solutions and Hints to selected exercises.

Exercises are given at the end of each chapter while figures and examples are provided throughout the whole text. The list of references contains books and selected papers, which are closely related to the topics considered in the book and may be helpful to the reader for advanced studies of convex analysis, its applications, and further extensions.

Since only elementary knowledge in linear algebra and basic calculus is required, this book can be used as a textbook for both undergraduate and graduate level courses in convex analysis and its applications. In fact, the authors have used these lecture notes for teaching such classes in their universities as well as while visiting some other schools. We hope that the book will make convex analysis more accessible to large groups of undergraduate and graduate students, researchers in different disciplines, and practitioners.

Boris S. Mordukhovich and Nguyen Mau Nam
December 2013

Acknowledgments

The first author acknowledges continued support by grants from the National Science Foundation and the second author acknowledges support by the Simons Foundation.

Boris S. Mordukhovich and Nguyen Mau Nam
December 2013

List of Symbols

$\mathbb{R}$	the real numbers
$\overline{\mathbb{R}} = (-\infty, \infty]$	the extended real line
$\mathbb{R}_+$	the nonnegative real numbers
$\mathbb{R}_>$	the positive real numbers
$\mathbb{N}$	the positive integers
$\operatorname{co} \Omega$	convex hull of Ω
$\operatorname{aff} \Omega$	affine hull of Ω
$\operatorname{int} \Omega$	interior of Ω
$\operatorname{ri} \Omega$	relative interior of Ω
$\operatorname{span} \Omega$	linear subspace generated by Ω
$\operatorname{cone} \Omega$	cone generated by Ω
K_Ω	convex cone generated by Ω
$\dim \Omega$	dimension of Ω
$\overline{\Omega}$	closure of Ω
$\operatorname{bd} \Omega$	boundary of Ω
$I\!B$	closed unit ball
$I\!B(\bar{x}; r)$	closed ball with center $\bar{x}$ and radius r
$\operatorname{dom} f$	domain of f
$\operatorname{epi} f$	epigraph of f
$\operatorname{gph} F$	graph of mapping F
$\mathcal{L}[a, b]$	line connecting a and b
$d(x; \Omega)$	distance from x to Ω
$\Pi(x; \Omega)$	projection of x to Ω
$\langle x, y \rangle$	inner product of x and y
A^*	adjoint/transpose of linear mapping/matrix A
$N(\bar{x}; \Omega)$	normal cone to Ω at $\bar{x}$
$\ker A$	kernel of linear mapping A
$D^*F(\bar{x}, \bar{y})$	coderivative to F at $(\bar{x}, \bar{y})$
$T(\bar{x}; \Omega)$	tangent cone to Ω at $\bar{x}$
F_∞	horizon/asymptotic cone of F
$\mathcal{T}_\Omega^F$	minimal time function defined by constant dynamic F and target set Ω
$L(x, \lambda)$	Lagrangian
ρ_F	Minkowski gauge of F

$\Pi_F(\cdot;\Omega)$	generalized projection defined by the minimal time function
$G_F(\bar{x},t)$	generalized ball defined by dynamic F with center $\bar{x}$ and radius t

CHAPTER 1

Convex Sets and Functions

This chapter presents definitions, examples, and basic properties of convex sets and functions in the Euclidean space $\mathbb{R}^n$ and also contains some related material.

1.1 PRELIMINARIES

We start with reviewing classical notions and properties of the Euclidean space $\mathbb{R}^n$. The proofs of the results presented in this section can be found in standard books on advanced calculus and linear algebra.

Let us denote by $\mathbb{R}^n$ the set of all $n-$tuples of real numbers $x = (x_1, \ldots, x_n)$. Then $\mathbb{R}^n$ is a linear space with the following operations:

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n),$$
$$\lambda(x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n),$$

where $(x_1, \ldots, x_n), (y_1, \ldots, y_n) \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$. The zero element of $\mathbb{R}^n$ and the number zero of $\mathbb{R}$ are often denoted by the same notation 0 if no confusion arises.

For any $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, we identify it with the column vector $x = [x_1, \ldots, x_n]^T$, where the symbol "T" stands for *vector transposition*. Given $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$, the *inner product* of x and y is defined by

$$\langle x, y\rangle := \sum_{i=1}^{n} x_i y_i.$$

The following proposition lists some important properties of the inner product in $\mathbb{R}^n$.

Proposition 1.1 *For $x, y, z \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$, we have:*

(i) $\langle x, x\rangle \geq 0$, *and* $\langle x, x\rangle = 0$ *if and only if* $x = 0$.
(ii) $\langle x, y\rangle = \langle y, x\rangle$.
(iii) $\langle \lambda x, y\rangle = \lambda\langle x, y\rangle$.
(iv) $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$.

The Euclidean *norm* of $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ is defined by

$$\|x\| := \sqrt{x_1^2 + \ldots + x_n^2}.$$

It follows directly from the definition that $\|x\| = \sqrt{\langle x, x\rangle}$.

Proposition 1.2 *For any $x, y \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$, we have:*

(i) $\|x\| \geq 0$, *and* $\|x\| = 0$ *if and only if* $x = 0$.
(ii) $\|\lambda x\| = |\lambda| \cdot \|x\|$.
(iii) $\|x + y\| \leq \|x\| + \|y\|$ (*the triangle inequality*).
(iv) $|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$ (*the Cauchy-Schwarz inequality*).

Using the Euclidean norm allows us to introduce the balls in $\mathbb{R}^n$, which can be used to define other topological notions in $\mathbb{R}^n$.

Definition 1.3 *The* CLOSED BALL *centered at $\bar{x}$ with radius $r \geq 0$ and the* CLOSED UNIT BALL *of* $\mathbb{R}^n$ *are defined, respectively, by*

$$I\!B(\bar{x}; r) := \big\{x \in \mathbb{R}^n \;\big|\; \|x - \bar{x}\| \leq r\big\} \quad \text{and} \quad I\!B := \big\{x \in \mathbb{R}^n \;\big|\; \|x\| \leq 1\big\}.$$

It is easy to see that $I\!B = I\!B(0; 1)$ and $I\!B(\bar{x}; r) = \bar{x} + rI\!B$.

Definition 1.4 *Let $\Omega \subset \mathbb{R}^n$. Then $\bar{x}$ is an* INTERIOR POINT *of Ω if there is $\delta > 0$ such that*

$$I\!B(\bar{x}; \delta) \subset \Omega.$$

The set of all interior points of Ω is denoted by $\operatorname{int} \Omega$. *Moreover, Ω is said to be* OPEN *if every point of Ω is its interior point.*

We get that Ω is open if and only if for every $\bar{x} \in \Omega$ there is $\delta > 0$ such that $I\!B(\bar{x}; \delta) \subset \Omega$. It is obvious that the empty set $\emptyset$ and the whole space $\mathbb{R}^n$ are open. Furthermore, any *open ball* $B(\bar{x}; r) := \{x \in \mathbb{R}^n \mid \|x - \bar{x}\| < r\}$ centered at $\bar{x}$ with radius r is open.

Definition 1.5 *A set $\Omega \subset \mathbb{R}^n$ is* CLOSED *if its complement $\Omega^c = \mathbb{R}^n \setminus \Omega$ is open in $\mathbb{R}^n$.*

It follows that the empty set, the whole space, and any ball $I\!B(\bar{x}; r)$ are closed in $\mathbb{R}^n$.

Proposition 1.6 **(i)** *The union of any collection of open sets in $\mathbb{R}^n$ is open.*

(ii) *The intersection of any finite collection of open sets in $\mathbb{R}^n$ is open.*
(iii) *The intersection of any collection of closed sets in $\mathbb{R}^n$ is closed.*
(iv) *The union of any finite collection of closed sets in $\mathbb{R}^n$ is closed.*

Definition 1.7 *Let $\{x_k\}$ be a sequence in $\mathbb{R}^n$. We say that $\{x_k\}$* CONVERGES *to $\bar{x}$ if $\|x_k - \bar{x}\| \to 0$ as $k \to \infty$. In this case we write*

$$\lim_{k\to\infty} x_k = \bar{x}.$$

This notion allows us to define the following important topological concepts for sets.

Definition 1.8 *Let Ω be a nonempty subset of $\mathbb{R}^n$. Then:*

(i) *The* CLOSURE *of Ω, denoted by $\overline{\Omega}$ or $\mathrm{cl}\,\Omega$, is the collection of limits of all convergent sequences belonging to Ω.*
(ii) *The* BOUNDARY *of Ω, denoted by $\mathrm{bd}\,\Omega$, is the set $\overline{\Omega} \setminus \mathrm{int}\,\Omega$.*

We can see that the closure of Ω is the intersection of all closed sets containing Ω and that the interior of Ω is the union of all open sets contained in Ω. It follows from the definition that $\bar{x} \in \overline{\Omega}$ if and only if for any $\delta > 0$ we have $I\!B(\bar{x};\delta) \cap \Omega \neq \emptyset$. Furthermore, $\bar{x} \in \mathrm{bd}\,\Omega$ if and only if for any $\delta > 0$ the closed ball $I\!B(\bar{x};\Omega)$ intersects both sets Ω and its complement Ω^c.

Definition 1.9 *Let $\{x_k\}$ be a sequence in $\mathbb{R}^n$ and let $\{k_\ell\}$ be a strictly increasing sequence of positive integers. Then the new sequence $\{x_{k_\ell}\}$ is called a* SUBSEQUENCE *of $\{x_k\}$.*

We say that a set Ω is *bounded* if it is contained in a ball centered at the origin with some radius $r > 0$, i.e., $\Omega \subset I\!B(0;r)$. Thus a sequence $\{x_k\}$ is bounded if there is $r > 0$ with

$$\|x_k\| \le r \quad \text{for all} \quad k \in \mathbb{N}\,.$$

The following important result is known as the *Bolzano-Weierstrass theorem*.

Theorem 1.10 Any bounded sequence in $\mathbb{R}^n$ contains a convergent subsequence.

The next concept plays a very significant role in analysis and optimization.

Definition 1.11 *We say that a set Ω is* COMPACT *in $\mathbb{R}^n$ if every sequence in Ω contains a subsequence converging to some point in Ω.*

The following result is a consequence of the Bolzano-Weierstrass theorem.

Theorem 1.12 A subset Ω of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.

For subsets Ω, Ω_1, and Ω_2 of $\mathbb{R}^n$ and for $\lambda \in \mathbb{R}$, we define the operations:

$$\Omega_1 + \Omega_2 := \big\{x + y \;\big|\; x \in \Omega_1,\ y \in \Omega_2\big\}, \quad \lambda\Omega := \big\{\lambda x \;\big|\; x \in \Omega\big\}.$$

The next proposition can be proved easily.

Proposition 1.13 *Let Ω_1 and Ω_2 be two subsets of $\mathbb{R}^n$.*

(i) *If Ω_1 is open or Ω_2 is open, then $\Omega_1 + \Omega_2$ is open.*
(ii) *If Ω_1 is closed and Ω_2 is compact, then $\Omega_2 + \Omega_2$ is closed.*

Recall now the notions of *bounds* for subsets of the real line.

Definition 1.14 *Let D be a subset of the real line. A number $m \in \mathbb{R}$ is a* LOWER BOUND *of D if we have*

$$x \geq m \ \text{ for all } \ x \in D.$$

If the set D has a lower bound, then it is BOUNDED BELOW*. Similarly, a number $M \in \mathbb{R}$ is an* UPPER BOUND *of D if*

$$x \leq M \ \text{ for all } \ x \in D,$$

and D is BOUNDED ABOVE *if it has an upper bound. Furthermore, we say that the set D is* BOUNDED *if it is simultaneously bounded below and above.*

Now we are ready to define the concepts of *infimum* and *supremum* of sets.

Definition 1.15 *Let $D \subset \mathbb{R}$ be nonempty and bounded below. The infimum of D, denoted by* $\inf D$*, is the greatest lower bound of D. When D is nonempty and bounded above, its supremum, denoted by* $\sup D$*, is the least upper bound of D. If D is not bounded below (resp. above), then we set* $\inf D := -\infty$ *(resp.* $\sup D := \infty$*). We also use the convention that* $\inf \emptyset := \infty$ *and* $\sup \emptyset := -\infty$.

The following fundamental axiom ensures that these notions are well-defined.

Completeness Axiom. *For every nonempty subset D of $\mathbb{R}$ that is bounded above, the least upper bound of D exists as a real number.*

Using the Completeness Axiom, it is easy to see that if a nonempty set is bounded below, then its greatest lower bound exists as a real number.

Throughout the book we consider for convenience *extended-real-valued* functions, which take values in $\overline{\mathbb{R}} := (-\infty, \infty]$. The usual conventions of extended arithmetics are that $a + \infty = \infty$ for any $a \in \mathbb{R}$, $\infty + \infty = \infty$, and $t \cdot \infty = \infty$ for $t > 0$.

Definition 1.16 *Let $f : \Omega \to \overline{\mathbb{R}}$ be an extended-real-valued function and let $\bar{x} \in \Omega$ with $f(\bar{x}) < \infty$. Then f is* CONTINUOUS *at $\bar{x}$ if for any $\epsilon > 0$ there is $\delta > 0$ such that*

$$|f(x) - f(\bar{x})| < \epsilon \ \text{ whenever } \ \|x - \bar{x}\| < \delta, \ x \in \Omega.$$

We say that f is continuous on Ω if it is continuous at every point of Ω.

It is obvious from the definition that if $f : \Omega \to \overline{\mathbb{R}}$ is continuous at $\bar{x}$ (with $f(\bar{x}) < \infty$), then it is finite on the intersection of Ω and a ball centered at $\bar{x}$ with some radius $r > 0$. Furthermore, $f : \Omega \to \overline{\mathbb{R}}$ is continuous at $\bar{x}$ (with $f(\bar{x}) < \infty$) if and only if for every sequence $\{x_k\}$ in Ω converging to $\bar{x}$ the sequence $\{f(x_k)\}$ converges to $f(\bar{x})$.

Definition 1.17 *Let $f : \Omega \to \overline{\mathbb{R}}$ and let $\bar{x} \in \Omega$ with $f(\bar{x}) < \infty$. We say that f has a* LOCAL MINIMUM *at $\bar{x}$ relative to Ω if there is $\delta > 0$ such that*

$$f(x) \geq f(\bar{x}) \ \textit{for all} \ \ x \in I\!B(\bar{x}; \delta) \cap \Omega.$$

We also say that f has a GLOBAL/ABSOLUTE MINIMUM *at $\bar{x}$ relative to Ω if*

$$f(x) \geq f(\bar{x}) \ \textit{for all} \ \ x \in \Omega.$$

The notions of *local and global maxima* can be defined similarly.

Finally in this section, we formulate a fundamental result of mathematical analysis and optimization known as the *Weierstrass existence theorem*.

Theorem 1.18 Let $f : \Omega \to \mathbb{R}$ be a continuous function, where Ω is a nonempty, compact subset of $\mathbb{R}^n$. Then there exist $\bar{x} \in \Omega$ and $\bar{u} \in \Omega$ such that

$$f(\bar{x}) = \inf\big\{f(x) \mid x \in \Omega\big\} \ \text{ and } \ f(\bar{u}) = \sup\big\{f(x) \mid x \in \Omega\big\}.$$

In Section 4.1 we present some "unilateral" versions of Theorem 1.18.

1.2 CONVEX SETS

We start the study of convexity with sets and then proceed to functions. Geometric ideas play an underlying role in convex analysis, its extensions, and applications. Thus we implement the geometric approach in this book.

Given two elements a and b in $\mathbb{R}^n$, define the *interval/line segment*

$$[a, b] := \big\{\lambda a + (1 - \lambda)b \mid \lambda \in [0, 1]\big\}.$$

Note that if $a = b$, then this interval reduces to a singleton $[a, b] = \{a\}$.

Definition 1.19 *A subset Ω of $\mathbb{R}^n$ is* CONVEX *if $[a, b] \subset \Omega$ whenever $a, b \in \Omega$. Equivalently, Ω is convex if $\lambda a + (1 - \lambda)b \in \Omega$ for all $a, b \in \Omega$ and $\lambda \in [0, 1]$.*

Given $\omega_1, \ldots, \omega_m \in \mathbb{R}^n$, the element $x = \sum_{i=1}^m \lambda_i \omega_i$, where $\sum_{i=1}^m \lambda_i = 1$ and $\lambda_i \geq 0$ for some $m \in \mathbb{N}$, is called a *convex combination* of $\omega_1, \ldots, \omega_m$.

Proposition 1.20 *A subset Ω of $\mathbb{R}^n$ is convex if and only if it contains all convex combinations of its elements.*

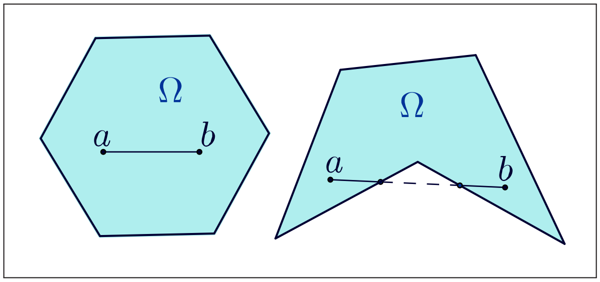

Figure 1.1: Convex set and nonconvex set.

Proof. The sufficient condition is trivial. To justify the necessity, we show by induction that any convex combination $x = \sum_{i=1}^{m} \lambda_i \omega_i$ of elements in Ω is an element of Ω. This conclusion follows directly from the definition for $m = 1, 2$. Fix now a positive integer $m \geq 2$ and suppose that every convex combination of $k \in \mathbb{N}$ elements from Ω, where $k \leq m$, belongs to Ω. Form the convex combination

$$y := \sum_{i=1}^{m+1} \lambda_i \omega_i, \ \sum_{i=1}^{m+1} \lambda_i = 1, \ \lambda_i \geq 0$$

and observe that if $\lambda_{m+1} = 1$, then $\lambda_1 = \lambda_2 = \ldots = \lambda_m = 0$, so $y = \omega_{m+1} \in \Omega$. In the case where $\lambda_{m+1} < 1$ we get the representations

$$\sum_{i=1}^{m} \lambda_i = 1 - \lambda_{m+1} \ \text{ and } \ \sum_{i=1}^{m} \frac{\lambda_i}{1 - \lambda_{m+1}} = 1,$$

which imply in turn the inclusion

$$z := \sum_{i=1}^{m} \frac{\lambda_i}{1 - \lambda_{m+1}} \omega_i \in \Omega.$$

It yields therefore the relationships

$$y = (1 - \lambda_{m+1}) \sum_{i=1}^{m} \frac{\lambda_i}{1 - \lambda_{m+1}} \omega_i + \lambda_{m+1} \omega_{m+1} = (1 - \lambda_{m+1}) z + \lambda_{m+1} \omega_{m+1} \in \Omega$$

and thus completes the proof of the proposition. $\square$

Proposition 1.21 *Let Ω_1 be a convex subset of $\mathbb{R}^n$ and let Ω_2 be a convex subset of $\mathbb{R}^p$. Then the Cartesian product $\Omega_1 \times \Omega_2$ is a convex subset of $\mathbb{R}^n \times \mathbb{R}^p$.*

Proof. Fix $a = (a_1, a_2), b = (b_1, b_2) \in \Omega_1 \times \Omega_2$, and $\lambda \in (0, 1)$. Then we have $a_1, b_1 \in \Omega_1$ and $a_2, b_2 \in \Omega_2$. The convexity of Ω_1 and Ω_2 gives us

$$\lambda a_i + (1-\lambda)b_i \in \Omega_i \ \text{ for } \ i = 1, 2,$$

which implies therefore that

$$\lambda a + (1-\lambda)b = \big(\lambda a_1 + (1-\lambda)b_1, \lambda a_2 + (1-\lambda)b_2\big) \in \Omega_1 \times \Omega_2.$$

Thus the Cartesian product $\Omega_1 \times \Omega_2$ is convex. □

Let us continue with the definition of affine mappings.

Definition 1.22 *A mapping $B : \mathbb{R}^n \to \mathbb{R}^p$ is* AFFINE *if there exist a linear mapping $A : \mathbb{R}^n \to \mathbb{R}^p$ and an element $b \in \mathbb{R}^p$ such that $B(x) = A(x) + b$ for all $x \in \mathbb{R}^n$.*

Every linear mapping is affine. Moreover, $B : \mathbb{R}^n \to \mathbb{R}^p$ is affine if and only if

$$B(\lambda x + (1-\lambda)y) = \lambda B(x) + (1-\lambda)B(y) \text{ for all } x, y \in \mathbb{R}^n \text{ and } \lambda \in \mathbb{R}\,.$$

Now we show that set convexity is preserved under *affine operations*.

Proposition 1.23 *Let $B : \mathbb{R}^n \to \mathbb{R}^p$ be an affine mapping. Suppose that Ω is a convex subset of $\mathbb{R}^n$ and Θ is a convex subset of $\mathbb{R}^p$. Then $B(\Omega)$ is a convex subset of $\mathbb{R}^p$ and $B^{-1}(\Theta)$ is a convex subset of $\mathbb{R}^n$.*

Proof. Fix any $a, b \in B(\Omega)$ and $\lambda \in (0, 1)$. Then $a = B(x)$ and $b = B(y)$ for $x, y \in \Omega$. Since Ω is convex, we have $\lambda x + (1-\lambda)y \in \Omega$. Then

$$\lambda a + (1-\lambda)b = \lambda B(x) + (1-\lambda)B(y) = B(\lambda x + (1-\lambda)y) \in B(\Omega),$$

which justifies the convexity of the image $B(\Omega)$.

Taking now any $x, y \in B^{-1}(\Theta)$ and $\lambda \in (0, 1)$, we get $B(x), B(y) \in \Theta$. This gives us

$$\lambda B(x) + (1-\lambda)B(y) = B\big(\lambda x + (1-\lambda)y\big) \in \Theta$$

by the convexity of Θ. Thus we have $\lambda x + (1-\lambda)y \in B^{-1}(\Theta)$, which verifies the convexity of the inverse image $B^{-1}(\Theta)$. □

Proposition 1.24 *Let $\Omega_1, \Omega_2 \subset \mathbb{R}^n$ be convex and let $\lambda \in \mathbb{R}$. Then both sets $\Omega_1 + \Omega_2$ and $\lambda\Omega_1$ are also convex in $\mathbb{R}^n$.*

Proof. It follows directly from the definitions. □

Next we proceed with *intersections* of convex sets.

Proposition 1.25 *Let $\{\Omega_\alpha\}_{\alpha\in I}$ be a collection of convex subsets of $\mathbb{R}^n$. Then $\bigcap_{\alpha\in I}\Omega_\alpha$ is also a convex subset of $\mathbb{R}^n$.*

Proof. Taking any $a, b \in \bigcap_{\alpha\in I}\Omega_\alpha$, we get that $a, b \in \Omega_\alpha$ for all $\alpha \in I$. The convexity of each Ω_α ensures that $\lambda a + (1-\lambda)b \in \Omega_\alpha$ for any $\lambda \in (0,1)$. Thus $\lambda a + (1-\lambda)b \in \bigcap_{\alpha\in I}\Omega_\alpha$ and the intersection $\bigcap_{\alpha\in I}\Omega_\alpha$ is convex. □

Definition 1.26 *Let Ω be a subset of $\mathbb{R}^n$. The* CONVEX HULL *of Ω is defined by*

$$\mathrm{co}\,\Omega := \bigcap\Big\{C \;\Big|\; C \text{ is convex and } \Omega \subset C\Big\}.$$

The next proposition follows directly from the definition and Proposition 1.25.

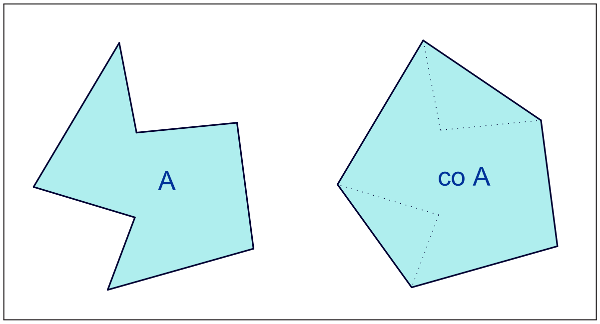

Figure 1.2: Nonconvex set and its convex hull.

Proposition 1.27 *The convex hull $\mathrm{co}\,\Omega$ is the smallest convex set containing Ω.*

Proof. The convexity of the set $\mathrm{co}\,\Omega \supset \Omega$ follows from Proposition 1.25. On the other hand, for any convex set C that contains Ω we clearly have $\mathrm{co}\,\Omega \subset C$, which verifies the conclusion. □

Proposition 1.28 *For any subset Ω of $\mathbb{R}^n$, its convex hull admits the representation*

$$\mathrm{co}\,\Omega = \Big\{\sum_{i=1}^{m}\lambda_i a_i \;\Big|\; \sum_{i=1}^{m}\lambda_i = 1,\ \lambda_i \ge 0,\ a_i \in \Omega,\ m \in \mathbb{N}\Big\}.$$

Proof. Denoting by C the right-hand side of the representation to prove, we obviously have $\Omega \subset C$. Let us check that the set C is convex. Take any $a, b \in C$ and get

$$a := \sum_{i=1}^{p} \alpha_i a_i, \quad b := \sum_{j=1}^{q} \beta_j b_j,$$

where $a_i, b_j \in \Omega$, $\alpha_i, \beta_j \geq 0$ with $\sum_{i=1}^{p} \alpha_i = \sum_{j=1}^{q} \beta_j = 1$, and $p, q \in \mathbb{N}$. It follows easily that for every number $\lambda \in (0, 1)$, we have

$$\lambda a + (1-\lambda)b = \sum_{i=1}^{p} \lambda\alpha_i a_i + \sum_{j=1}^{q} (1-\lambda)\beta_j b_j.$$

Then the resulting equality

$$\sum_{i=1}^{p} \lambda\alpha_i + \sum_{j=1}^{q} (1-\lambda)\beta_j = \lambda \sum_{i=1}^{p} \alpha_i + (1-\lambda) \sum_{j=1}^{q} \beta_j = 1$$

ensures that $\lambda a + (1-\lambda)b \in C$, and thus $\operatorname{co}\Omega \subset C$ by the definition of $\operatorname{co}\Omega$. Fix now any $a = \sum_{i=1}^{m} \lambda_i a_i \in C$ with $a_i \in \Omega \subset \operatorname{co}\Omega$ for $i = 1, \ldots, m$. Since the set $\operatorname{co}\Omega$ is convex, we conclude by Proposition 1.20 that $a \in \operatorname{co}\Omega$ and thus $\operatorname{co}\Omega = C$. □

Proposition 1.29 *The interior* $\operatorname{int}\Omega$ *and closure* $\overline{\Omega}$ *of a convex set* $\Omega \subset \mathbb{R}^n$ *are also convex.*

Proof. Fix $a, b \in \operatorname{int}\Omega$ and $\lambda \in (0, 1)$. Then find an open set V such that

$$a \in V \subset \Omega \quad \text{and so} \quad \lambda a + (1-\lambda)b \in \lambda V + (1-\lambda)b \subset \Omega.$$

Since $\lambda V + (1-\lambda)b$ is open, we get $\lambda a + (1-\lambda)b \in \operatorname{int}\Omega$, and thus the set $\operatorname{int}\Omega$ is convex.

To verify the convexity of $\overline{\Omega}$, we fix $a, b \in \overline{\Omega}$ and $\lambda \in (0, 1)$ and then find sequences $\{a_k\}$ and $\{b_k\}$ in Ω converging to a and b, respectively. By the convexity of Ω, the sequence $\{\lambda a_k + (1-\lambda)b_k\}$ lies entirely in Ω and converges to $\lambda a + (1-\lambda)b$. This ensures the inclusion $\lambda a + (1-\lambda)b \in \overline{\Omega}$ and thus justifies the convexity of the closure $\overline{\Omega}$. □

To proceed further, for any $a, b \in \mathbb{R}^n$, define the interval

$$[a, b) := \big\{\lambda a + (1-\lambda)b \;\big|\; \lambda \in (0, 1]\big\}.$$

We can also define the intervals $(a, b]$ and (a, b) in a similar way.

Lemma 1.30 *For a convex set* $\Omega \subset \mathbb{R}^n$ *with nonempty interior, take any* $a \in \operatorname{int}\Omega$ *and* $b \in \overline{\Omega}$. *Then* $[a, b) \subset \operatorname{int}\Omega$.

Proof. Since $b \in \overline{\Omega}$, for any $\epsilon > 0$, we have $b \in \Omega + \epsilon I\!B$. Take now $\lambda \in (0, 1]$ and let $x_\lambda := \lambda a + (1-\lambda)b$. Choosing $\epsilon > 0$ such that $a + \epsilon \dfrac{2-\lambda}{\lambda} I\!B \subset \Omega$ gives us

$$\begin{aligned} x_\lambda + \epsilon I\!B &= \lambda a + (1-\lambda)b + \epsilon I\!B \\ &\subset \lambda a + (1-\lambda)\big[\Omega + \epsilon I\!B\big] + \epsilon I\!B \\ &= \lambda a + (1-\lambda)\Omega + (1-\lambda)\epsilon I\!B + \epsilon I\!B \\ &\subset \lambda\Big[a + \epsilon\frac{2-\lambda}{\lambda} I\!B\Big] + (1-\lambda)\Omega \\ &\subset \lambda\Omega + (1-\lambda)\Omega \subset \Omega. \end{aligned}$$

This shows that $x_\lambda \in \operatorname{int}\Omega$ and thus verifies the inclusion $[a, b) \subset \operatorname{int}\Omega$. $\square$

Now we establish relationships between taking the interior and closure of convex sets.

Proposition 1.31 *Let $\Omega \subset \mathbb{R}^n$ be a convex set with nonempty interior. Then we have:*
(i) $\overline{\operatorname{int}\Omega} = \overline{\Omega}$ *and* **(ii)** $\operatorname{int}\Omega = \operatorname{int}\overline{\Omega}$.

Proof. **(i)** Obviously, $\overline{\operatorname{int}\Omega} \subset \overline{\Omega}$. For any $b \in \overline{\Omega}$ and $a \in \operatorname{int}\Omega$, define the sequence $\{x_k\}$ by

$$x_k := \frac{1}{k}a + \Big(1 - \frac{1}{k}\Big)b,\ k \in N.$$

Lemma 1.30 ensures that $x_k \in \operatorname{int}\Omega$. Since $x_k \to b$, we have $b \in \overline{\operatorname{int}\Omega}$ and thus verify **(i)**.

(ii) Since the inclusion $\operatorname{int}\Omega \subset \operatorname{int}\overline{\Omega}$ is obvious, it remains to prove the opposite inclusion $\operatorname{int}\overline{\Omega} \subset \operatorname{int}\Omega$. To proceed, fix any $b \in \operatorname{int}\overline{\Omega}$ and $a \in \operatorname{int}\Omega$. If $\epsilon > 0$ is sufficiently small, then

$$c := b + \epsilon(b-a) \in \overline{\Omega}, \text{ and hence } b = \frac{\epsilon}{1+\epsilon}a + \frac{1}{1+\epsilon}c \in (a, c) \subset \operatorname{int}\Omega,$$

which verifies that $\operatorname{int}\overline{\Omega} \subset \operatorname{int}\Omega$ and thus completes the proof. $\square$

1.3 CONVEX FUNCTIONS

This section collects basic facts about general (extended-real-valued) *convex functions* including their analytic and geometric characterizations, important properties as well as their specifications for particular subclasses. We also define *convex set-valued mappings* and use them to study a remarkable class of *optimal value functions* employed in what follows.

Definition 1.32 *Let $f : \Omega \to \overline{\mathbb{R}}$ be an extended-real-valued function defined on a convex set $\Omega \subset \mathbb{R}^n$. Then the function f is* CONVEX *on Ω if*

$$f\big(\lambda x + (1-\lambda)y\big) \le \lambda f(x) + (1-\lambda)f(y) \text{ for all } x, y \in \Omega \text{ and } \lambda \in (0, 1). \tag{1.1}$$

If the inequality in (1.1) *is strict for* $x \neq y$, *then* f *is* STRICTLY CONVEX *on* Ω.

Given a function $f : \Omega \to \overline{\mathbb{R}}$, the *extension* of f to $\mathbb{R}^n$ is defined by

$$\widetilde{f}(x) := \begin{cases} f(x) & \text{if } x \in \Omega, \\ \infty & \text{otherwise.} \end{cases}$$

Obviously, if f is convex on Ω, where Ω is a convex set, then $\widetilde{f}$ is convex on $\mathbb{R}^n$. Furthermore, if $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a convex function, then it is also convex on every convex subset of $\mathbb{R}^n$. This allows to consider without loss of generality extended-real-valued convex functions on the whole space $\mathbb{R}^n$.

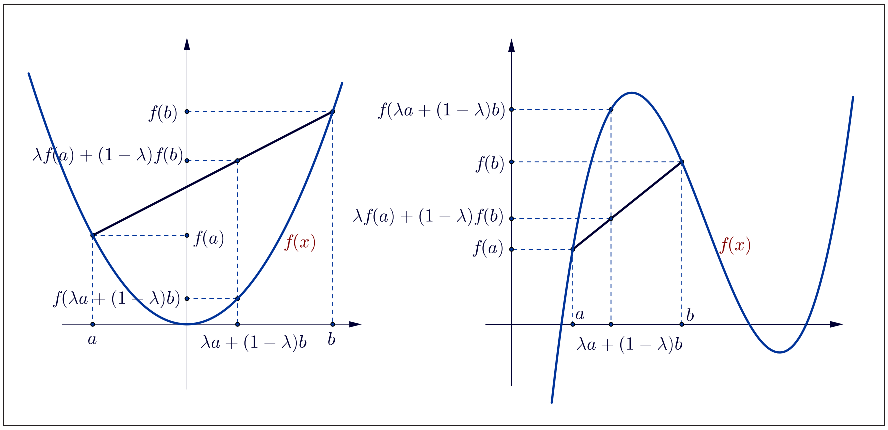

Figure 1.3: Convex function and nonconvex function.

Definition 1.33 *The* DOMAIN *and* EPIGRAPH *of* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *are defined, respectively, by*

$$\operatorname{dom} f := \{x \in \mathbb{R}^n \mid f(x) < \infty\} \quad \textit{and}$$

$$\operatorname{epi} f := \{(x,t) \in \mathbb{R}^n \times \mathbb{R} \mid x \in \mathbb{R}^n,\ t \geq f(x)\} = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} \mid x \in \operatorname{dom} f,\ t \geq f(x)\}.$$

Let us illustrate the convexity of functions by examples.

Example 1.34 The following functions are convex:

(i) $f(x) := \langle a, x\rangle + b$ for $x \in \mathbb{R}^n$, where $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$.
(ii) $g(x) := \|x\|$ for $x \in \mathbb{R}^n$.
(iii) $h(x) := x^2$ for $x \in \mathbb{R}$.

Indeed, the function f in **(i)** is convex since

$$\begin{aligned} f\big(\lambda x + (1-\lambda)y\big) &= \langle a, \lambda x + (1-\lambda)y\rangle + b = \lambda\langle a, x\rangle + (1-\lambda)\langle a, y\rangle + b \\ &= \lambda(\langle a, x\rangle + b) + (1-\lambda)(\langle a, y\rangle + b) \\ &= \lambda f(x) + (1-\lambda)f(y) \text{ for all } x, y \in \mathbb{R}^n \text{ and } \lambda \in (0,1). \end{aligned}$$

The function g in **(ii)** is convex since for $x, y \in \mathbb{R}^n$ and $\lambda \in (0,1)$, we have

$$g\big(\lambda x + (1-\lambda)y\big) = \|\lambda x + (1-\lambda)y\| \le \lambda\|x\| + (1-\lambda)\|y\| = \lambda g(x) + (1-\lambda)g(y),$$

which follows from the triangle inequality and the fact that $\|\alpha u\| = |\alpha| \cdot \|u\|$ whenever $\alpha \in \mathbb{R}$ and $u \in \mathbb{R}^n$. The convexity of the simplest quadratic function h in **(iii)** follows from a more general result for the quadratic function on $\mathbb{R}^n$ given in the next example.

Example 1.35 Let A be an $n \times n$ symmetric matrix. It is called *positive semidefinite* if $\langle Au, u\rangle \ge 0$ for all $u \in \mathbb{R}^n$. Let us check that A is positive semidefinite if and only if the function $f : \mathbb{R}^n \to \mathbb{R}$ defined by

$$f(x) := \frac{1}{2}\langle Ax, x\rangle, \quad x \in \mathbb{R}^n,$$

is convex. Indeed, a direct calculation shows that for any $x, y \in \mathbb{R}^n$ and $\lambda \in (0,1)$ we have

$$\lambda f(x) + (1-\lambda)f(y) - f\big(\lambda x + (1-\lambda)y\big) = \frac{1}{2}\lambda(1-\lambda)\langle A(x-y), x-y\rangle. \tag{1.2}$$

If the matrix A is positive semidefinite, then $\langle A(x-y), x-y\rangle \ge 0$, so the function f is convex by (1.2). Conversely, assuming the convexity of f and using equality (1.2) for $x = u$ and $y = 0$ verify that A is positive semidefinite.

The following characterization of convexity is known as the *Jensen inequality*.

Theorem 1.36 A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex if and only if for any numbers $\lambda_i \ge 0$ as $i = 1, \ldots, m$ with $\sum_{i=1}^m \lambda_i = 1$ and for any elements $x_i \in \mathbb{R}^n$, $i = 1, \ldots, m$, it holds that

$$f\Big(\sum_{i=1}^m \lambda_i x_i\Big) \le \sum_{i=1}^m \lambda_i f(x_i). \tag{1.3}$$

Proof. Since (1.3) immediately implies the convexity of f, we only need to prove that any convex function f satisfies the Jensen inequality (1.3). Arguing by induction and taking into account that for $m = 1$ inequality (1.3) holds trivially and for $m = 2$ inequality (1.3) holds by the definition of convexity, we suppose that it holds for an integer $m := k$ with $k \ge 2$. Fix numbers $\lambda_i \ge 0$, $i = 1, \ldots, k+1$, with $\sum_{i=1}^{k+1} \lambda_i = 1$ and elements $x_i \in \mathbb{R}^n$, $i = 1, \ldots, k+1$. We obviously have the relationship

$$\sum_{i=1}^k \lambda_i = 1 - \lambda_{k+1}.$$

If $\lambda_{k+1} = 1$, then $\lambda_i = 0$ for all $i = 1, \ldots, k$ and (1.3) obviously holds for $m := k + 1$ in this case. Supposing now that $0 \le \lambda_{k+1} < 1$, we get

$$\sum_{i=1}^{k} \frac{\lambda_i}{1 - \lambda_{k+1}} = 1$$

and by direct calculations based on convexity arrive at

$$\begin{aligned}
f\Big(\sum_{i=1}^{k+1} \lambda_i x_i\Big) &= f\Big[(1 - \lambda_{k+1})\frac{\sum_{i=1}^{k} \lambda_i x_i}{1 - \lambda_{k+1}} + \lambda_{k+1} x_{k+1}\Big] \\
&\le (1 - \lambda_{k+1}) f\Big(\frac{\sum_{i=1}^{k} \lambda_i x_i}{1 - \lambda_{k+1}}\Big) + \lambda_{k+1} f(x_{k+1}) \\
&= (1 - \lambda_{k+1}) f\Big(\sum_{i=1}^{k} \frac{\lambda_i}{1 - \lambda_{k+1}} x_i\Big) + \lambda_{k+1} f(x_{k+1}) \\
&\le (1 - \lambda_{k+1}) \sum_{i=1}^{k} \frac{\lambda_i}{1 - \lambda_{k+1}} f(x_i) + \lambda_{k+1} f(x_{k+1}) \\
&= \sum_{i=1}^{k+1} \lambda_i f(x_i).
\end{aligned}$$

This justifies inequality (1.3) and completes the proof of the theorem. □

The next theorem gives a *geometric characterization* of the function convexity via the convexity of the associated epigraphical set.

Theorem 1.37 A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex if and only if its epigraph epi f is a convex subset of the product space $\mathbb{R}^n \times \mathbb{R}$.

Proof. Assuming that f is convex, fix pairs $(x_1, t_1), (x_2, t_2) \in \text{epi}\, f$ and a number $\lambda \in (0, 1)$. Then we have $f(x_i) \le t_i$ for $i = 1, 2$. Thus the convexity of f ensures that

$$f\big(\lambda x_1 + (1 - \lambda) x_2\big) \le \lambda f(x_1) + (1 - \lambda) f(x_2) \le \lambda t_1 + (1 - \lambda) t_2.$$

This implies therefore that

$$\lambda(x_1, t_1) + (1 - \lambda)(x_2, t_2) = (\lambda x_1 + (1 - \lambda) x_2, \lambda t_1 + (1 - \lambda) t_2) \in \text{epi}\, f,$$

and thus the epigraph epi f is a convex subset of $\mathbb{R}^n \times \mathbb{R}$.

Conversely, suppose that the set epi f is convex and fix $x_1, x_2 \in \text{dom}\, f$ and a number $\lambda \in (0, 1)$. Then $(x_1, f(x_1)), (x_2, f(x_2)) \in \text{epi}\, f$. This tells us that

$$\big(\lambda x_1 + (1 - \lambda) x_2, \lambda f(x_1) + (1 - \lambda) f(x_2)\big) = \lambda\big(x_1, f(x_1)\big) + (1 - \lambda)\big(x_2, f(x_2)\big) \in \text{epi}\, f$$

and thus implies the inequality

$$f\big(\lambda x_1 + (1-\lambda)x_2\big) \le \lambda f(x_1) + (1-\lambda) f(x_2),$$

which justifies the convexity of the function f. □

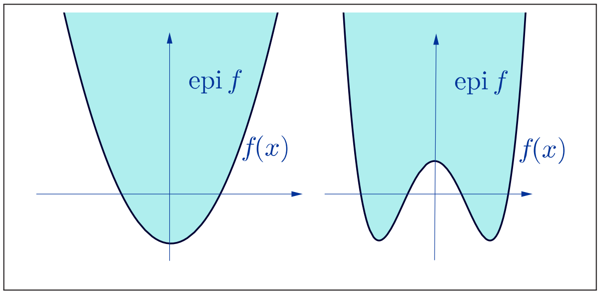

Figure 1.4: Epigraphs of convex function and nonconvex function.

Now we show that convexity is preserved under some important operations.

Proposition 1.38 *Let $f_i\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex functions for all $i = 1, \ldots, m$. Then the following functions are convex as well:*

(i) *The multiplication by scalars λf for any $\lambda > 0$.*
(ii) *The sum function $\sum_{i=1}^m f_i$.*
(iii) *The maximum function $\max_{1 \le i \le m} f_i$.*

Proof. The convexity of λf in **(i)** follows directly from the definition. It is sufficient to prove **(ii)** and **(iii)** for $m = 2$, since the general cases immediately follow by induction.

(ii) Fix any $x, y \in \mathbb{R}^n$ and $\lambda \in (0,1)$. Then we have

$$\begin{aligned}
\big(f_1 + f_2\big)\big(\lambda x + (1-\lambda)y\big) &= f_1\big(\lambda x + (1-\lambda)y\big) + f_2\big(\lambda x + (1-\lambda)y\big) \\
&\le \lambda f_1(x) + (1-\lambda) f_1(y) + \lambda f_2(x) + (1-\lambda) f_2(y) \\
&= \lambda (f_1 + f_2)(x) + (1-\lambda)(f_1 + f_2)(y),
\end{aligned}$$

which verifies the convexity of the sum function $f_1 + f_2$.

(iii) Denote $g := \max\{f_1, f_2\}$ and get for any $x, y \in \mathbb{R}^n$ and $\lambda \in (0,1)$ that

$$f_i\big(\lambda x + (1-\lambda)y\big) \le \lambda f_i(x) + (1-\lambda) f_i(y) \le \lambda g(x) + (1-\lambda) g(y)$$

for $i = 1, 2$. This directly implies that

$$g\big(\lambda x + (1-\lambda)y\big) = \max\big\{f_1\big(\lambda x + (1-\lambda)y\big), f_2\big(\lambda x + (1-\lambda)y\big)\big\} \le \lambda g(x) + (1-\lambda) g(y),$$

which shows that the maximum function $g(x) = \max\{f_1(x), f_2(x)\}$ is convex. □

The next result concerns the preservation of convexity under function compositions.

Proposition 1.39 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and let $\varphi : \mathbb{R} \to \overline{\mathbb{R}}$ be nondecreasing and convex on a convex set containing the range of the function f. Then the composition $\varphi \circ f$ is convex.*

Proof. Take any $x_1, x_2 \in \mathbb{R}^n$ and $\lambda \in (0, 1)$. Then we have by the convexity of f that

$$f\big(\lambda x_1 + (1-\lambda)x_2\big) \le \lambda f(x_1) + (1-\lambda) f(x_2).$$

Since φ is nondecreasing and it is also convex, it follows that

$$\begin{aligned}(\varphi \circ f)\big(\lambda x_1 + (1-\lambda)x_2\big) &= \varphi\big(f\big(\lambda x_1 + (1-\lambda)x_2\big)\big)\\ &\le \varphi\big(\lambda f(x_1) + (1-\lambda) f(x_2)\big)\\ &\le \lambda\varphi\big(f(x_1)\big) + (1-\lambda)\varphi\big(f(x_2)\big)\\ &= \lambda(\varphi \circ f)(x_1) + (1-\lambda)(\varphi \circ f)(x_2),\end{aligned}$$

which verifies the convexity of the composition $\varphi \circ f$. □

Now we consider the composition of a convex function and an affine mapping.

Proposition 1.40 *Let $B : \mathbb{R}^n \to \mathbb{R}^p$ be an affine mapping and let $f : \mathbb{R}^p \to \overline{\mathbb{R}}$ be a convex function. Then the composition $f \circ B$ is convex.*

Proof. Taking any $x, y \in \mathbb{R}^n$ and $\lambda \in (0, 1)$, we have

$$\begin{aligned}\big(f \circ B\big)\big(\lambda x + (1-\lambda)y\big) &= f(B(\lambda x + (1-\lambda)y)) = f\big(\lambda B(x) + (1-\lambda)B(y)\big)\\ &\le \lambda f\big(B(x)\big) + (1-\lambda) f\big(B(y)\big) = \lambda(f \circ B)(x) + (1-\lambda)(f \circ B)(y)\end{aligned}$$

and thus justify the convexity of the composition $f \circ B$. □

The following simple consequence of Proposition 1.40 is useful in applications.

Corollary 1.41 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function. For any $\bar{x}, d \in \mathbb{R}^n$, the function $\varphi\colon \mathbb{R} \to \overline{\mathbb{R}}$ defined by $\varphi(t) := f(\bar{x} + td)$ is convex as well. Conversely, if for every $\bar{x}, d \in \mathbb{R}^n$ the function φ defined above is convex, then f is also convex.*

Proof. Since $B(t) = \bar{x} + td$ is an affine mapping, the convexity of φ immediately follows from Proposition 1.40. To prove the converse implication, take any $x_1, x_2 \in \mathbb{R}^n$, $\lambda \in (0, 1)$ and let $\bar{x} := x_2$, $d := x_1 - x_2$. Since $\varphi(t) = f(\bar{x} + td)$ is convex, we have

$$\begin{aligned}f\big(\lambda x_1 + (1-\lambda)x_2\big) &= f\big(x_2 + \lambda(x_1 - x_2)\big) = \varphi(\lambda) = \varphi\big(\lambda(1) + (1-\lambda)(0)\big)\\ &\le \lambda\varphi(1) + (1-\lambda)\varphi(0) = \lambda f(x_1) + (1-\lambda) f(x_2),\end{aligned}$$

which verifies the convexity of the function f. □

The next proposition is trivial while useful in what follows.

Proposition 1.42 *Let $f : \mathbb{R}^n \times \mathbb{R}^p \to \overline{\mathbb{R}}$ be convex. For $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^p$, the functions $\varphi(y) := f(\bar{x}, y)$ and $\psi(x) := f(x, \bar{y})$ are also convex.*

Now we present an important extension of Proposition 1.38(iii).

Proposition 1.43 *Let $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ for $i \in I$ be a collection of convex functions with a nonempty index set I. Then the supremum function $f(x) := \sup_{i \in I} f_i(x)$ is convex.*

Proof. Fix $x_1, x_2 \in \mathbb{R}^n$ and $\lambda \in (0, 1)$. For every $i \in I$, we have

$$f_i\big(\lambda x_1 + (1-\lambda)x_2\big) \le \lambda f_i(x_1) + (1-\lambda) f_i(x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2),$$

which implies in turn that

$$f\big(\lambda x_1 + (1-\lambda)x_2\big) = \sup_{i \in I} f_i\big(\lambda x_1 + (1-\lambda)x_2\big) \le \lambda f(x_1) + (1-\lambda) f(x_2).$$

This justifies the convexity of the supremum function. $\square$

Our next intention is to characterize convexity of *smooth* functions of one variable. To proceed, we begin with the following lemma.

Lemma 1.44 *Given a convex function $f : \mathbb{R} \to \overline{\mathbb{R}}$, assume that its domain is an open interval I. For any $a, b \in I$ and $a < x < b$, we have the inequalities*

$$\frac{f(x) - f(a)}{x - a} \le \frac{f(b) - f(a)}{b - a} \le \frac{f(b) - f(x)}{b - x}.$$

Proof. Fix a, b, x as above and form the numbers $t := \dfrac{x-a}{b-a} \in (0, 1)$. Then

$$f(x) = f\big(a + (x-a)\big) = f\Big(a + \frac{x-a}{b-a}(b-a)\Big) = f\big(a + t(b-a)\big) = f\big(tb + (1-t)a\big).$$

This gives us the inequalities $f(x) \le t f(b) + (1-t) f(a)$ and

$$f(x) - f(a) \le t f(b) + (1-t) f(a) - f(a) = t\big[f(b) - f(a)\big] = \frac{x-a}{b-a}\big(f(b) - f(a)\big),$$

which can be equivalently written as

$$\frac{f(x) - f(a)}{x - a} \le \frac{f(b) - f(a)}{b - a}.$$

Similarly, we have the estimate

$$f(x) - f(b) \le tf(b) + (1-t)f(a) - f(b) = (t-1)\big[f(b) - f(a)\big] = \frac{x-b}{b-a}\Big(f(b) - f(a)\Big),$$

which finally implies that

$$\frac{f(b) - f(a)}{b-a} \le \frac{f(b) - f(x)}{b-x}$$

and thus completes the proof of the lemma. □

Theorem 1.45 Suppose that $f : \mathbb{R} \to \overline{\mathbb{R}}$ is differentiable on its domain, which is an open interval I. Then f is convex if and only if the derivative f' is nondecreasing on I.

Proof. Suppose that f is convex and fix $a < b$ with $a, b \in I$. By Lemma 1.44, we have

$$\frac{f(x) - f(a)}{x-a} \le \frac{f(b) - f(a)}{b-a}$$

for any $x \in (a, b)$. This implies by the derivative definition that

$$f'(a) \le \frac{f(b) - f(a)}{b-a}.$$

Similarly, we arrive at the estimate

$$\frac{f(b) - f(a)}{b-a} \le f'(b)$$

and conclude that $f'(a) \le f'(b)$, i.e., f' is a nondecreasing function.

To prove the converse, suppose that f' is nondecreasing on I and fix $x_1 < x_2$ with $x_1, x_2 \in I$ and $t \in (0, 1)$. Then

$$x_1 < x_t < x_2 \quad \text{for} \quad x_t := tx_1 + (1-t)x_2.$$

By the mean value theorem, we find c_1, c_2 such that $x_1 < c_1 < x_t < c_2 < x_2$ and

$$\begin{aligned} f(x_t) - f(x_1) &= f'(c_1)(x_t - x_1) = f'(c_1)(1-t)(x_2 - x_1), \\ f(x_t) - f(x_2) &= f'(c_2)(x_t - x_2) = f'(c_2)t(x_1 - x_2). \end{aligned}$$

This can be equivalently rewritten as

$$\begin{aligned} tf(x_t) - tf(x_1) &= f'(c_1)t(1-t)(x_2 - x_1), \\ (1-t)f(x_t) - (1-t)f(x_2) &= f'(c_2)t(1-t)(x_1 - x_2). \end{aligned}$$

Since $f'(c_1) \le f'(c_2)$, adding these equalities yields

$$f(x_t) \le tf(x_1) + (1-t)f(x_2),$$

which justifies the convexity of the function f. □

Corollary 1.46 *Let $f:\mathbb{R}\to\overline{\mathbb{R}}$ be twice differentiable on its domain, which is an open interval I. Then f is convex if and only if $f''(x)\geq 0$ for all $x\in I$.*

Proof. Since $f''(x)\geq 0$ for all $x\in I$ if and only if the derivative function f' is nondecreasing on this interval. Then the conclusion follows directly from Theorem 1.45. □

Example 1.47 Consider the function

$$f(x):=\begin{cases}\dfrac{1}{x} & \text{if } x>0,\\ \infty & \text{otherwise.}\end{cases}$$

To verify its convexity, we get that $f''(x)=\dfrac{2}{x^3}>0$ for all x belonging to the domain of f, which is $I=(0,\infty)$. Thus this function is convex on $\mathbb{R}$ by Corollary 1.46.

Next we define the notion of *set-valued mappings* (or *multifunctions*), which plays an important role in modern convex analysis, its extensions, and applications.

Definition 1.48 *We say that F is a* SET-VALUED MAPPING *between $\mathbb{R}^n$ and $\mathbb{R}^p$ and denote it by $F:\mathbb{R}^n\rightrightarrows\mathbb{R}^p$ if $F(x)$ is a subset of $\mathbb{R}^p$ for every $x\in\mathbb{R}^n$. The* DOMAIN *and* GRAPH *of F are defined, respectively, by*

$$\operatorname{dom} F:=\big\{x\in\mathbb{R}^n\;\big|\;F(x)\neq\emptyset\big\}\;\text{ and }\;\operatorname{gph} F:=\big\{(x,y)\in\mathbb{R}^n\times\mathbb{R}^p\;\big|\;y\in F(x)\big\}.$$

Any single-valued mapping $F:\mathbb{R}^n\to\mathbb{R}^p$ is a particular set-valued mapping where the set $F(x)$ is a singleton for every $x\in\mathbb{R}^n$. It is essential in the following definition that the mapping F is set-valued.

Definition 1.49 *Let $F:\mathbb{R}^n\rightrightarrows\mathbb{R}^p$ and let $\varphi:\mathbb{R}^n\times\mathbb{R}^p\to\overline{\mathbb{R}}$. The* OPTIMAL VALUE *or* MARGINAL FUNCTION *associated with F and φ is defined by*

$$\mu(x):=\inf\big\{\varphi(x,y)\;\big|\;y\in F(x)\big\},\;x\in\mathbb{R}^n. \tag{1.4}$$

Throughout this section we assume that $\mu(x)>-\infty$ for every $x\in\mathbb{R}^n$.

Proposition 1.50 *Assume that $\varphi:\mathbb{R}^n\times\mathbb{R}^p\to\overline{\mathbb{R}}$ is a convex function and that $F:\mathbb{R}^n\rightrightarrows\mathbb{R}^p$ is of convex graph. Then the optimal value function μ in (1.4) is convex.*

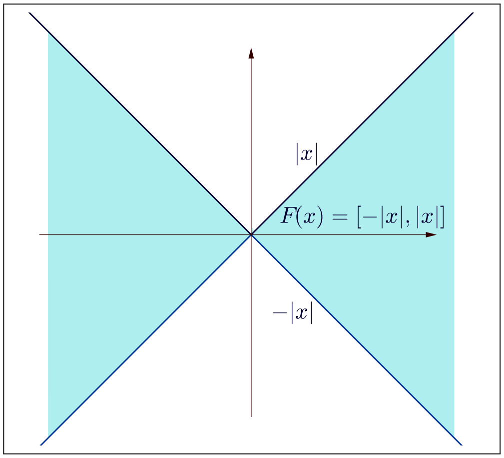

Figure 1.5: Graph of set-valued mapping.

Proof. Take $x_1, x_2 \in \operatorname{dom} \mu$, $\lambda \in (0,1)$. For any $\epsilon > 0$, find $y_i \in F(x_i)$ such that

$$\varphi(x_i, y_i) < \mu(x_i) + \epsilon \quad \text{for} \quad i = 1, 2.$$

It directly implies the inequalities

$$\lambda\varphi(x_1, y_1) < \lambda\mu(x_1) + \lambda\epsilon, \quad (1-\lambda)\varphi(x_2, y_2) < (1-\lambda)\mu(x_2) + (1-\lambda)\epsilon.$$

Summing up these inequalities and employing the convexity of φ yield

$$\begin{aligned}\varphi\big(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2\big) &\le \lambda\varphi(x_1, y_1) + (1-\lambda)\varphi(x_2, y_2)\\ &< \lambda\mu(x_1) + (1-\lambda)\mu(x_2) + \epsilon.\end{aligned}$$

Furthermore, the convexity of gph F gives us

$$\big(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2\big) = \lambda(x_1, y_1) + (1-\lambda)(x_2, y_2) \in \operatorname{gph} F,$$

and therefore $\lambda y_1 + (1-\lambda)y_2 \in F(\lambda x_1 + (1-\lambda)x_2)$. This implies that

$$\mu\big(\lambda x_1 + (1-\lambda)x_2\big) \le \varphi\big(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2\big) < \lambda\mu(x_1) + (1-\lambda)\mu(x_2) + \epsilon.$$

Letting finally $\epsilon \to 0$ ensures the convexity of the optimal value function μ. □

Using Proposition 1.50, we can verify convexity in many situations. For instance, given two convex functions $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$, $i = 1, 2$, let $\varphi(x, y) := f_1(x) + y$ and $F(x) := [f_2(x), \infty)$. Then the function φ is convex and set gph F = epi f_2 is convex as well, and hence we justify the convexity of the sum

$$\mu(x) = \inf_{y \in F(x)} \varphi(x, y) = f_1(x) + f_2(x).$$

Another example concerns compositions. Let $f : \mathbb{R}^p \to \overline{\mathbb{R}}$ be convex and let $B : \mathbb{R}^n \to \mathbb{R}^p$ be affine. Define $\varphi(x, y) := f(y)$ and $F(x) := \{B(x)\}$. Observe that φ is convex while F is of convex graph. Thus we have the convex composition

$$\mu(x) = \inf_{y \in F(x)} \varphi(x, y) = f\big(B(x)\big), \ x \in \mathbb{R}^n .$$

The examples presented above recover the results obtained previously by direct proofs. Now we establish via Proposition 1.50 the convexity of three new classes of functions.

Proposition 1.51 *Let $\varphi : \mathbb{R}^p \to \overline{\mathbb{R}}$ be convex and let $B : \mathbb{R}^p \to \mathbb{R}^n$ be affine. Consider the set-valued inverse image mapping $B^{-1} : \mathbb{R}^n \rightrightarrows \mathbb{R}^p$, define*

$$f(x) := \inf\big\{\varphi(y) \bigm| y \in B^{-1}(x)\big\}, \quad x \in \mathbb{R}^n,$$

and suppose that $f(x) > -\infty$ for all $x \in \mathbb{R}^n$. Then f is a convex function.

Proof. Let $\varphi(x, y) \equiv \varphi(y)$ and $F(x) := B^{-1}(x)$. Then the set

$$\operatorname{gph} F = \big\{(u, v) \in \mathbb{R}^n \times \mathbb{R}^p \bigm| B(v) = u\big\}$$

is obviously convex. Since we have the representation

$$f(x) = \inf_{y \in F(x)} \varphi(y), \quad x \in \mathbb{R}^n,$$

the convexity of f follows directly from Proposition 1.50. □

Proposition 1.52 *For convex functions $f_1, f_2 : \mathbb{R}^n \to \overline{\mathbb{R}}$, define the* INFIMAL CONVOLUTION

$$(f_1 \oplus f_2)(x) := \inf\big\{f_1(x_1) + f_2(x_2) \bigm| x_1 + x_2 = x\big\}$$

and suppose that $(f_1 \oplus f_2)(x) > -\infty$ for all $x \in \mathbb{R}^n$. Then $f_1 \oplus f_2$ is also convex.

Proof. Define $\varphi : \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$ by $\varphi(x_1, x_2) := f_1(x_1) + f_2(x_2)$ and $B : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ by $B(x_1, x_2) := x_1 + x_2$. We have

$$\inf\big\{\varphi(x_1, x_2) \bigm| (x_1, x_2) \in B^{-1}(x)\big\} = (f_1 \oplus f_2)(x) \ \text{ for all } \ x \in \mathbb{R}^n,$$

which implies the convexity of $(f_1 \oplus f_2)$ by Proposition 1.51. □

Definition 1.53 *A function $g : \mathbb{R}^p \to \overline{\mathbb{R}}$ is called* NONDECREASING COMPONENTWISE *if*

$$\big[x_i \le y_i \ \textit{for all} \ i = 1, \ldots, p\big] \Longrightarrow \big[g(x_1, \ldots, x_p) \le g(y_1, \ldots, y_p)\big].$$

Now we are ready to present the final consequence of Proposition 1.50 in this section that involves the composition.

Proposition 1.54 *Define $h : \mathbb{R}^n \to \mathbb{R}^p$ by $h(x) := (f_1(x), \ldots, f_p(x))$, where $f_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots, p$ are convex functions. Suppose that $g : \mathbb{R}^p \to \overline{\mathbb{R}}$ is convex and nondecreasing componentwise. Then the composition $g \circ h : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a convex function.*

Proof. Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^p$ be a set-valued mapping defined by

$$F(x) := [f_1(x), \infty) \times [f_2(x), \infty) \times \ldots \times [f_p(x), \infty).$$

Then the graph of F is represented by

$$\operatorname{gph} F = \big\{(x, t_1, \ldots, t_p) \in \mathbb{R}^n \times \mathbb{R}^p \ \big|\ t_i \ge f_i(x)\big\}.$$

Since all f_i are convex, the set $\operatorname{gph} F$ is convex as well. Define further $\varphi : \mathbb{R}^n \times \mathbb{R}^p \to \overline{\mathbb{R}}$ by $\varphi(x, y) := g(y)$ and observe, since g is increasing componentwise, that

$$\inf\big\{\varphi(x, y) \ \big|\ y \in F(x)\big\} = g\big(f_1(x), \ldots, f_p(x)\big) = (g \circ h)(x),$$

which ensures the convexity of the composition $g \circ h$ by Proposition 1.50. $\square$

1.4 RELATIVE INTERIORS OF CONVEX SETS

We begin this section with the definition and properties of *affine sets*. Given two elements a and b in $\mathbb{R}^n$, the line connecting them is

$$\mathcal{L}[a, b] := \big\{\lambda a + (1 - \lambda) b \ \big|\ \lambda \in \mathbb{R}\big\}.$$

Note that if $a = b$, then $\mathcal{L}[a, b] = \{a\}$.

Definition 1.55 *A subset Ω of $\mathbb{R}^n$ is* AFFINE *if for any $a, b \in \Omega$ we have $\mathcal{L}[a, b] \subset \Omega$.*

For instance, any point, line, and plane in $\mathbb{R}^3$ are affine sets. The empty set and the whole space are always affine. It follows from the definition that the intersection of any collection of affine sets is affine. This leads us to the construction of the *affine hull* of a set.

Definition 1.56 *The* AFFINE HULL *of a set $\Omega \subset \mathbb{R}^n$ is*

$$\operatorname{aff} \Omega := \bigcap \big\{C \ \big|\ C \ \textit{is affine and} \ \ \Omega \subset C\big\}.$$

An element x in $\mathbb{R}^n$ of the form

$$x = \sum_{i=1}^{m} \lambda_i \omega_i \text{ with } \sum_{i=1}^{m} \lambda_i = 1, \quad m \in \mathbb{N},$$

is called an *affine combination* of $\omega_1, \ldots, \omega_m$. The proof of the next proposition is straightforward and thus is omitted.

Proposition 1.57 *The following assertions hold:*

(i) *A set Ω is affine if and only if Ω contains all affine combinations of its elements.*
(ii) *Let Ω, Ω_1, and Ω_2 be affine subsets of $\mathbb{R}^n$. Then the sum $\Omega_1 + \Omega_2$ and the scalar product $\lambda\Omega$ for any $\lambda \in \mathbb{R}$ are also affine subsets of $\mathbb{R}^n$.*
(iii) *Let $B : \mathbb{R}^n \to \mathbb{R}^p$ be an affine mapping. If Ω is an affine subset of $\mathbb{R}^n$ and Θ is an affine subset of $\mathbb{R}^p$, then the image $B(\Omega)$ is an affine subset of $\mathbb{R}^p$ and the inverse image $B^{-1}(\Theta)$ is an affine subset of $\mathbb{R}^n$.*
(iv) *Given $\Omega \subset \mathbb{R}^n$, its affine hull is the smallest affine set containing Ω. We have*

$$\operatorname{aff} \Omega = \Big\{ \sum_{i=1}^{m} \lambda_i \omega_i \;\Big|\; \sum_{i=1}^{m} \lambda_i = 1,\ \omega_i \in \Omega,\ m \in \mathbb{N} \Big\}.$$

(v) *A set Ω is a (linear) subspace if and only if Ω is an affine set containing the origin.*

Next we consider relationships between affine sets and (linear) subspaces.

Lemma 1.58 *A nonempty subset Ω of $\mathbb{R}^n$ is affine if and only if $\Omega - \omega$ is a subspace of $\mathbb{R}^n$ for any $\omega \in \Omega$.*

Proof. Suppose that a nonempty set $\Omega \subset \mathbb{R}^n$ is affine. It follows from Proposition 1.57(v) that $\Omega - \omega$ is a subspace for any $\omega \in \Omega$. Conversely, fix $\omega \in \Omega$ and suppose that $\Omega - \omega$ is a subspace denoted by L. Then the set $\Omega = \omega + L$ is obviously affine. □

The preceding lemma leads to the following notion.

Definition 1.59 *An affine set Ω is* PARALLEL *to a subspace L if $\Omega = \omega + L$ for some $\omega \in \Omega$.*

The next proposition justifies the form of the parallel subspace.

Proposition 1.60 *Let Ω be a nonempty, affine subset of $\mathbb{R}^n$. Then it is parallel to the unique subspace L of $\mathbb{R}^n$ given by $L = \Omega - \Omega$.*

Proof. Given a nonempty, affine set Ω, fix $\omega \in \Omega$ and come up to the linear subspace $L := \Omega - \omega$ parallel to Ω. To justify the uniqueness of such L, take any $\omega_1, \omega_2 \in \Omega$ and any subspaces $L_1, L_2 \subset \mathbb{R}^n$ such that $\Omega = \omega_1 + L_1 = \omega_2 + L_2$. Then $L_1 = \omega_2 - \omega_1 + L_2$. Since $0 \in L_1$, we have $\omega_1 - \omega_2 \in L_2$. This implies that $\omega_2 - \omega_1 \in L_2$ and thus $L_1 = \omega_2 - \omega_1 + L_2 \subset L_2$. Similarly, we get $L_2 \subset L_1$, which justifies that $L_1 = L_2$.

It remains to verify the representation $L = \Omega - \Omega$. Let $\Omega = \omega + L$ with the unique subspace L and some $\omega \in \Omega$. Then $L = \Omega - \omega \subset \Omega - \Omega$. Fix any $x = u - \omega$ with $u, \omega \in \Omega$ and observe that $\Omega - \omega$ is a subspace parallel to Ω. Hence $\Omega - \omega = L$ by the uniqueness of L proved above. This ensures that $x \in \Omega - \omega = L$ and thus $\Omega - \Omega \subset L$. □

The uniqueness of the parallel subspace shows that the next notion is well defined.

Definition 1.61 *The* DIMENSION OF AN AFFINE SET $\emptyset \neq \Omega \subset \mathbb{R}^n$ *is the dimension of the linear subspace parallel to* Ω. *Furthermore, the* DIMENSION OF A CONVEX SET $\emptyset \neq \Omega \subset \mathbb{R}^n$ *is the dimension of its affine hull* aff Ω.

To proceed further, we need yet another definition important in what follows.

Definition 1.62 *The elements* $v_0, \ldots, v_m$ *in* $\mathbb{R}^n$, $m \geq 1$, *are* AFFINELY INDEPENDENT *if*

$$\Big[\sum_{i=0}^m \lambda_i v_i = 0,\ \sum_{i=0}^m \lambda_i = 0\Big] \Longrightarrow \big[\lambda_i = 0 \text{ for all } i = 0, \ldots, m\big].$$

It is easy to observe the following relationship with the linear independence.

Proposition 1.63 *The elements* $v_0, \ldots, v_m$ *in* $\mathbb{R}^n$ *are affinely independent if and only if the elements* $v_1 - v_0, \ldots, v_m - v_0$ *are linearly independent.*

Proof. Suppose that $v_0, \ldots, v_m$ are affinely independent and consider the system

$$\sum_{i=1}^m \lambda_i (v_i - v_0) = 0, \ \text{ i.e., } \ \lambda_0 v_0 + \sum_{i=1}^m \lambda_i v_i = 0,$$

where $\lambda_0 := -\sum_{i=1}^m \lambda_i$. Since the elements $v_0, \ldots, v_m$ are affinely independent and $\sum_{i=0}^m \lambda_i = 0$, we have that $\lambda_i = 0$ for all $i = 1, \ldots, m$. Thus $v_1 - v_0, \ldots, v_m - v_0$ are linearly independent. The proof of the converse statement is straightforward. □

Recall that the *span* of some set C, span C, is the linear subspace generated by C.

Lemma 1.64 *Let* $\Omega := \text{aff}\{v_0, \ldots, v_m\}$, *where* $v_i \in \mathbb{R}^n$ *for all* $i = 0, \ldots, m$. *Then the span of the set* $\{v_1 - v_0, \ldots, v_m - v_0\}$ *is the subspace parallel to* Ω.

Proof. Denote by L the subspace parallel to Ω. Then $\Omega - v_0 = L$ and therefore $v_i - v_0 \in L$ for all $i = 1, \ldots, m$. This gives

$$\operatorname{span}\{v_i - v_0 \mid i = 1, \ldots, m\} \subset L.$$

To prove the converse inclusion, fix any $v \in L$ and get $v + v_0 \in \Omega$. Thus we have

$$v + v_0 = \sum_{i=0}^{m} \lambda_i v_i, \quad \sum_{i=0}^{m} \lambda_i = 1.$$

This implies the relationship

$$v = \sum_{i=1}^{m} \lambda_i (v_i - v_0) \in \operatorname{span}\{v_i - v_0 \mid i = 1, \ldots, m\},$$

which justifies the converse inclusion and hence completes the proof. □

The proof of the next proposition is rather straightforward.

Proposition 1.65 *The elements $v_0, \ldots, v_m$ are affinely independent in $\mathbb{R}^n$ if and only if its affine hull $\Omega := \operatorname{aff}\{v_0, \ldots, v_m\}$ is m-dimensional.*

Proof. Suppose that $v_0, \ldots, v_m$ are affinely independent. Then Lemma 1.64 tells us that the subspace $L := \operatorname{span}\{v_i - v_0 \mid i = 1, \ldots, m\}$ is parallel to Ω. The linear independence of $v_1 - v_0, \ldots, v_m - v_0$ by Proposition 1.63 means that the subspace L is m-dimensional and so is Ω. The proof of the converse statement is also straightforward. □

Affinely independent systems lead us to the construction of *simplices*.

Definition 1.66 *Let $v_0, \ldots, v_m$ be affinely independent in $\mathbb{R}^n$. Then the set*

$$\Delta_m := \operatorname{co}\{v_i \mid i = 0, \ldots, m\}$$

is called an m-SIMPLEX in $\mathbb{R}^n$ with the vertices v_i, $i = 0, \ldots, m$.

An important role of simplex vertices is revealed by the following proposition.

Proposition 1.67 *Consider an m-simplex Δ_m with vertices v_i for $i = 0, \ldots, m$. For every $v \in \Delta_m$, there is a unique element $(\lambda_0, \ldots, \lambda_m) \in \mathbb{R}^{m+1}_+$ such that*

$$v = \sum_{i=0}^{m} \lambda_i v_i, \quad \sum_{i=0}^{m} \lambda_i = 1.$$

Proof. Let $(\lambda_0, \ldots, \lambda_m) \in \mathbb{R}^{m+1}_+$ and $(\mu_0, \ldots, \mu_m) \in \mathbb{R}^{m+1}_+$ satisfy

$$v = \sum_{i=0}^{m} \lambda_i v_i = \sum_{i=0}^{m} \mu_i v_i, \quad \sum_{i=0}^{m} \lambda_i = \sum_{i=0}^{m} \mu_i = 1.$$

This immediately implies the equalities

$$\sum_{i=0}^{m} (\lambda_i - \mu_i) v_i = 0, \quad \sum_{i=0}^{m} (\lambda_i - \mu_i) = 0.$$

Since $v_0, \ldots, v_m$ are affinely independent, we have $\lambda_i = \mu_i$ for $i = 0, \ldots, m$. □

Now we are ready to define a major notion of *relative interiors* of convex sets.

Definition 1.68 *Let Ω be a convex set. We say that an element $v \in \Omega$ belongs to the* RELATIVE INTERIOR $\operatorname{ri} \Omega$ *of Ω if there exists $\epsilon > 0$ such that $I\!B(v; \epsilon) \cap \operatorname{aff} \Omega \subset \Omega$.*

We begin the study of relative interiors with the following lemma.

Lemma 1.69 *Any linear mapping $A : \mathbb{R}^n \to \mathbb{R}^p$ is continuous.*

Proof. Let $\{e_1, \ldots, e_n\}$ be the standard orthonormal basis of $\mathbb{R}^n$ and let $v_i := A(e_i), i = 1, \ldots, n$. For any $x \in \mathbb{R}^n$ with $x = (x_1, \ldots, x_n)$, we have

$$A(x) = A\Big(\sum_{i=1}^{n} x_i e_i\Big) = \sum_{i=1}^{n} x_i A(e_i) = \sum_{i=1}^{n} x_i v_i.$$

Then the triangle inequality and the Cauchy-Schwarz inequality give us

$$\|A(x)\| \le \sum_{i=1}^{n} |x_i| \|v_i\| \le \sqrt{\sum_{i=1}^{n} |x_i|^2} \sqrt{\sum_{i=1}^{n} \|v_i\|^2} = M\|x\|,$$

where $M := \sqrt{\sum_{i=1}^{n} \|v_i\|^2}$. It follows furthermore that

$$\|A(x) - A(y)\| = \|A(x - y)\| \le M\|x - y\| \text{ for all } x, y \in \mathbb{R}^n,$$

which implies the continuity of the mapping A. □

The next proposition plays an essential role in what follows.

Proposition 1.70 *Let Δ_m be an m-simplex in $\mathbb{R}^n$ with some $m \ge 1$. Then $\operatorname{ri} \Delta_m \ne \emptyset$.*

Proof. Consider the vertices $v_0, \dots, v_m$ of the simplex Δ_m and denote

$$v := \frac{1}{m+1} \sum_{i=0}^{m} v_i .$$

We prove the proposition by showing that $v \in \operatorname{ri} \Delta_m$. Define

$$L := \operatorname{span}\{v_i - v_0 \mid i = 1, \dots, m\}$$

and observe that L is the m-dimensional subspace of $\mathbb{R}^n$ parallel to $\operatorname{aff} \Delta_m = \operatorname{aff}\{v_0, \dots, v_m\}$. It is easy to see that for every $x \in L$ there is a unique collection $(\lambda_0, \dots, \lambda_m) \in \mathbb{R}^{m+1}$ with

$$x = \sum_{i=0}^{m} \lambda_i v_i, \quad \sum_{i=0}^{m} \lambda_i = 0.$$

Consider the mapping $A : L \to \mathbb{R}^{m+1}$, which maps x to the corresponding coefficients $(\lambda_0, \dots, \lambda_m) \in \mathbb{R}^{m+1}$ as above. Then A is linear, and hence it is continuous by Lemma 1.69. Since $A(0) = 0$, we can choose $\delta > 0$ such that

$$\|A(u)\| < \frac{1}{m+1} \ \text{ whenever } \ \|u\| \le \delta.$$

Let us now show that $(v + \delta I\!B) \cap \operatorname{aff} \Delta_m \subset \Delta_m$, which means that $v \in \operatorname{ri} \Delta_m$. To proceed, fix any $x \in (v + \delta I\!B) \cap \operatorname{aff} \Delta_m$ and get that $x = v + u$ for some $u \in \delta I\!B$. Since $v, x \in \operatorname{aff} \Delta_m$ and $u = x - v$, we have $u \in L$. Denoting $A(u) := (\alpha_0, \dots, \alpha_m)$ gives us the representation $u = \sum_{i=0}^{m} \alpha_i v_i$ with $\sum_{i=0}^{m} \alpha_i = 0$ and the estimate

$$|\alpha_i| \le \|A(u)\| < \frac{1}{m+1} \ \text{ for all } \ i = 0, \dots, m.$$

Then implies in turn that

$$v + u = \sum_{i=0}^{m} (\frac{1}{m+1} + \alpha_i) v_i = \sum_{i=0}^{m} \mu_i v_i,$$

where $\mu_i := \dfrac{1}{m+1} + \alpha_i \ge 0$ for $i = 0, \dots, m$. Since $\sum_{i=0}^{m} \mu_i = 1$, this ensures that $x \in \Delta_m$. Thus $(v + \delta I\!B) \cap \operatorname{aff} \Delta_m \subset \Delta_m$ and therefore $v \in \operatorname{ri} \Delta_m$. □

Lemma 1.71 *Let Ω be a nonempty, convex set in $\mathbb{R}^n$ of dimension $m \ge 1$. Then there exist $m + 1$ affinely independent elements $v_0, \dots, v_m$ in Ω.*

Proof. Let $\Delta_k := \{v_0, \ldots, v_k\}$ be a $k-$simplex of maximal dimension contained in Ω. Then $v_0, \ldots, v_k$ are affinely independent. To verify now that $k = m$, form $K := \text{aff}\{v_0, \ldots, v_k\}$ and observe that $K \subset \text{aff}\,\Omega$ since $\{v_0, \ldots, v_k\} \subset \Omega$. The opposite inclusion also holds since we have $\Omega \subset K$. Justifying it, we argue by contradiction and suppose that there exists $w \in \Omega$ such that $w \notin K$. Then a direct application of the definition of affine independence shows that $v_0, \ldots, v_k, w$ are affinely independent being a subset of Ω, which is a contradiction. Thus $K = \text{aff}\,\Omega$, and hence we get $k = \dim K = \dim \text{aff }\Omega = \dim \Omega = m$. $\square$

The next is one of the most fundamental results of convex finite-dimensional geometry.

Theorem 1.72 Let Ω be a nonempty, convex set in $\mathbb{R}^n$. The following assertions hold:

(i) We always have $\text{ri}\,\Omega \neq \emptyset$.
(ii) We have $[a, b) \subset \text{ri}\,\Omega$ for any $a \in \text{ri}\,\Omega$ and $b \in \overline{\Omega}$.

Proof. **(i)** Let m be the dimension of Ω. Observe first that the case where $m = 0$ is trivial since in this case Ω is a singleton and $\text{ri}\,\Omega = \Omega$. Suppose that $m \geq 1$ and find $m + 1$ affinely independent elements $v_0, \ldots, v_m$ in Ω as in Lemma 1.71. Consider further the m-simplex

$$\Delta_m := \text{co}\big\{v_0, \ldots, v_m\big\}.$$

We can show that $\text{aff}\,\Delta_m = \text{aff}\,\Omega$. To complete the proof, take $v \in \text{ri}\,\Delta_m$, which exists by Proposition 1.70, and get for any small $\epsilon > 0$ that

$$I\!B(v, \epsilon) \cap \text{aff}\,\Omega = I\!B(v, \epsilon) \cap \text{aff}\,\Delta_m \subset \Delta_m \subset \Omega.$$

This verifies that $v \in \text{ri}\,\Omega$ by the definition of relative interior.

(ii) Let L be the subspace of $\mathbb{R}^n$ parallel to $\text{aff}\,\Omega$ and let $m := \dim L$. Then there is a bijective linear mapping $A : L \to \mathbb{R}^m$ such that both A and A^{-1} are continuous. Fix $x_0 \in \text{aff}\,\Omega$ and define the mapping $f : \text{aff}\,\Omega \to \mathbb{R}^m$ by $f(x) := A(x - x_0)$. It is easy to check that f is a bijective affine mapping and that both f and f^{-1} are continuous. We also see that $a \in \text{ri}\,\Omega$ if and only if $f(a) \in \text{int}\, f(\Omega)$, and that $b \in \overline{\Omega}$ if and only if $f(b) \in \overline{f(\Omega)}$. Then $[f(a), f(b)) \subset \text{int}\, f(\Omega)$ by Lemma 1.30. This shows that $[a, b) \subset \text{ri}\,\Omega$. $\square$

We conclude this section by the following properties of the relative interior.

Proposition 1.73 *Let Ω be a nonempty, convex subset of $\mathbb{R}^n$. For the convex sets* $\text{ri}\,\Omega$ *and* $\overline{\Omega}$, *we have that* **(i)** $\overline{\text{ri}\,\Omega} = \overline{\Omega}$ *and* **(ii)** $\text{ri}\,\Omega = \text{ri}\,\overline{\Omega}$.

Proof. Note that the convexity of $\operatorname{ri}\Omega$ follows from Theorem 1.72(ii) while the convexity of $\overline{\Omega}$ was proved in Proposition 1.29. To justify assertion **(i)** of this proposition, observe that the inclusion $\overline{\operatorname{ri}\Omega} \subset \overline{\Omega}$ is obvious. Fix $b \in \overline{\Omega}$, choose $a \in \operatorname{ri}\Omega$, and form the sequence

$$x_k := \frac{1}{k}a + \Big(1 - \frac{1}{k}\Big)b, \quad k \in \mathbb{N},$$

which converges to b as $k \to \infty$. Since $x_k \in \operatorname{ri}\Omega$ by Theorem 1.72(ii), we have $b \in \overline{\operatorname{ri}\Omega}$. Thus $\overline{\Omega} \subset \overline{\operatorname{ri}\Omega}$, which verifies **(i)**. The proof **(ii)** is similar to that of Proposition 1.31(ii). $\square$

1.5 THE DISTANCE FUNCTION

The last section of this chapter is devoted to the study of the class of distance functions for convex sets, which belongs to the most interesting and important subjects of convex analysis and its extensions. Functions of this type are intrinsically *nondifferentiable* while they naturally and frequently appear in analysis and applications.

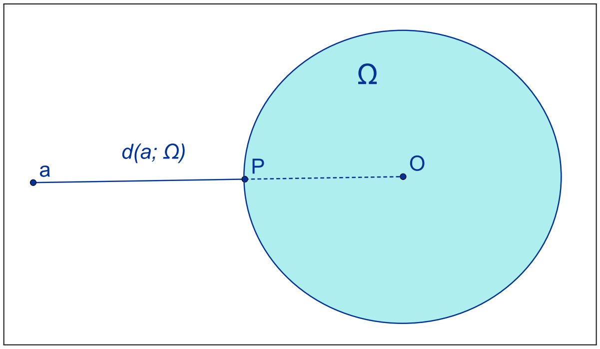

Figure 1.6: Distance function.

Given a set $\Omega \subset \mathbb{R}^n$, the *distance function* associated with Ω is defined by

$$d(x;\Omega) := \inf\big\{\|x - \omega\| \;\big|\; \omega \in \Omega\big\}. \tag{1.5}$$

Recall further that a mapping $f\colon \mathbb{R}^n \to \mathbb{R}^p$ is *Lipschitz continuous* with constant $\ell \geq 0$ on some set $C \subset \mathbb{R}^n$ if we have the estimate

$$\|f(x) - f(y)\| \leq \ell\|x - y\| \text{ for all } x, y \in C. \tag{1.6}$$

Note that the Lipschitz continuity of f in (1.6) specifies its continuity with a *linear rate*.

Proposition 1.74 *Let Ω be a nonempty subset of $\mathbb{R}^n$. The following hold:*

(i) *$d(x;\Omega) = 0$ if and only if $x \in \overline{\Omega}$.*

(ii) *The function $d(x;\Omega)$ is Lipschitz continuous with constant $\ell = 1$ on $\mathbb{R}^n$.*

Proof. **(i)** Suppose that $d(x;\Omega) = 0$. For each $k \in \mathbb{N}$, find $\omega_k \in \Omega$ such that

$$0 = d(x;\Omega) \leq \|x - \omega_k\| < d(x;\Omega) + \frac{1}{k} = \frac{1}{k},$$

which ensures that the sequence $\{\omega_k\}$ converges to x, and hence $x \in \overline{\Omega}$.

Conversely, let $x \in \overline{\Omega}$ and find a sequence $\{\omega_k\} \subset \Omega$ converging to x. Then

$$0 \leq d(x;\Omega) \leq \|x - \omega_k\| \ \text{ for all } \ k \in \mathbb{N},$$

which implies that $d(x;\Omega) = 0$ since $\|x - \omega_k\| \to 0$ as $k \to \infty$.

(ii) For any $\omega \in \Omega$, we have the estimate

$$d(x;\Omega) \leq \|x - \omega\| \leq \|x - y\| + \|y - \omega\|,$$

which implies in turn that

$$d(x;\Omega) \leq \|x - y\| + \inf\big\{\|y - \omega\| \ \big|\ \omega \in \Omega\big\} = \|x - y\| + d(y;\Omega).$$

Similarly, we get $d(y;\Omega) \leq \|y - x\| + d(x;\Omega)$ and thus $|d(x;\Omega) - d(y;\Omega)| \leq \|x - y\|$, which justifies by (1.6) the Lipschitz continuity of $d(x;\Omega)$ on $\mathbb{R}^n$ with constant $\ell = 1$. □

For each $x \in \mathbb{R}^n$, the *Euclidean projection* from x to Ω is defined by

$$\Pi(x;\Omega) := \big\{\omega \in \Omega \ \big|\ \|x - \omega\| = d(x;\Omega)\big\}. \tag{1.7}$$

Proposition 1.75 *Let Ω be a nonempty, closed subset of $\mathbb{R}^n$. Then for any $x \in \mathbb{R}^n$ the Euclidean projection $\Pi(x;\Omega)$ is nonempty.*

Proof. By definition (1.7), for each $k \in \mathbb{N}$ there exists $\omega_k \in \Omega$ such that

$$d(x;\Omega) \leq \|x - \omega_k\| < d(x;\Omega) + \frac{1}{k}.$$

It is clear that $\{\omega_k\}$ is a bounded sequence. Thus it has a subsequence $\{\omega_{k_\ell}\}$ that converges to ω. Since Ω is closed, $\omega \in \Omega$. Letting $\ell \to \infty$ in the inequality

$$d(x;\Omega) \leq \|x - \omega_{k_\ell}\| < d(x;\Omega) + \frac{1}{k_\ell},$$

we have $d(x;\Omega) = \|x - \omega\|$, which ensures that $\omega \in \Pi(x;\Omega)$. □

An interesting consequence of convexity is the following unique projection property.

Corollary 1.76 *If Ω is a nonempty, closed, convex subset of $\mathbb{R}^n$, then for each $x \in \mathbb{R}^n$ the Euclidean projection $\Pi(x;\Omega)$ is a singleton.*

Proof. The nonemptiness of the projection $\Pi(x;\Omega)$ follows from Proposition 1.75. To prove the uniqueness, suppose that $\omega_1, \omega_2 \in \Pi(x;\Omega)$ with $\omega_1 \neq \omega_2$. Then

$$\|x-\omega_1\| = \|x-\omega_2\| = d(x;\Omega).$$

By the classical parallelogram equality, we have that

$$2\|x-\omega_1\|^2 = \|x-\omega_1\|^2 + \|x-\omega_2\|^2 = 2\Big\|x - \frac{\omega_1+\omega_2}{2}\Big\|^2 + \frac{\|\omega_1-\omega_2\|^2}{2}.$$

This directly implies that

$$\Big\|x - \frac{\omega_1+\omega_2}{2}\Big\|^2 = \|x-\omega_1\|^2 - \frac{\|\omega_1-\omega_2\|^2}{4} < \|x-\omega_1\|^2 = \big[d(x;\Omega)\big]^2,$$

which is a contradiction due to the inclusion $\dfrac{\omega_1+\omega_2}{2} \in \Omega$. □

Now we show that the convexity of a nonempty, closed set and its distance function are equivalent. It is an easy exercise to show that the convexity of an arbitrary set Ω implies the convexity of its distance function.

Proposition 1.77 *Let Ω be a nonempty, closed subset of $\mathbb{R}^n$. Then the function $d(\cdot;\Omega)$ is convex if and only if the set Ω is convex.*

Proof. Suppose that Ω is convex. Taking $x_1, x_2 \in \mathbb{R}^n$ and $\omega_i := \Pi(x_i;\Omega)$, we have

$$\|x_i - \omega_i\| = d(x_i;\Omega) \;\text{ for }\; i = 1,2.$$

The convexity of Ω ensures that $\lambda\omega_1 + (1-\lambda)\omega_2 \in \Omega$ for any $\lambda \in (0,1)$. It yields

$$\begin{aligned} d\big(\lambda x_1 + (1-\lambda)x_2;\Omega\big) &\le \|\lambda x_1 + (1-\lambda)x_2 - [\lambda\omega_1 + (1-\lambda)\omega_2]\| \\ &\le \lambda\|x_1-\omega_1\| + (1-\lambda)\|x_2-\omega_2\| \\ &= \lambda d(x_1;\Omega_1) + (1-\lambda)d(x_2;\Omega_2), \end{aligned}$$

which implies therefore the convexity of the distance function $d(\cdot;\Omega)$ by

$$d\big(\lambda x_1 + (1-\lambda)x_2;\Omega\big) \le \lambda d(x_1;\Omega) + (1-\lambda)d(x_2;\Omega).$$

To prove the converse implication, suppose that $d(\cdot;\Omega)$ is convex and fix any $\omega_i \in \Omega$ for $i = 1, 2$ and $\lambda \in (0,1)$. Then we have

$$d\big(\lambda\omega_1 + (1-\lambda)\omega_2;\Omega\big) \le \lambda d(\omega_1;\Omega) + (1-\lambda)d(\omega_2;\Omega) = 0.$$

Since Ω is closed, this yields $\lambda\omega_1 + (1-\lambda)\omega_2 \in \Omega$ and so justifies the convexity of Ω. □

Next we characterize the Euclidean projection to convex sets in $\mathbb{R}^n$. In the proposition below and in what follows we often identify the projection $\Pi(x;\Omega)$ with its unique element if Ω is a nonempty, closed, convex set.

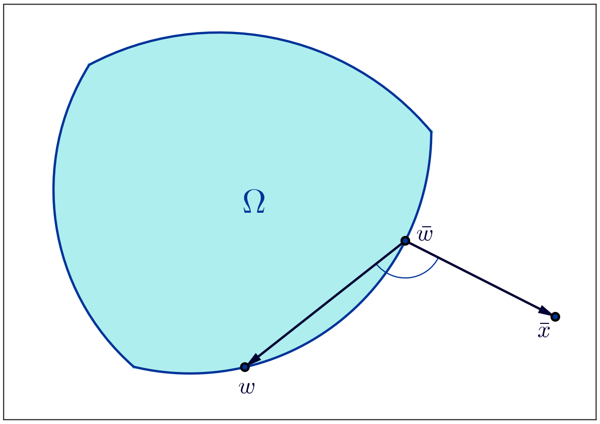

Figure 1.7: Euclidean projection.

Proposition 1.78 *Let Ω be a nonempty, convex subset of $\mathbb{R}^n$ and let $\bar{\omega} \in \Omega$. Then we have $\bar{\omega} \in \Pi(\bar{x};\Omega)$ if and only if*

$$\langle \bar{x} - \bar{\omega}, \omega - \bar{\omega}\rangle \le 0 \ \textit{for all} \ \ \omega \in \Omega. \tag{1.8}$$

Proof. Take $\bar{\omega} \in \Pi(\bar{x};\Omega)$ and get for any $\omega \in \Omega$, $\lambda \in (0,1)$ that $\bar{\omega} + \lambda(\omega - \bar{\omega}) \in \Omega$. Thus

$$\begin{aligned} \|\bar{x} - \bar{\omega}\|^2 = \big[d(\bar{x};\Omega)\big]^2 &\le \|\bar{x} - \big[\bar{\omega} + \lambda(\omega - \bar{\omega})\|^2 \\ &= \|\bar{x} - \bar{\omega}\|^2 - 2\lambda\langle \bar{x} - \bar{\omega}, \omega - \bar{\omega}\rangle + \lambda^2\|\omega - \bar{\omega}\|^2. \end{aligned}$$

This readily implies that

$$2\langle \bar{x} - \bar{\omega}, \omega - \bar{\omega}\rangle \le \lambda\|\omega - \bar{\omega}\|^2.$$

Letting $\lambda \to 0^+$, we arrive at (1.8).

To verify the converse, suppose that (1.8) holds. Then for any $\omega \in \Omega$ we get

$$\begin{aligned}\|\bar{x}-\omega\|^2 &= \|\bar{x}-\bar{\omega}+\bar{\omega}-\omega\|^2 \\ &= \|\bar{x}-\bar{\omega}\|^2+\|\bar{\omega}-\omega\|^2+2\langle \bar{x}-\bar{\omega},\bar{\omega}-\omega\rangle \\ &= \|\bar{x}-\bar{\omega}\|^2+\|\bar{\omega}-\omega\|^2-2\langle \bar{x}-\bar{\omega},\omega-\bar{\omega}\rangle \geq \|\bar{x}-\bar{\omega}\|^2.\end{aligned}$$

Thus $\|\bar{x}-\bar{\omega}\| \leq \|\bar{x}-\omega\|$ for all $\omega \in \Omega$, which implies $\bar{\omega} \in \Pi(\bar{x};\Omega)$ and completes the proof. □

We know from the above that for any nonempty, closed, convex set Ω in $\mathbb{R}^n$ the Euclidean projection $\Pi(x;\Omega)$ is a singleton. Now we show that the projection mapping is in fact *nonexpansive*, i.e., it satisfies the following Lipschitz property.

Proposition 1.79 *Let Ω be a nonempty, closed, convex subset of $\mathbb{R}^n$. Then for any elements $x_1, x_2 \in \mathbb{R}^n$ we have the estimate*

$$\big\|\Pi(x_1;\Omega)-\Pi(x_2;\Omega)\big\|^2 \leq \big\langle \Pi(x_1;\Omega)-\Pi(x_2;\Omega), x_1-x_2\big\rangle.$$

In particular, it implies the Lipschitz continuity of the projection with constant $\ell = 1$:

$$\big\|\Pi(x_1;\Omega)-\Pi(x_2;\Omega)\big\| \leq \|x_1-x_2\| \ \textit{for all} \ x_1, x_2 \in \mathbb{R}^n .$$

Proof. It follows from the preceding proposition that

$$\big\langle \Pi(x_2;\Omega)-\Pi(x_1;\Omega), x_1-\Pi(x_1;\Omega)\big\rangle \leq 0 \ \text{ for all } \ x_1, x_2 \in \mathbb{R}^n .$$

Changing the roles of x_1, x_2 in the inequality above and summing them up give us

$$\big\langle \Pi(x_1;\Omega)-\Pi(x_2;\Omega), x_2-x_1+\Pi(x_1;\Omega)-\Pi(x_2;\Omega)\big\rangle \leq 0.$$

This implies the first estimate in the proposition. Finally, the nonexpansive property of the Euclidean projection follows directly from

$$\begin{aligned}\big\|\Pi(x_1;\Omega)-\Pi(x_2;\Omega)\big\|^2 &\leq \big\langle \Pi(x_1;\Omega)-\Pi(x_2;\Omega), x_1-x_2\big\rangle \\ &\leq \|\Pi(x_1;\Omega)-\Pi(x_2;\Omega)\|\cdot\|x_1-x_2\|\end{aligned}$$

for all $x_1, x_2 \in \mathbb{R}^n$, which completes the proof of the proposition. □

1.6 EXERCISES FOR CHAPTER 1

Exercise 1.1 Let Ω_1 and Ω_2 be nonempty, closed, convex subsets of $\mathbb{R}^n$ such that Ω_1 is bounded or Ω_2 is bounded. Show that $\Omega_1-\Omega_2$ is a nonempty, closed, convex set. Give an example of two nonempty, closed, convex sets Ω_1, Ω_2 for which $\Omega_1-\Omega_2$ is not closed.

Exercise 1.2 Let Ω be a subset of $\mathbb{R}^n$. We say that Ω is a *cone* if $\lambda x \in \Omega$ whenever $\lambda \geq 0$ and $x \in \Omega$. Show that the following are equivalent:

(i) Ω is a convex cone.
(ii) $x + y \in \Omega$ whenever $x, y \in \Omega$, and $\lambda x \in \Omega$ whenever $x \in \Omega$ and $\lambda \geq 0$.

Exercise 1.3 **(i)** Let Ω be a nonempty, convex set that contains 0 and let $0 \leq \lambda_1 \leq \lambda_2$. Show that $\lambda_1 \Omega \subset \lambda_2 \Omega$.
(ii) Let Ω be a nonempty, convex set and let $\alpha, \beta \geq 0$. Show that $\alpha\Omega + \beta\Omega \subset (\alpha + \beta)\Omega$.

Exercise 1.4 **(i)** Let Ω_i for $i = 1, \ldots, m$ be nonempty, convex sets in $\mathbb{R}^n$. Show that $x \in \operatorname{co} \bigcup_{i=1}^m \Omega_i$ if and only if there exist elements $\omega_i \in \Omega_i$ and $\lambda_i \geq 0$ for $i = 1, \ldots, m$ with $\sum_{i=1}^m \lambda_i = 1$ such that $x = \sum_{i=1}^m \lambda_i \omega_i$.
(ii) Let Ω_i for $i = 1, \ldots, m$ be nonempty, convex cones in $\mathbb{R}^n$. Show that

$$\sum_{i=1}^m \Omega_i = \operatorname{co}\{\bigcup_{i=1}^m \Omega_i\}.$$

Exercise 1.5 Let Ω be a nonempty, convex cone in $\mathbb{R}^n$. Show that Ω is a linear subspace of $\mathbb{R}^n$ if and only if $\Omega = -\Omega$.

Exercise 1.6 Show that the following functions are convex on $\mathbb{R}^n$:

(i) $f(x) = \alpha\|x\|$, where $\alpha \geq 0$.
(ii) $f(x) = \|x - a\|^2$, where $a \in \mathbb{R}^n$.
(iii) $f(x) = \|Ax - b\|$, where A is an $p \times n$ matrix and $b \in \mathbb{R}^p$.
(iv) $f(x) = \|x\|^q$, where $q \geq 1$.

Exercise 1.7 Show that the following functions are convex on the given domains:

(i) $f(x) = e^{ax},\ x \in \mathbb{R}$, where a is a constant.
(ii) $f(x) = x^q$, $x \in [0, \infty)$, where $q \geq 1$ is a constant.
(iii) $f(x) = -\ln(x)$, $x \in (0, \infty)$.
(iv) $f(x) = x\ln(x)$, $x \in (0, \infty)$.

Exercise 1.8 **(i)** Give an example of a function $f : \mathbb{R}^n \times \mathbb{R}^p \to \overline{\mathbb{R}}$, which is convex with respect to each variable but not convex with respect to both.
(ii) Let $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ for $i = 1, 2$ be convex functions. Can we conclude that the *minimum function* $\min\{f_1, f_2\}(x) := \min\{f_1(x), f_2(x)\}$ is convex?

Exercise 1.9 Give an example showing that the product of two real-valued convex functions is not necessarily convex.

Exercise 1.10 Verify that the set $\{(x,t) \in \mathbb{R}^n \times \mathbb{R} \mid t \geq \|x\|\}$ is closed and convex.

Exercise 1.11 The *indicator function* associated with a set Ω is defined by

$$\delta(x;\Omega) := \begin{cases} 0 & \text{if } x \in \Omega, \\ \infty & \text{otherwise.} \end{cases}$$

(i) Calculate $\delta(\cdot;\Omega)$ for $\Omega = \emptyset$, $\Omega = \mathbb{R}^n$, and $\Omega = [-1,1]$.
(ii) Show that the set Ω is convex set if and only if its indicator function $\delta(\cdot;\Omega)$ is convex.

Exercise 1.12 Show that if $f : \mathbb{R} \to [0,\infty)$ is a convex function, then its q-power $f^q(x) := (f(x))^q$ is also convex for any $q \geq 1$.

Exercise 1.13 Let $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ for $i = 1,2$ be convex functions. Define

$$\Omega_1 := \{(x,\lambda_1,\lambda_2) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \mid \lambda_1 \geq f_1(x)\},$$
$$\Omega_2 := \{(x,\lambda_1,\lambda_2) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \mid \lambda_2 \geq f_2(x)\}.$$

(i) Show that the sets Ω_1, Ω_2 are convex.
(ii) Define the set-valued mapping $F\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^2$ by $F(x) := [f_1(x),\infty) \times [f_2(x),\infty)$ and verify that the graph of F is $\Omega_1 \cap \Omega_2$.

Exercise 1.14 We say that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is *positively homogeneous* if $f(\alpha x) = \alpha f(x)$ for all $\alpha > 0$, and that f is *subadditive* if $f(x+y) \leq f(x) + f(y)$ for all $x, y \in \mathbb{R}^n$. Show that a positively homogeneous function is convex if and only if it is subadditive.

Exercise 1.15 Given a nonempty set $\Omega \subset \mathbb{R}^n$, define

$$K_\Omega := \{\lambda x \mid \lambda \geq 0,\ x \in \Omega\} = \bigcup_{\lambda \geq 0} \lambda\Omega.$$

(i) Show that K_Ω is a cone.
(ii) Show that K_Ω is the smallest cone containing Ω.
(iii) Show that if Ω is convex, then the cone K_Ω is convex as well.

Exercise 1.16 Let $\varphi : \mathbb{R}^n \times \mathbb{R}^p \to \overline{\mathbb{R}}$ be a convex function and let K be a nonempty, convex subset of $\mathbb{R}^p$. Suppose that for each $x \in \mathbb{R}^n$ the function $\varphi(x,\cdot)$ is bounded below on K. Verify that the function on $\mathbb{R}^n$ defined by $f(x) := \inf\{\varphi(x,y) \mid y \in K\}$ is convex.

Exercise 1.17 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function.

(i) Show that for every $\alpha \in \mathbb{R}$ the *level set* $\{x \in \mathbb{R}^n \mid f(x) \leq \alpha\}$ is convex.
(ii) Let $\Omega \subset \overline{\mathbb{R}}$ be a convex set. Is it true in general that the *inverse image* $f^{-1}(\Omega)$ is a convex subset of $\mathbb{R}^n$?

Exercise 1.18 Let C be a convex subset of $\mathbb{R}^{n+1}$.

(i) For $x \in \mathbb{R}^n$, define $F(x) := \{\lambda \in \mathbb{R} \mid (x, \lambda) \in C\}$. Show that F is a set-valued mapping with convex graph and give an explicit formula for its graph.
(ii) For $x \in \mathbb{R}^n$, define the function

$$f_C(x) := \inf\big\{\lambda \in \mathbb{R} \,\big|\, (x, \lambda) \in C\big\}. \tag{1.9}$$

Find an explicit formula for f_C when C is the closed unit ball of $\mathbb{R}^2$.
(iii) For the function f_C defined in **(ii)**, show that if $f_C(x) > -\infty$ for all $x \in \mathbb{R}^n$, then f_C is a convex function.

Exercise 1.19 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be bounded from below. Define its *convexification*

$$(\mathrm{co}\, f)(x) := \inf\Big\{\sum_{i=1}^{m} \lambda_i f(x_i) \,\Big|\, \lambda_i \ge 0,\ \sum_{i=1}^{m} \lambda_i = 1,\ \sum_{i=1}^{m} \lambda_i x_i = x,\ m \in \mathbb{N}\Big\}. \tag{1.10}$$

Verify that (1.10) is convex with $\mathrm{co}\, f = f_C$, $C := \mathrm{co}\,(\mathrm{epi}\, f)$, and f_C defined in (1.9).

Exercise 1.20 We say that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is *quasiconvex* if

$$f\big(\lambda x + (1-\lambda)y\big) \le \max\big\{f(x), f(y)\big\} \text{ for all } x, y \in \mathbb{R}^n,\ \lambda \in (0,1).$$

(i) Show that a function f is quasiconvex if and only if for any $\alpha \in \mathbb{R}$ the level set $\{x \in \mathbb{R}^n \mid f(x) \le \alpha\}$ is a convex set.
(ii) Show that any convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is quasiconvex. Give an example demonstrating that the converse is not true.

Exercise 1.21 Let a be a nonzero element in $\mathbb{R}^n$ and let $b \in \mathbb{R}$. Show that

$$\Omega := \big\{x \in \mathbb{R}^n \,\big|\, \langle a, x\rangle = b\big\}$$

is an affine set with $\dim \Omega = 1$.

Exercise 1.22 Let the set $\{v_1, \ldots, v_m\}$ consist of affinely independent elements and let $v \notin \mathrm{aff}\{v_1, \ldots, v_m\}$. Show that $v_1, \ldots, v_m, v$ are affinely independent.

Exercise 1.23 Suppose that Ω is a convex subset of $\mathbb{R}^n$ with $\dim \Omega = m$, $m \ge 1$. Let the set $\{v_1, \ldots, v_m\} \subset \Omega$ consist of affinely independent elements and let

$$\Delta_m := \mathrm{co}\,\big\{v_1, \ldots, v_m\big\}.$$

Show that $\mathrm{aff}\, \Omega = \mathrm{aff}\, \Delta_m = \mathrm{aff}\big\{v_1, \ldots, v_m\big\}$.

Exercise 1.24 Let Ω be a nonempty, convex subset of $\mathbb{R}^n$.

(i) Show that $\operatorname{aff}\Omega = \operatorname{aff}\overline{\Omega}$.
(ii) Show that for any $\bar{x} \in \operatorname{ri}\Omega$ and $x \in \overline{\Omega}$ there exists $t > 0$ such that $\bar{x} + t(\bar{x} - x) \in \Omega$.
(iii) Prove Proposition 1.73(ii).

Exercise 1.25 Let $\Omega_1, \Omega_2 \subset \mathbb{R}^n$ be convex sets with $\operatorname{ri}\Omega_1 \cap \operatorname{ri}\Omega_2 \neq \emptyset$. Show that:

(i) $\overline{\Omega_1 \cap \Omega_2} = \overline{\Omega}_1 \cap \overline{\Omega}_2$.
(ii) $\operatorname{ri}(\Omega_1 \cap \Omega_2) = \operatorname{ri}\Omega_1 \cap \operatorname{ri}\Omega_2$.

Exercise 1.26 Let $\Omega_1, \Omega_2 \subset \mathbb{R}^n$ be convex sets with $\overline{\Omega}_1 = \overline{\Omega}_2$. Show that $\operatorname{ri}\Omega_1 = \operatorname{ri}\Omega_2$.

Exercise 1.27 **(i)** Let $B : \mathbb{R}^n \to \mathbb{R}^p$ be an affine mapping and let Ω be a convex subset of $\mathbb{R}^n$. Prove the equality

$$B(\operatorname{ri}\Omega) = \operatorname{ri} B(\Omega).$$

(ii) Let Ω_1 and Ω_2 be convex subsets of $\mathbb{R}^n$. Show that $\operatorname{ri}(\Omega_1 - \Omega_2) = \operatorname{ri}\Omega_1 - \operatorname{ri}\Omega_2$.

Exercise 1.28 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function. Show that:

(i) $\operatorname{aff}(\operatorname{epi} f) = \operatorname{aff}(\operatorname{dom} f) \times \mathbb{R}$.
(ii) $\operatorname{ri}(\operatorname{epi} f) = \big\{(x, \lambda) \;\big|\; x \in \operatorname{ri}(\operatorname{dom} f),\ \lambda > f(x)\big\}$.

Exercise 1.29 Find the explicit formulas for the distance function $d(x; \Omega)$ and the Euclidean projection $\Pi(x; \Omega)$ in the following cases:

(i) Ω is the closed unit ball of $\mathbb{R}^n$.
(ii) $\Omega := \big\{(x_1, x_2) \in \mathbb{R}^2 \;\big|\; |x_1| + |x_2| \le 1\big\}$.
(iii) $\Omega := [-1, 1] \times [-1, 1]$.

Exercise 1.30 Find the formulas for the projection $\Pi(x; \Omega)$ in the following cases:

(i) $\Omega := \big\{x \in \mathbb{R}^n \;\big|\; \langle a, x\rangle = b\big\}$, where $0 \neq a \in \mathbb{R}^n$ and $b \in \mathbb{R}$.
(ii) $\Omega := \big\{x \in \mathbb{R}^n \;\big|\; Ax = b\big\}$, where A is an $m \times n$ matrix with $\operatorname{rank} A = m$ and $b \in \mathbb{R}^m$.
(iii) Ω is the nonnegative orthant $\Omega := \mathbb{R}^n_+$.

Exercise 1.31 For $a_i \le b_i$ with $i = 1, \ldots, n$, define

$$\Omega := \big\{x = (x_1, \ldots, x_n) \in \mathbb{R}^n \;\big|\; a_i \le x \le b_i \;\text{ for all }\; i = 1, \ldots, n\big\}.$$

Show that the projection $\Pi(x; \Omega)$ has the following representation:

$$\Pi(x; \Omega) = \big\{u = (u_1, \ldots, u_n) \in \mathbb{R}^n \;\big|\; u_i = \max\{a_i, \min\{b, x_i\}\} \;\text{ for }\; i \in \{1, \ldots, n\}\big\}.$$

CHAPTER 2

Subdifferential Calculus

In this chapter we present basic results of generalized differentiation theory for convex functions and sets. A geometric approach of variational analysis based on convex separation for extremal systems of sets allows us to derive general calculus rules for normals to sets and subgradients of functions. We also include into this chapter some related results on Lipschitz continuity of convex functions and convex duality. Note that some extended versions of the presented calculus results can be found in exercises for this chapter, with hints to their proofs given at the end of the book.

2.1 CONVEX SEPARATION

We start this section with convex separation theorems, which play a fundamental role in convex analysis and particularly in deriving calculus rules by the geometric approach. The first result concerns the *strict separation* of a nonempty, closed, convex set and an out-of-set point.

Proposition 2.1 *Let Ω be a nonempty, closed, convex subset of $\mathbb{R}^n$ and let $\bar{x} \notin \Omega$. Then there is a nonzero element $v \in \mathbb{R}^n$ such that*

$$\sup\big\{\langle v, x\rangle \bigm| x \in \Omega\big\} < \langle v, \bar{x}\rangle. \tag{2.1}$$

Proof. Denote $\bar{w} := \Pi(\bar{x}; \Omega)$ and let $v := \bar{x} - \bar{w} \neq 0$. Applying Proposition 1.78 gives us

$$\langle v, x - \bar{w}\rangle = \langle \bar{x} - \bar{w}, x - \bar{w}\rangle \leq 0 \text{ for any } x \in \Omega.$$

It follows furthermore that

$$\langle v, x - (\bar{x} - v)\rangle = \langle v, x - \bar{w}\rangle \leq 0,$$

which implies in turn that $\langle v, x\rangle \leq \langle v, \bar{x}\rangle - \|v\|^2$ and thus verifies (2.1). □

In the setting of Proposition 2.1 we choose a number $b \in \mathbb{R}$ such that

$$\sup\big\{\langle v, x\rangle \bigm| x \in \Omega\big\} < b < \langle v, \bar{x}\rangle$$

and define the function $A(x) := \langle v, x\rangle$ for which $A(\bar{x}) > b$ and $A(x) < b$ for all $x \in \Omega$. Then we say that $\bar{x}$ and Ω are *strictly separated by the hyperplane* $L := \{x \in \mathbb{R}^n \mid A(x) = b\}$.

The next theorem extends the result of Proposition 2.1 to the case of two nonempty, closed, convex sets at least one of which is bounded.

Theorem 2.2 Let Ω_1 and Ω_2 be nonempty, closed, convex subsets of $\mathbb{R}^n$ with $\Omega_1 \cap \Omega_2 = \emptyset$. If Ω_1 is bounded or Ω_2 is bounded, then there is a nonzero element $v \in \mathbb{R}^n$ such that

$$\sup\{\langle v, x\rangle \mid x \in \Omega_1\} < \inf\{\langle v, y\rangle \mid y \in \Omega_2\}.$$

Proof. Denote $\Omega := \Omega_1 - \Omega_2$. Then Ω is a nonempty, closed, convex set and $0 \notin \Omega$. Applying Proposition 2.1 to Ω and $\bar{x} = 0$, we have

$$\gamma := \sup\{\langle v, x\rangle \mid x \in \Omega\} < 0 = \langle v, \bar{x}\rangle \text{ with some } 0 \neq v \in \mathbb{R}^n.$$

For any $x \in \Omega_1$ and $y \in \Omega_2$, we have $\langle v, x - y\rangle \leq \gamma$, and so $\langle v, x\rangle \leq \gamma + \langle v, y\rangle$. Therefore,

$$\sup\{\langle v, x\rangle \mid x \in \Omega_1\} \leq \gamma + \inf\{\langle v, y\rangle \mid y \in \Omega_2\} < \inf\{\langle v, y\rangle \mid y \in \Omega_2\},$$

which completes the proof of the theorem. □

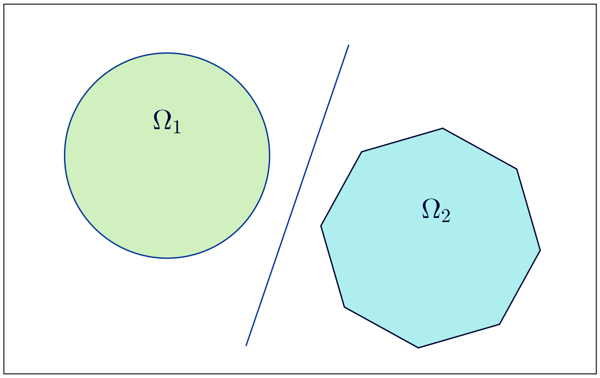

Figure 2.1: Illustration of convex separation.

Next we show that two nonempty, closed, convex sets with empty intersection in $\mathbb{R}^n$ can be separated, while not strictly in general, if both sets are unbounded.

Corollary 2.3 *Let Ω_1 and Ω_2 be nonempty, closed, convex subsets of $\mathbb{R}^n$ such that $\Omega_1 \cap \Omega_2 = \emptyset$. Then they are* SEPARATED *in the sense that there is a nonzero element $v \in \mathbb{R}^n$ with*

$$\sup\{\langle v, x\rangle \mid x \in \Omega_1\} \leq \inf\{\langle v, y\rangle \mid y \in \Omega_2\}. \tag{2.2}$$

Proof. For every fixed $k \in \mathbb{N}$, we define the set $\Theta_k := \Omega_2 \cap I\!B(0; k)$ and observe that it is a closed, bounded, convex set, which is also nonempty if k is sufficiently large. Employing Theorem 2.2 allows us to find $0 \neq u_k \in \mathbb{R}^n$ such that

$$\langle u_k, x \rangle < \langle u_k, y \rangle \ \text{ for all } \ x \in \Omega_1, \ y \in \Theta_k.$$

Denote $v_k := \dfrac{u_k}{\|u_k\|}$ and assume without loss of generality that $v_k \to v \neq 0$ as $k \to \infty$. Fix further $x \in \Omega_1$ and $y \in \Omega_2$ and take $k_0 \in \mathbb{N}$ such that $y \in \Theta_k$ for all $k \geq k_0$. Then

$$\langle v_k, x \rangle < \langle v_k, y \rangle \ \text{ whenever } \ k \geq k_0.$$

Letting $k \to \infty$, we have $\langle v, x \rangle \leq \langle v, y \rangle$, which justifies the claim of the corollary. □

Now our intention is to establish a general separation theorem for convex sets that may *not be closed*. To proceed, we verify first the following useful result.

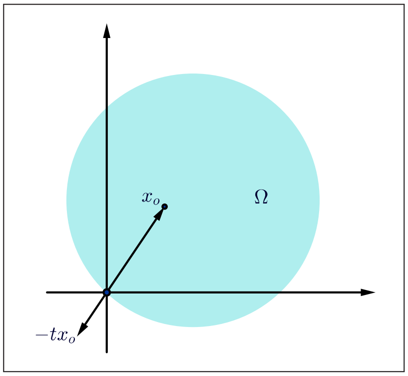

Figure 2.2: Geometric illustration of Lemma 2.4.

Lemma 2.4 *Let Ω be a nonempty, convex subset of $\mathbb{R}^n$ such that $0 \in \overline{\Omega} \setminus \Omega$. Then there is a sequence $x_k \to 0$ as $k \to \infty$ with $x_k \notin \overline{\Omega}$ for all $k \in \mathbb{N}$.*

Proof. Recall that $\operatorname{ri} \Omega \neq \emptyset$ by Theorem 1.72(i). Fix $x_0 \in \operatorname{ri} \Omega$ and show that $-tx_0 \notin \overline{\Omega}$ for all $t > 0$; see Figure 2.2. Indeed, suppose the contrary that $-tx_0 \in \overline{\Omega}$ and get by Theorem 1.72(ii) that

$$0 = \frac{t}{1+t} x_0 + \frac{1}{1+t}(-tx_0) \in \operatorname{ri} \Omega \subset \Omega,$$

a contradiction. Letting then $x_k := -\dfrac{x_0}{k}$ gives us $x_k \notin \overline{\Omega}$ for every k and $x_k \to 0$. □

Note that Lemma 2.4 may not hold for nonconvex sets. As an example, consider the set $\Omega := \mathbb{R}^n \setminus \{0\}$ for which $0 \in \overline{\Omega} \setminus \Omega$ while the conclusion of the Lemma 2.4 fails.

Theorem 2.5 Let Ω_1 and Ω_2 be nonempty, convex subsets of $\mathbb{R}^n$ with $\Omega_1 \cap \Omega_2 = \emptyset$. Then they are separated in the sense of (2.2) with some $v \neq 0$.

Proof. Consider the set $\Omega := \Omega_1 - \Omega_2$ for which $0 \notin \Omega$. We split the proof of the theorem into the following two cases:

Case 1: $0 \notin \overline{\Omega}$. In this case we apply Proposition 2.1 and find $v \neq 0$ such that

$$\langle v, z \rangle < 0 \quad \text{for all} \quad z \in \overline{\Omega}.$$

This gives us $\langle v, x \rangle < \langle v, y \rangle$ whenever $x \in \Omega_1$ and $y \in \Omega_2$, which justifies the statement.

Case 2: $0 \in \overline{\Omega}$. In this case we employ Lemma 2.4 and find a sequence $x_k \to 0$ with $x_k \notin \overline{\Omega}$. Then Proposition 2.1 produces a sequence $\{v_k\}$ of nonzero elements such that

$$\langle v_k, z \rangle < \langle v_k, x_k \rangle \text{ for all } z \in \overline{\Omega}.$$

Without loss of generality, we get (dividing by $\|v_k\|$ and extracting a convergent subsequence) that $v_k \to v$ and $\|v_k\| = \|v\| = 1$. This gives us by letting $k \to \infty$ that

$$\langle v, z \rangle \leq 0 \quad \text{for all} \quad z \in \overline{\Omega}.$$

To complete the proof, we just repeat the arguments of Case 1. $\square$

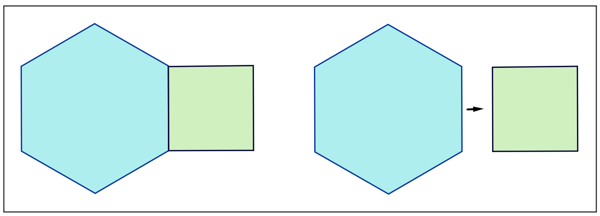

Figure 2.3: Extremal system of sets.

Theorem 2.5 ensures the separation of convex sets, which do not intersect. To proceed now with separation of convex sets which may have common points as in Figure 2.3, we introduce the notion of *set extremality* borrowed from *variational analysis* that deals with both convex and nonconvex sets while emphasizing *intrinsic optimality/extremality* structures of problems under consideration; see, e.g., [14] for more details.

Definition 2.6 *Given two nonempty, convex sets $\Omega_1, \Omega_2 \subset \mathbb{R}^n$, we say that the set system $\{\Omega_1, \Omega_2\}$ is* EXTREMAL *if for any $\epsilon > 0$ there is $a \in \mathbb{R}^n$ such that $\|a\| < \epsilon$ and*

$$(\Omega_1 - a) \cap \Omega_2 = \emptyset.$$

The next proposition gives us an important example of extremal systems of sets.

Proposition 2.7 *Let Ω be a nonempty, convex subset of $\mathbb{R}^n$ and let $\bar{x} \in \Omega$ be a boundary point of Ω. Then the sets Ω and $\{\bar{x}\}$ form an extremal system.*

Proof. Since $\bar{x} \in \text{bd}\,\Omega$, for any number $\epsilon > 0$ we get $I\!B(\bar{x}; \epsilon/2) \cap \Omega^c \neq \emptyset$. Thus choosing $b \in I\!B(\bar{x}; \epsilon/2) \cap \Omega^c$ and denoting $a := b - \bar{x}$ yield $\|a\| < \epsilon$ and $(\Omega - a) \cap \{\bar{x}\} = \emptyset$. $\square$

The following result is a separation theorem for extremal systems of convex sets. It admits a far-going extension to nonconvex settings, which is known as the *extremal principle* and is at the heart of variational analysis and its applications [14]. It is worth mentioning that, in contrast to general settings of variational analysis, we do *not* impose here any *closedness* assumptions on the sets in question. This will allow us to derive in the subsequent sections of this chapter various calculus rules for normals to convex sets and subgradients of convex functions by using a simple variational device without imposing closedness and lower semicontinuity assumptions conventional in variational analysis.

Theorem 2.8 Let Ω_1 and Ω_2 be two nonempty, convex subsets of $\mathbb{R}^n$. Then Ω_1 and Ω_2 form an extremal system if and only if there is a nonzero element $v \in \mathbb{R}^n$ which separates the sets Ω_1 and Ω_2 in the sense of (2.2).

Proof. Suppose that Ω_1 and Ω_2 form an extremal system of convex sets in $\mathbb{R}^n$. Then there is a sequence $\{a_k\}$ in $\mathbb{R}^n$ with $a_k \to 0$ and

$$(\Omega_1 - a_k) \cap \Omega_2 = \emptyset, \quad k \in \mathbb{N}.$$

For every $k \in \mathbb{N}$, apply Theorem 2.5 to the convex sets $\Omega_1 - a_k$ and Ω_2 and find a sequence of nonzero elements $u_k \in \mathbb{R}^n$ satisfying

$$\langle u_k, x - a_k \rangle \le \langle u_k, y \rangle \text{ for all } x \in \Omega_1 \text{ and } y \in \Omega_2, \quad k \in \mathbb{N}.$$

The normalization $v_k := \dfrac{u_k}{\|u_k\|}$ with $\|v_k\| = 1$ gives us

$$\langle v_k, x - a_k \rangle \le \langle v_k, y \rangle \text{ for all } x \in \Omega_1 \text{ and } y \in \Omega_2. \tag{2.3}$$

Thus we find $v \in \mathbb{R}^n$ with $\|v\| = 1$ such that $v_k \to v$ as $k \to \infty$ along a subsequence. Passing to the limit in (2.3) verifies the validity of (2.2).

For the converse implication, suppose that there is a nonzero element $v \in \mathbb{R}^n$ which separates the sets Ω_1 and Ω_2 in the sense of (2.2). Then it is not hard to see that

$$\Big(\Omega_1 - \frac{1}{k}v\Big) \cap \Omega_2 = \emptyset \ \text{ for every } \ k \in \mathbb{N}\,.$$

Thus Ω_1 and Ω_2 form an extremal system. □

2.2 NORMALS TO CONVEX SETS

In this section we apply the separation results obtained above (particularly, the extremal version in Theorem 2.8) to the study of generalized normals to convex sets and derive basic rules of the *normal cone calculus*.

Definition 2.9 *Let $\Omega \subset \mathbb{R}^n$ be a convex set with $\bar{x} \in \Omega$. The* NORMAL CONE *to Ω at $\bar{x}$ is*

$$N(\bar{x};\Omega) := \big\{v \in \mathbb{R}^n \big|\ \langle v, x - \bar{x}\rangle \le 0 \ \textit{for all} \ \ x \in \Omega\big\}. \tag{2.4}$$

By convention, we put $N(\bar{x};\Omega) := \emptyset$ for $\bar{x} \notin \Omega$.

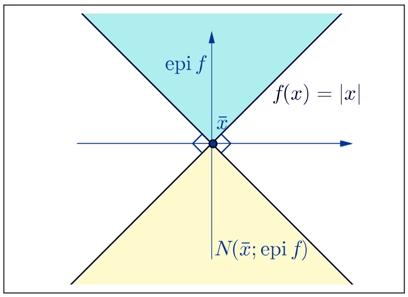

Figure 2.4: Epigraph of the absolute value function and the normal cone to it at the origin.

We first list some elementary properties of the normal cone defined above.

Proposition 2.10 *Let $\bar{x} \in \Omega$ for a convex subset Ω of $\mathbb{R}^n$. Then we have:*

(i) *$N(\bar{x};\Omega)$ is a closed, convex cone containing the origin.*
(ii) *If $\bar{x} \in \operatorname{int}\Omega$, then $N(\bar{x};\Omega) = \{0\}$.*

Proof. To verify **(i)**, fix $v_i \in N(\bar{x};\Omega)$ and $\lambda_i \ge 0$ for $i = 1, 2$. We get by the definition that $0 \in N(\bar{x};\Omega)$ and $\langle v_i, x - \bar{x}\rangle \le 0$ for all $x \in \Omega$ implying

$$\langle \lambda_1 v_1 + \lambda_2 v_2, x - \bar{x}\rangle \le 0 \ \text{ whenever } \ x \in \Omega.$$

Thus $\lambda_1 v_1 + \lambda_2 v_2 \in N(\bar{x}; \Omega)$, and hence $N(\bar{x}; \Omega)$ is a convex cone. To check its closedness, fix a sequence $\{v_k\} \subset N(\bar{x}; \Omega)$ that converges to v. Then passing to the limit in

$$\langle v_k, x - \bar{x} \rangle \le 0 \ \text{ for all } \ x \in \Omega \ \text{ as } \ k \to \infty$$

yields $\langle v, x - \bar{x} \rangle \le 0$ whenever $x \in \Omega$, and so $v \in N(\bar{x}; \Omega)$. This completes the proof of **(i)**.

To check **(ii)**, pick $v \in N(\bar{x}; \Omega)$ with $\bar{x} \in \operatorname{int} \Omega$ and find $\delta > 0$ such that $\bar{x} + \delta I\!B \subset \Omega$. Thus for every $e \in I\!B$ we have the inequality

$$\langle v, \bar{x} + \delta e - \bar{x} \rangle \le 0,$$

which shows that $\delta \langle v, e \rangle \le 0$, and hence $\langle v, e \rangle \le 0$. Choose $t > 0$ so small that $tv \in I\!B$. Then $\langle v, tv \rangle \le 0$ and therefore $t\|v\|^2 = 0$, i.e., $v = 0$. $\square$

We start calculus rules with deriving rather simple while very useful results.

Proposition 2.11 **(i)** *Let Ω_1 and Ω_2 be nonempty, convex subsets of $\mathbb{R}^n$ and $\mathbb{R}^p$, respectively. For $(\bar{x}_1, \bar{x}_2) \in \Omega_1 \times \Omega_2$, we have*

$$N\big((\bar{x}_1, \bar{x}_2); \Omega_1 \times \Omega_2\big) = N(\bar{x}_1; \Omega_1) \times N(\bar{x}_2; \Omega_2).$$

(ii) *Let Ω_1 and Ω_2 be convex subsets of $\mathbb{R}^n$ with $\bar{x}_i \in \Omega_i$ for $i = 1, 2$. Then*

$$N(\bar{x}_1 + \bar{x}_2; \Omega_1 + \Omega_2) = N(\bar{x}_1; \Omega_1) \cap N(\bar{x}_2; \Omega_2).$$

Proof. To verify **(i)**, fix $(v_1, v_2) \in N((\bar{x}_1, \bar{x}_2); \Omega_1 \times \Omega_2)$ and get by the definition that

$$\langle (v_1, v_2), (x_1, x_2) - (\bar{x}_1, \bar{x}_2) \rangle = \langle v_1, x_1 - \bar{x}_1 \rangle + \langle v_2, x_2 - \bar{x}_2 \rangle \le 0 \tag{2.5}$$

whenever $(x_1, x_2) \in \Omega_1 \times \Omega_2$. Putting $x_2 := \bar{x}_2$ in (2.5) gives us

$$\langle v_1, x_1 - \bar{x}_1 \rangle \le 0 \ \text{ for all } \ x_1 \in \Omega_1,$$

which means that $v_1 \in N(\bar{x}_1; \Omega_1)$. Similarly, we obtain $v_2 \in N(\bar{x}_2; \Omega_2)$ and thus justify the inclusion "$\subset$" in (2.5). The opposite inclusion is obvious.

Let us now verify **(ii)**. Fix $v \in N(\bar{x}_1 + \bar{x}_2; \Omega_1 + \Omega_2)$ and get by the definition that

$$\langle v, x_1 + x_2 - (\bar{x}_1 + \bar{x}_2) \rangle \le 0 \ \text{ whenever } \ x_1 \in \Omega_1, \ x_2 \in \Omega_2.$$

Putting there $x_1 := \bar{x}_1$ and $x_2 := \bar{x}_2$ gives us $v \in N(\bar{x}_1; \Omega_1) \cap N(\bar{x}_2; \Omega_2)$. The opposite inclusion in **(ii)** is also straightforward. $\square$

Consider further a linear mapping $A : \mathbb{R}^n \to \mathbb{R}^p$ and associate it with a $p \times n$ matrix as usual. The *adjoint mapping* $A^* : \mathbb{R}^p \to \mathbb{R}^n$ is defined by

$$\langle Ax, y\rangle = \langle x, A^*y\rangle \text{ for all } x \in \mathbb{R}^n,\ y \in \mathbb{R}^p,$$

which corresponds to the *matrix transposition*. The following calculus result describes normals to solution sets for systems of linear equations.

Proposition 2.12 *Let $B\colon \mathbb{R}^n \to \mathbb{R}^p$ be defined by $B(x) := Ax + b$, where $A : \mathbb{R}^n \to \mathbb{R}^p$ is a linear mapping and $b \in \mathbb{R}^p$. Given $c \in \mathbb{R}$, consider the solution set*

$$\Omega := \{x \in \mathbb{R}^n \mid Bx = c\}.$$

For any $\bar{x} \in \Omega$ with $B(\bar{x}) = c$, we have

$$N(\bar{x}; \Omega) = \operatorname{im} A^* = \{v \in \mathbb{R}^n \mid v = A^*y,\ y \in \mathbb{R}^p\}. \tag{2.6}$$

Proof. Take $v \in N(\bar{x}; \Omega)$ and get $\langle v, x - \bar{x}\rangle \le 0$ for all x such that $Ax + b = c$. Fixing any $u \in \ker A$, we see that $A(\bar{x} - u) = A\bar{x} = c - b$, i.e., $A(\bar{x} - u) + b = c$, which yields $\langle v, u\rangle \ge 0$ whenever $u \in \ker A$. Since $-u \in \ker A$, it follows that $\langle v, u\rangle = 0$ for all $u \in \ker A$. Arguing by contradiction, suppose that there is no $y \in \mathbb{R}^p$ with $v = A^*y$. Thus $v \notin W := A^*\mathbb{R}^p$, where the set $W \subset \mathbb{R}^n$ is obviously nonempty, closed, and convex. Employing Proposition 2.1 gives us a nonzero element $\bar{u} \in \mathbb{R}^n$ satisfying

$$\sup\{\langle \bar{u}, w\rangle \mid w \in W\} < \langle \bar{u}, v\rangle,$$

and implying that $0 < \langle \bar{u}, v\rangle$ since $0 \in W$. Furthermore

$$t\langle \bar{u}, A^*y\rangle = \langle \bar{u}, A^*(ty)\rangle < \langle \bar{u}, v\rangle \text{ for all } t \in \mathbb{R},\ y \in \mathbb{R}^p.$$

Hence $\langle \bar{u}, A^*y\rangle = 0$, which shows that $\langle A\bar{u}, y\rangle = 0$ whenever $y \in \mathbb{R}^p$, i.e., $A\bar{u} = 0$. Thus $\bar{u} \in \ker A$ while $\langle v, \bar{u}\rangle > 0$. This contradiction justifies the inclusion "$\subset$" in (2.6).

To verify the opposite inclusion in (2.6), fix any $v \in \mathbb{R}^n$ with $v = A^*y$ for some $y \in \mathbb{R}^p$. For any $x \in \Omega$, we have

$$\langle v, x - \bar{x}\rangle = \langle A^*y, x - \bar{x}\rangle = \langle y, Ax - A\bar{x}\rangle = 0,$$

which means that $v \in N(\bar{x}; \Omega)$ and thus completes the proof of the proposition. □

The following proposition allows us to characterize the separation of convex (generally non-closed) sets in terms of their normal cones.

Proposition 2.13 *Let Ω_1 and Ω_2 be convex subsets of $\mathbb{R}^n$ with $\bar{x} \in \Omega_1 \cap \Omega_2$. Then the following assertions are equivalent:*

(i) $\{\Omega_1, \Omega_2\}$ *forms an extremal system of two convex sets in* $\mathbb{R}^n$.
(ii) *The sets* Ω_1 *and* Ω_2 *are separated in the sense of* (2.2) *with some* $v \neq 0$.
(iii) $N(\bar{x}; \Omega_1) \cap \big[-N(\bar{x}; \Omega_2)\big] \neq \{0\}$.

Proof. The equivalence of **(i)** and **(ii)** follows from Theorem 2.8. Suppose that **(ii)** is satisfied. Since $\bar{x} \in \Omega_2$, we have

$$\sup\big\{\langle v, x\rangle \;\big|\; x \in \Omega_1\big\} \leq \langle v, \bar{x}\rangle,$$

and thus $\langle v, x - \bar{x}\rangle \leq 0$ for all $x \in \Omega_1$, i.e., $v \in N(\bar{x}; \Omega_1)$. Similarly, we can verify the inclusion $-v \in N(\bar{x}; \Omega_2)$, and hence arrive at **(iii)**.

Now suppose that **(iii)** is satisfied and let $0 \neq v \in N(\bar{x}; \Omega_1) \cap \big[-N(\bar{x}; \Omega_2)\big]$. Then (2.2) follows directly from the definition of the normal cone. Indeed, for any $x \in \Omega_1$ and $y \in \Omega_2$ one has

$$\langle v, x - \bar{x}\rangle \leq 0 \text{ and } \langle -v, y - \bar{x}\rangle \leq 0.$$

This implies $\langle v, x - \bar{x}\rangle \leq 0 \leq \langle v, y - \bar{x}\rangle$, and hence $\langle v, x\rangle \leq \langle v, y\rangle$. $\square$

The following corollary provides a normal cone characterization of boundary points.

Corollary 2.14 *Let* Ω *be a nonempty, convex subset of* $\mathbb{R}^n$ *and let* $\bar{x} \in \Omega$. *Then* $\bar{x} \in \operatorname{bd}\Omega$ *if and only if the normal cone to* Ω *is nontrivial, i.e.,* $N(\bar{x}; \Omega) \neq \{0\}$.

Proof. The "only if" part follows from Proposition 2.10(ii). To justify the "if" part, take $\bar{x} \in \operatorname{bd}\Omega$ and consider the sets $\Omega_1 := \Omega$ and $\Omega_2 := \{\bar{x}\}$, which form an extremal system by Proposition 2.7. Note that $N(\bar{x}; \Omega_2) = \mathbb{R}^n$. Thus the claimed result is a direct consequence of Proposition 2.13 since $N(\bar{x}; \Omega_1) \cap [-N(\bar{x}; \Omega_2)] = N(\bar{x}; \Omega_1) \neq \{0\}$. $\square$

Now we are ready to establish a major result of the normal cone calculus for convex sets providing a relationship between normals to set intersections and normals to each set in the intersection. Its proof is also based on *convex separation* in *extremal* systems of sets. Note that this approach allows us to capture general convex sets, which may not be closed.

Theorem 2.15 Let Ω_1 and Ω_2 be convex subsets of $\mathbb{R}^n$ with $\bar{x} \in \Omega_1 \cap \Omega_2$. Then for any $v \in N(\bar{x}; \Omega_1 \cap \Omega_2)$ there exist $\lambda \geq 0$ and $v_i \in N(\bar{x}; \Omega_i)$, $i = 1, 2$, such that

$$\lambda v = v_1 + v_2, \ (\lambda, v_1) \neq (0, 0). \tag{2.7}$$

Proof. Fix $v \in N(\bar{x}; \Omega_1 \cap \Omega_2)$ and get by the definition that

$$\langle v, x - \bar{x}\rangle \leq 0 \ \text{ for all } \ x \in \Omega_1 \cap \Omega_2.$$

Denote $\Theta_1 := \Omega_1 \times [0, \infty)$ and $\Theta_2 := \{(x, \lambda) \mid x \in \Omega_2,\ \lambda \le \langle v, x - \bar{x}\rangle\}$. These convex sets form an extremal system since we obviously have $(\bar{x}, 0) \in \Theta_1 \cap \Theta_2$ and

$$\Theta_1 \cap (\Theta_2 - (0, a)) = \emptyset \ \text{ whenever } \ a > 0.$$

By Theorem 2.8, find $0 \ne (w, \gamma) \in \mathbb{R}^n \times \mathbb{R}$ such that

$$\langle w, x\rangle + \lambda_1 \gamma \le \langle w, y\rangle + \lambda_2 \gamma \ \text{ for all } \ (x, \lambda_1) \in \Theta_1,\ (y, \lambda_2) \in \Theta_2. \tag{2.8}$$

It is easy to see that $\gamma \le 0$. Indeed, assuming $\gamma > 0$ and taking into account that $(\bar{x}, k) \in \Theta_1$ when $k > 0$ and $(\bar{x}, 0) \in \Theta_2$, we get

$$\langle w, \bar{x}\rangle + k\gamma \le \langle w, \bar{x}\rangle,$$

which leads to a contradiction. There are two possible cases to consider.

Case 1: $\gamma = 0$. In this case we have $w \ne 0$ and

$$\langle w, x\rangle \le \langle w, y\rangle \ \text{ for all } \ x \in \Omega_1,\ y \in \Omega_2.$$

Then following the proof of Proposition 2.13 gives us $w \in N(\bar{x}; \Omega_1) \cap [-N(\bar{x}; \Omega_2)]$, and thus (2.7) holds for $\lambda = 0$, $v_1 = w$, and $v_2 = -w$.

Case 2: $\gamma < 0$. In this case we let $\mu := -\gamma > 0$ and deduce from (2.8), by using $(x, 0) \in \Theta_1$ when $x \in \Omega_1$, and $(\bar{x}, 0) \in \Theta_2$, that

$$\langle w, x\rangle \le \langle w, \bar{x}\rangle \ \text{ for all } \ x \in \Omega_1.$$

This implies that $w \in N(\bar{x}; \Omega_1)$, and hence $\dfrac{w}{\mu} \in N(\bar{x}; \Omega_1)$. To proceed further, we get from (2.8), due to $(\bar{x}, 0) \in \Theta_1$ and $(y, \alpha) \in \Theta_2$ for all $y \in \Omega_2$ with $\alpha := \langle v, y - \bar{x}\rangle$, that

$$\Big\langle w, \bar{x}\Big\rangle \le \Big\langle w, y\Big\rangle + \gamma\langle v, y - \bar{x}\rangle \ \text{ whenever } \ y \in \Omega_2.$$

Dividing both sides by γ gives us

$$\Big\langle \frac{w}{\gamma}, \bar{x}\Big\rangle \ge \Big\langle \frac{w}{\gamma}, y\Big\rangle + \langle v, y - \bar{x}\rangle \ \text{ whenever } \ y \in \Omega_2.$$

This clearly implies the inequality

$$\Big\langle \frac{w}{\gamma} + v, y - \bar{x}\Big\rangle \le 0 \ \text{ for all } \ y \in \Omega_2,$$

and thus $\dfrac{w}{\gamma} + v = -\dfrac{w}{\mu} + v \in N(\bar{x}; \Omega_2)$. Letting finally $v_1 := \dfrac{w}{\mu} \in N(\bar{x}; \Omega_1)$ and $v_2 := -\dfrac{w}{\mu} + v \in N(\bar{x}; \Omega_2)$ gives us $v = v_1 + v_2$, which ensures the validity of (2.7) with $\lambda = 1$. $\square$

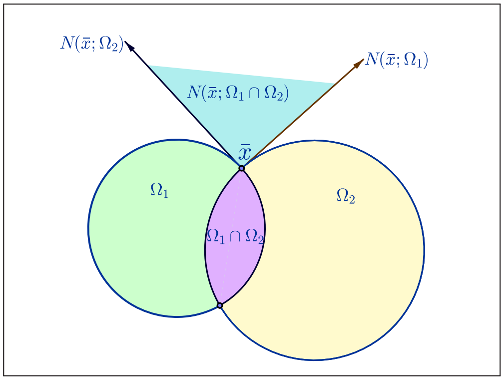

Figure 2.5: Normal cone to set intersection.

Observe that representation (2.7) does not provide a calculus rule for representing normals to set intersections unless $\lambda \neq 0$. However, it is easy to derive such an *intersection rule* directly from Theorem 2.15 while imposing the following *basic qualification condition*, which plays a crucial role in the subsequent calculus results; see Figure 2.5.

Corollary 2.16 *Let Ω_1 and Ω_2 be convex subsets of $\mathbb{R}^n$ with $\bar{x} \in \Omega_1 \cap \Omega_2$. Assume that the basic qualification condition* (BQC) *is satisfied:*

$$N(\bar{x};\Omega_1) \cap \big[-N(\bar{x};\Omega_2)\big] = \{0\}. \tag{2.9}$$

Then we have the normal cone intersection rule

$$N(\bar{x};\Omega_1 \cap \Omega_2) = N(\bar{x};\Omega_1) + N(\bar{x};\Omega_2). \tag{2.10}$$

Proof. For any $v \in N(\bar{x};\Omega_1 \cap \Omega_2)$, we find by Theorem 2.15 a number $\lambda \geq 0$ and elements $v_i \in N(\bar{x};\Omega_i)$ as $i = 1, 2$ such that the conditions in (2.7) hold. Assuming that $\lambda = 0$ in (2.7) gives us that $0 \neq v_1 = -v_2 \in N(\bar{x};\Omega_1) \cap [-N(\bar{x};\Omega_2)]$, which contradicts BQC (2.9). Thus $\lambda > 0$ and $v = v_1/\lambda + v_2/\lambda \in N(\bar{x};\Omega_1) + N(\bar{x};\Omega_2)$. Hence the inclusion "$\subset$" in (2.10) is satisfied. The opposite inclusion in (2.10) can be proved easily using the definition of the normal cone. Indeed, fix any $v \in N(\bar{x};\Omega_1) + N(\bar{x};\Omega_2)$ with $v = v_1 + v_2$, where $v_i \in N(\bar{x};\Omega_i)$ for $i = 1, 2$. For any $x \in \Omega_1 \cap \Omega_2$, one has

$$\langle v, x - \bar{x}\rangle = \langle v_1 + v_2, x - \bar{x}\rangle = \langle v_1, x - \bar{x}\rangle + \langle v_2, x - \bar{x}\rangle \leq 0,$$

which implies $v \in N(\bar{x};\Omega_1 \cap \Omega_2)$ and completes the proof of the corollary. $\square$

The following example shows that the intersection rule (2.10) does not hold generally without the validity of the basic qualification condition (2.9).

Example 2.17 Let $\Omega_1 := \{(x,\lambda) \in \mathbb{R}^2 \mid \lambda \geq x^2\}$ and let $\Omega_2 := \{(x,\lambda) \in \mathbb{R}^2 \mid \lambda \leq -x^2\}$. Then for $\bar{x} = (0,0)$ we have $N(\bar{x};\Omega_1 \cap \Omega_2) = \mathbb{R}^2$ while $N(\bar{x};\Omega_1) + N(\bar{x};\Omega_2) = \{0\} \times \mathbb{R}$.

Next we present an easily verifiable condition ensuring the validity of BQC (2.9). More general results related to relative interior are presented in Exercise 2.13 and its solution.

Corollary 2.18 *Let $\Omega_1, \Omega_2 \subset \mathbb{R}^n$ be such that* $\operatorname{int}\Omega_1 \cap \Omega_2 \neq \emptyset$ *or* $\operatorname{int}\Omega_2 \cap \Omega_1 \neq \emptyset$. *Then for any $\bar{x} \in \Omega_1 \cap \Omega_2$ both* BQC (2.9) *and intersection rule* (2.10) *are satisfied.*

Proof. Consider for definiteness the case where $\operatorname{int}\Omega_1 \cap \Omega_2 \neq \emptyset$. Let us fix $\bar{x} \in \Omega_1 \cap \Omega_2$ and show that BQC (2.9) holds, which ensures the intersection rule (2.10) by Corollary 2.16. Arguing by contradiction, suppose the opposite

$$N(\bar{x};\Omega_1) \cap \big[-N(\bar{x};\Omega_2)\big] \neq \{0\}$$

and then find by Proposition 2.13 a nonzero element $v \in \mathbb{R}^n$ that separates the sets Ω_1 and Ω_2 as in (2.2). Fixing now $\bar{u} \in \operatorname{int}\Omega_1 \cap \Omega_2$ gives us a number $\delta > 0$ such that $\bar{u} + \delta I\!B \subset \Omega_1$. It ensures by (2.2) that

$$\langle v, x\rangle \leq \langle v, \bar{u}\rangle \ \text{ for all } \ x \in \bar{u} + \delta I\!B.$$

This implies in turn the inequality

$$\langle v, \bar{u}\rangle + \delta\langle v, e\rangle \leq \langle v, \bar{u}\rangle \ \text{ whenever } \ e \in I\!B$$

and tells us that $\langle v, e\rangle \leq 0$ for all $e \in I\!B$. Choosing finally $e := \dfrac{v}{\|v\|}$, we arrive at a contradiction and thus complete the proof of this corollary. □

Employing induction arguments allows us to derive a counterpart of Corollary 2.16 (and, similarly, of Corollary 2.18) for finitely many sets.

Corollary 2.19 *Let $\Omega_i \subset \mathbb{R}^n$ for $i = 1,\ldots,m$ be convex sets and let $\bar{x} \in \bigcap_{i=1}^m \Omega_i$. Assume the validity of the following qualification condition:*

$$\Big[\sum_{i=1}^m v_i = 0,\ v_i \in N(\bar{x};\Omega_i)\Big] \Longrightarrow \big[v_i = 0 \ \textit{for all} \ i = 1,\ldots,m\big].$$

Then we have the intersection formula

$$N\Big(\bar{x};\bigcap_{i=1}^m \Omega_i\Big) = \sum_{i=1}^m N(\bar{x};\Omega_i).$$

2.3 LIPSCHITZ CONTINUITY OF CONVEX FUNCTIONS

In this section we study the fundamental notion of Lipschitz continuity for convex functions. This notion can be treated as "*continuity at a linear rate*" being highly important for many aspects of convex and variational analysis and their applications.

Definition 2.20 *A function* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *is* Lipschitz continuous *on a set* $\Omega \subset \operatorname{dom} f$ *if there is a constant* $\ell \geq 0$ *such that*

$$|f(x) - f(y)| \leq \ell \|x - y\| \ \textit{for all} \ x, y \in \Omega. \tag{2.11}$$

We say that f *is* locally Lipschitz continuous *around* $\bar{x} \in \operatorname{dom} f$ *if there are constants* $\ell \geq 0$ *and* $\delta > 0$ *such that* (2.11) *holds with* $\Omega = I\!B(\bar{x}; \delta)$.

Here we give several verifiable conditions that ensure Lipschitz continuity and characterize its localized version for general convex functions. One of the most effective characterizations of local Lipschitz continuity, holding for nonconvex functions as well, is given via the following notion of the *singular* (or *horizon*) *subdifferential*, which was largely underinvestigated in convex analysis and was properly recognized only in the framework of variational analysis; see [14, 27]. Note that, in contrast to the (usual) subdifferential of convex analysis defined in the next section, the singular subdifferential does not reduce to the classical derivative for differentiable functions.

Definition 2.21 *Let* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be a convex function and let* $\bar{x} \in \operatorname{dom} f$. *The* singular subdifferential *of the function* f *at* $\bar{x}$ *is defined by*

$$\partial^\infty f(\bar{x}) := \big\{ v \in \mathbb{R}^n \,\big|\, (v, 0) \in N\big((\bar{x}, f(\bar{x})); \operatorname{epi} f\big)\big\}. \tag{2.12}$$

The following example illustrates the calculation of $\partial^\infty f(\bar{x})$ based on definition (2.12).

Example 2.22 Consider the function

$$f(x) := \begin{cases} 0 & \text{if } |x| \leq 1, \\ \infty & \text{otherwise.} \end{cases}$$

Then $\operatorname{epi} f = [-1, 1] \times [0, \infty)$ and for $\bar{x} = 1$ we have

$$N\big((\bar{x}, f(\bar{x})); \operatorname{epi} f\big) = [0, \infty) \times (-\infty, 0],$$

which shows that $\partial^\infty f(\bar{x}) = [0, \infty)$; see Figure 2.6.

The next proposition significantly simplifies the calculation of $\partial^\infty f(\bar{x})$.

Proposition 2.23 *For a convex function* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *with* $\bar{x} \in \operatorname{dom} f$, *we have*

$$\partial^\infty f(\bar{x}) = N(\bar{x}; \operatorname{dom} f).$$

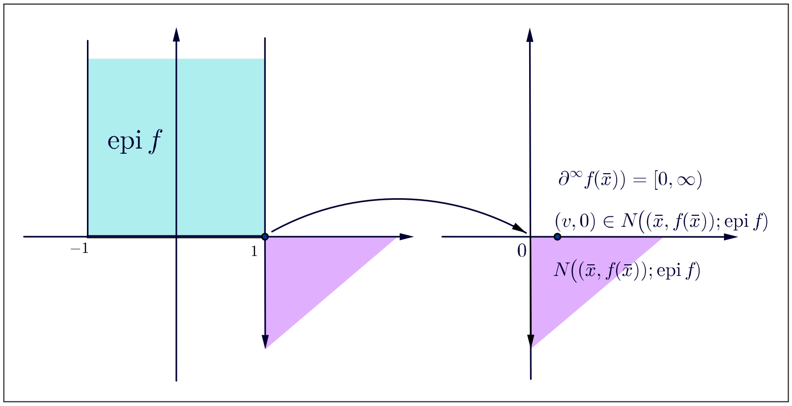

Figure 2.6: Singular subgradients.

Proof. Fix any $v \in \partial^\infty f(\bar{x})$ with $x \in \text{dom } f$ and observe that $(x, f(x)) \in \text{epi } f$. Then using (2.12) and the normal cone construction (2.4) give us

$$\langle v, x - \bar{x} \rangle = \langle v, x - \bar{x} \rangle + 0\big(f(x) - f(\bar{x})\big) \le 0,$$

which shows that $v \in N(\bar{x}; \text{dom } f)$. Conversely, suppose that $v \in N(\bar{x}; \text{dom } f)$ and fix any $(x, \lambda) \in \text{epi } f$. Then we have $f(x) \le \lambda$, and so $x \in \text{dom } f$. Thus

$$\langle v, x - \bar{x} \rangle + 0(\lambda - f(\bar{x})) = \langle v, x - \bar{x} \rangle \le 0,$$

which implies that $(v, 0) \in N((\bar{x}, f(\bar{x})); \text{epi } f)$, i.e., $v \in \partial^\infty f(\bar{x})$. $\square$

To proceed with the study of Lipschitz continuity of convex functions, we need the following two lemmas, which are of their own interest.

Lemma 2.24 *Let $\{e_i \mid i = 1, \ldots, n\}$ be the standard orthonormal basis of $\mathbb{R}^n$. Denote*

$$A := \big\{\bar{x} \pm \epsilon e_i \;\big|\; i = 1, \ldots, n\big\}, \quad \epsilon > 0.$$

Then the following properties hold:

(i) $\bar{x} + \gamma e_i \in \text{co } A$ *for* $|\gamma| \le \epsilon$ *and* $i = 1, \ldots, n$.
(ii) $I\!B(\bar{x}; \epsilon/n) \subset \text{co } A$.

Proof. **(i)** For $|\gamma| \le \epsilon$, find $t \in [0,1]$ with $\gamma = t(-\epsilon) + (1-t)\epsilon$. Then $\bar{x} \pm \epsilon e_i \in A$ gives us

$$\bar{x} + \gamma e_i = t(\bar{x} - \epsilon e_i) + (1-t)(\bar{x} + \epsilon e_i) \in \text{co } A.$$

(ii) For $x \in I\!B(\bar{x};\epsilon/n)$, we have $x = \bar{x} + (\epsilon/n)u$ with $\|u\| \le 1$. Represent u via $\{e_i\}$ by

$$u = \sum_{i=1}^{n} \lambda_i e_i,$$

where $|\lambda_i| \le \sqrt{\sum_{i=1}^{n} \lambda_i^2} = \|u\| \le 1$ for every i. This gives us

$$x = \bar{x} + \frac{\epsilon}{n}u = \bar{x} + \sum_{i=1}^{n} \frac{\epsilon\lambda_i}{n} e_i = \sum_{i=1}^{n} \frac{1}{n}\Big(\bar{x} + \epsilon\lambda_i e_i\Big).$$

Denoting $\gamma_i := \epsilon\lambda_i$ yields $|\gamma_i| \le \epsilon$. It follows from **(i)** that $\bar{x} + \epsilon\lambda_i e_i = \bar{x} + \gamma_i e_i \in \text{co}\, A$, and thus $x \in \text{co}\, A$ since it is a convex combination of elements in $\text{co}\, A$. $\square$

Lemma 2.25 *If a convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is bounded above on $I\!B(\bar{x};\delta)$ for some element $\bar{x} \in$ $\text{dom}\, f$ and number $\delta > 0$, then f is bounded on $I\!B(\bar{x};\delta)$.*

Proof. Denote $m := f(\bar{x})$ and take $M > 0$ with $f(x) \le M$ for all $x \in I\!B(\bar{x};\delta)$. Picking any $u \in I\!B(\bar{x};\delta)$, consider the element $x := 2\bar{x} - u$. Then $x \in I\!B(\bar{x};\delta)$ and

$$m = f(\bar{x}) = f\big(\frac{x+u}{2}\big) \le \frac{1}{2}f(x) + \frac{1}{2}f(u),$$

which shows that $f(u) \ge 2m - f(x) \ge 2m - M$ and thus f is bounded on $I\!B(\bar{x};\delta)$. $\square$

The next major result shows that the local boundedness of a convex function implies its local Lipschitz continuity.

Theorem 2.26 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex with $\bar{x} \in \text{dom}\, f$. If f is bounded above on $I\!B(\bar{x};\delta)$ for some $\delta > 0$, then f is Lipschitz continuous on $I\!B(\bar{x};\delta/2)$.

Proof. Fix $x, y \in I\!B(\bar{x};\delta/2)$ with $x \ne y$ and consider the element

$$u := x + \frac{\delta}{2\|x-y\|}\Big(x - y\Big).$$

Since $e := \dfrac{x-y}{\|x-y\|} \in I\!B$, we have $u = x + \dfrac{\delta}{2}e \in \bar{x} + \dfrac{\delta}{2}I\!B + \dfrac{\delta}{2}I\!B \subset \bar{x} + \delta I\!B$. Denoting further $\alpha := \dfrac{\delta}{2\|x-y\|}$ gives us $u = x + \alpha(x-y)$, and thus

$$x = \frac{1}{\alpha+1}u + \frac{\alpha}{\alpha+1}y.$$

It follows from the convexity of f that

$$f(x) \leq \frac{1}{\alpha+1} f(u) + \frac{\alpha}{\alpha+1} f(y),$$

which implies in turn the inequalities

$$f(x) - f(y) \leq \frac{1}{\alpha+1}(f(u) - f(y)) \leq 2M \frac{1}{\alpha+1} = 2M \frac{2\|x-y\|}{\delta + 2\|x-y\|} \leq \frac{4M}{\delta}\|x-y\|$$

with $M := \sup\{|f(x)| \mid x \in I\!B(\bar{x};\delta)\} < \infty$ by Lemma 2.25. Interchanging the role of x and y above, we arrive at the estimate

$$|f(x) - f(y)| \leq \frac{4M}{\delta}\|x-y\|$$

and thus verify the Lipschitz continuity of f on $I\!B(\bar{x};\delta/2)$. $\square$

The following two consequences of Theorem 2.26 ensure the *automatic* local Lipschitz continuity of a convex function on appropriate sets.

Corollary 2.27 *Any extended-real-valued convex function* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *is locally Lipschitz continuous on in the interior of its domain* $\operatorname{int}(\operatorname{dom} f)$.

Proof. Pick $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$ and choose $\epsilon > 0$ such that $\bar{x} \pm \epsilon e_i \in \operatorname{dom} f$ for every i. Considering the set A from Lemma 2.24, we get $I\!B(\bar{x};\epsilon/n) \subset \operatorname{co} A$ by assertion (ii) of this lemma. Denote $M := \max\{f(a) \mid a \in A\} < \infty$ over the finite set A. Using the representation

$$x = \sum_{i=1}^{m} \lambda_i a_i \ \text{ with } \ \lambda_i \geq 0 \ \sum_{i=1}^{m} \lambda_i = 1, \ a_i \in A$$

for any $x \in I\!B(\bar{x};\epsilon/n)$ shows that

$$f(x) \leq \sum_{i=1}^{m} \lambda_i f(a_i) \leq \sum_{i=1}^{m} \lambda_i M = M,$$

and so f is bounded above on $I\!B(\bar{x};\epsilon/n)$. Then Theorem 2.26 tells us that f is Lipschitz continuous on $I\!B(\bar{x};\epsilon/2n)$ and thus it is locally Lipschitz continuous on $\operatorname{int}(\operatorname{dom} f)$. $\square$

It immediately follows from Corollary 2.27 that any *finite* convex function on $\mathbb{R}^n$ is always locally Lipschitz continuous on the whole space.

Corollary 2.28 *If* $f : \mathbb{R}^n \to \mathbb{R}$ *is convex, then it is locally Lipschitz continuous on* $\mathbb{R}^n$.

The final result of this section provides several *characterizations* of the local Lipschitz continuity of extended-real-valued convex functions.

Theorem 2.29 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex with $\bar{x} \in \operatorname{dom} f$. The following are equivalent:

(i) f is continuous at $\bar{x}$.
(ii) $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$.
(iii) f is locally Lipschitz continuous around $\bar{x}$.
(iv) $\partial^\infty f(\bar{x}) = \{0\}$.

Proof. To verify **(i)⟹(ii)**, by the continuity of f at $\bar{x}$, for any $\epsilon > 0$ find $\delta > 0$ such that

$$|f(x) - f(\bar{x})| < \epsilon \text{ whenever } x \in I\!B(\bar{x};\delta).$$

Since $f(\bar{x}) < \infty$, this implies that $I\!B(\bar{x};\delta) \subset \operatorname{dom} f$, and so $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$.

Implication **(ii)⟹(iii)** follows directly from Corollary 2.27 while **(iii)⟹(i)** is obvious. Thus conditions **(i), (ii)**, and **(iii)** are equivalent. To show next that **(ii)⟹(iv)**, take $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$ for which $N(\bar{x};\operatorname{dom} f) = \{0\}$. Then Proposition 2.23 yields $\partial^\infty f(\bar{x}) = \{0\}$.

To verify finally **(iv)⟹(ii)**, we argue by contradiction and suppose that $\bar{x} \notin \operatorname{int}(\operatorname{dom} f)$. Then $\bar{x} \in \operatorname{bd}(\operatorname{dom} f)$ and it follows from Corollary 2.14 that $N(\bar{x};\operatorname{dom} f) \neq \{0\}$, which contradicts the condition $\partial^\infty f(\bar{x}) = \{0\}$ by Proposition 2.23. □

2.4 SUBGRADIENTS OF CONVEX FUNCTIONS

In this section we define and start studying the concept of *subdifferential* (or the collection of *subgradients*) for convex extended-real-valued functions. This notion of generalized derivative is one of the most fundamental in analysis, with a variety of important applications. The revolutionary idea behind the subdifferential concept, which distinguishes it from other notions of generalized derivatives in mathematics, is its *set-valuedness* as the indication of nonsmoothness of a function around the reference individual point. It not only provides various advantages and flexibility but also creates certain difficulties in developing its calculus and applications. The major techniques of subdifferential analysis revolve around *convex separation* of sets.

Definition 2.30 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function and let $\bar{x} \in \operatorname{dom} f$. An element $v \in \mathbb{R}^n$ is called a* SUBGRADIENT *of f at $\bar{x}$ if*

$$\langle v, x - \bar{x}\rangle \le f(x) - f(\bar{x}) \text{ for all } x \in \mathbb{R}^n. \tag{2.13}$$

The collection of all the subgradients of f at $\bar{x}$ is called the SUBDIFFERENTIAL *of the function at this point and is denoted by $\partial f(\bar{x})$.*

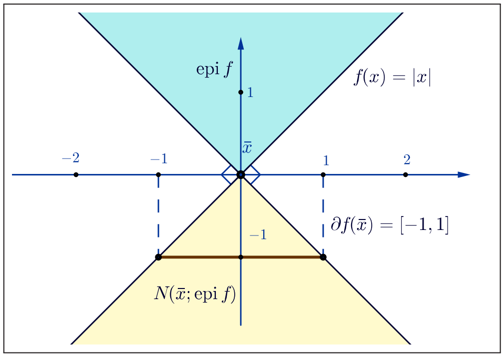

Figure 2.7: Subdifferential of the absolute value function at the origin.

Note that it suffices to take only $x \in \text{dom } f$ in the subgradient definition (2.13). Observe also that $\partial f(\bar{x}) \neq \emptyset$ under more general conditions; see, e.g., Proposition 2.47 as well as other results in this direction given in [26].

The next proposition shows that the subdifferential $\partial f(\bar{x})$ can be equivalently defined geometrically via the normal cone to the epigraph of f at $(\bar{x}, f(\bar{x}))$.

Proposition 2.31 *For any convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and any $\bar{x} \in \text{dom } f$, we have*

$$\partial f(\bar{x}) = \big\{ v \in \mathbb{R}^n \;\big|\; (v, -1) \in N\big((\bar{x}, f(\bar{x})); \text{epi } f\big) \big\}. \tag{2.14}$$

Proof. Fix $v \in \partial f(\bar{x})$ and $(x, \lambda) \in \text{epi } f$. Since $\lambda \geq f(x)$, we deduce from (2.13) that

$$\begin{aligned} \big\langle (v, -1), (x, \lambda) - (\bar{x}, f(\bar{x})) \big\rangle &= \langle v, x - \bar{x} \rangle - (\lambda - f(\bar{x})) \\ &\leq \langle v, x - \bar{x} \rangle - (f(x) - f(\bar{x})) \leq 0. \end{aligned}$$

This readily implies that $(v, -1) \in N((\bar{x}, f(\bar{x})); \text{epi } f)$.

To verify the opposite inclusion in (2.14), take an arbitrary element $v \in \mathbb{R}^n$ such that $(v, -1) \in N((\bar{x}, f(\bar{x})); \text{epi } f)$. For any $x \in \text{dom } f$, we have $(x, f(x)) \in \text{epi } f$. Thus

$$\langle v, x - \bar{x} \rangle - (f(x) - f(\bar{x})) = \big\langle (v, -1), (x, f(x)) - (\bar{x}, f(\bar{x})) \big\rangle \leq 0,$$

which shows that $v \in \partial f(\bar{x})$ and so justifies representation (2.14). $\square$

Along with representing the subdifferential via the normal cone in (2.14), there is the following simple way to express the normal cone to a set via the subdifferential (as well as the singular subdifferential (2.12)) of the extended-real-indicator function.

Example 2.32 Given a nonempty, convex set $\Omega \subset \mathbb{R}^n$, define its *indicator function* by

$$f(x) = \delta(x; \Omega) := \begin{cases} 0 & \text{if } x \in \Omega, \\ \infty & \text{otherwise.} \end{cases} \tag{2.15}$$

Then epi $f = \Omega \times [0, \infty)$, and we have by Proposition 2.11(i) that

$$N\big((\bar{x}, f(\bar{x})); \text{epi } f\big) = N(\bar{x}; \Omega) \times (-\infty, 0],$$

which readily implies the relationships

$$\partial f(\bar{x}) = \partial^\infty f(\bar{x}) = N(\bar{x}; \Omega). \tag{2.16}$$

A remarkable feature of convex functions is that every local minimizer for a convex function gives also the global/absolute minimum to the function.

Proposition 2.33 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ convex and let $\bar{x} \in \operatorname{dom} f$ be a local minimizer of f. Then f attains its global minimum at this point.*

Proof. Since $\bar{x}$ is a local minimizer of f, there is $\delta > 0$ such that

$$f(u) \geq f(\bar{x}) \text{ for all } u \in I\!B(\bar{x}; \delta).$$

Fix $x \in \mathbb{R}^n$ and construct a sequence of $x_k := (1 - k^{-1})\bar{x} + k^{-1}x$ as $k \in \mathbb{N}$. Thus we have $x_k \in I\!B(\bar{x}; \delta)$ when k is sufficiently large. It follows from the convexity of f that

$$f(\bar{x}) \leq f(x_k) \leq (1 - k^{-1}) f(\bar{x}) + k^{-1} f(x),$$

which readily implies that $k^{-1} f(\bar{x}) \leq k^{-1} f(x)$, and hence $f(\bar{x}) \leq f(x)$ for all $x \in \mathbb{R}^n$. □

Next we recall the classical notion of *derivative* for functions on $\mathbb{R}^n$, which we understand throughout the book in the sense of *Fréchet* unless otherwise stated; see Section 3.1.

Definition 2.34 *We say that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is (Fréchet)* DIFFERENTIABLE *at $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$ if there exists an element $v \in \mathbb{R}^n$ such that*

$$\lim_{x \to \bar{x}} \frac{f(x) - f(\bar{x}) - \langle v, x - \bar{x} \rangle}{\|x - \bar{x}\|} = 0.$$

In this case the element v is uniquely defined and is denoted by $\nabla f(\bar{x}) := v$.

The Fermat stationary rule of classical analysis tells us that if $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\bar{x}$ and attains its local minimum at this point, then $\nabla f(\bar{x}) = 0$. The following proposition gives a subdifferential counterpart of this result for general convex functions. Its proof is a direct consequence of the definitions and Proposition 2.33.

Proposition 2.35 *Let* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be convex and let* $\bar{x} \in \operatorname{dom} f$. *Then* f *attains its local/global minimum at* $\bar{x}$ *if and only if* $0 \in \partial f(\bar{x})$.

Proof. Suppose that f attains its global minimum at $\bar{x}$. Then

$$f(\bar{x}) \le f(x) \text{ for all } x \in \mathbb{R}^n,$$

which can be rewritten as

$$0 = \langle 0, x - \bar{x} \rangle \le f(x) - f(\bar{x}) \text{ for all } x \in \mathbb{R}^n .$$

The definition of the subdifferential shows that this is equivalent to $0 \in \partial f(\bar{x})$. □

Now we show that the subdifferential (2.13) is indeed a singleton for differentiable functions reducing to the classical derivative/gradient at the reference point and clarifying the notion of differentiability in the case of convex functions.

Proposition 2.36 *Let* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be convex and differentiable at* $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$. *Then we have* $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$ *and*

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \le f(x) - f(\bar{x}) \textit{ for all } x \in \mathbb{R}^n . \tag{2.17}$$

Proof. It follows from the differentiability of f at $\bar{x}$ that for any $\epsilon > 0$ there is $\delta > 0$ with

$$-\epsilon\|x - \bar{x}\| \le f(x) - f(\bar{x}) - \langle \nabla f(\bar{x}), x - \bar{x} \rangle \le \epsilon\|x - \bar{x}\| \text{ whenever } \|x - \bar{x}\| < \delta. \tag{2.18}$$

Consider further the convex function

$$\varphi(x) := f(x) - f(\bar{x}) - \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \epsilon\|x - \bar{x}\|, \quad x \in \mathbb{R}^n,$$

and observe that $\varphi(x) \ge \varphi(\bar{x}) = 0$ for all $x \in I\!B(\bar{x}; \delta)$. The convexity of φ ensures that $\varphi(x) \ge \varphi(\bar{x})$ for all $x \in \mathbb{R}^n$. Thus

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \le f(x) - f(\bar{x}) + \epsilon\|x - \bar{x}\| \text{ whenever } x \in \mathbb{R}^n,$$

which yields (2.17) by letting $\epsilon \downarrow 0$.

It follows from (2.17) that $\nabla f(\bar{x}) \in \partial f(\bar{x})$. Picking now $v \in \partial f(\bar{x})$, we get

$$\langle v, x - \bar{x} \rangle \le f(x) - f(\bar{x}).$$

Then the second part of (2.18) gives us that

$$\langle v - \nabla f(\bar{x}), x - \bar{x} \rangle \le \epsilon\|x - \bar{x}\| \text{ whenever } \|x - \bar{x}\| < \delta.$$

Finally, we observe that $\|v - \nabla f(\bar{x})\| \le \epsilon$, which yields $v = \nabla f(\bar{x})$ since $\epsilon > 0$ was chosen arbitrarily. Thus $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$. □

Corollary 2.37 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a strictly convex function. Then the differentiability of f at $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$ implies the strict inequality*

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle < f(x) - f(\bar{x}) \ \ \textit{whenever} \ \ x \neq \bar{x}. \tag{2.19}$$

Proof. Since f is convex, we get from Proposition 2.36 that

$$\langle \nabla f(\bar{x}), u - \bar{x} \rangle \le f(u) - f(\bar{x}) \ \text{ for all } \ u \in \mathbb{R}^n .$$

Fix $x \neq \bar{x}$ and let $u := (x + \bar{x})/2$. It follows from the above that

$$\Big\langle \nabla f(\bar{x}), \frac{x + \bar{x}}{2} - \bar{x} \Big\rangle \le f\Big(\frac{x + \bar{x}}{2}\Big) - f(\bar{x}) < \frac{1}{2} f(x) + \frac{1}{2} f(\bar{x}) - f(\bar{x}).$$

Thus we arrive at the strict inequality (2.19) and complete the proof. □

A simple while important example of a nonsmooth convex function is the norm function on $\mathbb{R}^n$. Here is the calculation of its subdifferential.

Example 2.38 Let $p(x) := \|x\|$ be the Euclidean norm function on $\mathbb{R}^n$. Then we have

$$\partial p(x) = \begin{cases} I\!B & \text{if } \ x = 0, \\ \Big\{ \dfrac{x}{\|x\|} \Big\} & \text{otherwise.} \end{cases}$$

To verify this, observe first that the Euclidean norm function p is differentiable at any nonzero point with $\nabla p(x) = \dfrac{x}{\|x\|}$ as $x \neq 0$. It remains to calculate its subdifferential at $x = 0$. To proceed by definition (2.13), we have that $v \in \partial p(0)$ if and only if

$$\langle v, x \rangle = \langle v, x - 0 \rangle \le p(x) - p(0) = \|x\| \ \text{ for all } \ x \in \mathbb{R}^n .$$

Letting $x = v$ gives us $\langle v, v \rangle \le \|v\|$, which implies that $\|v\| \le 1$, i.e., $v \in I\!B$. Now take $v \in I\!B$ and deduce from the Cauchy-Schwarz inequality that

$$\langle v, x - 0 \rangle = \langle v, x \rangle \le \|v\| \cdot \|x\| \le \|x\| = p(x) - p(0) \ \text{ for all } \ x \in \mathbb{R}^n$$

and thus $v \in \partial p(0)$, which shows that $\partial p(0) = I\!B$.

The next theorem calculates the subdifferential of the *distance function*.

Theorem 2.39 Let Ω be a nonempty, closed, convex subset of $\mathbb{R}^n$. Then we have

$$\partial d(\bar{x}; \Omega) = \begin{cases} N(\bar{x}; \Omega) \cap I\!B & \text{if } \ \bar{x} \in \Omega, \\ \Big\{ \dfrac{\bar{x} - \Pi(\bar{x}; \Omega)}{d(\bar{x}; \Omega)} \Big\} & \text{otherwise.} \end{cases}$$

Proof. Consider the case where $\bar{x} \in \Omega$ and fix any element $w \in \partial d(\bar{x};\Omega)$. It follows from the definition of the subdifferential that

$$\langle w, x-\bar{x}\rangle \le d(x;\Omega) - d(\bar{x};\Omega) = d(x;\Omega) \text{ for all } x \in \mathbb{R}^n. \tag{2.20}$$

Since the distance function $d(\cdot;\Omega)$ is Lipschitz continuous with constant $\ell = 1$, we have

$$\langle w, x-\bar{x}\rangle \le \|x-\bar{x}\| \text{ for all } x \in \mathbb{R}^n.$$

This implies that $\|w\| \le 1$, i.e., $w \in I\!B$. It also follows from (2.20) that

$$\langle w, x-\bar{x}\rangle \le 0 \text{ for all } x \in \Omega.$$

Thus $w \in N(\bar{x};\Omega)$, which verifies that $w \in N(\bar{x};\Omega) \cap I\!B$.

To prove the opposite inclusion, fix any $w \in N(\bar{x};\Omega) \cap I\!B$. Then $\|w\| \le 1$ and

$$\langle w, u-\bar{x}\rangle \le 0 \text{ for all } u \in \Omega.$$

Thus for every $x \in \mathbb{R}^n$ and every $u \in \Omega$ we have

$$\begin{aligned}\langle w, x-\bar{x}\rangle &= \langle w, x-u+u-\bar{x}\rangle = \langle w, x-u\rangle + \langle w, u-\bar{x}\rangle \\ &\le \langle w, x-u\rangle \le \|w\|\cdot\|x-u\| \le \|x-u\|.\end{aligned}$$

Since u is taken arbitrarily in Ω, this implies $\langle w, x-\bar{x}\rangle \le d(x;\Omega) = d(x;\Omega) - d(\bar{x};\Omega)$, and hence $w \in \partial d(\bar{x};\Omega)$.

Suppose now that $\bar{x} \notin \Omega$, take $\bar{z} := \Pi(\bar{x};\Omega) \in \Omega$, and fix any $w \in \partial d(\bar{x};\Omega)$. Then

$$\begin{aligned}\langle w, x-\bar{x}\rangle &\le d(x;\Omega) - d(\bar{x};\Omega) = d(x;\Omega) - \|\bar{x}-\bar{z}\| \\ &\le \|x-\bar{z}\| - \|\bar{x}-\bar{z}\| \text{ for all } x \in \mathbb{R}^n.\end{aligned}$$

Denoting $p(x) := \|x-\bar{z}\|$, we have

$$\langle w, x-\bar{x}\rangle \le p(x) - p(\bar{x}) \text{ for all } x \in \mathbb{R}^n,$$

which ensures that $w \in \partial p(\bar{x}) = \left\{\dfrac{\bar{x}-\bar{z}}{\|\bar{x}-\bar{z}\|}\right\}$. Let us show that $w = \dfrac{\bar{x}-\bar{z}}{\|\bar{x}-\bar{z}\|}$ is a subgradient of $d(\cdot;\Omega)$ at $\bar{x}$. Indeed, for any $x \in \mathbb{R}^n$ denote $p_x := \Pi(x;\Omega)$ and get

$$\begin{aligned}\langle w, x-\bar{x}\rangle &= \langle w, x-\bar{z}\rangle + \langle w, \bar{z}-\bar{x}\rangle = \langle w, x-\bar{z}\rangle - \|\bar{x}-\bar{z}\| \\ &= \langle w, x-p_x\rangle + \langle w, p_x-\bar{z}\rangle - \|\bar{x}-\bar{z}\|.\end{aligned}$$

Since $\bar{z} = \Pi(\bar{x};\Omega)$, it follows from Proposition 1.78 that $\langle \bar{x}-\bar{z}, p_x-\bar{z}\rangle \le 0$, and so we have $\langle w, p_x-\bar{z}\rangle \le 0$. Using the fact that $\|w\| = 1$ and the Cauchy-Schwarz inequality gives us

$$\begin{aligned}\langle w, x-\bar{x}\rangle &= \langle w, x-p_x\rangle + \langle w, p_x-\bar{z}\rangle - \|\bar{x}-\bar{z}\| \\ &\le \|w\|\cdot\|x-p_x\| - \|\bar{x}-\bar{z}\| = \|x-p_x\| - \|\bar{x}-\bar{z}\| \\ &= d(x;\Omega) - d(\bar{x};\Omega) \text{ for all } x \in \mathbb{R}^n.\end{aligned}$$

Thus we arrive at $w \in \partial d(\bar{x};\Omega)$ and complete the proof of the theorem. $\square$

We conclude this section by providing first-order and second-order differential characterizations of convexity for differentiable and twice continuously differentiable (C^2) functions, respectively, on given convex regions of $\mathbb{R}^n$.

Theorem 2.40 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a differentiable function on its domain D, which is an open convex set. Then f is convex if and only if

$$\langle \nabla f(u), x - u\rangle \le f(x) - f(u) \text{ for all } x, u \in D. \tag{2.21}$$

Proof. The "only if" part follows from Proposition 2.36. To justify the converse, suppose that (2.21) holds and then fix any $x_1, x_2 \in D$ and $t \in (0,1)$. Denoting $x_t := tx_1 + (1-t)x_2$, we have $x_t \in D$ by the convexity of D. Then

$$\langle \nabla f(x_t), x_1 - x_t\rangle \le f(x_1) - f(x_t), \quad \langle \nabla f(x_t), x_2 - x_t\rangle \le f(x_2) - f(x_t).$$

It follows furthermore that

$$\begin{aligned} &t\langle \nabla f(x_t), x_1 - x_t\rangle \le tf(x_1) - tf(x_t) \text{ and}\\ &(1-t)\langle \nabla f(x_t), x_2 - x_t\rangle \le (1-t)f(x_2) - (1-t)f(x_t). \end{aligned}$$

Summing up these inequalities, we arrive at

$$0 \le tf(x_1) + (1-t)f(x_2) - f(x_t),$$

which ensures that $f(x_t) \le tf(x_1) + (1-t)f(x_2)$, and so verifies the convexity of f. □

Lemma 2.41 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex and twice continuously differentiable on an open subset V of its domain containing $\bar{x}$. Then we have*

$$\langle \nabla^2 f(\bar{x})u, u\rangle \ge 0 \textit{ for all } u \in \mathbb{R}^n,$$

where $\nabla^2 f(\bar{x})$ is the Hessian matrix of f at $\bar{x}$.

Proof. Let $A := \nabla^2 f(\bar{x})$, which is symmetric matrix. Then

$$\lim_{h \to 0} \frac{f(\bar{x} + h) - f(\bar{x}) - \langle \nabla f(\bar{x}), h\rangle - \frac{1}{2}\langle Ah, h\rangle}{\|h\|^2} = 0. \tag{2.22}$$

It follows from (2.22) that for any $\epsilon > 0$ there is $\delta > 0$ such that

$$-\epsilon\|h\|^2 \le f(\bar{x} + h) - f(\bar{x}) - \langle \nabla f(\bar{x}), h\rangle - \frac{1}{2}\langle Ah, h\rangle \le \epsilon\|h\|^2 \text{ for all } \|h\| \le \delta.$$

By Proposition 2.36, we readily have

$$f(\bar{x}+h)-f(\bar{x})-\langle\nabla f(\bar{x}),h\rangle\geq 0.$$

Combining the above inequalities ensures that

$$-\epsilon\|h\|^2\leq\frac{1}{2}\langle Ah,h\rangle\ \text{whenever}\ \ \|h\|\leq\delta. \tag{2.23}$$

Observe further that for any $0\neq u\in\mathbb{R}^n$ the element $h:=\delta\dfrac{u}{\|u\|}$ satisfies $\|h\|\leq\delta$ and, being substituted into (2.23), gives us the estimate

$$-\epsilon\delta^2\leq\frac{1}{2}\delta^2\Big\langle A\frac{u}{\|u\|},\frac{u}{\|u\|}\Big\rangle.$$

It shows therefore (since the case where $u=0$ is trivial) that

$$-2\epsilon\|u\|^2\leq\langle Au,u\rangle\ \ \text{whenever}\ \ u\in\mathbb{R}^n,$$

which implies by letting $\epsilon\downarrow 0$ that $0\leq\langle Au,u\rangle$ for all $u\in\mathbb{R}^n$. □

Now we are ready to justify the aforementioned second-order characterization of convexity for twice continuously differentiable functions.

Theorem 2.42 Let $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ be a function twice continuously differentiable on its domain D, which is an open convex subset of $\mathbb{R}^n$. Then f is convex if and only if $\nabla^2 f(\bar{x})$ is positive semidefinite for every $\bar{x}\in D$.

Proof. Taking Lemma 2.41 into account, we only need to verify that if $\nabla^2 f(\bar{x})$ is positive semidefinite for every $\bar{x}\in D$, then f is convex. To proceed, for any $x_1,x_2\in D$ define $x_t:=tx_1+(1-t)x_2$ and consider the function

$$\varphi(t):=f\big(tx_1+(1-t)x_2\big)-tf(x_1)-(1-t)f(x_2),\quad t\in\mathbb{R}.$$

It is clear that φ is well defined on an open interval I containing $(0,1)$. Then $\varphi'(t)=\langle\nabla f(x_t),x_1-x_2\rangle-f(x_1)+f(x_2)$ and $\varphi''(t)=\langle\nabla^2 f(x_t)(x_1-x_2),x_1-x_2\rangle\geq 0$ for every $t\in I$ since $\nabla^2 f(x_t)$ is positive semidefinite. It follows from Corollary 1.46 that φ is convex on I. Since $\varphi(0)=\varphi(1)=0$, for any $t\in(0,1)$ we have

$$\varphi(t)=\varphi\big(t(1)+(1-t)0\big)\leq t\varphi(1)+(1-t)\varphi(0)=0,$$

which implies in turn the inequality

$$f\big(tx_1+(1-t)x_2\big)\leq tf(x_1)+(1-t)f(x_2),$$

This justifies that the function f is convex on its domain and thus on $\mathbb{R}^n$. □

2.5 BASIC CALCULUS RULES

Here we derive basic calculus rules for the convex subdifferential (2.13) and its singular counterpart (2.12). First observe the obvious subdifferential rule for scalar multiplication.

Proposition 2.43 *For a convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ with $\bar{x} \in \operatorname{dom} f$, we have*

$$\partial(\alpha f)(\bar{x}) = \alpha \partial f(\bar{x}) \ \ \textit{and} \ \ \partial^\infty(\alpha f)(\bar{x}) = \alpha \partial^\infty f(\bar{x}) \ \ \textit{whenever} \ \ \alpha > 0.$$

Let us now establish the fundamental *subdifferential sum rules* for both constructions (2.13) and (2.12) that play a crucial role in convex analysis. We give a geometric proof for this result reducing it to the intersection rule for normals to sets, which in turn is based on convex separation of extremal set systems. Observe that our variational approach does not require any closedness/lower semicontinuity assumptions on the functions in question.

Theorem 2.44 Let $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ for $i = 1, 2$ be convex functions and let $\bar{x} \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$. Impose the singular subdifferential qualification condition

$$\partial^\infty f_1(\bar{x}) \cap \big[-\partial^\infty f_2(\bar{x}) \big] = \{0\}. \tag{2.24}$$

Then we have the subdifferential sum rules

$$\partial(f_1 + f_2)(\bar{x}) = \partial f_1(\bar{x}) + \partial f_2(\bar{x}), \quad \partial^\infty(f_1 + f_2)(\bar{x}) = \partial^\infty f_1(\bar{x}) + \partial^\infty f_2(\bar{x}). \tag{2.25}$$

Proof. For definiteness we verify the first rule in (2.25); the proof of the second result is similar. Fix an arbitrary subgradient $v \in \partial(f_1 + f_2)(\bar{x})$ and define the sets

$$\Omega_1 := \{(x, \lambda_1, \lambda_2) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \mid \lambda_1 \geq f_1(x)\},$$
$$\Omega_2 := \{(x, \lambda_1, \lambda_2) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \mid \lambda_2 \geq f_2(x)\}.$$

Let us first verify the inclusion

$$(v, -1, -1) \in N\big((\bar{x}, f_1(\bar{x}), f_2(\bar{x})); \Omega_1 \cap \Omega_2\big). \tag{2.26}$$

For any $(x, \lambda_1, \lambda_2) \in \Omega_1 \cap \Omega_2$, we have $\lambda_1 \geq f_1(x)$ and $\lambda_2 \geq f_2(x)$. Then

$$\begin{aligned}&\langle v, x - \bar{x}\rangle + (-1)(\lambda_1 - f_1(\bar{x})) + (-1)(\lambda_2 - f_2(\bar{x}))\\ &\leq \langle v, x - \bar{x}\rangle - \big(f_1(x) + f_2(x) - f_1(\bar{x}) - f_2(\bar{x})\big) \leq 0,\end{aligned}$$

where the last estimate holds due to $v \in \partial(f_1 + f_2)(\bar{x})$. Thus (2.26) is satisfied.

To proceed with employing the intersection rule in (2.26), let us first check that the assumed qualification condition (2.24) yields

$$N\big((\bar{x}, f_1(\bar{x}), f_2(\bar{x})); \Omega_1\big) \cap \big[-N\big((\bar{x}, f_1(\bar{x}), f_2(\bar{x})); \Omega_2\big)\big] = \{0\}, \tag{2.27}$$

which is the basic qualification condition (2.9) needed for the application of Corollary 2.16 in the setting of (2.26). Indeed, we have from $\Omega_1 = \text{epi}\, f_1 \times \mathbb{R}$ that

$$N\big((\bar{x}, f_1(\bar{x}), f_2(\bar{x})); \Omega_1\big) = N\big((\bar{x}, f_1(\bar{x})); \text{epi}\, f_1\big) \times \{0\}$$

and observe similarly the representation

$$N\big((\bar{x}, f_1(\bar{x}), f_2(\bar{x})); \Omega_2\big) = \big\{(v, 0, \gamma) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \,\big|\, (v, \gamma) \in N\big((\bar{x}, f_2(\bar{x})); \text{epi}\, f_2\big)\big\}.$$

Further, fix any element

$$(v, \gamma_1, \gamma_2) \in N\big((\bar{x}, f_1(\bar{x}), f_2(\bar{x})); \Omega_1\big) \cap \big[-N\big((\bar{x}, f_1(\bar{x}), f_2(\bar{x})); \Omega_2\big)\big].$$

Then $\gamma_1 = \gamma_2 = 0$, and hence

$$(v, 0) \in N\big((\bar{x}, f_1(\bar{x})); \text{epi}\, f_1\big) \text{ and } (-v, 0) \in N\big((\bar{x}, f_2(\bar{x})); \text{epi}\, f_2\big).$$

It follows by the definition of the singular subdifferential that

$$v \in \partial^\infty f_1(\bar{x}) \cap \big[-\partial^\infty f_2(\bar{x})\big] = \{0\}.$$

It yields $(v, \gamma_1, \gamma_2) = (0, 0, 0) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}$. Applying Corollary 2.16 in (2.26) shows that

$$N\big((\bar{x}, f_1(\bar{x}), f_2(\bar{x})); \Omega_1 \cap \Omega_2\big) = N\big((\bar{x}, f_1(\bar{x}), f_2(\bar{x})); \Omega_1\big) + N\big((\bar{x}, f_1(\bar{x}), f_2(\bar{x})); \Omega_2\big),$$

which implies in turn that

$$(v, -1, -1) = (v_1, -\gamma_1, 0) + (v_2, 0, -\gamma_2)$$

with $(v_1, -\gamma_1) \in N((\bar{x}, f_1(\bar{x})); \text{epi}\, f_1)$ and $(v_2, -\gamma_2) \in N((\bar{x}, f_1(\bar{x})); \text{epi}\, f_2)$. Then we get $\gamma_1 = \gamma_2 = 1$ and $v = v_1 + v_2$, where $v_i \in \partial f_i(\bar{x})$ for $i = 1, 2$. This verifies the inclusion "$\subset$" in the first subdifferential sum rule of (2.25).

The opposite inclusion therein follows easily from the definition of the subdifferential. Indeed, any $v \in \partial f_1(\bar{x}) + \partial f_2(\bar{x})$ can be represented as $v = v_1 + v_2$, where $v_i \in \partial f_i(\bar{x})$ for $i = 1, 2$. Then we have

$$\begin{aligned}\langle v, x - \bar{x}\rangle &= \langle v_1, x - \bar{x}\rangle + \langle v_2, x - \bar{x}\rangle \\ &\le f_1(x) - f_1(\bar{x}) + f_2(x) - f_2(\bar{x}) = (f_1 + f_1)(x) - (f_1 + f_2)(\bar{x})\end{aligned}$$

for all $x \in \mathbb{R}^n$, which readily implies that $v \in \partial(f_1 + f_2)(\bar{x})$ and therefore completes the proof of the theorem. $\square$

Note that by Proposition 2.23 the qualification condition (2.24) can be written as

$$N(\bar{x}; \text{dom}\, f_1) \cap \big[-N(\bar{x}; \text{dom}\, f_2)\big] = \{0\}.$$

Let us present some useful consequences of Theorem 2.44. The sum rule for the (basic) subdifferential (2.13) in the following corollary is known as the *Moreau-Rockafellar theorem*.

Corollary 2.45 *Let $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ for $i = 1, 2$ be convex functions such that there exists $u \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$ for which f_1 is continuous at u or f_2 is continuous at u. Then*

$$\partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x) \tag{2.28}$$

whenever $x \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$. Consequently, if both functions f_i are finite-valued on $\mathbb{R}^n$, then the sum rule (2.28) holds for all $x \in \mathbb{R}^n$.

Proof. Suppose for definiteness that f_1 is continuous at $u \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$, which implies that $u \in \operatorname{int}(\operatorname{dom} f_1) \cap \operatorname{dom} f_2$. Then the proof of Corollary 2.18 shows that

$$N(x; \operatorname{dom} f_1) \cap \big[-N(x; \operatorname{dom} f_2)\big] = \{0\} \;\text{ whenever }\; x \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2.$$

Thus the conclusion follows directly from Theorem 2.44. □

The next corollary for sums of finitely many functions can be derived from Theorem 2.44 by induction, where the validity of the qualification condition in the next step of induction follows from the singular subdifferential sum rule in (2.25).

Corollary 2.46 *Consider finitely many convex functions $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ for $i = 1, \ldots, m$. The following assertions hold:*

(i) *Let $\bar{x} \in \bigcap_{i=1}^m \operatorname{dom} f_i$ and let the implication*

$$\Big[\sum_{i=1}^m v_i = 0,\; v_i \in \partial^\infty f_i(\bar{x})\Big] \Longrightarrow \big[v_i = 0 \;\text{ for all }\; i = 1, \ldots, m\big] \tag{2.29}$$

hold. Then we have the subdifferential sum rules

$$\partial\Big(\sum_{i=1}^m f_i\Big)(\bar{x}) = \sum_{i=1}^m \partial f_i(\bar{x}), \quad \partial^\infty\Big(\sum_{i=1}^m f_i\Big)(\bar{x}) = \sum_{i=1}^m \partial f_i(\bar{x}).$$

(ii) *Suppose that there is $u \in \bigcap_{i=1}^m \operatorname{dom} f_i$ such that all (except possibly one) functions f_i are continuous at u. Then we have the equality*

$$\partial\Big(\sum_{i=1}^m f_i\Big)(x) = \sum_{i=1}^m \partial f_i(x)$$

for any common point of the domains $x \in \bigcap_{i=1}^m \operatorname{dom} f_i$.

The next result not only ensures the subdifferentiability of convex functions, but also justifies the compactness of its subdifferential at interior points of the domain.

Proposition 2.47 *For any convex function* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *and any* $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$, *we have that the subgradient set* $\partial f(\bar{x})$ *is nonempty and compact in* $\mathbb{R}^n$.

Proof. Let us first verify the boundedness of the subdifferential. It follows from Corollary 2.27 that f is locally Lipschitz continuous around $\bar{x}$ with some constant ℓ, i.e., there exists a positive number η such that

$$|f(x) - f(y)| \le \ell\|x - y\| \ \text{ for all } \ x, y \in I\!B(\bar{x}; \eta).$$

Thus for any subgradient $v \in \partial f(\bar{x})$ and any element $h \in \mathbb{R}^n$ with $\|h\| \le \delta$ we have

$$\langle v, h\rangle \le f(\bar{x} + h) - f(\bar{x}) \le \ell\|h\|.$$

This implies the bound $\|v\| \le \ell$. The closedness of $\partial f(\bar{x})$ follows directly from the definition, and so we get the compactness of the subgradient set.

It remains to verify that $\partial f(\bar{x}) \ne \emptyset$. Since $(\bar{x}, f(\bar{x})) \in \operatorname{bd}(\operatorname{epi} f)$, we deduce from Corollary 2.14 that $N((\bar{x}, f(\bar{x})); \operatorname{epi} f) \ne \{0\}$. Take such an element $(0, 0) \ne (v, -\gamma) \in N((\bar{x}, f(\bar{x})); \operatorname{epi} f)$ and get by the definition that

$$\langle v, x - \bar{x}\rangle - \gamma(\lambda - f(\bar{x})) \le 0 \ \text{ for all } \ (x, \lambda) \in \operatorname{epi} f.$$

Using this inequality with $x := \bar{x}$ and $\lambda = f(\bar{x}) + 1$ yields $\gamma \ge 0$. If $\gamma = 0$, then $v \in \partial^\infty f(\bar{x}) = \{0\}$ by Theorem 2.29, which is a contradiction. Thus $\gamma > 0$ and $(v/\gamma, -1) \in N((\bar{x}, f(\bar{x})); \operatorname{epi} f)$. Hence $v/\gamma \in \partial f(\bar{x})$, which completes the proof of the proposition. $\square$

Given $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and any $\alpha \in \mathbb{R}$, define the *level set*

$$\Omega_\alpha := \big\{x \in \mathbb{R}^n \ \big|\ f(x) \le \alpha\big\}.$$

Proposition 2.48 *Let* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be convex and let* $\bar{x} \in \Omega_\alpha$ *with* $f(\bar{x}) = \alpha$ *for some* $\alpha \in \mathbb{R}$. *If* $0 \notin \partial f(\bar{x})$, *then the normal cone to the level set* Ω_α *at* $\bar{x}$ *is calculated by*

$$N(\bar{x}; \Omega_\alpha) = \partial^\infty f(\bar{x}) \cup \mathbb{R}_> \partial f(\bar{x}), \tag{2.30}$$

where $\mathbb{R}_>$ *denotes the set of all positive real numbers.*

Proof. To proceed, consider the set $\Theta := \mathbb{R}^n \times \{\alpha\} \subset \mathbb{R}^{n+1}$ and observe that

$$\Omega_\alpha \times \{\alpha\} = \text{epi } f \cap \Theta.$$

Let us check the relationship

$$N\big((\bar{x}, \alpha); \text{epi } f\big) \cap \big[-N\big((\bar{x}, \alpha); \Theta\big)\big] = \{0\}, \tag{2.31}$$

i.e., the qualification condition (2.9) holds for the sets epi f and Θ. Indeed, take $(v, -\lambda)$ from the set on left-hand side of (2.31) and deduce from the structure of Θ that $v = 0$ and that $\lambda \geq 0$. If we suppose that $\lambda > 0$, then

$$\frac{1}{\lambda}(v, -\lambda) = (0, -1) \in N\big((\bar{x}, \alpha); \text{epi } f\big),$$

and so $0 \in \partial f(\bar{x})$, which contradicts the assumption of this proposition. Thus we have $(v, -\lambda) = (0, 0)$, and then applying the intersection rule of Corollary 2.16 gives us

$$N(\bar{x}, \Omega_\alpha) \times \mathbb{R} = N\big((\bar{x}, \alpha); \Omega_\alpha \times \{\alpha\}\big) = N\big((\bar{x}, \alpha); \text{epi } f\big) + (\{0\} \times \mathbb{R}).$$

Fix now any $v \in N(\bar{x}; \Omega_\alpha)$ and get $(v, 0) \in N(\bar{x}, \Omega_\alpha) \times \mathbb{R}$. This implies the existence of $\lambda \geq 0$ such that $(v, -\lambda) \in N((\bar{x}, \alpha); \text{epi } f)$ and $(v, 0) = (v, -\lambda) + (0, \lambda)$. If $\lambda = 0$, then $v \in \partial^\infty f(\bar{x})$. In the case when $\lambda > 0$ we obtain $v \in \lambda \partial f(\bar{x})$ and thus verify the inclusion "$\subset$" in (2.30).

Let us finally prove the opposite inclusion in (2.30). Since $\Omega_\alpha \subset \text{dom } f$, we get

$$\partial^\infty f(\bar{x}) = N(\bar{x}; \text{dom } f) \subset N(\bar{x}; \Omega_\alpha).$$

Take any $v \in \mathbb{R}_> \partial f(\bar{x})$ and find $\lambda > 0$ such that $v \in \lambda \partial f(\bar{x})$. Taking any $x \in \Omega_\alpha$ gives us $f(x) \leq \alpha = f(\bar{x})$. Therefore, we have

$$\langle \lambda^{-1} v, x - \bar{x} \rangle \leq f(x) - f(\bar{x}) \leq 0,$$

which implies that $v \in N(\bar{x}; \Omega_\alpha)$ and thus completes the proof. □

Corollary 2.49 *In the setting of Proposition 2.48 suppose that f is continuous at $\bar{x}$ and that $0 \notin \partial f(\bar{x})$. Then we have the representation*

$$N(\bar{x}; \Omega_\alpha) = \mathbb{R}_+ \partial f(\bar{x}) = \bigcup_{\alpha \geq 0} \alpha \partial f(\bar{x}), \quad \alpha \in \mathbb{R}.$$

Proof. Since f is continuous at $\bar{x}$, it follows from Theorem 2.29 that $\partial^\infty f(\bar{x}) = \{0\}$. Then the result of this corollary follows directly from Proposition 2.48. □

We now obtain some other calculus rules for subdifferentiation of convex function. For brevity, let us only give results for the basic subdifferential (2.13) leaving their singular subdifferential counterparts as exercises; see Section 2.11.

To proceed with deriving chain rules, we present first the following geometric lemma.

Lemma 2.50 *Let $B : \mathbb{R}^n \to \mathbb{R}^p$ be an affine mapping given by $B(x) = A(x) + b$, where $A : \mathbb{R}^n \to \mathbb{R}^p$ is a linear mapping and $b \in \mathbb{R}^p$. For any $(\bar{x}, \bar{y}) \in \operatorname{gph} B$, we have*

$$N\big((\bar{x}, \bar{y}); \operatorname{gph} B\big) = \big\{(u, v) \in \mathbb{R}^n \times \mathbb{R}^p \bigm| u = -A^* v\big\}.$$

Proof. It is clear that the set $\operatorname{gph} B$ is convex and $(u, v) \in N((\bar{x}, \bar{y}) \operatorname{gph} B)$ if and only if

$$\langle u, x - \bar{x}\rangle + \langle v, B(x) - B(\bar{x})\rangle \le 0 \;\text{ whenever }\; x \in \mathbb{R}^n. \tag{2.32}$$

It follows directly from the definitions that

$$\begin{aligned}\langle u, x - \bar{x}\rangle + \langle v, B(x) - B(\bar{x})\rangle &= \langle u, x - \bar{x}\rangle + \langle v, A(x) - A(\bar{x})\rangle\\ &= \langle u, x - \bar{x}\rangle + \langle A^* v, x - \bar{x}\rangle\\ &= \langle u + A^* v, x - \bar{x}\rangle,\end{aligned}$$

which implies the equivalence of (2.32) to $\langle u + A^* v, x - \bar{x}\rangle \le 0$, and so to $u = -A^* v$. □

Theorem 2.51 Let $f : \mathbb{R}^p \to \overline{\mathbb{R}}$ be a convex function and let $B : \mathbb{R}^n \to \mathbb{R}^p$ be as in Lemma 2.50 with $B(\bar{x}) \in \operatorname{dom} f$ for some $\bar{x} \in \mathbb{R}^p$. Denote $\bar{y} := B(\bar{x})$ and assume that

$$\ker A^* \cap \partial^\infty f(\bar{y}) = \{0\}. \tag{2.33}$$

Then we have the subdifferential chain rule

$$\partial(f \circ B)(\bar{x}) = A^*\big(\partial f(\bar{y})\big) = \big\{A^* v \bigm| v \in \partial f(\bar{y})\big\}. \tag{2.34}$$

Proof. Fix $v \in \partial(f \circ B)(\bar{x})$ and form the subsets of $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}$ by

$$\Omega_1 := \operatorname{gph} B \times \mathbb{R} \quad\text{and}\quad \Omega_2 := \mathbb{R}^n \times \operatorname{epi} f.$$

It follows from the definition of the subdifferential and the normal cone that $(v, 0, -1) \in N((\bar{x}, \bar{y}, \bar{z}); \Omega_1 \cap \Omega_2)$, where $\bar{z} := f(\bar{y})$. Indeed, fix any $(x, y, \lambda) \in \Omega_1 \cap \Omega_2$. Then $y = B(x)$ and $\lambda \ge f(y)$, and so $\lambda \ge f(B(x))$. Thus

$$\langle v, x - \bar{x}\rangle + 0(y - \bar{y}) + (-1)(\lambda - \bar{z}) \le \langle v, x - \bar{x}\rangle - [f(B(x)) - f(B(\bar{x}))] \le 0.$$

Employing the intersection rule from Corollary 2.19 under (2.33) gives us

$$(v, 0, -1) \in N\big((\bar{x}, \bar{y}, \bar{z}); \Omega_1\big) + N\big((\bar{x}, \bar{y}, \bar{z}); \Omega_2\big),$$

which reads that $(v, 0, -1) = (v, -w, 0) + (0, w, -1)$ with $(v, -w) \in N((\bar{x}, \bar{y}); \operatorname{gph} B)$ and $(w, -1) \in N((\bar{y}, \bar{z}); \operatorname{epi} f)$. Then we get

$$v = A^* w \quad \text{and} \quad w \in \partial f(\bar{y}),$$

which implies in turn that $v \in A^*(\partial f(\bar{y}))$ and thus verifies the inclusion "$\subset$" in (2.34). The opposite inclusion follows directly from the definition of the subdifferential. $\square$

In the corollary below we present some conditions that guarantee that the qualification (2.33) in Theorem 2.51 is satisfied.

Corollary 2.52 *In the setting of Theorem 2.51 suppose that either A is surjective or f is continuous at the point $\bar{y} \in \mathbb{R}^n$; the latter assumption is automatic when f is finite-valued. Then we have the subdifferential chain rule (2.34).*

Proof. Let us check that the qualification condition (2.33) holds under the assumptions of this corollary. The surjectivity of A gives us $\ker A^* = \{0\}$, and in the second case we have $\bar{y} \in \operatorname{int}(\operatorname{dom} f)$, so $\partial^\infty f(\bar{y}) = \{0\}$. The validity of (2.33) is obvious. $\square$

The last consequence of Theorem 2.51 presents a chain rule for normals to sets.

Corollary 2.53 *Let $\Omega \subset \mathbb{R}^p$ be a convex set and let $B : \mathbb{R}^n \to \mathbb{R}^p$ be an affine mapping as in Lemma 2.50 satisfying $\bar{y} := B(\bar{x}) \in \Omega$. Then we have*

$$N(\bar{x}; B^{-1}(\Omega)) = A^*(N(\bar{y}; \Omega))$$

under the validity of the qualification condition

$$\ker A^* \cap N(\bar{y}; \Omega) = \{0\}.$$

Proof. This follows from Theorem 2.51 with $f(x) = \delta(x; \Omega)$. $\square$

Given $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ for $i = 1, \ldots, m$, consider the *maximum function*

$$f(x) := \max_{i=1,\ldots,m} f_i(x), \quad x \in \mathbb{R}^n, \tag{2.35}$$

and for $\bar{x} \in \mathbb{R}^n$ define the set

$$I(\bar{x}) := \big\{ i \in \{1, \ldots, m\} \;\big|\; f_i(\bar{x}) = f(\bar{x}) \big\}.$$

In the proposition below we assume that each function f_i for $i = 1, \ldots, m$ is continuous at the reference point. A more general version follows afterward.

Proposition 2.54 *Let $f_i\colon \mathbb{R}^n \to \overline{\mathbb{R}}$, $i = 1, \ldots, m$, be convex functions. Take any point $\bar{x} \in \bigcap_{i=1}^m \operatorname{dom} f_i$ and assume that each f_i is continuous at $\bar{x}$. Then we have the maximum rule*

$$\partial\big(\max f_i\big)(\bar{x}) = \operatorname{co} \bigcup_{i \in I(\bar{x})} \partial f_i(\bar{x}).$$

Proof. Let $f(x)$ be the maximum function defined in (2.35). We obviously have

$$\operatorname{epi} f = \bigcap_{i=1}^m \operatorname{epi} f_i.$$

Since $f_i(\bar{x}) < f(\bar{x}) =: \bar{\alpha}$ for any $i \notin I(\bar{x})$, there exist a neighborhood U of $\bar{x}$ and $\delta > 0$ such that $f_i(x) < \alpha$ whenever $(x, \alpha) \in U \times (\bar{\alpha} - \delta, \bar{\alpha} + \delta)$. It follows that $(\bar{x}, \bar{\alpha}) \in \operatorname{int} \operatorname{epi} f_i$, and so $N((\bar{x}, \bar{\alpha}); \operatorname{epi} f_i) = \{(0,0)\}$ for such indices i.

Let us show that the qualification condition from Corollary 2.19 is satisfied for the sets $\Omega_i := \operatorname{epi} f_i$, $i = 1, \ldots, m$, at $(\bar{x}, \bar{\alpha})$. Fix $(v_i, -\lambda_i) \in N\big((\bar{x}, \bar{\alpha}); \operatorname{epi} f_i\big)$ with

$$\sum_{i=1}^m (v_i, \lambda_i) = 0.$$

Then $(v_i, -\lambda_i) = (0,0)$ for $i \notin I(\bar{x})$, and hence

$$\sum_{i \in I(\bar{x})} (v_i, -\lambda_i) = 0.$$

Note that $f_i(\bar{x}) = f(\bar{x}) = \bar{\alpha}$ for $i \in I(\bar{x})$. The proof of Proposition 2.47 tells us that $\lambda_i \geq 0$ for $i \in I(\bar{x})$. Thus the equality $\sum_{i \in I(\bar{x})} \lambda_i = 0$ yields $\lambda_i = 0$, and so $v_i \in \partial^\infty f_i(\bar{x}) = \{0\}$ for every $i \in I(\bar{x})$ by Theorem 2.29. This verifies that $(v_i, -\lambda_i) = (0,0)$ for $i = 1, \ldots, m$. It follows furthermore that

$$N\big((\bar{x}, f(\bar{x})); \operatorname{epi} f\big) = \sum_{i=1}^m N\big((\bar{x}, \bar{\alpha}); \operatorname{epi} f_i\big) = \sum_{i \in I(\bar{x})} N\big((\bar{x}, f_i(\bar{x})); \operatorname{epi} f_i\big).$$

For any $v \in \partial f(\bar{x})$, we have $(v, -1) \in N((\bar{x}, f(\bar{x})); \operatorname{epi} f)$, which allows us to find $(v_i, -\lambda_i) \in N((\bar{x}, f_i(\bar{x})); \operatorname{epi} f_i)$ for $i \in I(\bar{x})$ such that

$$(v, -1) = \sum_{i \in I(\bar{x})} (v_i, -\lambda_i).$$

This yields $\sum_{i \in I(\bar{x})} \lambda_i = 1$ and $v = \sum_{i \in I(\bar{x})} v_i$. If $\lambda_i = 0$, then $v_i \in \partial^\infty f_i(\bar{x}) = \{0\}$. Hence we assume without loss of generality that $\lambda_i > 0$ for every $i \in I(\bar{x})$. Since $N((\bar{x}, f_i(\bar{x})); \operatorname{epi} f_i)$ is a cone, it gives us $(v_i/\lambda_i, -1) \in N((\bar{x}, f_i(\bar{x})); \text{epi } f_i)$, and so $v_i \in \lambda_i \partial f_i(\bar{x})$. We get the representation $v = \sum_{i \in I(\bar{x})} \lambda_i u_i$, where $u_i \in \partial f_i(\bar{x})$ and $\sum_{i \in I(\bar{x})} \lambda_i = 1$. Thus

$$v \in \text{co} \bigcup_{i \in I(\bar{x})} \partial f_i(\bar{x}).$$

Note that the opposite inclusion follows directly from

$$\partial f_i(\bar{x}) \subset \partial f(\bar{x}) \text{ for all } i \in I(\bar{x}),$$

which in turn follows from the definition. Indeed, for any $i \in I(\bar{x})$ and $v \in \partial f_i(\bar{x})$ we have

$$\langle v, x - \bar{x} \rangle \le f_i(x) - f_i(\bar{x}) = f_i(x) - f(\bar{x}) \le f(x) - f(\bar{x}) \quad \text{whenever } x \in \mathbb{R}^n,$$

which implies that $v \in \partial f(\bar{x})$ and thus completes the proof. □

To proceed with the general case, define

$$\Lambda(\bar{x}) := \Big\{(\lambda_1, \ldots, \lambda_m) \;\Big|\; \lambda_i \ge 0, \; \sum_{i=1}^m \lambda_i = 1, \; \lambda_i\big(f_i(\bar{x}) - f(\bar{x})\big) = 0\Big\}$$

and observe from the definition of $\Lambda(\bar{x})$ hat

$$\Lambda(\bar{x}) = \Big\{(\lambda_1, \ldots, \lambda_m) \;\Big|\; \lambda_i \ge 0, \; \sum_{i \in I(\bar{x})} \lambda_i = 1, \; \lambda_i = 0 \text{ for } i \notin I(\bar{x})\Big\}.$$

Theorem 2.55 Let $f_i \colon \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex functions for $i = 1, \ldots, m$. Take any point $\bar{x} \in \bigcap_{i=1}^m \text{dom } f_i$ and assume that each f_i is continuous at $\bar{x}$ for any $i \notin I(\bar{x})$ and that the qualification condition

$$\Big[\sum_{i \in I(\bar{x})} v_i = 0, \; v_i \in \partial^\infty f_i(\bar{x})\Big] \Longrightarrow \big[v_i = 0 \; \text{ for all } \; i \in I(\bar{x})\big]$$

holds. Then we have the maximum rule

$$\partial\big(\max f_i\big)(\bar{x}) = \bigcup \Big\{ \sum_{i \in I(\bar{x})} \lambda_i \circ \partial f_i(\bar{x}) \;\Big|\; (\lambda_1, \ldots, \lambda_m) \in \Lambda(\bar{x}) \Big\},$$

where $\lambda_i \circ \partial f_i(\bar{x}) := \lambda_i \partial f_i(\bar{x})$ if $\lambda_i > 0$, and $\lambda_i \circ \partial f_i(\bar{x}) := \partial^\infty f_i(\bar{x})$ if $\lambda_i = 0$.

Proof. Let $f(x)$ be the maximum function from (2.35). Repeating the first part of the proof of Proposition 2.54 tells us that $N((\bar{x}, f(\bar{x})); \text{epi } f_i) = \{(0,0)\}$ for $i \notin I(\bar{x})$. The imposed qualification condition allows us to employ the intersection rule from Corollary 2.19 to get

$$N\big((\bar{x}, f(\bar{x})); \text{epi } f\big) = \sum_{i=1}^m N\big((\bar{x}, f(\bar{x})); \text{epi } f_i\big) = \sum_{i \in I(\bar{x})} N\big((\bar{x}, f_i(\bar{x})); \text{epi } f_i\big). \tag{2.36}$$

For any $v \in \partial f(\bar{x})$, we have $(v, -1) \in N((\bar{x}, f(\bar{x})); \operatorname{epi} f)$, and thus find by using (2.36) such pairs $(v_i, -\lambda_i) \in N((\bar{x}, f_i(\bar{x})); \operatorname{epi} f_i)$ for $i \in I(\bar{x})$ that

$$(v, -1) = \sum_{i \in I(\bar{x})} (v_i, -\lambda_i).$$

This shows that $\sum_{i \in I(\bar{x})} \lambda_i = 1$ and $v = \sum_{i \in I(\bar{x})} v_i$ with $v_i \in \lambda_i \circ \partial f_i(\bar{x})$ and $\lambda_i \geq 0$ for all $i \in I(\bar{x})$. Therefore, we arrive at

$$v \in \bigcup \Big\{ \sum_{i \in I(\bar{x})} \lambda_i \circ \partial f_i(\bar{x}) \ \Big| \ (\lambda_1, \ldots, \lambda_m) \in \Lambda(\bar{x}) \Big\},$$

which justifies the inclusion "$\subset$" in the maximum rule. The opposite inclusion can also be verified easily by using representations (2.36). $\square$

2.6 SUBGRADIENTS OF OPTIMAL VALUE FUNCTIONS

This section is mainly devoted to the study of the class of *optimal value/marginal functions* defined in (1.4). To proceed, we first introduce and examine the notion of *coderivatives* for set-valued mappings, which is borrowed from variational analysis and plays a crucial role in many aspects of the theory and applications of set-valued mappings; in particular, those related to subdifferentiation of optimal value functions considered below.

Definition 2.56 *Let $F: \mathbb{R}^n \rightrightarrows \mathbb{R}^p$ be a set-valued mapping with convex graph and let $(\bar{x}, \bar{y}) \in \operatorname{gph} F$. The* CODERIVATIVE *of F at $(\bar{x}, \bar{y})$ is the set-valued mapping $D^*F(\bar{x}, \bar{y}): \mathbb{R}^p \rightrightarrows \mathbb{R}^n$ with the values*

$$D^*F(\bar{x}, \bar{y})(v) := \big\{ u \in \mathbb{R}^n \ \big| \ (u, -v) \in N\big((\bar{x}, \bar{y}); \operatorname{gph} F\big) \big\}, \quad v \in \mathbb{R}^p. \tag{2.37}$$

It follows from (2.37) and definition (2.4) of the normal cone to convex sets that the coderivative values of convex-graph mappings can be calculated by

$$D^*F(\bar{x}, \bar{y})(v) = \Big\{ u \in \mathbb{R}^n \ \Big| \ \langle u, \bar{x} \rangle - \langle v, \bar{y} \rangle = \max_{(x,y) \in \operatorname{gph} F} \big[\langle u, x \rangle - \langle v, y \rangle \big] \Big\}.$$

Now we present explicit calculations of the coderivative in some particular settings.

Proposition 2.57 *Let $F: \mathbb{R}^n \to \mathbb{R}^p$ be a single-valued mapping given by $F(x) := Ax + b$, where $A: \mathbb{R}^n \to \mathbb{R}^p$ is a linear mapping, and where $b \in \mathbb{R}^p$. For any pair $(\bar{x}, \bar{y})$ such that $\bar{y} = A\bar{x} + b$, we have the representation*

$$D^*F(\bar{x}, \bar{y})(v) = \{A^*v\} \ \text{ whenever } \ v \in \mathbb{R}^p.$$

Proof. Lemma 2.50 tells us that

$$N\big((\bar{x},\bar{y});\operatorname{gph} F\big) = \big\{(u,v) \in \mathbb{R}^n \times \mathbb{R}^p \;\big|\; u = -A^* v\big\}.$$

Thus $u \in D^*F(\bar{x},\bar{y})(v)$ if and only if $u = A^*v$. $\square$

Proposition 2.58 *Given a convex function* $f:\mathbb{R}^n \to \overline{\mathbb{R}}$ *with* $\bar{x} \in \operatorname{dom} f$*, define the set-valued mapping* $F:\mathbb{R}^n \rightrightarrows \mathbb{R}$ *by*

$$F(x) := \big[f(x),\infty\big).$$

Denoting $\bar{y} := f(\bar{x})$*, we have*

$$D^*F(\bar{x},\bar{y})(\lambda) = \begin{cases} \lambda\partial f(\bar{x}) & \text{if } \lambda > 0,\\ \partial^\infty f(\bar{x}) & \text{if } \lambda = 0,\\ \emptyset & \text{if } \lambda < 0. \end{cases}$$

In particular, if f *is finite-valued, then* $D^*F(\bar{x},\bar{y})(\lambda) = \lambda\partial f(\bar{x})$ *for all* $\lambda \ge 0$.

Proof. It follows from the definition of F that

$$\operatorname{gph} F = \big\{(x,\lambda) \in \mathbb{R}^n \times \mathbb{R} \;\big|\; \lambda \in F(x)\big\} = \big\{(x,\lambda) \in \mathbb{R}^n \times \mathbb{R} \;\big|\; \lambda \ge f(x)\big\} = \operatorname{epi} f.$$

Thus the inclusion $v \in D^*F(\bar{x},\bar{y})(\lambda)$ is equivalent to $(v,-\lambda) \in N((\bar{x},\bar{y});\operatorname{epi} f)$. If $\lambda > 0$, then $(v/\lambda,-1) \in N((\bar{x},\bar{y});\operatorname{epi} f)$, which implies $v/\lambda \in \partial f(\bar{x})$, and hence $v \in \lambda\partial f(\bar{x})$. The formula in the case where $\lambda = 0$ follows from the definition of the singular subdifferential. If $\lambda < 0$, then $D^*F(\bar{x},\bar{y})(\lambda) = \emptyset$ because $\lambda \ge 0$ whenever $(v,-\lambda) \in N((\bar{x},\bar{y});\operatorname{epi} f)$ by the proof of Proposition 2.47. To verify the last part of the proposition, observe that in this case $D^*F(\bar{x},\bar{y})(0) = \partial^\infty f(\bar{x}) = \{0\}$ by Proposition 2.23, so we arrive at the conclusion. $\square$

The next simple example is very natural and useful in what follows.

Example 2.59 Define $F:\mathbb{R}^n \rightrightarrows \mathbb{R}^p$ by $F(x) = K$ for every $x \in \mathbb{R}^n$, where $K \subset \mathbb{R}^p$ is a nonempty, convex set. Then $\operatorname{gph} F = \mathbb{R}^n \times K$, and Proposition 2.11(i) tells us that $N((\bar{x},\bar{y});\operatorname{gph} F) = \{0\} \times N(\bar{y};K)$ for any $(\bar{x},\bar{y}) \in \mathbb{R}^n \times K$. Thus

$$D^*F(\bar{x},\bar{y})(v) = \begin{cases} \{0\} & \text{if } -v \in N(\bar{y};K),\\ \emptyset & \text{otherwise.} \end{cases}$$

In particular, for the case where $K = \mathbb{R}^p$ we have

$$D^*F(\bar{x},\bar{y})(v) = \begin{cases} \{0\} & \text{if } v = 0,\\ \emptyset & \text{otherwise.} \end{cases}$$

Now we derive a formula for calculating the subdifferential of the *optimal value/marginal function* (1.4) given in the form

$$\mu(x) := \inf\{\varphi(x,y) \mid y \in F(x)\}$$

via the coderivative of the set-valued mapping F and the subdifferential of the function φ. First we derive the following useful estimate.

Lemma 2.60 *Let $\mu(\cdot)$ be the optimal value function* (1.4) *generated by a convex-graph mapping $F: \mathbb{R}^n \rightrightarrows \mathbb{R}^p$ and a convex function $\varphi: \mathbb{R}^n \times \mathbb{R}^p \to \overline{\mathbb{R}}$. Suppose that $\mu(x) > -\infty$ for all $x \in \mathbb{R}^n$, fix some $\bar{x} \in \operatorname{dom}\mu$, and consider the solution set*

$$S(\bar{x}) := \{\bar{y} \in F(\bar{x}) \mid \mu(\bar{x}) = \varphi(\bar{x},\bar{y})\}.$$

If $S(\bar{x}) \neq \emptyset$, then for any $\bar{y} \in S(\bar{x})$ we have

$$\bigcup_{(u,v)\in\partial\varphi(\bar{x},\bar{y})} \big[u + D^*F(\bar{x},\bar{y})(v)\big] \subset \partial\mu(\bar{x}). \tag{2.38}$$

Proof. Pick w from the left-hand side of (2.38) and find $(u,v) \in \partial\varphi(\bar{x},\bar{y})$ with

$$w - u \in D^*F(\bar{x},\bar{y})(v).$$

It gives us $(w-u,-v) \in N((\bar{x},\bar{y}); \operatorname{gph} F)$ and thus

$$\langle w-u, x-\bar{x}\rangle - \langle v, y-\bar{y}\rangle \le 0 \ \text{ for all } \ (x,y) \in \operatorname{gph} F,$$

which shows that whenever $y \in F(x)$ we have

$$\langle w, x-\bar{x}\rangle \le \langle u, x-\bar{x}\rangle + \langle v, y-\bar{y}\rangle \le \varphi(x,y) - \varphi(\bar{x},\bar{y}) = \varphi(x,y) - \mu(\bar{x}).$$

This implies in turn the estimate

$$\langle w, x-\bar{x}\rangle \le \inf_{y\in F(x)} \varphi(x,y) - \mu(\bar{x}) = \mu(x) - \mu(\bar{x}),$$

which justifies the inclusion $w \in \partial\mu(\bar{x})$, and hence completes the proof of (2.38). □

Theorem 2.61 Consider the optimal value function (1.4) under the assumptions of Lemma 2.60. For any $\bar{y} \in S(\bar{x})$, we have the equality

$$\partial\mu(\bar{x}) = \bigcup_{(u,v)\in\partial\varphi(\bar{x},\bar{y})} \big[u + D^*F(\bar{x},\bar{y})(v)\big] \tag{2.39}$$

provided the validity of the qualification condition

$$\partial^\infty\varphi(\bar{x},\bar{y}) \cap \big[-N((\bar{x},\bar{y}); \operatorname{gph} F)\big] = \{(0,0)\}, \tag{2.40}$$

which holds, in particular, when φ is continuous at $(\bar{x},\bar{y})$.

Proof. Taking Lemma 2.60 into account, it is only required to verify the inclusion "$\subset$" in (2.39). To proceed, pick $w \in \partial\mu(\bar{x})$ and $\bar{y} \in S(\bar{x})$. For any $x \in \mathbb{R}^n$, we have

$$\begin{aligned}\langle w, x-\bar{x}\rangle &\le \mu(x)-\mu(\bar{x}) = \mu(x)-\varphi(\bar{x},\bar{y})\\ &\le \varphi(x,y)-\varphi(\bar{x},\bar{y}) \ \text{ for all } \ y \in F(x).\end{aligned}$$

This implies that, whenever $(x,y) \in \mathbb{R}^n \times \mathbb{R}^p$, the following inequality holds:

$$\langle w, x-\bar{x}\rangle + \langle 0, y-\bar{y}\rangle \le \varphi(x,y)+\delta\big((x,y);\operatorname{gph} F\big) - \big[\varphi(\bar{x},\bar{y})+\delta\big((\bar{x},\bar{y});\operatorname{gph} F\big)\big].$$

Denote further $f(x,y) := \varphi(x,y)+\delta((x,y);\operatorname{gph} F)$ and deduce from the subdifferential sum rule in Theorem 2.44 under the qualification condition (2.40) that the inclusion

$$(w,0) \in \partial f(\bar{x},\bar{y}) = \partial\varphi(\bar{x},\bar{y}) + N\big((\bar{x},\bar{y});\operatorname{gph} F\big)$$

is satisfied. This shows that $(w,0) = (u_1,v_1)+(u_2,v_2)$ with $(u_1,v_1) \in \partial\varphi(\bar{x},\bar{y})$ and $(u_2,v_2) \in N((\bar{x},\bar{y});\operatorname{gph} F)$. This yields $v_2 = -v_1$, so $(u_2,-v_1) \in N((\bar{x},\bar{y});\operatorname{gph} F)$. Finally, we get $u_2 \in D^*F(\bar{x},\bar{y})(v_1)$ and therefore

$$w = u_1+u_2 \in u_1 + D^*F(\bar{x},\bar{y})(v_1),$$

which completes the proof of the theorem. $\square$

Let us present several important consequences of Theorem 2.61. The first one concerns the following *chain rule* for convex compositions.

Corollary 2.62 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and let $\varphi : \mathbb{R} \to \overline{\mathbb{R}}$ be a nondecreasing convex function. Take $\bar{x} \in \mathbb{R}^n$ and denote $\bar{y} := f(\bar{x})$ assuming that $\bar{y} \in \operatorname{dom}\varphi$. Then*

$$\partial(\varphi\circ f)(\bar{x}) = \bigcup_{\lambda\in\partial\varphi(\bar{y})} \lambda\partial f(\bar{x})$$

under the assumption that either $\partial^\infty\varphi(\bar{y}) = \{0\}$ or $0 \notin \partial f(\bar{x})$.

Proof. The composition $\varphi\circ f$ is a convex function by Proposition 1.39. Define the set-valued mapping $F(x) := [f(x),\infty)$ and observe that

$$(\varphi\circ f)(x) = \inf_{y\in F(x)} \varphi(y)$$

since φ is nondecreasing. Note each of the assumptions $\partial^\infty\varphi(\bar{y}) = \{0\}$ and $0 \notin \partial f(\bar{x})$ ensure the validity of the qualification condition (2.40). It follows from Theorem 2.61 that

$$\partial(\varphi\circ f)(\bar{x}) = \bigcup_{\lambda\in\partial\varphi(\bar{y})} D^*F(\bar{x},\bar{y})(\lambda).$$

Taking again into account that φ is nondecreasing yields $\lambda \geq 0$ for every $\lambda \in \partial\varphi(\bar{y})$. This tells us by Proposition 2.58 that

$$\partial(\varphi \circ f)(\bar{x}) = \bigcup_{\lambda \in \partial\varphi(\bar{y})} D^*F(\bar{x}, \bar{y})(\lambda) = \bigcup_{\lambda \in \partial\varphi(\bar{y})} \lambda \partial f(\bar{x}),$$

which verifies the claimed chain rule of this corollary. □

The next corollary addresses calculating subgradients of the optimal value functions over parameter-independent constraints.

Corollary 2.63 *Let $\varphi : \mathbb{R}^n \times \mathbb{R}^p \to \overline{\mathbb{R}}$, where the function $\varphi(x, \cdot)$ is bounded below on $\mathbb{R}^p$ for every $x \in \mathbb{R}^n$. Define*

$$\mu(x) := \inf\big\{\varphi(x, y) \;\big|\; y \in \mathbb{R}^p\big\}, \quad S(x) := \big\{y \in \mathbb{R}^p \;\big|\; \varphi(x, y) = \mu(x)\big\}.$$

For $\bar{x} \in \operatorname{dom} \mu$ and for every $\bar{y} \in S(\bar{x})$, we have

$$\partial\mu(\bar{x}) = \big\{u \in \mathbb{R}^n \;\big|\; (u, 0) \in \partial\varphi(\bar{x}, \bar{y})\big\}.$$

Proof. Follows directly from Theorem 2.61 for $F(x) \equiv \mathbb{R}^p$ and Example 2.59. Note that the qualification condition (2.40) is satisfied since $N((\bar{x}, \bar{y}); \operatorname{gph} F) = \{(0, 0)\}$. □

The next corollary specifies the general result for problems with affine constraints.

Corollary 2.64 *Let $B : \mathbb{R}^p \to \mathbb{R}^n$ be given by $B(y) := Ay + b$ via a linear mapping $A : \mathbb{R}^p \to \mathbb{R}^n$ and an element $b \in \mathbb{R}^n$ and let $\varphi : \mathbb{R}^p \to \overline{\mathbb{R}}$ be a convex function bounded below on the inverse image set $B^{-1}(x)$ for all $x \in \mathbb{R}^n$. Define*

$$\mu(x) := \inf\big\{\varphi(y) \;\big|\; B(y) = x\big\}, \quad S(x) := \big\{y \in B^{-1}(x) \;\big|\; \varphi(y) = \mu(x)\big\}$$

and fix $\bar{x} \in \mathbb{R}^n$ with $\mu(\bar{x}) < \infty$ and $S(\bar{x}) \neq \emptyset$. For any $\bar{y} \in S(\bar{x})$, we have

$$\partial\mu(\bar{x}) = (A^*)^{-1}\big(\partial\varphi(\bar{y})\big).$$

Proof. The setting of this corollary corresponds to Theorem 2.61 with $F(x) = B^{-1}(x)$ and $\varphi(x, y) \equiv \varphi(y)$. Then we have

$$N\big((\bar{x}, \bar{y}); \operatorname{gph} F\big) = \big\{(u, v) \in \mathbb{R}^n \times \mathbb{R}^p \;\big|\; -A^*u = v\big\},$$

which therefore implies the coderivative representation

$$D^*F(\bar{x}, \bar{y})(v) = \big\{u \in \mathbb{R}^n \;\big|\; A^*u = v\big\} = (A^*)^{-1}(v), \quad v \in \mathbb{R}^p.$$

It is easy to see that the qualification condition (2.40) is automatic in this case. This yields

$$\partial\mu(\bar{x}) = \bigcup_{v\in\partial\varphi(\bar{y})} \big[D^*F(\bar{x},\bar{y})(v)\big] = (A^*)^{-1}(\partial\varphi(\bar{y})),$$

and thus we complete the proof. □

Finally in this section, we present a useful application of Corollary 2.64 to deriving a calculus rule for subdifferentiation of *infimal convolutions* important in convex analysis.

Corollary 2.65 *Given convex functions $f_1, f_2\colon \mathbb{R}^n \to \overline{\mathbb{R}}$, consider their infimal convolution*

$$(f_1 \oplus f_2)(x) := \inf\big\{f_1(x_1) + f_2(x_2) \bigm| x_1 + x_2 = x\big\}, \quad x \in \mathbb{R}^n,$$

and suppose that $(f_1 \oplus f_2)(x) > -\infty$ for all $x \in \mathbb{R}^n$. Fix $\bar{x} \in \operatorname{dom}(f_1 \oplus f_2)$ and $\bar{x}_1, \bar{x}_2 \in \mathbb{R}^n$ satisfying $\bar{x} = \bar{x}_1 + \bar{x}_2$ and $(f_1 \oplus f_2)(\bar{x}) = f_1(\bar{x}_1) + f_2(\bar{x}_2)$. Then we have

$$\partial(f_1 \oplus f_2)(\bar{x}) = \partial f_1(\bar{x}_1) \cap \partial f_2(\bar{x}_2).$$

Proof. Define $\varphi\colon \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$ and $A\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ by

$$\varphi(x_1, x_2) := f_1(x_1) + f_2(x_2), \quad A(x_1, x_2) := x_1 + x_2.$$

It is easy to see that $A^*v = (v, v)$ for any $v \in \mathbb{R}^n$ and that $\partial\varphi(\bar{x}_1, \bar{x}_2) = (\partial f_1(\bar{x}_1), \partial f_2(\bar{x}_2))$ for $(\bar{x}_1, \bar{x}_2) \in \mathbb{R}^n \times \mathbb{R}^n$. Then $v \in \partial(f_1 \oplus f_2)(\bar{x})$ if and only if we have $A^*v = (v, v) \in \partial\varphi(\bar{x}_1, \bar{x}_2)$, which means that $v \in \partial f_1(\bar{x}_1) \cap \partial f_2(\bar{x}_2)$. Thus the claimed calculus result follows from Corollary 2.64. □

2.7 SUBGRADIENTS OF SUPPORT FUNCTIONS

This short section is devoted to the study of subgradient properties for an important class of nonsmooth convex functions known as *support functions* of convex sets.

Definition 2.66 *Given a nonempty (not necessarily convex) subset Ω of $\mathbb{R}^n$, the* SUPPORT FUNCTION *of Ω is defined by*

$$\sigma_\Omega(x) := \sup\big\{\langle x, \omega\rangle \bigm| \omega \in \Omega\big\}. \tag{2.41}$$

First we present some properties of (2.41) that can be deduced from the definition.

Proposition 2.67 *The following assertions hold:*

(i) *For any nonempty subset Ω of $\mathbb{R}^n$, the support function σ_Ω is subadditive, positively homogeneous, and hence convex.*

(ii) *For any nonempty subsets $\Omega_1, \Omega_2 \subset \mathbb{R}^n$ and any $x \in \mathbb{R}^n$, we have*

$$\sigma_{\Omega_1+\Omega_2}(x) = \sigma_{\Omega_1}(x) + \sigma_{\Omega_2}(x), \quad \sigma_{\Omega_1 \cup \Omega_2}(x) = \max\big\{\sigma_{\Omega_1}(x), \sigma_{\Omega_2}(x)\big\}.$$

The next result calculates subgradients of support functions via convex separation.

Theorem 2.68 Let Ω be a nonempty, closed, convex subset of $\mathbb{R}^n$ and let

$$S(x) := \big\{p \in \Omega \bigm| \langle x, p\rangle = \sigma_\Omega(x)\big\}.$$

For any $\bar{x} \in \operatorname{dom}\sigma_\Omega$, we have the subdifferential formula

$$\partial\sigma_\Omega(\bar{x}) = S(\bar{x}).$$

Proof. Fix $p \in \partial\sigma_\Omega(\bar{x})$ and get from definition (2.13) that

$$\langle p, x - \bar{x}\rangle \le \sigma_\Omega(x) - \sigma_\Omega(\bar{x}) \ \text{ whenever } \ x \in \mathbb{R}^n. \tag{2.42}$$

Taking this into account, it follows for any $u \in \mathbb{R}^n$ that

$$\langle p, u\rangle = \langle p, u + \bar{x} - \bar{x}\rangle \le \sigma_\Omega(u + \bar{x}) - \sigma_\Omega(\bar{x}) \le \sigma_\Omega(u) + \sigma_\Omega(\bar{x}) - \sigma_\Omega(\bar{x}) = \sigma_\Omega(u). \tag{2.43}$$

Assuming that $p \notin \Omega$, we use the separation result of Proposition 2.1 and find $u \in \mathbb{R}^n$ with

$$\sigma_\Omega(u) = \sup\big\{\langle u, \omega\rangle \bigm| \omega \in \Omega\big\} < \langle p, u\rangle,$$

which contradicts (2.43) and so verifies $p \in \Omega$. From (2.43) we also get $\langle p, \bar{x}\rangle \le \sigma_\Omega(\bar{x})$. Letting $x = 0$ in (2.42) gives us $\sigma_\Omega(\bar{x}) \le \langle p, \bar{x}\rangle$. Hence $\langle p, \bar{x}\rangle = \sigma_\Omega(\bar{x})$ and thus $p \in S(\bar{x})$.

Now fix $p \in S(\bar{x})$ and get from $p \in \Omega$ and $\langle p, \bar{x}\rangle = \sigma_\Omega(\bar{x})$ that

$$\langle p, x - \bar{x}\rangle = \langle p, x\rangle - \langle p, \bar{x}\rangle \le \sigma_\Omega(x) - \sigma_\Omega(\bar{x}) \ \text{ for all } \ x \in \mathbb{R}^n.$$

This shows that $p \in \partial\sigma_\Omega(\bar{x})$ and thus completes the proof. □

Note that Theorem 2.68 immediately implies that

$$\partial\sigma_\Omega(0) = S(0) = \big\{\omega \in \Omega \bigm| \langle 0, \omega\rangle = \sigma_\Omega(0) = 0\big\} = \Omega \tag{2.44}$$

for any nonempty, closed, convex set $\Omega \subset \mathbb{R}^n$.

The last result of this section justifies a certain *monotonicity* property of support functions with respect to set inclusions.

Proposition 2.69 *Let Ω_1 and Ω_2 be a nonempty, closed, convex subsets of $\mathbb{R}^n$. Then the inclusion $\Omega_1 \subset \Omega_2$ holds if and only if $\sigma_{\Omega_1}(v) \le \sigma_{\Omega_2}(v)$ for all $v \in \mathbb{R}^n$.*

Proof. Taking $\Omega_1 \subset \Omega_2$ as in the proposition, for any $v \in \mathbb{R}^n$ we have

$$\sigma_{\Omega_1}(v) = \sup\big\{\langle v, x\rangle \;\big|\; x \in \Omega_1\big\} \le \sup\big\{\langle v, x\rangle \;\big|\; x \in \Omega_2\big\} = \sigma_{\Omega_2}(v).$$

Conversely, suppose that $\sigma_{\Omega_1}(v) \le \sigma_{\Omega_2}(v)$ whenever $v \in \mathbb{R}^n$. Since $\sigma_{\Omega_1}(0) = \sigma_{\Omega_2}(0) = 0$, it follows from definition (2.13) of the subdifferential that

$$\partial\sigma_{\Omega_1}(0) \subset \partial\sigma_{\Omega_2}(0),$$

which yields $\Omega_1 \subset \Omega_2$ by formula (2.44). □

2.8 FENCHEL CONJUGATES

Many important issues of convex analysis and its applications (in particular, to optimization) are based on *duality*. The following notion plays a crucial role in duality considerations.

Definition 2.70 *Given a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ (not necessarily convex), its* Fenchel conjugate *$f^* : \mathbb{R}^n \to [-\infty, \infty]$ is*

$$f^*(v) := \sup\big\{\langle v, x\rangle - f(x) \;\big|\; x \in \mathbb{R}^n\big\} = \sup\big\{\langle v, x\rangle - f(x) \;\big|\; x \in \operatorname{dom} f\big\}. \tag{2.45}$$

Note that $f^*(v) = -\infty$ is allowed in (2.45) while $f^*(v) > -\infty$ for all $v \in \mathbb{R}^n$ if $\operatorname{dom} f \neq \emptyset$. It follows directly from the definitions that for any nonempty set $\Omega \subset \mathbb{R}^n$ the conjugate to its indicator function is the support function of Ω:

$$\delta^*_\Omega(v) = \sup\big\{\langle v, x\rangle \;\big|\; x \in \Omega\big\} = \sigma_\Omega(v), \quad v \in \Omega. \tag{2.46}$$

The next two propositions can be easily verified.

Proposition 2.71 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function, not necessarily convex, with $\operatorname{dom} f \neq \emptyset$. Then its Fenchel conjugate f^* is convex on $\mathbb{R}^n$.*

Proof. Function (2.45) is convex as the supremum of a family of affine functions. □

Proposition 2.72 *Let $f, g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be such that $f(x) \le g(x)$ for all $x \in \mathbb{R}^n$. Then we have $f^*(v) \ge g^*(v)$ for all $v \in \mathbb{R}^n$.*

Proof. For any fixed $v \in \mathbb{R}^n$, it follows from (2.45) that

$$\langle v, x\rangle - f(x) \ge \langle v, x\rangle - g(x), \quad x \in \mathbb{R}^n.$$

This readily implies the relationships

$$f^*(v) = \sup\big\{\langle v, x\rangle - f(x) \;\big|\; x \in \mathbb{R}^n\big\} \ge \sup\big\{\langle v, x\rangle - g(x) \;\big|\; x \in \mathbb{R}^n\big\} = g^*(v)$$

for all $v \in \mathbb{R}^n$, and therefore $f^* \ge g^*$ on $\mathbb{R}^n$. □

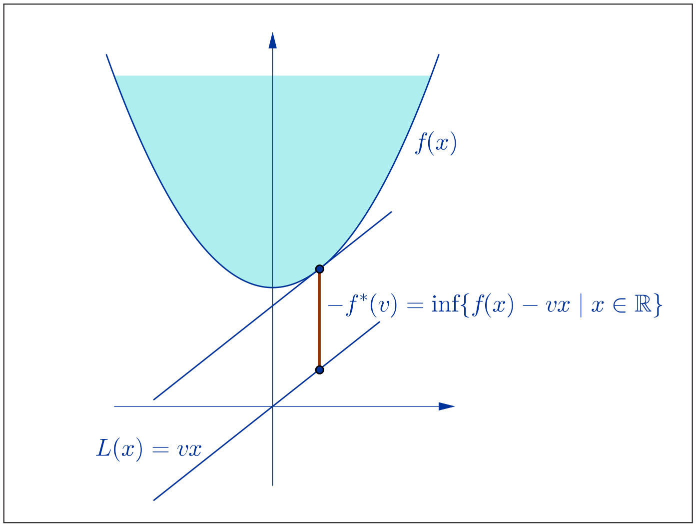

Figure 2.8: Fenchel conjugate.

The following two examples illustrate the calculation of conjugate functions.

Example 2.73 **(i)** Given $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$, consider the affine function

$$f(x) := \langle a, x \rangle + b, \quad x \in \mathbb{R}^n .$$

Then it can be seen directly from the definition that

$$f^*(v) = \begin{cases} -b & \text{if } v = a, \\ \infty & \text{otherwise.} \end{cases}$$

(ii) Given any $p > 1$, consider the power function

$$f(x) := \begin{cases} \dfrac{x^p}{p} & \text{if } x \geq 0, \\ \infty & \text{otherwise.} \end{cases}$$

For any $v \in \mathbb{R}$, the conjugate of this function is given by

$$f^*(v) = \sup\Big\{vx - \frac{x^p}{p} \;\Big|\; x \geq 0\Big\} = -\inf\Big\{\frac{x^p}{p} - vx \;\Big|\; x \geq 0\Big\}.$$

It is clear that $f^*(v) = 0$ if $v \leq 0$ since in this case $vx - p^{-1}x^p \leq 0$ when $x \geq 0$. Considering the case of $v > 0$, we see that function $\psi_v(x) := p^{-1}x^p - vx$ is convex and differentiable on

$(0,\infty)$ with $\psi_v'(x) = x^{p-1} - v$. Thus $\psi_v'(x) = 0$ if and only if $x = v^{1/(p-1)}$, and so ψ_v attains its minimum at $x = v^{1/(p-1)}$. Therefore, the conjugate function is calculated by

$$f^*(v) = \Big(1 - \frac{1}{p}\Big) v^{p/(p-1)}, \quad v \in \mathbb{R}^n .$$

Taking q from $q^{-1} = 1 - p^{-1}$, we express the conjugate function as

$$f^*(v) = \begin{cases} 0 & \text{if } v \le 0, \\ \dfrac{v^q}{q} & \text{otherwise.} \end{cases}$$

Note that the calculation in Example 2.73(ii) shows that

$$vx \le \frac{x^p}{p} + \frac{v^q}{q} \quad \text{for any} \;\; x, v \ge 0.$$

The first assertion of the next proposition demonstrates that such a relationship in a more general setting. To formulate the second assertion below, we define the *biconjugate* of f as the conjugate of f^*, i.e., $f^{**}(x) := (f^*)^*(x)$.

Proposition 2.74 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function with* dom $f \ne \emptyset$. *Then we have:*

(i) $\langle v, x\rangle \le f(x) + f^*(v)$ *for all* $x, v \in \mathbb{R}^n$.

(ii) $f^{**}(x) \le f(x)$ *for all* $x \in \mathbb{R}^n$.

Proof. Observe first that **(i)** is obvious if $f(x) = \infty$. If $x \in \operatorname{dom} f$, we get from (2.45) that $f^*(v) \ge \langle v, x\rangle - f(x)$, which verifies **(i)**. It implies in turn that

$$\sup\big\{\langle v, x\rangle - f^*(v) \;\big|\; v \in \mathbb{R}^n\big\} \le f(x) \;\; \text{for all} \;\; x, v \in \mathbb{R}^n,$$

which thus verifies **(ii)** and completes the proof. $\square$

The following important result reveals a close relationship between subgradients and Fenchel conjugates of convex functions.

Theorem 2.75 For any convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and any $\bar{x} \in \operatorname{dom} f$, we have that $v \in \partial f(\bar{x})$ if and only if

$$f(\bar{x}) + f^*(v) = \langle v, \bar{x}\rangle. \tag{2.47}$$

Proof. Taking any $v \in \partial f(\bar{x})$ and using definition (2.13) gives us

$$f(\bar{x}) + \langle v, x\rangle - f(x) \le \langle v, \bar{x}\rangle \ \text{ for all } \ x \in \mathbb{R}^n .$$

This readily implies the inequality

$$f(\bar{x}) + f^*(v) = f(\bar{x}) + \sup\big\{\langle v, x\rangle - f(x) \,\big|\, x \in \mathbb{R}^n\big\} \le \langle v, \bar{x}\rangle.$$

Since the opposite inequality holds by Proposition 2.74(i), we arrive at (2.47).

Conversely, suppose that $f(\bar{x}) + f^*(v) = \langle v, \bar{x}\rangle$. Applying Proposition 2.74(i), we get the estimate $f^*(v) \ge \langle v, x\rangle - f(x)$ for every $x \in \mathbb{R}^n$. This shows that $v \in \partial f(\bar{x})$. □

The result obtained allows us to find conditions ensuring that the biconjugate f^{**} of a convex function agrees with the function itself.

Proposition 2.76 *Let $\bar{x} \in \operatorname{dom} f$ for a convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$. Suppose that $\partial f(\bar{x}) \neq \emptyset$. Then we have the equality $f^{**}(\bar{x}) = f(\bar{x})$.*

Proof. By Proposition 2.74(ii) it suffices to verify the opposite inequality therein. Fix $v \in \partial f(\bar{x})$ and get $\langle v, \bar{x}\rangle = f(\bar{x}) + f^*(v)$ by the preceding theorem. This shows that

$$f(\bar{x}) = \langle v, \bar{x}\rangle - f^*(v) \le \sup\big\{\langle \bar{x}, v\rangle - f^*(v) \,\big|\, v \in \mathbb{R}^n\big\} = f^{**}(\bar{x}),$$

which completes the proof of this proposition. □

Taking into account Proposition 2.47, we arrive at the following corollary.

Corollary 2.77 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function and let $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$. Then we have the equality $f^{**}(\bar{x}) = f(\bar{x})$.*

Finally in this section, we prove a necessary and sufficient condition for the validity of the biconjugacy equality $f = f^{**}$ known as the *Fenchel-Moreau theorem*.

Theorem 2.78 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function with $\operatorname{dom} f \neq \emptyset$ and let $\mathcal{A}$ be the set of all affine functions of the form $\varphi(x) = \langle a, x\rangle + b$ for $x \in \mathbb{R}^n$, where $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Denote

$$\mathcal{A}(f) := \big\{\varphi \in \mathcal{A} \,\big|\, \varphi(x) \le f(x) \ \text{ for all } \ x \in \mathbb{R}^n\big\}.$$

Then $\mathcal{A}(f) \neq \emptyset$ whenever $\operatorname{epi} f$ is closed and convex. Moreover, the following are equivalent:

(i) $\operatorname{epi} f$ is closed and convex.
(ii) $f(x) = \sup_{\varphi \in \mathcal{A}(f)} \varphi(x)$ for all $x \in \mathbb{R}^n$.
(iii) $f^{**}(x) = f(x)$ for all $x \in \mathbb{R}^n$.

Proof. Let us first show that $\mathcal{A}(f) \neq \emptyset$. Fix any $x_0 \in \operatorname{dom} f$ and choose $\lambda_0 < f(x_0)$. Then $(x_0, \lambda_0) \notin \operatorname{epi} f$. By Proposition 2.1, there exist $(\bar{v}, \bar{\gamma}) \in \mathbb{R}^n \times \mathbb{R}$ and $\epsilon > 0$ such that

$$\langle \bar{v}, x \rangle + \bar{\gamma}\lambda < \langle \bar{v}, x_0 \rangle + \bar{\gamma}\lambda_0 - \epsilon \ \text{ whenever } \ (x, \lambda) \in \operatorname{epi} f. \tag{2.48}$$

Since $(x_0, f(x_0) + \alpha) \in \operatorname{epi} f$ for all $\alpha \geq 0$, we get

$$\bar{\gamma}(f(\bar{x}) + \alpha) < \bar{\gamma}\lambda_0 - \epsilon \ \text{ whenever } \ \alpha \geq 0.$$

This implies $\bar{\gamma} < 0$ since if not, we can let $\alpha \to \infty$ and arrive at a contradiction. For any $x \in \operatorname{dom} f$, it follows from (2.48) as $(x, f(x)) \in \operatorname{epi} f$ that

$$\langle \bar{v}, x \rangle + \bar{\gamma} f(x) < \langle \bar{v}, x_0 \rangle + \bar{\gamma}\lambda_0 - \epsilon \ \text{ for all } \ x \in \operatorname{dom} f.$$

This allows us to conclude that

$$f(x) > \langle \frac{\bar{v}}{\bar{\gamma}}, x_0 - x \rangle + \lambda_0 - \frac{\epsilon}{\bar{\gamma}} \ \text{ if } \ x \in \operatorname{dom} f.$$

Define now $\varphi(x) := \langle \frac{\bar{v}}{\bar{\gamma}}, x_0 - x \rangle + \lambda_0 - \frac{\epsilon}{\bar{\gamma}}$. Then $\varphi \in \mathcal{A}(f)$, and so $\mathcal{A}(f) \neq \emptyset$.

Let us next prove that **(i)$\Longrightarrow$(ii)**. By definition we need to show that for any $\lambda_0 < f(x_0)$ there is $\varphi \in \mathcal{A}(f)$ such that $\lambda_0 < \varphi(x_0)$. Since $(x_0, \lambda_0) \notin \operatorname{epi} f$, we apply again Proposition 2.1 to obtain (2.48). In the case where $x_0 \in \operatorname{dom} f$, it was proved above that $\varphi \in \mathcal{A}(f)$. Moreover, we have $\varphi(x_0) = \lambda_0 - \frac{\epsilon}{\bar{\gamma}} > \lambda_0$ since $\bar{\gamma} < 0$. Consider now the case where $x_0 \notin \operatorname{dom} f$. It follows from (2.48) by taking any $x \in \operatorname{dom} f$ and letting $\lambda \to \infty$ that $\bar{\gamma} \leq 0$. If $\bar{\gamma} < 0$, we can apply the same procedure and arrive at the conclusion. Hence we only need to consider the case where $\bar{\gamma} = 0$. In this case

$$\langle \bar{v}, x - x_0 \rangle + \epsilon < 0 \ \text{ whenever } \ x \in \operatorname{dom} f.$$

Since $\mathcal{A}(f) \neq \emptyset$, choose $\varphi_0 \in \mathcal{A}(f)$ and define

$$\varphi_k(x) := \varphi_0(x) + k(\langle \bar{v}, x - x_0 \rangle + \epsilon), \ k \in \mathbb{N}.$$

It is obvious that $\varphi_k \in \mathcal{A}(f)$ and $\varphi_k(x_0) = \varphi_0(x_0) + k\epsilon > \lambda_0$ for large k. This justifies **(ii)**.

Let us now verify implication **(ii)$\Longrightarrow$(iii)**. Fix any $\varphi \in \mathcal{A}(f)$. Then $\varphi \leq f$, and hence $\varphi^{**} \leq f^{**}$. Applying Proposition 2.76 ensures that $\varphi = \varphi^{**} \leq f^{**}$. It follows therefore that

$$f(x) = \sup \{\varphi(x) \mid \varphi \in \mathcal{A}(f)\} \leq f^{**} \ \text{ for every } \ x \in \mathbb{R}^n.$$

The opposite inequality $f^{**} \leq f$ holds by Proposition 2.74(ii), and thus $f^{**} = f$.

The last implication **(iii)$\Longrightarrow$(i)** is obvious because the set $\operatorname{epi} g^*$ is always closed and convex for any function $g : \mathbb{R}^n \to \overline{\mathbb{R}}$. $\square$

2.9 DIRECTIONAL DERIVATIVES

Our next topic is *directional differentiability* of convex functions and its relationships with subdifferentiation. In contrast to classical analysis, directional derivative constructions in convex analysis are *one-sided* and related to directions with no classical plus-minus symmetry.

Definition 2.79 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be an extended-real-valued function and let $\bar{x} \in \operatorname{dom} f$. The* DIRECTIONAL DERIVATIVE *of the function f at the point $\bar{x}$ in the direction $d \in \mathbb{R}^n$ is the following limit—if it exists as either a real number or $\pm\infty$:*

$$f'(\bar{x}; d) := \lim_{t \to 0^+} \frac{f(\bar{x} + td) - f(\bar{x})}{t}. \tag{2.49}$$

Note that construction (2.49) is sometimes called the *right* directional derivative f at $\bar{x}$ in the direction d. Its *left* counterpart is defined by

$$f'_-(\bar{x}; d) := \lim_{t \to 0^-} \frac{f(\bar{x} + td) - f(\bar{x})}{t}.$$

It is easy to see from the definitions that

$$f'_-(\bar{x}; d) = -f'(\bar{x}; -d) \ \text{ for all } \ d \in \mathbb{R}^n,$$

and thus properties of the left directional derivative $f'_-(\bar{x}; d)$ reduce to those of the right one (2.49), which we study in what follows.

Lemma 2.80 *Given a convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ with $\bar{x} \in \operatorname{dom} f$ and given $d \in \mathbb{R}^n$, define*

$$\varphi(t) := \frac{f(\bar{x} + td) - f(\bar{x})}{t}, \quad t > 0.$$

Then the function φ is nondecreasing on $(0, \infty)$.

Proof. Fix any numbers $0 < t_1 < t_2$ and get the representation

$$\bar{x} + t_1 d = \frac{t_1}{t_2}\Big(\bar{x} + t_2 d\Big) + \Big(1 - \frac{t_1}{t_2}\Big)\bar{x}.$$

It follows from the convexity of f that

$$f(\bar{x} + t_1 d) \le \frac{t_1}{t_2} f(\bar{x} + t_2 d) + \Big(1 - \frac{t_1}{t_2}\Big) f(\bar{x}),$$

which implies in turn the inequality

$$\varphi(t_1) = \frac{f(\bar{x} + t_1 d) - f(\bar{x})}{t_1} \le \frac{f(\bar{x} + t_2 d) - f(\bar{x})}{t_2} = \varphi(t_2).$$

This verifies that φ is nondecreasing on $(0, \infty)$. □

The next proposition establishes the directional differentiability of convex functions.

Proposition 2.81 *For any convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and any $\bar{x} \in \operatorname{dom} f$, the directional derivative $f'(\bar{x}; d)$ (and hence its left counterpart) exists in every direction $d \in \mathbb{R}^n$. Furthermore, it admits the representation via the function φ is defined in Lemma 2.80:*

$$f'(\bar{x}; d) = \inf_{t>0} \varphi(t), \quad d \in \mathbb{R}^n$$

Proof. Lemma 2.80 tells us that the function φ is nondecreasing. Thus we have

$$f'(\bar{x}; d) = \lim_{t \to 0+} \frac{f(\bar{x} + td) - f(\bar{x})}{t} = \lim_{t \to 0^+} \varphi(t) = \inf_{t>0} \varphi(t),$$

which verifies the results claimed in the proposition. □

Corollary 2.82 *If $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a convex function, then $f'(\bar{x}; d)$ is a real number for any $\bar{x} \in$* int(dom f) *and $d \in \mathbb{R}^n$.*

Proof. It follows from Theorem 2.29 that f is locally Lipschitz continuous around $\bar{x}$, i.e., there is $\ell \geq 0$ such that

$$\left| \frac{f(\bar{x} + td) - f(\bar{x})}{t} \right| \leq \frac{\ell t \|d\|}{t} = \ell \|d\| \text{ for all small } t > 0,$$

which shows that $|f'(\bar{x}; d)| \leq \ell \|d\| < \infty$. □

To establish relationships between directional derivatives and subgradients of general convex functions, we need the following useful observation.

Lemma 2.83 *For any convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $\bar{x} \in \operatorname{dom} f$, we have*

$$f'(\bar{x}; d) \leq f(\bar{x} + d) - f(\bar{x}) \text{ whenever } d \in \mathbb{R}^n.$$

Proof. Using Lemma 2.80, we have for the function φ therein that

$$\varphi(t) \leq \varphi(1) = f(\bar{x} + d) - f(\bar{x}) \text{ for all } t \in (0, 1),$$

which justifies the claimed property due to $f'(\bar{x}; d) = \inf_{t>0} \varphi(t) \leq \varphi(1)$. □

Theorem 2.84 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex with $\bar{x} \in \operatorname{dom} f$. The following are equivalent:

(i) $v \in \partial f(\bar{x})$.
(ii) $\langle v, d \rangle \leq f'(\bar{x}; d)$ for all $d \in \mathbb{R}^n$.
(iii) $f'_-(\bar{x}; d) \leq \langle v, d \rangle \leq f'(\bar{x}; d)$ for all $d \in \mathbb{R}^n$.

Proof. Picking any $v \in \partial f(\bar{x})$ and $t > 0$, we get

$$\langle v, td \rangle \le f(\bar{x} + td) - f(\bar{x}) \text{ whenever } d \in \mathbb{R}^n,$$

which verifies the implication **(i)**$\Longrightarrow$**(ii)** by taking the limit as $t \to 0^+$. Assuming now that assertion **(ii)** holds, we get by Lemma 2.83 that

$$\langle v, d \rangle \le f'(\bar{x}; d) \le f(\bar{x} + d) - f(\bar{x}) \text{ for all } d \in \mathbb{R}^n .$$

It ensures by definition (2.13) that $v \in \partial f(\bar{x})$, and thus assertions **(i)** and **(ii)** are equivalent.

It is obvious that **(iii)** yields **(ii)**. Conversely, if **(ii)** is satisfied, then for $d \in \mathbb{R}^n$ we have $\langle v, -d \rangle \le f'(\bar{x}; -d)$, and thus

$$f'_-(\bar{x}; d) = -f'(\bar{x}; -d) \le \langle v, d \rangle \text{ for any } d \in \mathbb{R}^n .$$

This justifies the validity of **(iii)** and completes the proof of the theorem. □

Let us next list some properties of (2.49) as a function of the direction.

Proposition 2.85 *For any convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ with $\bar{x} \in \operatorname{dom} f$, we define the directional function $\psi(d) := f'(\bar{x}; d)$, which satisfies the following properties:*

(i) $\psi(0) = 0$.
(ii) $\psi(d_1 + d_2) \le \psi(d_1) + \psi(d_2)$ *for all $d_1, d_2 \in \mathbb{R}^n$ provided that the right-hand side is well-defined.*
(iii) $\psi(\alpha d) = \alpha\psi(d)$ *whenever $d \in \mathbb{R}^n$ and $\alpha > 0$.*
(iv) *If furthermore $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$, then ψ is finite on $\mathbb{R}^n$.*

Proof. It is straightforward to deduce properties **(i)–(iii)** directly from definition (2.49). For instance, **(ii)** is satisfied due to the relationships

$$\begin{aligned}
\psi(d_1 + d_2) &= \lim_{t \to 0^+} \frac{f(\bar{x} + t(d_1 + d_2)) - f(\bar{x})}{t} \\
&= \lim_{t \to 0^+} \frac{f\left(\dfrac{\bar{x} + 2td_1 + \bar{x} + 2td_2}{2}\right) - f(\bar{x})}{t} \\
&\le \lim_{t \to 0^+} \frac{f(\bar{x} + 2td_1) - f(\bar{x})}{2t} + \lim_{t \to 0^+} \frac{f(\bar{x} + 2td_2) - f(\bar{x})}{2t} = \psi(d_1) + \psi(d_2).
\end{aligned}$$

Thus it remains to check that $\psi(d)$ is finite for every $d \in \mathbb{R}^n$ when $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$. To proceed, choose $\alpha > 0$ so small that $\bar{x} + \alpha d \in \operatorname{dom} f$. It follows from Lemma 2.83 that

$$\psi(\alpha d) = f'(\bar{x}; \alpha d) \le f(\bar{x} + \alpha d) - f(\bar{x}) < \infty.$$

Employing **(iii)** gives us $\psi(d) < \infty$. Further, we have from **(i)** and **(ii)** that

$$0 = \psi(0) = \psi\big(d + (-d)\big) \le \psi(d) + \psi(-d), \quad d \in \mathbb{R}^n,$$

which implies that $\psi(d) \ge -\psi(-d)$. This yields $\psi(d) > -\infty$ and so verifies **(iv)**. □

To derive the second major result of this section, yet another lemma is needed.

Lemma 2.86 *We have the following relationships between a convex function $f:\mathbb{R}^n \to \overline{\mathbb{R}}$ and the directional function ψ defined in Proposition 2.85 where $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$:*

(i) $\partial f(\bar{x}) = \partial\psi(0)$.
(ii) $\psi^*(v) = \delta_\Omega(v)$ *for all* $v \in \mathbb{R}^n$, *where* $\Omega := \partial\psi(0)$.

Proof. It follows from Theorem 2.84 that $v \in \partial f(\bar{x})$ if and only if

$$\langle v, d - 0\rangle = \langle v, d\rangle \le f'(\bar{x}; d) = \psi(d) = \psi(d) - \psi(0), \quad d \in \mathbb{R}^n.$$

This is equivalent to $v \in \partial\psi(0)$, and hence **(i)** holds.

To justify **(ii)**, let us first show that $\psi^*(v) = 0$ for all $v \in \Omega = \partial\psi(0)$. Indeed, we have

$$\psi^*(v) = \sup\big\{\langle v, d\rangle - \psi(d) \bigm| d \in \mathbb{R}^n\big\} \ge \langle v, 0\rangle - \psi(0) = 0.$$

Picking now any $v \in \partial\psi(0)$ gives us

$$\langle v, d\rangle = \langle v, d - 0\rangle \le \psi(d) - \psi(0) = \psi(d), \quad d \in \mathbb{R}^n,$$

which implies therefore that

$$\psi^*(v) = \sup\big\{\langle v, d\rangle - \psi(d) \bigm| d \in \mathbb{R}^n\big\} \le 0$$

and so ensures the validity of $\psi^*(v) = 0$ for any $v \in \partial\psi(0)$.

It remains to verify that $\psi^*(v) = \infty$ if $v \notin \partial\psi(0)$. For such an element v, find $d_0 \in \mathbb{R}^n$ with $\langle v, d_0\rangle > \psi(d_0)$. Since ψ is positively homogeneous by Proposition 2.85, it follows that

$$\begin{aligned}\psi^*(v) &= \sup\big\{\langle v, d\rangle - \psi(d) \bigm| d \in \mathbb{R}^n\big\} \ge \sup_{t>0}(\langle v, td_0\rangle - \psi(td_0))\\ &= \sup_{t>0} t(\langle v, d_0\rangle - \psi(d_0)) = \infty,\end{aligned}$$

which completes the proof of the lemma. □

Now we are ready establish a major relationship between the directional derivative and the subdifferential of an arbitrary convex function.

Theorem 2.87 Given a convex function $f:\mathbb{R}^n \to \overline{\mathbb{R}}$ and a point $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$, we have

$$f'(\bar{x}; d) = \max\big\{\langle v, d\rangle \bigm| v \in \partial f(\bar{x})\big\} \text{ for any } d \in \mathbb{R}^n.$$

Proof. It follows from Proposition 2.76 that

$$f'(\bar{x};d) = \psi(d) = \psi^{**}(d), \quad d \in \mathbb{R}^n,$$

by the properties of the function $\psi(d) = f'(\bar{x};d)$ from Proposition 2.85. Note that ψ is a finite convex function in this setting. Employing now Lemma 2.86 tells us that $\psi^*(v) = \delta_\Omega(v)$, where $\Omega = \partial\psi(0) = \partial f(\bar{x})$. Hence we have by (2.46) that

$$\psi^{**}(d) = \delta^*_\Omega(d) = \sup\big\{\langle v, d\rangle \;\big|\; v \in \Omega\big\}.$$

Since the set $\Omega = \partial f(\bar{x})$ is compact by Proposition 2.47, we complete the proof. $\square$

2.10 SUBGRADIENTS OF SUPREMUM FUNCTIONS

Let T be a nonempty subset of $\mathbb{R}^p$ and let $g : T \times \mathbb{R}^n \to \mathbb{R}$. For convenience, we also use the notation $g_t(x) := g(t, x)$. The *supremum function* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ for g_t over T is

$$f(x) := \sup_{t \in T} g(t, x) = \sup_{t \in T} g_t(x), \quad x \in \mathbb{R}^n. \tag{2.50}$$

If the supremum in (2.50) is attained (this happens, in particular, when the index set T is compact and $g(\cdot, x)$ is continuous), then (2.50) reduces to the *maximum function*, which can be written in form (2.35) when T is a finite set.

The main goal of this section is to calculate the subdifferential (2.13) of (2.50) when the functions g_t are convex. Note that in this case the supremum function (2.50) is also convex by Proposition 1.43. In what follows we assume without mentioning it that *the functions* $g_t\colon \mathbb{R}^n \to \mathbb{R}$ *are convex for all* $t \in T$.

For any point $\bar{x} \in \mathbb{R}^n$, define the *active index set*

$$S(\bar{x}) := \big\{t \in T \;\big|\; g_t(\bar{x}) = f(\bar{x})\big\}, \tag{2.51}$$

which may be empty if the supremum in (2.50) is not attained for $x = \bar{x}$. We first present a simple lower subdifferential estimate for (2.66).

Proposition 2.88 *Let* $\operatorname{dom} f \neq \emptyset$ *for the supremum function* (2.50) *over an arbitrary index set* T. *For any* $\bar{x} \in \operatorname{dom} f$, *we have the inclusion*

$$\operatorname{clco} \bigcup_{t \in S(\bar{x})} \partial g_t(\bar{x}) \subset \partial f(\bar{x}). \tag{2.52}$$

Proof. Inclusion (2.52) obviously holds if $S(\bar{x}) = \emptyset$. Supposing now that $S(\bar{x}) \neq \emptyset$, fix $t \in S(\bar{x})$ and $v \in \partial g_t(\bar{x})$. Then we get $g_t(\bar{x}) = f(\bar{x})$ and therefore

$$\langle v, x - \bar{x}\rangle \le g_t(x) - g_t(\bar{x}) = g_t(x) - f(\bar{x}) \le f(x) - f(\bar{x}),$$

which shows that $v \in \partial f(\bar{x})$. Since the subgradient set $\partial f(\bar{x})$ is a closed and convex, we conclude that inclusion (2.52) holds in this case. $\square$

Further properties of the supremum/maximum function (2.50) involve the following useful extensions of continuity. Its lower version will be studied and applied in Chapter 4.

Definition 2.89 *Let Ω be a nonempty subset of $\mathbb{R}^n$. A function $f : \Omega \to \mathbb{R}$ is* UPPER SEMICONTINUOUS AT $\bar{x} \in \Omega$ *if for every sequence $\{x_k\}$ in Ω converging to $\bar{x}$ it holds*

$$\limsup_{k\to\infty} f(x_k) \le f(\bar{x}).$$

We say that f is UPPER SEMICONTINUOUS ON Ω *if it is upper semicontinuous at every point of this set.*

Proposition 2.90 *Let T be a nonempty, compact subset of $\mathbb{R}^p$ and let $g(\cdot; x)$ be upper semicontinuous on T for every $x \in \mathbb{R}^n$. Then f in (2.50) is a finite-valued convex function.*

Proof. We need to check that $f(x) < \infty$ for every $x \in \mathbb{R}^n$. Suppose by contradiction that $f(\bar{x}) = \infty$ for some $\bar{x} \in \mathbb{R}^n$. Then there exists a sequence $\{t_k\} \subset T$ such that $g(t_k, \bar{x}) \to \infty$. Since T is compact, assume without loss of generality that $t_k \to \bar{t} \in T$. By the upper semicontinuity of $g(\cdot, \bar{x})$ on T we have

$$\infty = \limsup_{k\to\infty} g(t_k, \bar{x}) \le g(\bar{t}, \bar{x}) < \infty,$$

which is a contradiction that completes the proof of the proposition. $\square$

Proposition 2.91 *Let the index set $\emptyset \ne T \subset \mathbb{R}^p$ be compact and let $g(\cdot; x)$ be upper semicontinuous for every $x \in \mathbb{R}^n$. Then the active index set $S(\bar{x}) \subset T$ is nonempty and compact for every $\bar{x} \in \mathbb{R}^n$. Furthermore, we have the compactness in $\mathbb{R}^n$ of the convex hull*

$$C := \operatorname{co} \bigcup_{t \in S(\bar{x})} \partial g_t(\bar{x}).$$

Proof. Fix $\bar{x} \in \mathbb{R}^n$ and get from Proposition 2.90 that $f(\bar{x}) < \infty$ for the supremum function (2.50). Consider an arbitrary sequence $\{t_k\} \subset T$ with $g(t_k, \bar{x}) \to f(\bar{x})$ as $k \to \infty$. By the compactness of T we find $\bar{t} \in T$ such that $t_k \to \bar{t}$ along some subsequence. It follows from the assumed upper semicontinuity of g that

$$f(\bar{x}) = \limsup_{k\to\infty} g(t_k, \bar{x}) \le g(\bar{t}, \bar{x}) \le f(\bar{x}),$$

which ensures that $\bar{t} \in S(\bar{x})$ by the construction of S in (2.51), and thus $S(\bar{x}) \ne \emptyset$. Since $g(t, \bar{x}) \le f(\bar{x})$, for any $t \in T$ we get the representation

$$S(\bar{x}) = \big\{t \in T \;\big|\; g(t, \bar{x}) \ge f(\bar{x})\big\}$$

implying the closedness of $S(\bar{x})$ (and hence its compactness since $S(\bar{x}) \subset T$), which is an immediate consequence of the upper semicontinuity of $g(\cdot;\bar{x})$.

It remains to show that the subgradient set

$$Q := \bigcup_{t \in S(\bar{x})} \partial g_t(\bar{x})$$

is compact in $\mathbb{R}^n$; this immediately implies the claimed compactness of $C = \operatorname{co} Q$, since the convex hull of a compact set is compact (*show it*, or see Corollary 3.7 in the next chapter). To proceed with verifying the compactness of Q, we recall first that $Q \subset \partial f(\bar{x})$ by Proposition 2.88. This implies that the set Q is bounded in $\mathbb{R}^n$, since the subgradient set of the finite-valued convex function is compact by Proposition 2.47. To check the closedness of Q, take a sequence $\{v_k\} \subset Q$ converging to some $\bar{v}$ and for each $k \in \mathbb{N}$ and find $t_k \in S(\bar{x})$ such that $v_k \in \partial g_{t_k}(\bar{x})$. Since $S(\bar{x})$ is compact, we may assume that $t_k \to \bar{t} \in S(\bar{x})$. Then $g(t_k,\bar{x}) = g(\bar{t},\bar{x}) = f(\bar{x})$ and

$$\langle v_k, x - \bar{x}\rangle \le g(t_k,x) - g(t_k,\bar{x}) = g(t_k,x) - g(\bar{t},\bar{x}) \ \text{ for all } \ x \in \mathbb{R}^n$$

by definition (2.13). This gives us the relationships

$$\langle \bar{v}, x - \bar{x}\rangle = \limsup_{k\to\infty}\langle v_k, x - \bar{x}\rangle \le \limsup_{k\to\infty} g(t_k,x) - g(\bar{t},\bar{x}) \le g(\bar{t},x) - g(\bar{t},\bar{x}) = g_{\bar{t}}(x) - g_{\bar{t}}(\bar{x}),$$

which verify that $\bar{v} \in \partial g_{\bar{t}}(\bar{x}) \subset Q$ and thus complete the proof. □

The next result is of its own interest while being crucial for the main subdifferential result of this section.

Theorem 2.92 In the setting of Proposition 2.91 we have

$$f'(\bar{x};v) \le \sigma_C(v) \ \text{ for all } \ \bar{x} \in \mathbb{R}^n, \ v \in \mathbb{R}^n$$

via the support function of the set C defined therein.

Proof. It follows from Proposition 2.81 and Proposition 2.82 that $-\infty < f'(\bar{x};v) = \inf_{\lambda>0}\varphi(\lambda) < \infty$, where the function φ is defined in Lemma 2.80 and is proved to be nondecreasing on $(0,\infty)$. Thus there is a strictly decreasing sequence $\lambda_k \downarrow 0$ as $k \to \infty$ with

$$f'(\bar{x};v) = \lim_{k\to\infty}\frac{f(\bar{x}+\lambda_k v) - f(\bar{x})}{\lambda_k}, \quad k \in \mathbb{N}.$$

For every k, we select $t_k \in T$ such that $f(\bar{x}+\lambda_k v) = g(t_k,\bar{x}+\lambda_k v)$. The compactness of T allows us to find $\bar{t} \in T$ such that $t_k \to \bar{t} \in T$ along some subsequence. Let us show that $\bar{t} \in S(\bar{x})$. Since $g(t_k,\cdot)$ is convex and $\lambda_k < 1$ for large k, we have

$$g(t_k,\bar{x}+\lambda_k v) = g\big(t_k,\lambda_k(\bar{x}+v) + (1-\lambda_k)\bar{x}\big) \le \lambda_k g(t_k,\bar{x}+v) + (1-\lambda_k)g(t_k,\bar{x}), \quad (2.53)$$

which implies by the continuity of f (any finite-valued convex function is continuous) and the upper semicontinuity of $g(\cdot;x)$ that

$$f(\bar{x}) = \lim_{k\to\infty} f(\bar{x}+\lambda_k v) = \lim_{k\to\infty} g(t_k, \bar{x}+\lambda_k v) \le g(\bar{t}, \bar{x}).$$

This yields $f(\bar{x}) = g(\bar{t}, \bar{x})$, and so $\bar{t} \in S(\bar{x})$. Moreover, for any $\epsilon > 0$ and large k, we get

$$g(t_k, \bar{x}+v) \le \gamma + \epsilon \;\text{ with }\; \gamma := g(\bar{t}, \bar{x}+v)$$

by the upper semicontinuity of g. It follows from (2.53) that

$$g(t_k, \bar{x}) \ge \frac{g(t_k, \bar{x}+\lambda_k v) - \lambda_k(\gamma+\epsilon)}{1-\lambda_k} \to f(\bar{x}),$$

which tells us that $\lim_{k\to\infty} g(t_k, \bar{x}) = f(\bar{x})$ due to

$$f(\bar{x}) \le \liminf_{k\to\infty} g(t_k, \bar{x}) \le \limsup_{k\to\infty} g(t_k, \bar{x}) \le g(\bar{t}, \bar{x}) = f(\bar{x}).$$

Fix further any $\lambda > 0$ and, taking into account that $\lambda_k < \lambda$ for all k sufficiently large, deduce from Lemma 2.80 that

$$\begin{aligned}
\frac{f(\bar{x}+\lambda_k v) - f(\bar{x})}{\lambda_k} &= \frac{g(t_k, \bar{x}+\lambda_k v) - g(\bar{t}, \bar{x})}{\lambda_k} \\
&\le \frac{g(t_k, \bar{x}+\lambda_k v) - g(t_k, \bar{x})}{\lambda_k} \\
&\le \frac{g(t_k, \bar{x}+\lambda v) - g(t_k, \bar{x})}{\lambda}.
\end{aligned}$$

This ensures in turn the estimates

$$\begin{aligned}
\limsup_{k\to\infty} \frac{f(\bar{x}+\lambda_k v) - f(\bar{x})}{\lambda_k} &\le \limsup_{k\to\infty} \frac{g(t_k, \bar{x}+\lambda v) - g(t_k, \bar{x})}{\lambda} \\
&\le \frac{g(\bar{t}, \bar{x}+\lambda v) - g(\bar{t}, \bar{x})}{\lambda} \\
&= \frac{g_{\bar{t}}(\bar{x}+\lambda v) - g_{\bar{t}}(\bar{x})}{\lambda},
\end{aligned}$$

which implies the relationships

$$f'(\bar{x}; v) \le \inf_{\lambda>0} \frac{g_{\bar{t}}(\bar{x}+\lambda v) - g_{\bar{t}}(\bar{x})}{\lambda} = g'_{\bar{t}}(\bar{x}; v).$$

Using finally Theorem 2.87 and the inclusion $\partial g_{\bar{t}}(\bar{x}) \subset C$ from Proposition 2.88 give us

$$f'(\bar{x}; v) \le g'_{\bar{t}}(\bar{x}; v) = \sup_{w\in\partial g_{\bar{t}}(\bar{x})} \langle w, v\rangle \le \sup_{w\in C} \langle w, v\rangle = \sigma_C(v),$$

and justify therefore the statement of the theorem. $\square$

Now we are ready to establish the main result of this section.

Theorem 2.93 Let f be defined in (2.50), where the index set $\emptyset \neq T \subset \mathbb{R}^p$ is compact, and where the function $g(t,\cdot)$ is convex on $\mathbb{R}^n$ for every $t \in T$. Suppose in addition that the function $g(\cdot,x)$ is upper semicontinuous on T for every $x \in \mathbb{R}^n$. Then we have

$$\partial f(\bar{x}) = \mathrm{co} \bigcup_{t \in S(\bar{x})} \partial g_t(\bar{x}).$$

Proof. Taking any $w \in \partial f(\bar{x})$, we deduce from Theorem 2.87 and Theorem 2.92 that

$$\langle w, v \rangle \le f'(\bar{x}, v) \le \sigma_C(v) \quad \text{for all} \quad v \in \mathbb{R}^n,$$

where C is defined in Proposition 2.91. Thus $w \in \partial \sigma_C(0) = C$ by formula (2.44) for the set $\Omega = C$, which is a nonempty, closed, convex subset of $\mathbb{R}^n$. The opposite inclusion follows from Proposition 2.88. $\square$

2.11 EXERCISES FOR CHAPTER 2

Exercise 2.1 **(i)** Let $\Omega \subset \mathbb{R}^n$ be a nonempty, closed, convex cone and let $\bar{x} \notin \Omega$. Show that there exists a nonzero element $v \in \mathbb{R}^n$ such that $\langle v, x \rangle \le 0$ for all $x \in \Omega$ and $\langle v, \bar{x} \rangle > 0$.
(ii) Let Ω be a subspace of $\mathbb{R}^n$ and let $\bar{x} \notin \Omega$. Show that there exists a nonzero element $v \in \mathbb{R}^n$ such that $\langle v, x \rangle = 0$ for all $x \in \Omega$ and $\langle v, \bar{x} \rangle > 0$.

Exercise 2.2 Let Ω_1 and Ω_2 be two nonempty, convex subsets of $\mathbb{R}^n$ satisfying the condition $\mathrm{ri}\,\Omega_1 \cap \mathrm{ri}\,\Omega_2 = \emptyset$. Show that there is a nonzero element $v \in \mathbb{R}^n$ which separates the sets Ω_1 and Ω_2 in the sense of (2.2).

Exercise 2.3 Two nonempty, convex subsets of $\mathbb{R}^n$ are called *properly separated* if there exists a nonzero element $v \in \mathbb{R}^n$ such that

$$\sup\big\{\langle v, x \rangle \;\big|\; x \in \Omega_1\big\} \le \inf\big\{\langle v, y \rangle \;\big|\; y \in \Omega_2\big\},$$

$$\inf\big\{\langle v, x \rangle \;\big|\; x \in \Omega_1\big\} < \sup\big\{\langle v, y \rangle \;\big|\; y \in \Omega_2\big\}.$$

Prove that Ω_1 and Ω_2 are properly separated if and only if $\mathrm{ri}\,\Omega_1 \cap \mathrm{ri}\,\Omega_2 = \emptyset$.

Exercise 2.4 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function attaining its minimum at some point $\bar{x} \in \mathrm{dom}\, f$. Show that the sets $\{\Omega_1, \Omega_2\}$ with $\Omega_1 := \mathrm{epi}\, f$ and $\Omega_2 := \mathbb{R}^n \times \{f(\bar{x})\}$ form an extremal system.

Exercise 2.5 Let Ω_i for $i = 1, 2$ be nonempty, convex subsets of $\mathbb{R}^n$ with at least one point in common and satisfy

$$\operatorname{int}\Omega_1 \neq \emptyset \text{ and } \operatorname{int}\Omega_1 \cap \Omega_2 = \emptyset.$$

Show that there are elements v_1, v_2 in $\mathbb{R}^n$, not equal to zero simultaneously, and real numbers α_1, α_2 such that

$$\begin{aligned}&\langle v_i, x\rangle \leq \alpha_i \text{ for all } x \in \Omega_i,\ i = 1, 2,\\ &v_1 + v_2 = 0, \text{ and } \alpha_1 + \alpha_2 = 0.\end{aligned}$$

Exercise 2.6 Let $\Omega := \{x \in \mathbb{R}^n \mid x \leq 0\}$. Show that

$$N(\bar{x}; \Omega) = \{v \in \mathbb{R}^n \mid v \geq 0,\ \langle v, \bar{x}\rangle = 0\} \text{ for any } \bar{x} \in \Omega.$$

Exercise 2.7 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function differentiable at $\bar{x}$. Show that

$$N\big((\bar{x}, f(\bar{x})); \operatorname{epi} f\big) = \{\lambda\big(\nabla f(\bar{x}), -1\big) \mid \lambda \geq 0\}.$$

Exercise 2.8 Use induction to prove Corollary 2.19.

Exercise 2.9 Calculate the subdifferentials and the singular subdifferentials of the following convex functions at every point of their domains:

(i) $f(x) = |x-1| + |x-2|, \quad x \in \mathbb{R}.$
(ii) $f(x) = e^{|x|}, \quad x \in \mathbb{R}.$
(iii) $f(x) = \begin{cases} x^2 - 1 & \text{if } |x| \leq 1,\\ \infty & \text{otherwise on } \mathbb{R}.\end{cases}$
(iv) $f(x) = \begin{cases} -\sqrt{1-x^2} & \text{if } |x| \leq 1,\\ \infty & \text{otherwise on } \mathbb{R}.\end{cases}$
(v) $f(x_1, x_2) = |x_1| + |x_2|, \quad (x_1, x_2) \in \mathbb{R}^2.$
(vi) $f(x_1, x_2) = \max\{|x_1|, |x_2|\}, \quad (x_1, x_2) \in \mathbb{R}^2.$
(vii) $f(x) = \max\{\langle a, x\rangle - b, 0\}, \quad x \in \mathbb{R}^n$, where $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$.

Exercise 2.10 Let $f : \mathbb{R} \to \mathbb{R}$ be a nondecreasing convex function and let $\bar{x} \in \mathbb{R}$. Show that $v \geq 0$ for $v \in \partial f(\bar{x})$.

Exercise 2.11 Use induction to prove Corollary 2.46.

Exercise 2.12 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function. Give an alternative proof for the part of Proposition 2.47 stating that $\partial f(\bar{x}) \neq \emptyset$ for any element $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$. Does the conclusion still hold if $\bar{x} \in \operatorname{ri}(\operatorname{dom} f)$?

Exercise 2.13 Let Ω_1 and Ω_2 be nonempty, convex subsets of $\mathbb{R}^n$. Suppose that $\operatorname{ri}\Omega_1 \cap \operatorname{ri}\Omega_2 \neq \emptyset$. Prove that

$$N(\bar{x};\Omega_1 \cap \Omega_2) = N(\bar{x};\Omega_1) + N(\bar{x};\Omega_2) \ \text{ for all } \ \bar{x} \in \Omega_1 \cap \Omega_2.$$

Exercise 2.14 Let $f_1, f_2\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex functions satisfying the condition

$$\operatorname{ri}(\operatorname{dom} f_1) \cap \operatorname{ri}(\operatorname{dom} f_2) \neq \emptyset.$$

Prove that for all $\bar{x} \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$ we have

$$\partial(f_1 + f_2)(\bar{x}) = \partial f_1(\bar{x}) + \partial f_2(\bar{x}), \quad \partial^\infty(f_1 + f_2)(\bar{x}) = \partial^\infty f_1(\bar{x}) + \partial^\infty f_2(\bar{x}).$$

Exercise 2.15 In the setting of Theorem 2.51 show that

$$N((\bar{x},\bar{y},\bar{z});\Omega_1) \cap [-N((\bar{x},\bar{y},\bar{z});\Omega_2)] = \{0\}$$

under the qualification condition (2.33).

Exercise 2.16 Calculate the singular subdifferential of the composition $f \circ B$ in the setting of Theorem 2.51.

Exercise 2.17 Consider the composition $g \circ h$ in the setting of Proposition 1.54 and take any $\bar{x} \in \operatorname{dom}(g \circ h)$. Prove that

$$\partial(g \circ h)(\bar{x}) = \Big\{ \sum_{i=1}^{p} \lambda_i v_i \ \Big|\ (\lambda_1,\ldots,\lambda_p) \in \partial g(\bar{y}),\ v_i \in \partial f_i(\bar{x}),\ i = 1,\ldots,p \Big\}$$

under the validity of the qualification condition

$$\Big[0 \in \sum_{i=1}^{p} \lambda_i \partial f_i(\bar{x}),\ (\lambda_1,\ldots,\lambda_p) \in \partial^\infty g(\bar{y}) \Big] \Longrightarrow \big[\lambda_i = 0 \ \text{ for } \ i = 1,\ldots,p\big].$$

Exercise 2.18 Calculate the singular subdifferential of the optimal value functions in the setting of Theorem 2.61.

Exercise 2.19 Calculate the support functions for the following sets:

(i) $\Omega_1 := \big\{ x = (x_1,\ldots,x_n) \in \mathbb{R}^n \ \big|\ \|x\|_1 := \sum_{i=1}^n |x_i| \le 1 \big\}$.
(ii) $\Omega_2 := \big\{ x = (x_1,\ldots,x_n) \in \mathbb{R}^n \ \big|\ \|x\|_\infty := \max_{i=1}^n |x_i| \le 1 \big\}$.

Exercise 2.20 Find the Fenchel conjugate for each of the following functions on $\mathbb{R}$:

(i) $f(x) = e^x$.
(ii) $f(x) = \begin{cases} -\ln(x) & x > 0, \\ \infty & x \le 0. \end{cases}$
(iii) $f(x) = ax^2 + bx + c$, where $a > 0$.
(iv) $f(x) = \delta_{I\!B}(x)$, where $I\!B = [-1, 1]$.
(v) $f(x) = |x|$.

Exercise 2.21 Let Ω be a nonempty subset of $\mathbb{R}^n$. Find the biconjugate function δ_Ω^{**}.

Exercise 2.22 Prove the following statements for the listed functions on $\mathbb{R}^n$:

(i) $(\lambda f)^*(v) = \lambda f^*\left(\dfrac{v}{\lambda}\right)$, where $\lambda > 0$.
(ii) $(f + \lambda)^*(v) = f^*(v) - \lambda$, where $\lambda \in \mathbb{R}$.
(iii) $f_u^*(v) = f^*(v) + \langle u, v\rangle$, where $u \in \mathbb{R}^n$ and $f_u(x) := f(x - u)$.

Exercise 2.23 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex and positively homogeneous function. Show that $f^*(v) = \delta_\Omega(v)$ for all $v \in \mathbb{R}^n$, where $\Omega := \partial f(0)$.

Exercise 2.24 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function with $\operatorname{dom} f = \emptyset$. Show that $f^{**}(x) = f(x)$ for all $x \in \mathbb{R}^n$.

Exercise 2.25 Let $f : \mathbb{R} \to \mathbb{R}$ be convex. Prove that $\partial f(x) = [f'_-(x), f'_+(x)]$, where $f'_-(x)$ and $f'_+(x)$ stand for the left and right derivative of f at x, respectively.

Exercise 2.26 Define the function $f : \mathbb{R} \to \overline{\mathbb{R}}$ by

$$f(x) := \begin{cases} -\sqrt{1 - x^2} & \text{if } |x| \le 1, \\ \infty & \text{otherwise.} \end{cases}$$

Show that f is a convex function and calculate its directional derivatives $f'(-1; d)$ and $f'(1; d)$ in the direction $d = 1$.

Exercise 2.27 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex and let $\bar{x} \in \operatorname{dom} f$. For an element $d \in \mathbb{R}^n$, consider the function φ defined in Lemma 2.80 and show that φ is nondecreasing on $(-\infty, 0)$.

Exercise 2.28 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex and let $\bar{x} \in \operatorname{dom} f$. Prove the following:

(i) $f'_-(\bar{x}; d) \le f'(\bar{x}; d)$ for all $d \in \mathbb{R}^n$.
(ii) $f(\bar{x}) - f(\bar{x} - d) \le f'_-(\bar{x}; d) \le f'(\bar{x}; d) \le f(\bar{x} + d) - f(\bar{x})$, $d \in \mathbb{R}^n$.

Exercise 2.29 Prove the assertions of Proposition 2.67.

Exercise 2.30 Let $\Omega \subset \mathbb{R}^n$ be a nonempty, compact set. Using Theorem 2.93, calculate the subdifferential of the support function $\partial\sigma_\Omega(0)$.

Exercise 2.31 Let Ω be a nonempty, compact subset of $\mathbb{R}^n$. Define the function

$$\mu_\Omega(x) := \sup\{\|x-\omega\| \bigm| \omega \in \Omega\}, \quad x \in \mathbb{R}^n .$$

Prove that μ_Ω is a convex function and find its subdifferential $\partial\mu_\Omega(x)$ at any $x \in \mathbb{R}^n$.

Exercise 2.32 Using the representation

$$\|x\| = \sup\big\{\langle v, x\rangle \bigm| \|v\| \le 1\big\}$$

and Theorem 2.93, calculate the subdifferential $\partial p(\bar{x})$ of the norm function $p(x) := \|x\|$ at any point $\bar{x} \in \mathbb{R}^n$.

Exercise 2.33 Let $d(x;\Omega)$ be the distance function (1.5) associated with some nonempty, closed, convex subset Ω of $\mathbb{R}^n$.

(i) Prove the representation

$$d(x;\Omega) = \sup\big\{\langle x, v\rangle - \sigma_\Omega(v) \bigm| \|v\| \le 1\big\}$$

via the support function of Ω.
(ii) Calculate the subdifferential $\partial d(\bar{x};\Omega)$ at any point $\bar{x} \in \mathbb{R}^n$.

CHAPTER 3

Remarkable Consequences of Convexity

This chapter is devoted to remarkable topics of convex analysis that, together with subdifferential calculus and related results of Chapter 2, constitute the major achievements of this discipline most important for many applications—first of all to convex optimization.

3.1 CHARACTERIZATIONS OF DIFFERENTIABILITY

In this section we define another classical notion of differentiability, the Gâteaux derivative, and show that for convex functions it can be characterized, together with the Fréchet one, via a single-element subdifferential of convex analysis.

Definition 3.1 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be finite around $\bar{x}$. We say that f is* Gâteaux differentiable *at $\bar{x}$ if there is $v \in \mathbb{R}^n$ such that for any $d \in \mathbb{R}^n$ we have*

$$\lim_{t\to 0} \frac{f(\bar{x}+td) - f(\bar{x}) - t\langle v, d\rangle}{t} = 0.$$

Then we call v the Gâteaux derivative *of f at $\bar{x}$ and denote by $f'_G(\bar{x})$.*

It follows from the definition and the equality $f'_-(\bar{x}; d) = -f'(\bar{x}; -d)$ that $f'_G(\bar{x}) = v$ if and only if the directional derivative $f'(\bar{x}; d)$ exists and is *linear* with respect to d with $f'(\bar{x}; d) = \langle v, d\rangle$ for all $d \in \mathbb{R}^n$; see Exercise 3.1 and its solution. Moreover, if f is Gâteaux differentiable at $\bar{x}$, then

$$\langle f'_G(\bar{x}), d\rangle = \lim_{t\to 0} \frac{f(\bar{x}+td) - f(\bar{x})}{t} \text{ for all } d \in \mathbb{R}^n.$$

Proposition 3.2 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function (not necessarily convex) which is locally Lipschitz continuous around $\bar{x} \in \operatorname{dom} f$. Then the following assertions are equivalent:*

(i) *f is Fréchet differentiable at $\bar{x}$.*
(ii) *f is Gâteaux differentiable at $\bar{x}$.*

Proof. Let f be Fréchet differentiable at $\bar{x}$ and let $v := \nabla f(\bar{x})$. For a fixed $0 \neq d \in \mathbb{R}^n$, taking into account that $\|\bar{x} + td - \bar{x}\| = |t| \cdot \|d\| \to 0$ as $t \to 0$, we have

$$\lim_{t \to 0} \frac{f(\bar{x} + td) - f(\bar{x}) - t\langle v, d\rangle}{t} = \|d\| \lim_{t \to 0} \frac{f(\bar{x} + td) - f(\bar{x}) - \langle v, td\rangle}{t\|d\|} = 0,$$

which implies the Gâteaux differentiablity of f at $\bar{x}$ with $f'_G(\bar{x}) = \nabla f(\bar{x})$. Note that the local Lipschitz continuity of f around $\bar{x}$ is not required in this implication.

Assuming now that f is Gâteaux differentiable at $\bar{x}$ with $v := f'_G(\bar{x})$, let us show that

$$\lim_{h \to 0} \frac{f(\bar{x} + h) - f(\bar{x}) - \langle v, h\rangle}{\|h\|} = 0.$$

Suppose by contradiction that it does not hold. Then we can find $\epsilon_0 > 0$ and a sequence of $h_k \to 0$ with $h_k \neq 0$ as $k \in \mathbb{N}$ such that

$$\frac{\left|f(\bar{x} + h_k) - f(\bar{x}) - \langle v, h_k\rangle\right|}{\|h_k\|} \geq \epsilon_0. \tag{3.1}$$

Denote $t_k := \|h_k\|$ and $d_k := \dfrac{h_k}{\|h_k\|}$. Then $t_k \to 0$, and we can assume without loss of generality that $d_k \to d$ with $\|d\| = 1$. This gives us the estimate

$$\epsilon_0\|d_k\| \leq \frac{\left|f(\bar{x} + t_k d_k) - f(\bar{x}) - \langle v, t_k d_k\rangle\right|}{t_k}.$$

Let $\ell \geq 0$ be a Lipschitz constant of f around $\bar{x}$. The triangle inequality ensures that

$$\begin{aligned}
\epsilon_0\|d_k\| &\leq \frac{\left|f(\bar{x} + t_k d_k) - f(\bar{x}) - \langle v, t_k d_k\rangle\right|}{t_k} \\
&\leq \frac{\left|f(\bar{x} + t_k d_k) - f(\bar{x} + t_k d)\right| + \left|\langle v, t_k d\rangle - \langle v, t_k d_k\rangle\right|}{t_k} + \frac{\left|f(\bar{x} + t_k d) - f(\bar{x}) - \langle v, t_k d\rangle\right|}{t_k} \\
&\leq \ell\|d_k - d\| + \|v\| \cdot \|d_k - d\| + \frac{\left|f(\bar{x} + t_k d) - f(\bar{x}) - \langle v, t_k d\rangle\right|}{t_k} \to 0 \text{ as } k \to \infty.
\end{aligned}$$

This contradicts (3.1) since $\epsilon_0\|d_k\| \to \epsilon_0$, so f is Fréchet differentiable at $\bar{x}$. □

Now we are ready to derive a nice subdifferential characterization of differentiability.

Theorem 3.3 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function with $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$. The following assertions are equivalent:

(i) f is Fréchet differentiable at $\bar{x}$.
(ii) f is Gâteaux differentiable at $\bar{x}$.
(iii) $\partial f(\bar{x})$ is a singleton.

Proof. It follows from Theorem 2.29 that f is locally Lipschitz continuous around $\bar{x}$, and hence **(i)** and **(ii)** are equivalent by Proposition 3.2. Suppose now that **(ii)** is satisfied and let $v := f'_G(\bar{x})$. Then $f'_-(\bar{x}; d) = f'(\bar{x}; d) = \langle v, d \rangle$ for all $d \in \mathbb{R}^n$. Observe that $\partial f(\bar{x}) \neq \emptyset$ since $\bar{x} \in \text{int}(\text{dom}\, f)$. Fix any $w \in \partial f(\bar{x})$. It follows from Theorem 2.84 that

$$f'_-(\bar{x}; d) \leq \langle w, d \rangle \leq f'(\bar{x}; d) \text{ whenever } d \in \mathbb{R}^n,$$

and thus $\langle w, d \rangle = \langle v, d \rangle$, so $\langle w - v, d \rangle = 0$ for all d. Plugging there $d = w - v$ gives us $w = v$ and thus justifies that $\partial f(\bar{x})$ is a singleton.

Suppose finally that **(iii)** is satisfied with $\partial f(\bar{x}) = \{v\}$ and then show that **(ii)** holds. By Theorem 2.87 we have the representation

$$f'(\bar{x}; d) = \sup_{w \in \partial f(\bar{x})} \langle w, d \rangle = \langle v, d \rangle,$$

for all $d \in \mathbb{R}^n$. This implies that f is Gâteaux and hence Fréchet differentiable at $\bar{x}$. □

3.2 CARATHÉODORY THEOREM AND FARKAS LEMMA

We begin with preliminary results on conic and convex hulls of sets that lead us to one of the most remarkable results of convex geometry established by Constantin Carathéodory in 1911. Recall that a set Ω is called a *cone* if $\lambda x \in \Omega$ for all $\lambda \geq 0$ and $x \in \Omega$. It follows from the definition that $0 \in \Omega$ if Ω is a nonempty cone. Given a nonempty set $\Omega \subset \mathbb{R}^n$, we say that K_Ω is the *convex cone generated by* Ω, or the *convex conic hull* of Ω, if it is the intersection of all the convex cones containing Ω.

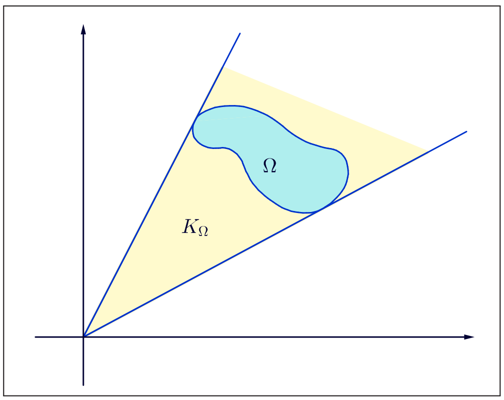

Figure 3.1: Convex cone generated by a set.

Proposition 3.4 *Let Ω be a nonempty subset of $\mathbb{R}^n$. Then the convex cone generated by Ω admits the following representation:*

$$K_\Omega = \Big\{ \sum_{i=1}^{m} \lambda_i a_i \;\Big|\; \lambda_i \geq 0,\ a_i \in \Omega,\ m \in \mathbb{N} \Big\}. \tag{3.2}$$

If, in particular, the set Ω is convex, then $K_\Omega = \mathbb{R}_+ \Omega$.

Proof. Denote by C the set on the right-hand side of (3.2) and observe that it is a convex cone which contains Ω. It remains to show that $C \subset K$ for any convex cone K containing Ω. To see it, fix such a cone and form the combination $x = \sum_{i=1}^{m} \lambda_i a_i$ with any $\lambda_i \geq 0$, $a_i \in \Omega$, and $m \in \mathbb{N}$. If $\lambda_i = 0$ for all i, then $x = 0 \in K$. Otherwise, suppose that $\lambda_i > 0$ for some i and denote $\lambda := \sum_{i=1}^{m} \lambda_i > 0$. This tells us that

$$x = \lambda \Big(\sum_{i=1}^{m} \frac{\lambda_i}{\lambda} a_i \Big) \in K,$$

which yields $C \subset K$, so $K_\Omega = C$. The last assertion of the proposition is obvious. $\square$

Proposition 3.5 *For any nonempty set $\Omega \subset \mathbb{R}^n$ and any $x \in K_\Omega \setminus \{0\}$, we have*

$$x = \sum_{i=1}^{m} \lambda_i a_i \;\; \textit{with} \;\; \lambda_i > 0,\ a_i \in \Omega \;\; \textit{as} \;\; i = 1, \ldots, m \;\; \textit{and} \;\; m \leq n.$$

Proof. Let $x \in K_\Omega \setminus \{0\}$. It follows from Proposition 3.4 that $x = \sum_{i=1}^{m} \mu_i a_i$ with some $\mu_i > 0$ and $a_i \in \Omega$ as for $i = 1, \ldots, m$ and $m \in \mathbb{N}$. If the elements $a_1, \ldots, a_m$ are linearly dependent, then there are numbers $\gamma_i \in \mathbb{R}$, not all zeros, such that

$$\sum_{i=1}^{m} \gamma_i a_i = 0.$$

We can assume that $I := \{ i = 1, \ldots, m \mid \gamma_i > 0 \} \neq \emptyset$ and get for any $\epsilon > 0$ that

$$x = \sum_{i=1}^{m} \mu_i a_i = \sum_{i=1}^{m} \mu_i a_i - \epsilon \Big(\sum_{i=1}^{m} \gamma_i a_i \Big) = \sum_{i=1}^{m} \big(\mu_i - \epsilon \gamma_i \big) a_i.$$

Let $\epsilon := \min \Big\{ \frac{\mu_i}{\gamma_i} \;\Big|\; i \in I \Big\} = \frac{\mu_{i_0}}{\gamma_{i_0}}$, where $i_0 \in I$, and denote $\beta_i := \mu_i - \epsilon \gamma_i$ for $i = 1, \ldots, m$. Then we have the relationships

$$x = \sum_{i=1}^{m} \beta_i a_i \;\; \text{with} \;\; \beta_{i_0} = 0 \;\; \text{and} \;\; \beta_i \geq 0 \;\; \text{for any} \;\; i = 1, \ldots, m.$$

Continuing this procedure, after a finite number of steps we represent x as a positive linear combination of linearly independent elements $\{a_j \mid j \in J\}$, where $J \subset \{1, \ldots, m\}$. Thus the index set J cannot contain more than n elements. $\square$

The following remarkable result is known as the *Carathéodory theorem*. It relates convex hulls of sets with the dimension of the space; see an illustration in Figure 3.2.

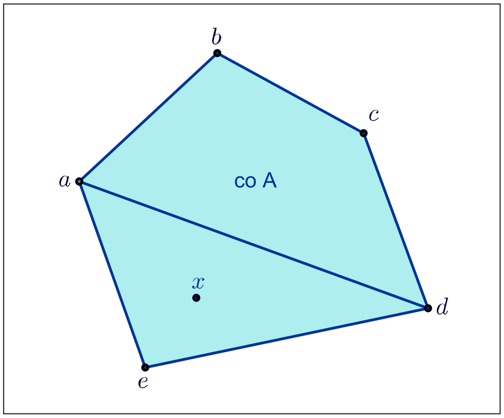

Figure 3.2: Carathéodory theorem.

Theorem 3.6 Let Ω be a nonempty subset of $\mathbb{R}^n$. Then every $x \in \operatorname{co} \Omega$ can be represented as a convex combination of no more than $n + 1$ elements of the set Ω.

Proof. Denoting $B := \{1\} \times \Omega \subset \mathbb{R}^{n+1}$, it is easy to observe that $\operatorname{co} B = \{1\} \times \operatorname{co} \Omega$ and that $\operatorname{co} B \subset K_B$. Take now any $x \in \operatorname{co} \Omega$. Then $(1, x) \in \operatorname{co} B$ and by Proposition 3.5 find $\lambda_i \geq 0$ and $(1, a_i) \in B$ for $i = 0, \ldots, m$ with $m \leq n$ such that

$$(1, x) = \sum_{i=0}^{m} \lambda_i (1, a_i).$$

This implies that $x = \sum_{i=0}^{m} \lambda_i a_i$ with $\sum_{i=0}^{m} \lambda_i = 1$, $\lambda_i \geq 0$, and $m \leq n$. $\square$

The next useful assertion is a direct consequence of Theorem 3.6.

Corollary 3.7 *Suppose that a subset Ω of $\mathbb{R}^n$ is compact. Then the set* $\operatorname{co} \Omega$ *is also compact.*

Now we present two propositions, which lead us to another remarkable result known as the *Farkas lemma*; see Theorem 3.10.

Proposition 3.8 *Let $a_1, \ldots, a_m$ be linearly dependent elements in $\mathbb{R}^n$, where $a_i \neq 0$ for all $i = 1, \ldots, m$. Define the sets*

$$\Omega := \{a_1, \ldots, a_m\} \text{ and } \Omega_i := \Omega \setminus \{a_i\}, \quad i = 1, \ldots, m.$$

Then the convex cones generated by these sets are related as

$$K_\Omega = \bigcup_{i=1}^{m} K_{\Omega_i}. \tag{3.3}$$

Proof. Obviously, $\cup_{i=1}^m K_{\Omega_i} \subset K_\Omega$. To prove the converse inclusion, pick $x \in K_\Omega$ and get

$$x = \sum_{i=1}^{m} \mu_i a_i \text{ with } \mu_i \geq 0, \quad i = 1, \ldots, m.$$

If $x = 0$, then $x \in \cup_{i=1}^m K_{\Omega_i}$ obviously. Otherwise, the linear dependence of a_i allows us to find $\gamma_i \in \mathbb{R}$ not all zeros such that

$$\sum_{i=1}^{m} \gamma_i a_i = 0.$$

Following the proof of Proposition 3.5, we represent x as

$$x = \sum_{i=1, i \neq i_0}^{m} \lambda_i a_i \text{ with some } i_0 \in \{1, \ldots, m\}.$$

This shows that $x \in \cup_{i=1}^m K_{\Omega_i}$ and thus verifies that equality (3.3) holds. □

Proposition 3.9 *Let $a_1, \ldots, a_m$ be elements of $\mathbb{R}^n$. Then the convex cone K_Ω generated by the set $\Omega := \{a_1, \ldots, a_m\}$ is closed in $\mathbb{R}^n$.*

Proof. Assume first that the elements $a_1, \ldots, a_m$ are linearly independent and take any sequence $\{x_k\} \subset K_\Omega$ converging to x. By the construction of K_Ω, find numbers $\alpha_{ki} \geq 0$ for each $i = 1, \ldots, m$ and $k \in \mathbb{N}$ such that

$$x_k = \sum_{i=1}^{m} \alpha_{ki} a_i.$$

Letting $\alpha_k := (\alpha_{k1}, \ldots, \alpha_{km}) \in \mathbb{R}^m$ and arguing by contradiction, it is easy to check that the sequence $\{\alpha_k\}$ is bounded. This allows us to suppose without loss of generality that $\alpha_k \to (\alpha_1, \ldots, \alpha_m)$ as $k \to \infty$ with $\alpha_i \geq 0$ for all $i = 1, \ldots, m$. Thus we get by passing to the limit that $x_k \to x = \sum_{i=1}^m \alpha_i a_i \in K_\Omega$, which verifies the closedness of K_Ω.

Considering next arbitrary elements $a_1, \ldots, a_m$, assume without loss of generality that $a_i \neq 0$ for all $i = 1, \ldots, m$. By the above, it remains to examine the case where $a_1, \ldots, a_m$ are linearly dependent. Then by Proposition 3.8 we have the cone representation (3.3), where each set Ω_i contains $m-1$ elements. If at least one of the sets Ω_i consists of linearly dependent elements, we can represent the corresponding cone K_{Ω_i} in a similar way via the sets of $m-2$ elements. This gives us after a finite number of steps that

$$K_\Omega = \bigcup_{j=1}^{p} K_{C_j},$$

where each set $C_j \subset \Omega$ contains only linear independent elements, and so the cones K_{C_j} are closed. Thus the cone K_Ω under consideration is closed as well. □

Now we are ready to derive the fundamental result of convex analysis and optimization established by Gyula Farkas in 1894.

Theorem 3.10 Let $\Omega := \{a_1, \ldots, a_m\}$ with $a_1, \ldots, a_m \in \mathbb{R}^n$ and let $b \in \mathbb{R}^n$. The following assertions are equivalent:

(i) $b \in K_\Omega$.
(ii) For any $x \in \mathbb{R}^n$, we have the implication

$$\big[\langle a_i, x\rangle \le 0,\ i = 1, \ldots, m\big] \Longrightarrow \big[\langle b, x\rangle \le 0\big].$$

Proof. To verify $(\mathbf{i}) \Longrightarrow (\mathbf{ii})$, take $b \in K_\Omega$ and find $\lambda_i \ge 0$ for $i = 1, \ldots, m$ such that

$$b = \sum_{i=1}^{m} \lambda_i a_i.$$

Then, given $x \in \mathbb{R}^n$, the inequalities $\langle a_i, x\rangle \le 0$ for all $i = 1, \ldots, m$ imply that

$$\langle b, x\rangle = \sum_{i=1}^{m} \lambda_i \langle a_i, x\rangle \le 0,$$

which is **(ii)**. To prove the converse implication, suppose that $b \notin K_\Omega$ and show that **(ii)** does not hold in this case. Indeed, since the set K_Ω is closed and convex, the strict separation result of Proposition 2.1 yields the existence of $\bar{x} \in \mathbb{R}^n$ satisfying

$$\sup\big\{\langle u, \bar{x}\rangle \;\big|\; u \in K_\Omega\big\} < \langle b, \bar{x}\rangle.$$

Since K_Ω is a cone, we have $0 = \langle 0, \bar{x}\rangle < \langle b, \bar{x}\rangle$ and $tu \in K_\Omega$ for all $t > 0$, $u \in K_\Omega$. Thus

$$t \sup\big\{\langle u, \bar{x}\rangle \;\big|\; u \in K_\Omega\big\} < \langle b, \bar{x}\rangle \quad \text{whenever} \quad t > 0.$$

Dividing both sides of this inequality by t and letting $t \to \infty$ give us

$$\sup \left\{ \langle u, \bar{x} \rangle \;\middle|\; u \in K_\Omega \right\} \leq 0.$$

It follows that $\langle a_i, \bar{x} \rangle \leq 0$ for all $i = 1, \ldots, m$ while $\langle b, \bar{x} \rangle > 0$. This is a contradiction, which therefore completes the proof of the theorem. □

3.3 RADON THEOREM AND HELLY THEOREM

In this section we present a celebrated theorem of convex geometry discovered by Edward Helly in 1913. Let us start with the following simple lemma.

Lemma 3.11 *Consider arbitrary elements $w_1, \ldots, w_m$ in $\mathbb{R}^n$ with $m \geq n + 2$. Then these elements are affinely dependent.*

Proof. Form the elements $w_2 - w_1, \ldots, w_m - w_1$ in $\mathbb{R}^n$ and observe that these elements are linearly dependent since $m - 1 > n$. Thus the elements $w_1, \ldots, w_m$ are affinely dependent by Proposition 1.63. □

We continue with a theorem by Johann Radon established in 1923 and used by himself to give a simple proof of the Helly theorem.

Theorem 3.12 Consider the elements $w_1, \ldots, w_m$ of $\mathbb{R}^n$ with $m \geq n + 2$ and let $I := \{1, \ldots, m\}$. Then there exist two nonempty, disjoint subsets I_1 and I_2 of I with $I = I_1 \cup I_2$ such that for $\Omega_1 := \{w_i \mid i \in I_1\}$, $\Omega_2 := \{w_i \mid i \in I_2\}$ one has

$$\operatorname{co} \Omega_1 \cap \operatorname{co} \Omega_2 \neq \emptyset.$$

Proof. Since $m \geq n + 2$, it follows from Lemma 3.11 that the elements $w_1, \ldots, w_m$ are affinely dependent. Thus there are real numbers $\lambda_1, \ldots, \lambda_m$, not all zeros, such that

$$\sum_{i=1}^{m} \lambda_i w_i = 0, \quad \sum_{i=1}^{m} \lambda_i = 0.$$

Consider the index sets $I_1 := \{i = 1, \ldots, m \mid \lambda_i \geq 0\}$ and $I_2 := \{i = 1, \ldots, m \mid \lambda_i < 0\}$. Since $\sum_{i=1}^{m} \lambda_i = 0$, both sets I_1 and I_2 are nonempty and $\sum_{i \in I_1} \lambda_i = -\sum_{i \in I_2} \lambda_i$. Denoting $\lambda := \sum_{i \in I_1} \lambda_i$, we have the equalities

$$\sum_{i \in I_1} \lambda_i w_i = -\sum_{i \in I_2} \lambda_i w_i \quad \text{and} \quad \sum_{i \in I_1} \frac{\lambda_i}{\lambda} w_i = \sum_{i \in I_2} \frac{-\lambda_i}{\lambda} w_i.$$

Define $\Omega_1 := \{w_i \mid i \in I_1\}$, $\Omega_2 := \{w_i \mid i \in I_2\}$ and observe that $\operatorname{co}\Omega_1 \cap \operatorname{co}\Omega_2 \neq \emptyset$ since

$$\sum_{i \in I_1} \frac{\lambda_i}{\lambda} w_i = \sum_{i \in I_2} \frac{-\lambda_i}{\lambda} w_i \in \operatorname{co}\Omega_1 \cap \operatorname{co}\Omega_2.$$

This completes the proof of the Radon theorem. □

Now we are ready to formulate and prove the fundamental Helly theorem.

Theorem 3.13 Let $\mathcal{O} := \{\Omega_1, \ldots, \Omega_m\}$ be a collection of convex sets in $\mathbb{R}^n$ with $m \geq n+1$. Suppose that the intersection of any subcollection of $n+1$ sets from $\mathcal{O}$ is nonempty. Then we have $\bigcap_{i=1}^m \Omega_i \neq \emptyset$.

Proof. Let us prove the theorem by using induction with respect to the number of elements in $\mathcal{O}$. The conclusion of the theorem is obvious if $m = n+1$. Suppose that the conclusion holds for any collection of m convex sets of $\mathbb{R}^n$ with $m \geq n+1$. Let $\{\Omega_1, \ldots, \Omega_m, \Omega_{m+1}\}$ be a collection of convex sets in $\mathbb{R}^n$ such that the intersection of any subcollection of $n+1$ sets is nonempty. For $i = 1, \ldots, m+1$, define

$$\Theta_i := \bigcap_{j=1, j \neq i}^{m+1} \Omega_j$$

and observe that $\Theta_i \subset \Omega_j$ whenever $j \neq i$ and $i, j = 1, \ldots, m+1$. It follows from the induction hypothesis that $\Theta_i \neq \emptyset$, and so we can pick $w_i \in \Theta_i$ for every $i = 1, \ldots, m+1$. Applying the above Radon theorem for the elements $w_1, \ldots w_{m+1}$, we find two nonempty, disjoint subsets index sets I_1 and I_2 of $I := \{1, \ldots, m+1\}$ with $I = I_1 \cup I_2$ such that for $W_1 := \{w_i \mid i \in I_1\}$ and $W_2 := \{w_i \mid i \in I_2\}$ one has $\operatorname{co} W_1 \cap \operatorname{co} W_2 \neq \emptyset$. This allows us to pick an element $w \in \operatorname{co} W_1 \cap \operatorname{co} W_2$. We finally show that $w \in \cap_{i=1}^{m+1} \Omega_i$ and thus arrive at the conclusion. Since $i \neq j$ for $i \in I_1$ and $j \in I_2$, it follows that $\Theta_i \subset \Omega_j$ whenever $i \in I_1$ and $j \in I_2$. Fix any $i = 1, \ldots, m+1$ and consider the case where $i \in I_1$. Then $w_j \in \Theta_j \subset \Omega_i$ for every $j \in I_2$. Hence $w \in \operatorname{co} W_2 = \operatorname{co}\{w_j \mid j \in I_2\} \subset \Omega_i$ due the convexity of Ω_i. It shows that $w \in \Omega_i$ for every $i \in I_1$. Similarly, we can verify that $w \in \Omega_i$ for every $i \in I_2$ and thus complete the proof of the theorem. □

3.4 TANGENTS TO CONVEX SETS

In Chapter 2 we studied the normal cone to convex sets and learned that the very definition and major properties of it were related to convex separation. In this section we define the tangent cone to convex sets independently while observing that it can be constructed via *duality* from the normal cone to the set under consideration.

We start with cones. Given a nonempty, convex cone $C \subset \mathbb{R}^n$, the *polar* of C is

$$C^\circ := \{v \in \mathbb{R}^n \mid \langle v, c \rangle \leq 0 \text{ for all } c \in C\}.$$

Proposition 3.14 *For any nonempty, convex cone $C \subset \mathbb{R}^n$, we have*

$$(C^\circ)^\circ = \text{cl}\, C.$$

Proof. The inclusion $C \subset (C^\circ)^\circ$ follows directly from the definition, and hence $\text{cl}\, C \subset (C^\circ)^\circ$ since $(C^\circ)^\circ$ is a closed set. To verify the converse inclusion, take any $c \in (C^\circ)^\circ$ and suppose by contradiction that $c \notin \text{cl}\, C$. Since C is a cone, by convex separation from Proposition 2.1, find $v \neq 0$ such that

$$\langle v, x\rangle \le 0 \text{ for all } x \in C \text{ and } \langle v, c\rangle > 0,$$

which yields $v \in C^\circ$ thus contradicts to the fact that $c \in (C^\circ)^\circ$. □

Definition 3.15 *Let Ω be a nonempty, convex set. The* TANGENT CONE *to Ω at $\bar{x} \in \Omega$ is*

$$T(\bar{x}; \Omega) := \text{cl}\, K_{\{\Omega - \bar{x}\}} = \text{cl}\, \mathbb{R}^+(\Omega - \bar{x}).$$

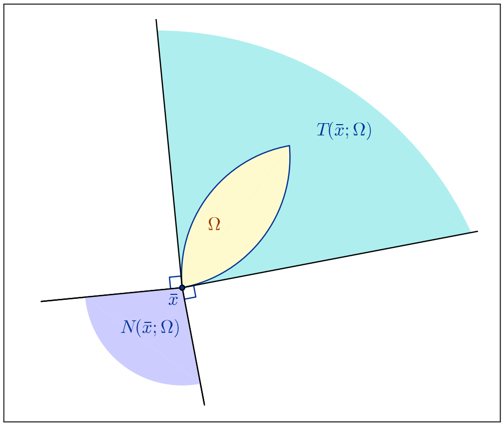

Figure 3.3: Tangent cone and normal cone.

Now we get the *duality correspondence* between normals and tangents to convex sets; see a demonstration in Figure 3.3.

Theorem 3.16 Let Ω be a convex set and let $\bar{x} \in \Omega$. Then we have the equalities

$$N(\bar{x}; \Omega) = \left[T(\bar{x}; \Omega)\right]^\circ \text{ and } T(\bar{x}; \Omega) = \left[N(\bar{x}; \Omega)\right]^\circ. \tag{3.4}$$

Proof. To verify first the inclusion $N(\bar{x};\Omega) \subset [T(\bar{x};\Omega)]^\circ$, pick any $v \in N(\bar{x};\Omega)$ and get by definition that $\langle v, x-\bar{x}\rangle \le 0$ for all $x \in \Omega$. This implies that

$$\big\langle v, t(x-\bar{x})\big\rangle \le 0 \text{ whenever } t \ge 0 \text{ and } x \in \Omega.$$

Fix $w \in T(\bar{x};\Omega)$ and find $t_k \ge 0$ and $x_k \in \Omega$ with $t_k(x_k - \bar{x}) \to w$ as $k \to \infty$. Then

$$\langle v, w\rangle = \lim_{k\to\infty} \langle v, t_k(x_k-\bar{x})\rangle \le 0 \text{ and thus } v \in \big[T(\bar{x};\Omega)\big]^\circ.$$

To verify the opposite inclusion in (3.4), fix any $v \in [T(\bar{x};\Omega)]^\circ$ and deduce from the relations $\langle v, w\rangle \le 0$ as $w \in T(\bar{x};\Omega)$ and $x - \bar{x} \in T(\bar{x};\Omega)$ as $x \in \Omega$ that

$$\langle v, x-\bar{x}\rangle \le 0 \text{ for all } x \in \Omega.$$

This yields $w \in N(\bar{x};\Omega)$ and thus completes the proof of the first equality in (3.4). The second equality in (3.4) follows from the first one due to

$$\big[N(\bar{x};\Omega)\big]^\circ = \big[T(\bar{x};\Omega)\big]^{\circ\circ} = \operatorname{cl} T(\bar{x};\Omega)$$

by Proposition 3.14 and the closedness of $T(\bar{x};\Omega)$ by Definition 3.15. □

The next result, partly based on the Farkas lemma from Theorem 3.10, provides effective representations of the tangent and normal cones to linear inequality systems.

Theorem 3.17 Let $a_i \in \mathbb{R}^n$, let $b_i \in \mathbb{R}$ for $i = 1,\ldots,m$, and let

$$\Omega := \big\{x \in \mathbb{R}^n \;\big|\; \langle a_i, x\rangle \le b_i\big\}.$$

Suppose that $\bar{x} \in \Omega$ and define the active index set $I(\bar{x}) := \{i = 1,\ldots,m \mid \langle a_i,\bar{x}\rangle = b_i\}$. Then we have the relationships

$$T(\bar{x};\Omega) = \big\{v \in \mathbb{R}^n \;\big|\; \langle a_i, v\rangle \le 0 \text{ for all } i \in I(\bar{x})\big\}, \tag{3.5}$$

$$N(\bar{x};\Omega) = \operatorname{cone}\big\{a_i \;\big|\; i \in I(\bar{x})\big\}. \tag{3.6}$$

Here we use the convention that $\operatorname{cone}\emptyset := \{0\}$.

Proof. To justify (3.5), consider the convex cone

$$K := \big\{v \in \mathbb{R}^n \;\big|\; \langle a_i, v\rangle \le 0 \text{ for all } i \in I(\bar{x})\big\}.$$

It is easy to see that for any $x \in \Omega$, any $i \in I(\bar{x})$, and any $t \ge 0$ we have

$$\big\langle a_i, t(x-\bar{x})\big\rangle = t\big(\langle a_i, x\rangle - \langle a_i,\bar{x}\rangle\big) \le t(b_i - b_i) = 0,$$

and hence $\mathbb{R}^+(\Omega - \bar{x}) \subset K$. Thus it follows from the closedness of K that

$$T(\bar{x}; \Omega) = \text{cl}\, \mathbb{R}^+(\Omega - \bar{x}) \subset K.$$

To verify the opposite inclusion in (3.5), observe that $\langle a_i, \bar{x}\rangle < b_i$ for $i \notin I(\bar{x})$. Using this and picking an arbitrary element $v \in K$ give us

$$\langle a_i, \bar{x} + tv\rangle \le b_i \quad \text{for all} \quad i \notin I(\bar{x}) \quad \text{and small} \quad t > 0.$$

Since $\langle a_i, v\rangle \le 0$ for all $i \in I(\bar{x})$, it tells us that $\langle a_i, \bar{x} + tv\rangle \le b_i$ as $i = 1, \ldots, m$ yielding

$$\bar{x} + tv \in \Omega, \quad \text{or} \quad v \in \frac{1}{t}\big(\Omega - \bar{x}\big) \subset T(\bar{x}; \Omega).$$

This justifies the tangent cone representation (3.5).

Next we prove (3.6) for the normal cone. The first equality in (3.4) ensures that

$$w \in N(\bar{x}; \Omega) \quad \text{if and only if} \quad \langle w, v\rangle \le 0 \quad \text{for all} \quad v \in T(\bar{x}; \Omega).$$

Then we see from (3.5) that $w \in N(\bar{x}; \Omega)$ if and only if

$$\big[\langle a_i, v\rangle \le 0 \quad \text{for all} \quad i \in I(\bar{x})\big] \Longrightarrow \big[\langle w, v\rangle \le 0\big]$$

whenever $v \in \mathbb{R}^n$. The Farkas lemma from Theorem 3.10 tells us that $w \in N(\bar{x}; \Omega)$ if and only if $w \in \text{cone}\{a_i \mid i \in I(\bar{x})\}$, and thus (3.6) holds. □

3.5 MEAN VALUE THEOREMS

The mean value theorem for differentiable functions is one of the central results of classical real analysis. Now we derive its subdifferential counterparts for general convex functions.

Theorem 3.18 Let $f\colon \mathbb{R} \to \overline{\mathbb{R}}$ be convex, let $a < b$, and let the interval $[a, b]$ belong to the domain of f, which is an open interval in $\mathbb{R}$. Then there is $c \in (a, b)$ such that

$$\frac{f(b) - f(a)}{b - a} \in \partial f(c). \tag{3.7}$$

Proof. Define the real-valued function $g\colon [a, b] \to \mathbb{R}$ by

$$g(x) := f(x) - \Big[\frac{f(b) - f(a)}{b - a}(x - a) + f(a)\Big]$$

for which $g(a) = g(b)$. This implies, by the convexity and hence continuity of g on its open domain, that g has a local minimum at some $c \in (a, b)$. The function

$$h(x) := -\Big[\frac{f(b) - f(a)}{b - a}(x - a) + f(a)\Big]$$

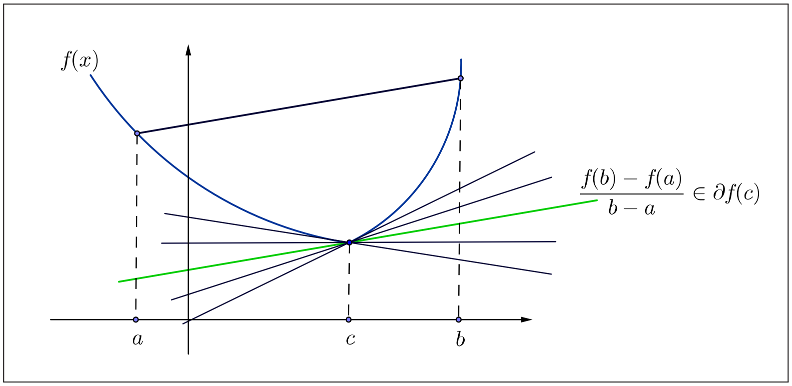

Figure 3.4: Subdifferential mean value theorem.

is obviously differentiable at c, and hence

$$\partial h(c) = \{h'(c)\} = \Big\{-\frac{f(b)-f(a)}{b-a}\Big\}.$$

The subdifferential Fermat rule and the subdifferential sum rule from Chapter 2 yield

$$0 \in \partial g(c) = \partial f(c) - \Big\{\frac{f(b)-f(a)}{b-a}\Big\},$$

which implies (3.7) and completes the proof of the theorem. □

To proceed further with deriving a counterpart of the mean value theorem for convex functions on $\mathbb{R}^n$ (Theorem 3.20), we first present the following lemma.

Lemma 3.19 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function and let D be the domain of f, which is an open subset of $\mathbb{R}^n$. Given $a, b \in D$ with $a \neq b$, define the function $\varphi\colon \mathbb{R} \to \overline{\mathbb{R}}$ by*

$$\varphi(t) := f\big(tb + (1-t)a\big), \quad t \in \mathbb{R}. \tag{3.8}$$

Then for any number $t_0 \in (0,1)$ we have

$$\partial\varphi(t_0) = \big\{\langle v, b-a\rangle \;\big|\; v \in \partial f(c_0)\big\} \text{ with } c_0 := t_0 a + (1-t_0)b.$$

Proof. Define the vector-valued functions $A\colon \mathbb{R} \to \mathbb{R}^n$ and $B\colon \mathbb{R} \to \mathbb{R}^n$ by

$$A(t) := (b-a)t \text{ and } B(t) := tb + (1-t)a = t(b-a) + a = A(t) + a, \quad t \in \mathbb{R}.$$

It is clear that $\varphi(t) = (f \circ B)(t)$ and that the adjoint mapping to A is $A^*v = \langle v, b-a\rangle$, $v \in \mathbb{R}^n$. By the subdifferential chain rule from Theorem 2.51, we have

$$\partial\varphi(t_0) = A^*\partial f(c_0) = \big\{\langle v, b-a\rangle \;\big|\; v \in \partial f(c_0)\big\},$$

which verifies the result claimed in the lemma. □

Theorem 3.20 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function in the setting of Lemma 3.19. Then there exists an element $c \in (a, b)$ such that

$$f(b) - f(a) \in \big\langle \partial f(c), b-a\big\rangle := \big\{\langle v, b-a\rangle \;\big|\; v \in \partial f(c)\big\}.$$

Proof. Consider the function $\varphi\colon \mathbb{R} \to \overline{\mathbb{R}}$ from (3.8) for which $\varphi(0) = f(a)$ and $\varphi(1) = f(b)$. Applying first Theorem 3.18 and then Lemma 3.19 to φ, we find $t_0 \in (0,1)$ such that

$$f(b) - f(a) = \varphi(1) - \varphi(0) \in \partial\varphi(t_0) = \big\{\langle v, b-a\rangle \;\big|\; v \in \partial f(c)\big\}$$

with $c = t_0 a + (1-t_0)b$. This justifies our statement. □

3.6 HORIZON CONES

This section concerns behavior of unbounded convex sets at *infinity*.

Definition 3.21 *Given a nonempty, closed, convex subset F of $\mathbb{R}^n$ and a point $x \in F$, the* HORIZON CONE *of F at x is defined by the formula*

$$F_\infty(x) := \big\{d \in \mathbb{R}^n \;\big|\; x + td \in F \ \ \textit{for all}\ \ t > 0\big\}.$$

Note that the horizon cone is also known in the literature as the *asymptotic cone* of F at the point in question. Another equivalent definition of $F_\infty(x)$ is given by

$$F_\infty(x) = \bigcap_{t>0} \frac{F - x}{t},$$

which clearly implies that $F_\infty(x)$ is a closed, convex cone in $\mathbb{R}^n$. The next proposition shows that $F_\infty(x)$ is the same for any $x \in F$, and so it can be simply denoted by F_∞.

Proposition 3.22 *Let F be a nonempty, closed, convex subset of $\mathbb{R}^n$. Then we have the equality $F_\infty(x_1) = F_\infty(x_2)$ for any $x_1, x_2 \in F$.*

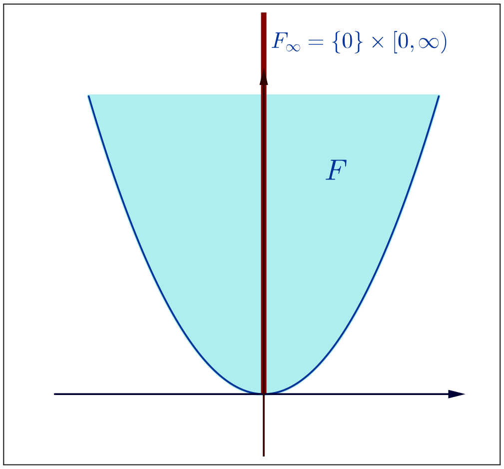

Figure 3.5: Horizon cone.

Proof. It suffices to verify that $F_\infty(x_1) \subset F_\infty(x_2)$ whenever $x_1, x_2 \in F$. Taking any direction $d \in F_\infty(x_1)$ and any number $t > 0$, we show that $x_2 + td \in F$. Consider the sequence

$$x_k := \frac{1}{k}\Big(x_1 + ktd\Big) + \Big(1 - \frac{1}{k}\Big)\, x_2, \; k \in \mathbb{N}\,.$$

Then $x_k \in F$ for every k because $d \in F_\infty(x_1)$ and F is convex. We also have $x_k \to x_2 + td$, and so $x_2 + td \in F$ since F is closed. Thus $d \in F_\infty(x_2)$. □

Corollary 3.23 *Let F be a closed, convex subset of $\mathbb{R}^n$ that contains the origin. Then*

$$F_\infty = \bigcap_{t>0} tF.$$

Proof. It follows from the construction of F_∞ and Proposition 3.22 that

$$F_\infty = F_\infty(0) = \bigcap_{t>0} \frac{F - 0}{t} = \bigcap_{t>0} \frac{1}{t} F = \bigcap_{t>0} tF,$$

which therefore justifies the claim. □

Proposition 3.24 *Let F be a nonempty, closed, convex subset of $\mathbb{R}^n$. For a given element $d \in \mathbb{R}^n$, the following assertions are equivalent:*

(i) $d \in F_\infty$.
(ii) *There are sequences $\{t_k\} \subset [0, \infty)$ and $\{f_k\} \subset F$ such that $t_k \to 0$ and $t_k f_k \to d$.*

Proof. To verify **(i)**$\Longrightarrow$**(ii)**, take $d \in F_\infty$ and fix $\bar{x} \in F$. It follows from the definition that

$$\bar{x} + kd \in F \quad \text{for all} \quad k \in \mathbb{N}.$$

For each $k \in \mathbb{N}$, this allows us to find $f_k \in F$ such that

$$\bar{x} + kd = f_k, \quad \text{or equivalently,} \quad \frac{1}{k}\bar{x} + d = \frac{1}{k} f_k.$$

Letting $t_k := \dfrac{1}{k}$, we see that $t_k f_k \to d$ as $k \to \infty$.

To prove **(ii)**$\Longrightarrow$**(i)**, suppose that there are sequences $\{t_k\} \subset [0, \infty)$ and $\{f_k\} \subset F$ with $t_k \to 0$ and $t_k f_k \to d$. Fix $x \in F$ and verify that $d \in F_\infty$ by showing that

$$x + td \in F \quad \text{for all} \quad t > 0.$$

Indeed, for any fixed $t > 0$ we have $0 \leq t \cdot t_k < 1$ when k is sufficiently large. Thus

$$(1 - t \cdot t_k)x + t \cdot t_k f_k \to x + td \quad \text{as} \quad k \to \infty.$$

It follows from the convexity of F that every element $(1 - t \cdot t_k)x + t \cdot t_k f_k$ belongs to F. Hence $x + td \in F$ by the closedness, and thus we get $d \in F_\infty$. $\square$

Finally in this section, we present the expected characterization of the set boundedness in terms of its horizon cone.

Theorem 3.25 Let F be a nonempty, closed, convex subset of $\mathbb{R}^n$. Then the set F is bounded if and only if its horizon cone is trivial, i.e., $F_\infty = \{0\}$.

Proof. Suppose F is bounded and take any $d \in F_\infty$. By Proposition 3.24, there exist sequences $\{t_k\} \subset [0, \infty)$ with $t_k \to 0$ and $\{f_k\} \subset F$ such that $t_k f_k \to d$ as $k \to \infty$. It follows from the boundedness of F that $t_k f_k \to 0$, which shows that $d = 0$.

To prove the converse implication, suppose by contradiction that F is unbounded while $F_\infty = \{0\}$. Then there is a sequence $\{x_k\} \subset F$ with $\|x_k\| \to \infty$. This allows us to form the sequence of (unit) directions

$$d_k := \frac{x_k}{\|x_k\|}, \quad k \in \mathbb{N}.$$

It ensures without loss of generality that $d_k \to d$ as $k \to \infty$ with $\|d\| = 1$. Fix any $x \in F$ and observe that for all $t > 0$ and $k \in \mathbb{N}$ sufficiently large we have

$$u_k := \left(1 - \frac{t}{\|x_k\|}\right)x + \frac{t}{\|x_k\|}x_k \in F$$

by the convexity of F. Furthermore, $x + td \in F$ since $u_k \to x + td$ and F is closed. Thus $d \in F_\infty$, which is a contradiction that completes the proof of the theorem. $\square$

3.7 MINIMAL TIME FUNCTIONS AND MINKOWSKI GAUGE

The main object of this and next sections is a remarkable class of convex functions known as the *minimal time function*. They are relatively new in convex analysis and highly important in applications. Some of their recent applications to *location problems* are given in Chapter 4. We also discuss relationships between minimal time functions and *Minkowski gauge* functions well-recognized in convex analysis and optimization.

Definition 3.26 *Let F be a nonempty, closed, convex subset of $\mathbb{R}^n$ and let $\Omega \subset \mathbb{R}^n$ be a nonempty set, not necessarily convex. The* MINIMAL TIME FUNCTION *associated with the sets F and Ω is defined by*

$$\mathcal{T}_\Omega^F(x) := \inf\big\{t \geq 0 \;\big|\; (x + tF) \cap \Omega \neq \emptyset\big\}. \tag{3.9}$$

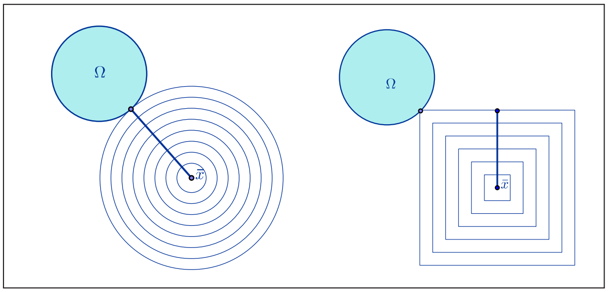

Figure 3.6: Minimal time functions.

The name comes from the following interpretation: (3.9) signifies the minimal time needed for the point x to reach the *target set* Ω along the *constant dynamics* F; see Figure 3.6. Note that when $F = I\!B$, the closed unit ball in $\mathbb{R}^n$, the minimal time function (3.9) reduces to the distance function $d(x;\Omega)$ considered in Section 2.5. In the case where $F = \{0\}$, the minimal time function (3.9) is the indicator function of Ω. If $\Omega = \{0\}$, we get $\mathcal{T}_\Omega^F(x) = \rho_F(-x)$, where ρ_F is the Minkowski gauge defined later in this section. The minimal time function (3.9) also covers another interesting class of functions called the *directional minimal time function*, which corresponds to the case where F is a nonzero vector.

In this section we study general properties of the minimal time function (3.9) used in what follows. These properties are certainly of independent interest.

Observe that $\mathcal{T}_\Omega^F(x) = 0$ if $x \in \Omega$, and that $\mathcal{T}_\Omega^F(x)$ is finite if there exists $t \geq 0$ with

$$(x + tF) \cap \Omega \neq \emptyset,$$

which holds, in particular, when $0 \in \operatorname{int} F$. If no such t exists, then $\mathcal{T}_\Omega^F(x) = \infty$. Thus the minimal time function (3.9) is an extended-real-valued function in general.

The sets F and Ω in (3.9) are not necessarily bounded. The next theorem reveals behavior of $\mathcal{T}_{\Omega}^{F}(x)$ in the case where Ω is bounded, while F is not necessarily bounded. It is an exercise to formulate and prove related results for the case where F is bounded, while Ω is not necessarily bounded.

Theorem 3.27 Let F be a nonempty, closed, convex subset of $\mathbb{R}^n$ and let $\Omega \subset \mathbb{R}^n$ be a nonempty, closed set. Assume that Ω is bounded. Then

$$\mathcal{T}_{\Omega}^{F}(x) = 0 \text{ implies } x \in \Omega - F_{\infty}.$$

The converse holds if we assume additionally that $0 \in F$.

Proof. Suppose that $\mathcal{T}_{\Omega}^{F}(x) = 0$ and find $t_k \to 0$, $t_k \geq 0$, such that

$$(x + t_k F) \cap \Omega \neq \emptyset, \quad k \in \mathbb{N}.$$

This gives us $q_k \in \Omega$ and $f_k \in F$ with $x + t_k f_k = q_k$ for each k. By the boundedness and closedness of Ω, we can select a subsequence of $\{q_k\}$ (no relabeling) that converges to some $q \in \Omega$, and thus $t_k f_k \to q - x$. It follows from Proposition 3.24 that $q - x \in F_{\infty}$, which justifies $x \in \Omega - F_{\infty}$.

To prove the opposite implication, pick any $x \in \Omega - F_{\infty}$ and find $q \in \Omega$ and $d \in F_{\infty}$ such that $x = q - d$. Since $0 \in F$ and $d \in F_{\infty}$, we get from Corollary 3.23 the inclusions $k(q - x) = kd \in F$ for all $k \in \mathbb{N}$. This shows that

$$q - x \in \frac{1}{k}F \text{ and so } \left(x + \frac{1}{k}F\right) \cap \Omega \neq \emptyset, \quad k \in \mathbb{N}.$$

The latter ensures that $0 \leq \mathcal{T}_{\Omega}^{F}(x) \leq \dfrac{1}{k}$ for all $k \in \mathbb{N}$, which holds only when $\mathcal{T}_{\Omega}^{F}(x) = 0$ and thus completes the proof of the theorem. $\square$

The following example illustrates the result of Theorem 3.27.

Example 3.28 Let $F = \mathbb{R} \times [-1, 1] \subset \mathbb{R}^2$ and let Ω be the disk centered at $(1, 0)$ with radius $r = 1$. Using the representation of F_{∞} in Corollary 3.23, we get $F_{\infty} = \mathbb{R} \times \{0\}$ and $\Omega - F_{\infty} = \mathbb{R} \times [-1, 1]$. Thus $\mathcal{T}_{\Omega}^{F}(x) = 0$ if and only if $x \in \mathbb{R} \times [-1, 1]$.

The next result characterizes the *convexity* of the minimal time function (3.9).

Proposition 3.29 *Let F be a nonempty, closed, convex subset of $\mathbb{R}^n$. If $\Omega \subset \mathbb{R}^n$ is nonempty and convex, then the minimal time function* (3.9) *is convex. If furthermore F is bounded and Ω is closed, then the converse holds as well.*

Proof. Take any $x_1, x_2 \in \operatorname{dom} \mathcal{T}_\Omega^F$ and let $\lambda \in (0,1)$. We now show that

$$\mathcal{T}_\Omega^F\big(\lambda x_1 + (1-\lambda)x_2\big) \le \lambda \mathcal{T}_\Omega^F(x_1) + (1-\lambda)\mathcal{T}_\Omega^F(x_2) \tag{3.10}$$

provided that both sets F and Ω are convex. Denote $\gamma_i := \mathcal{T}_\Omega^F(x_i)$ for $i = 1,2$ and, given any $\epsilon > 0$, select numbers t_i so that

$$\gamma_i \le t_i < \gamma_i + \epsilon \ \text{ and } \ (x_i + t_i F) \cap \Omega \neq \emptyset, \quad i = 1,2.$$

It follows from the convexity of F and Ω that

$$\big[\lambda x_1 + (1-\lambda)x_2 + \big(\lambda t_1 + (1-\lambda)t_2\big)F\big] \cap \Omega \neq \emptyset,$$

which implies in turn the estimates

$$\mathcal{T}_\Omega^F\big(\lambda x_1 + (1-\lambda)x_2\big) \le \lambda t_1 + (1-\lambda)t_2 \le \lambda \mathcal{T}_\Omega^F(x_1) + (1-\lambda)\mathcal{T}_\Omega^F(x_2) + \epsilon.$$

Since ϵ is arbitrary, we arrive at (3.10) and thus verifies the convexity of $\mathcal{T}_\Omega^F$.

Conversely, suppose that $\mathcal{T}_\Omega^F$ is convex provided that F is bounded and that Ω is closed. Then the level set

$$\big\{x \in \mathbb{R}^n \bigm| \mathcal{T}_\Omega^F(x) \le 0\big\}$$

is definitely convex. We can easily show set reduces to the target set Ω, which justifies the convexity of the latter. □

Definition 3.30 *Let F be a nonempty, closed, convex subset of $\mathbb{R}^n$. The* Minkowski gauge *associated with the set F is defined by*

$$\rho_F(x) := \inf\big\{t \ge 0 \bigm| x \in tF\big\}, \quad x \in \mathbb{R}^n. \tag{3.11}$$

The following theorem summarizes the main properties of Minkowski gauge functions.

Theorem 3.31 Let F be a closed, convex set that contains the origin. The Minkowski gauge ρ_F is an extended-real-valued function, which is positively homogeneous and subadditive. Furthermore, $\rho_F(x) = 0$ if and only if $x \in F_\infty$.

Proof. We first prove that ρ_F is subadditive meaning that

$$\rho_F(x_1 + x_2) \le \rho_F(x_1) + \rho_F(x_2) \ \text{ for all } \ x_1, x_2 \in \mathbb{R}^n. \tag{3.12}$$

This obviously holds if the right-hand side of (3.12) is equal to infinity, i.e., we have that $x_1 \notin \operatorname{dom} \rho_F$ or $x_2 \notin \operatorname{dom} \rho_F$. Suppose now that $x_1, x_2 \in \operatorname{dom} \rho_F$ and fix $\epsilon > 0$. Then there are numbers $t_1, t_2 \ge 0$ such that

$$\rho_F(x_1) \le t_1 < \rho_F(x_1) + \epsilon, \ \rho_F(x_2) \le t_2 < \rho_F(x_2) + \epsilon \ \text{ and } \ x_1 \in t_1 F, \ x_2 \in t_2 F.$$

It follows from the convexity of F that $x_1 + x_2 \in t_1 F + t_2 F \subset (t_1 + t_2)F$, and so

$$\rho_F(x_1 + x_2) \le t_1 + t_2 < \rho_F(x_1) + \rho_F(x_2) + 2\epsilon,$$

which justifies (3.12) since $\epsilon > 0$ was chosen arbitrarily. To check the positive homogeneous property of ρ_F, we take any $\alpha > 0$ and conclude that

$$\begin{aligned}\rho_F(\alpha x) &= \inf\{t \ge 0 \mid \alpha x \in tF\} = \inf\Big\{t \ge 0 \;\Big|\; x \in \frac{t}{\alpha}F\Big\}\\ &= \alpha \inf\Big\{\frac{t}{\alpha} \ge 0 \;\Big|\; x \in \frac{t}{\alpha}F\Big\} = \alpha\rho_F(x).\end{aligned}$$

It remains to verify the last statement of the theorem. Observe that $\rho_F(x) = \mathcal{T}_\Omega^{-F}(x)$ is a special case of the minimal time function with $\Omega = \{0\}$. Since $0 \in F$, Theorem 3.27 tells us that $\rho_F(x) = 0$ if and only if $x \in \Omega - (-F)_\infty = F_\infty$, which completes the proof. $\square$

We say for simplicity that $\varphi\colon \mathbb{R}^n \to \mathbb{R}$ is *ℓ–Lipschitz continuous* on $\mathbb{R}^n$ if it is Lipschitz continuous on $\mathbb{R}^n$ with Lipschitz constant $\ell \ge 0$. The next result calculates a constant ℓ of Lipschitz continuity of the Minkowski gauge on $\mathbb{R}^n$.

Proposition 3.32 *Let F be a closed, convex set that contains the origin as an interior point. Then the Minkowski gauge ρ_F is ℓ-Lipschitz continuous on $\mathbb{R}^n$ with the constant*

$$\ell := \inf\Big\{\frac{1}{r} \;\Big|\; I\!B(0;r) \subset F,\ r > 0\Big\}. \tag{3.13}$$

In particular, we have $\rho_F(x) \le \ell\|x\|$ for all $x \in \mathbb{R}^n$.

Proof. The condition $0 \in \operatorname{int} F$ ensures that $\ell < \infty$ for the constant ℓ from (3.13). Take any $r > 0$ satisfying $I\!B(0;r) \subset F$ and get from the definitions that

$$\rho_F(x) = \inf\{t \ge 0 \mid x \in tF\} \le \inf\{t \ge 0 \mid x \in tI\!B(0;r)\} = \inf\{t \ge 0 \mid \|x\| \le rt\} = \frac{\|x\|}{r},$$

which implies that $\rho_F(x) \le \ell\|x\|$. Then the subadditivity of ρ_F gives us that

$$\rho_F(x) - \rho_F(y) \le \rho_F(x - y) \le \ell\|x - y\| \ \text{ for all } \ x, y \in \mathbb{R}^n,$$

and thus ρ_F is ℓ–Lipschitz continuous on $\mathbb{R}^n$ with constant (3.13). $\square$

Next we relate the minimal time function to the Minkowski gauge.

Theorem 3.33 Let F be a nonempty, closed, convex set. Then for any $\Omega \ne \emptyset$ we have

$$\mathcal{T}_\Omega^F(x) = \inf\{\rho_F(\omega - x) \mid \omega \in \Omega\}, \quad x \in \mathbb{R}^n. \tag{3.14}$$

Proof. Consider first the case where $\mathcal{T}^F_\Omega(x) = \infty$. In this case we have

$$\{t \geq 0 \mid (x + tF) \cap \Omega \neq \emptyset\} = \emptyset,$$

which implies that $\{t \geq 0 \mid \omega - x \in tF\} = \emptyset$ for every $\omega \in \Omega$. Thus $\rho_F(\omega - x) = \infty$ as $\omega \in \Omega$, and hence the right-hand side of (3.14) is also infinity.

Considering now the case where $\mathcal{T}^F_\Omega(x) < \infty$, fix any $t \geq 0$ such that $(x + tF) \cap \Omega \neq \emptyset$. This gives us $f \in F$ and $\omega \in \Omega$ with $x + tf = \omega$. Thus $\omega - x \in tF$, and so $\rho_F(\omega - x) \leq t$, which yields $\inf_{\omega\in\Omega} \rho_F(\omega - x) \leq t$. It follows from definition (3.9) that

$$\inf_{\omega\in\Omega} \rho_F(\omega - x) \leq \mathcal{T}^F_\Omega(x) < \infty.$$

To verify the opposite inequality in (3.14), denote $\gamma := \inf_{\omega\in\Omega} \rho_F(\omega - x)$ and for any $\epsilon > 0$ find $\omega \in \Omega$ with $\rho_F(\omega - x) < \gamma + \epsilon$. By definition of the Minkowski gauge (3.11), we get $t \geq 0$ such that $t < \gamma + \epsilon$ and $\omega - x \in tF$. This gives us $\omega \in x + tF$ and therefore

$$\mathcal{T}^F_\Omega(x) \leq t < \gamma + \epsilon.$$

Since ϵ was chosen arbitrarily, we arrive at

$$\mathcal{T}^F_\Omega(x) \leq \gamma = \inf_{\omega\in\Omega} \rho_F(\omega - x),$$

which implies the remaining inequality in (3.14) and completes the proof. □

As a consequence of the obtained result and Proposition 3.32, we establish now the $\ell-$Lipschitz continuity of minimal time function with the constant ℓ calculated in (3.13).

Corollary 3.34 *Let F be a closed, convex set that contains the origin as an interior point. Then for any nonempty set Ω the function $\mathcal{T}^F_\Omega$ is $\ell-$Lipschitz with constant* (3.13).

Proof. It follows from the subadditivity of ρ_F and the result of Proposition 3.32 that

$$\rho_F(\omega - x) \leq \rho_F(\omega - y) + \rho_F(y - x) \leq \rho_F(\omega - y) + \ell\|y - x\| \quad \text{for all} \quad \omega \in \Omega$$

whenever $x, y \in \mathbb{R}^n$. Then Theorem 3.33 tells us that

$$\mathcal{T}^F_\Omega(x) \leq \mathcal{T}^F_\Omega(y) + \ell\|x - y\|,$$

which implies the $\ell-$Lipschitz property of $\mathcal{T}^F_\Omega$ due to the arbitrary choice of x, y. □

Define next the *enlargement* of the target set Ω via minimal time function by

$$\Omega_r := \{x \in \mathbb{R}^n \mid \mathcal{T}^F_\Omega(x) \leq r\}, \quad r > 0. \tag{3.15}$$

Proposition 3.35 *Let F be a nonempty, closed, convex subset of $\mathbb{R}^n$ and let $\Omega \subset \mathbb{R}^n$ be a nonempty set. Then for any nonempty set Ω and any $x \notin \Omega_r$ with $\mathcal{T}_\Omega^F(x) < \infty$ we have*

$$\mathcal{T}_\Omega^F(x) = \mathcal{T}_{\Omega_r}^F(x) + r.$$

Proof. It follows from $\Omega \subset \Omega_r$ that $\mathcal{T}_{\Omega_r}^F(x) \leq \mathcal{T}_\Omega^F(x) < \infty$. Take any $t \geq 0$ with

$$(x + tF) \cap \Omega_r \neq \emptyset$$

and find $f_1 \in F$ and $u \in \Omega_r$ such that $x + tf_1 = u$. Since $u \in \Omega_r$, we have $\mathcal{T}_\Omega^F(u) \leq r$. For any $\epsilon > 0$, there is $s \geq 0$ with $s < r + \epsilon$ and $(u + sF) \cap \Omega \neq \emptyset$. This implies the existence of $\omega \in \Omega$ and $f_2 \in F$ such that $u + sf_2 = \omega$. Thus

$$\omega = u + sf_2 = (x + tf_1) + sf_2 \in x + (t + s)F$$

since F is convex. This gives us

$$\mathcal{T}_\Omega^F(x) \leq t + s \leq t + r + \epsilon \quad \text{and so} \quad \mathcal{T}_\Omega^F(x) \leq t + r$$

by the arbitrary choice of $\epsilon > 0$. This allows us to conclude that

$$\mathcal{T}_\Omega^F(x) \leq \mathcal{T}_{\Omega_r}^F + r.$$

To verify the opposite inequality, denote $\gamma := \mathcal{T}_\Omega^F(x)$ and observe that $r < \gamma$ by $x \notin \Omega_r$. For any $\epsilon > 0$, find $t \in [0, \infty)$, $f \in F$, and $\omega \in \Omega$ with $\gamma \leq t < \gamma + \epsilon$ and $x + tf = \omega$. Hence

$$\omega = x + tf = x + (t - r)f + rf \in x + (t - r)f + rF \quad \text{and so} \quad \mathcal{T}_\Omega^F(x + (t - r)f) \leq r,$$

which shows that $x + (t - r)f \in \Omega_r$. We also see that $x + (t - r)f \in x + (t - r)F$. Thus $\mathcal{T}_{\Omega_r}^F(x) \leq t - r \leq \gamma - r + \epsilon$, which verifies that

$$r + \mathcal{T}_{\Omega_r}^F \leq \gamma = \mathcal{T}_\Omega^F(x)$$

since $\epsilon > 0$ was chosen arbitrarily. This completes the proof. $\square$

Proposition 3.36 *Let F be a nonempty, closed, convex subset of $\mathbb{R}^n$. Then for any nonempty set Ω, any $x \in \operatorname{dom} \mathcal{T}_\Omega^F$, any $t \geq 0$, and any $f \in F$ we have*

$$\mathcal{T}_\Omega^F(x - tf) \leq \mathcal{T}_\Omega^F(x) + t.$$

Proof. Fix $\epsilon > 0$ and select $s \geq 0$ so that

$$\mathcal{T}_\Omega^F(x) \leq s < \mathcal{T}_\Omega^F(x) + \epsilon \quad \text{and} \quad (x + sF) \cap \Omega \neq \emptyset.$$

Then $(x - tf + tF + sF) \cap \Omega \neq \emptyset$, and hence $(x - tf + (t + s)F) \cap \Omega \neq \emptyset$. It shows that

$$\mathcal{T}_\Omega^F(x - tf) \leq t + s \leq t + \mathcal{T}_\Omega^F(x) + \epsilon,$$

which implies the conclusion of the proposition by letting $\epsilon \to 0^+$. $\square$

3.8 SUBGRADIENTS OF MINIMAL TIME FUNCTIONS

Observing that the minimal time function (3.9) is *intrinsically nonsmooth*, we study here its generalized differentiation properties in the case where the target set Ω is convex. In this case (since the constant dynamics F is always assumed convex) function (3.9) is convex as well by Proposition 3.29. The results below present precise formulas for calculating the subdifferential $\partial\mathcal{T}_{\Omega}^{F}(\bar{x})$ depending on the position of $\bar{x}$ with respect to the set $\Omega - F_\infty$.

Theorem 3.37 Let both sets $0 \in F \subset \mathbb{R}^n$ and $\Omega \subset \mathbb{R}^n$ be nonempty, closed, convex. For any $\bar{x} \in \Omega - F_\infty$, the subdifferential of $\mathcal{T}_{\Omega}^{F}$ at $\bar{x}$ is calculated by

$$\partial\mathcal{T}_{\Omega}^{F}(\bar{x}) = N(\bar{x}; \Omega - F_\infty) \cap C^*, \tag{3.16}$$

where $C^* := \{v \in \mathbb{R}^n \mid \sigma_F(-v) \le 1\}$ is defined via (2.41). In particular, we have

$$\partial\mathcal{T}_{\Omega}^{F}(\bar{x}) = N(\bar{x}; \Omega) \cap C^* \ \text{ for } \ \bar{x} \in \Omega. \tag{3.17}$$

Proof. Fix any $v \in \partial\mathcal{T}_{\Omega}^{F}(\bar{x})$. Using (2.13) and the fact that $\mathcal{T}_{\Omega}^{F}(x) = 0$ for all $x \in \Omega - F_\infty$, we get $v \in N(\bar{x}; \Omega - F_\infty)$. Write $\bar{x} \in \Omega - F_\infty$ as $\bar{x} = \bar{\omega} - d$ with $\bar{\omega} \in \Omega$ and $d \in F_\infty$. Fix any $f \in F$ and define $x_t := (\bar{x} + d) - tf = \bar{\omega} - tf$ as $t > 0$. Then $(x_t + tF) \cap \Omega \neq \emptyset$, and so

$$\langle v, x_t - \bar{x}\rangle \le \mathcal{T}_{\Omega}^{F}(x_t) \le t, \quad \text{or equivalently } \Big\langle v, \frac{d}{t} - f \Big\rangle \le 1 \ \text{ for all } \ t > 0.$$

By letting $t \to \infty$, it implies that $\langle v, -f\rangle \le 1$, and hence we get $v \in C^*$. This verifies the inclusion $\partial\mathcal{T}_{\Omega}^{F}(\bar{x}) \subset N(\bar{x}; \Omega - F_\infty) \cap C^*$.

To prove the opposite inclusion in (3.16), fix any $v \in N(\bar{x}; \Omega - F_\infty) \cap C^*$ and any $u \in \operatorname{dom} \mathcal{T}_{\Omega}^{F}$. For any $\epsilon > 0$, there exist $t \in [0, \infty)$, $\omega \in \Omega$, and $f \in F$ such that

$$\mathcal{T}_{\Omega}^{F}(u) \le t < \mathcal{T}_{\Omega}^{F}(u) + \epsilon \ \text{ and } \ u + tf = \omega.$$

Since $\Omega \subset \Omega - F_\infty$, we have the relationships

$$\langle v, u - \bar{x}\rangle = \langle v, \omega - \bar{x}\rangle + t\langle v, -f\rangle \le t < \mathcal{T}_{\Omega}^{F}(u) + \epsilon = \mathcal{T}_{\Omega}^{F}(u) - \mathcal{T}_{\Omega}^{F}(\bar{x}) + \epsilon,$$

which show that $v \in \partial\mathcal{T}_{\Omega}^{F}(\bar{x})$. In this way we get $N(\bar{x}; \Omega - F_\infty) \cap C^* \subset \partial\mathcal{T}_{\Omega}^{F}(\bar{x})$ and thus justify the subdifferential formula (3.16) in the general case where $\bar{x} \in \Omega - F_\infty$.

It remains to verify the simplified representation (3.17) in the case where $\bar{x} \in \Omega$. Since $\Omega \subset \Omega - F_\infty$, we obviously have the inclusion

$$N(\bar{x}; \Omega - F_\infty) \cap C^* \subset N(\bar{x}; \Omega) \cap C^*.$$

To show the converse inclusion, pick $v \in N(\bar{x}; \Omega) \cap C^*$ and fix $x \in \Omega - F_\infty$. Taking $\omega \in \Omega$ and $d \in F_\infty$ with $x = \omega - d$ gives us $td \in F$ since $0 \in F$, so $\langle v, -td\rangle \le 1$ for all $t > 0$. Thus $\langle v, -d\rangle \le 0$, which readily implies that

$$\langle v, x - \bar{x}\rangle = \langle v, \omega - d - \bar{x}\rangle = \langle v, \omega - \bar{x}\rangle + \langle v, -d\rangle \le 0.$$

In this way we arrive at $v \in N(\bar{x}; \Omega - F_\infty) \cap C^*$ and thus complete the proof. □

The next theorem provides the subdifferential calculation for minimal time functions in the cases complemented to those in Theorem 3.37, i.e., when $\bar{x} \notin \Omega - F_\infty$ and $\bar{x} \notin \Omega$.

Theorem 3.38 Let both sets $0 \in F \subset \mathbb{R}^n$ and $\Omega \subset \mathbb{R}^n$ be nonempty, closed, convex and let $\bar{x} \in \operatorname{dom} \mathcal{T}_\Omega^F$ for the minimal time function (3.9). Suppose that Ω is bounded. Then for $\bar{x} \notin \Omega - F_\infty$ we have

$$\partial \mathcal{T}_\Omega^F(\bar{x}) = N(\bar{x}; \Omega_r) \cap S^*, \tag{3.18}$$

where the enlargement Ω_r is defined in (3.15), and where

$$S^* := \big\{ v \in \mathbb{R}^n \;\big|\; \sigma_F(-v) = 1 \big\} \;\text{ with }\; r = \mathcal{T}_\Omega^F(\bar{x}) > 0.$$

Proof. Pick $v \in \partial \mathcal{T}_\Omega^F(\bar{x})$. Observe from the definition that

$$\langle v, x - \bar{x} \rangle \le \mathcal{T}_\Omega^F(x) - \mathcal{T}_\Omega^F(\bar{x}) \;\text{ for all }\; x \in \mathbb{R}^n. \tag{3.19}$$

Then $\mathcal{T}_\Omega^F(x) \le r = \mathcal{T}_\Omega^F(\bar{x})$, and hence $\langle v, x - \bar{x} \rangle \le 0$ for any $x \in \Omega_r$. This yields $v \in N(\bar{x}; \Omega_r)$. Using (3.19) for $x := \bar{x} - f$ with $f \in F$ and Proposition 3.36 ensures that $\sigma_F(-v) \le 1$. Fix $\epsilon \in (0, r)$ and select $t \in \mathbb{R}$, $f \in F$, and $\omega \in \Omega$ so that

$$r \le t < r + \epsilon^2 \;\text{ and }\; \omega = \bar{x} + tf.$$

We can write $\omega = \bar{x} + \epsilon f + (t - \epsilon) f$ and then see that $\mathcal{T}_\Omega^F(\bar{x} + \epsilon f) \le t - \epsilon$. Applying (3.19) to $x := \bar{x} + \epsilon f$ gives us the estimates

$$\langle v, \epsilon f \rangle \le \mathcal{T}_\Omega^F(\bar{x} + \epsilon f) - \mathcal{T}_\Omega^F(\bar{x}) \le t - \epsilon - r \le \epsilon^2 - \epsilon.$$

which imply in turn that

$$1 - \epsilon \le \langle -v, f \rangle \le \sigma_F(-v).$$

Letting now $\epsilon \downarrow 0$ tells us that $\sigma_F(-v) \ge 1$, i.e., $\sigma_F(-v) = 1$ in this case. Thus $v \in S^*$, which proves the inclusion $\partial \mathcal{T}_\Omega^F(\bar{x}) \subset N(\bar{x}; \Omega_r) \cap S^*$.

To verify the opposite inclusion in (3.18), take any $v \in N(\bar{x}; \Omega_r)$ with $\sigma_F(-v) = 1$ and show that (3.19) holds. It follows from Theorem 3.37 that $v \in \partial \mathcal{T}_{\Omega_r}^F(\bar{x})$, and so

$$\langle v, x - \bar{x} \rangle \le \mathcal{T}_{\Omega_r}^F(x) \;\text{ for all }\; x \in \mathbb{R}^n.$$

Fix $x \in \mathbb{R}^n$ and deduce from Proposition 3.35 that $\mathcal{T}_\Omega^F(x) - r = \mathcal{T}_{\Omega_r}^F(x)$ in the case $t := \mathcal{T}_\Omega^F(x) > r$, which ensures (3.19). Suppose now that $t \le r$ and for any $\epsilon > 0$ choose $f \in F$ such that $\langle v, -f \rangle > 1 - \epsilon$. Then Proposition 3.36 tells us that $\mathcal{T}_\Omega^F(x - (r - t) f) \le r$, and so $x - (r - t) f \in \Omega_r$. Since $v \in N(\bar{x}; \Omega_r)$, one has $\langle v, x - (r - t) f - \bar{x} \rangle \le 0$, which yields

$$\langle v, x - \bar{x} \rangle \le \langle v, f \rangle (r - t) \le (1 - \epsilon)(t - r) = (1 - \epsilon)\big(\mathcal{T}_\Omega^F(x) - \mathcal{T}_\Omega^F(\bar{x})\big).$$

Since $\epsilon > 0$ was arbitrarily chosen, we get (3.19), and hence verify that $v \in \partial \mathcal{T}_\Omega^F(\bar{x})$. $\square$

The next result addresses the same setting as in Theorem 3.38 while presenting another formula for calculating the subdifferential of the minimal time function that involves the subdifferential of the Minkowski gauge and generalized projections to the target set.

Theorem 3.39 Let both sets $0 \in F \subset \mathbb{R}^n$ and $\Omega \subset \mathbb{R}^n$ be nonempty, closed, convex and let $\bar{x} \in \operatorname{dom} \mathcal{T}_\Omega^F$. Suppose that Ω is bounded and that $\bar{x} \notin \Omega - F_\infty$. Then we have

$$\partial \mathcal{T}_\Omega^F(\bar{x}) = \big[-\partial \rho_F(\bar{\omega} - \bar{x})\big] \cap N(\bar{\omega}; \Omega) \tag{3.20}$$

for any $\bar{\omega} \in \Pi_F(\bar{x}; \Omega)$ via the generalized projector

$$\Pi_F(\bar{x}; \Omega) := \big\{\omega \in \Omega \;\big|\; \mathcal{T}_\Omega^F(\bar{x}) = \rho_F(\omega - \bar{x})\big\}. \tag{3.21}$$

Proof. Fixing any $v \in \partial \mathcal{T}_\Omega^F(\bar{x})$ and $\bar{\omega} \in \Pi_F(\bar{x}; \Omega)$, we get $\mathcal{T}_\Omega^F(\bar{x}) = \rho_F(\bar{\omega} - \bar{x})$ and

$$\langle v, x - \bar{x}\rangle \le \mathcal{T}_\Omega^F(x) - \mathcal{T}_\Omega^F(\bar{x}) \;\text{ for all }\; x \in \mathbb{R}^n. \tag{3.22}$$

It is easy to check the following relationships that hold for any $\omega \in \Omega$:

$$\begin{aligned}\langle v, \omega - \bar{\omega}\rangle = \langle v, (\omega - \bar{\omega} + \bar{x}) - \bar{x}\rangle &\le \mathcal{T}_\Omega^F(\omega - \bar{\omega} + \bar{x}) - \mathcal{T}_\Omega^F(\bar{x})\\ &\le \rho_F\big(\omega - (\omega - \bar{\omega} + \bar{x})\big) - \rho_F(\bar{\omega} - \bar{x}) = 0.\end{aligned}$$

This shows that $v \in N(\bar{\omega}; \Omega)$. Denote now $\widetilde{u} := \bar{\omega} - \bar{x}$ and for any $t \in (0,1)$ and $u \in \mathbb{R}^n$ get

$$\begin{aligned}\langle v, -t(u - \widetilde{u})\rangle &\le \mathcal{T}_\Omega^F\big(\bar{x} - t(u - \widetilde{u})\big) - \mathcal{T}_\Omega^F(\bar{x})\\ &\le \rho_F\big(\bar{\omega} - (\bar{x} - t(u - \widetilde{u}))\big) - \rho_F(\bar{\omega} - \bar{x})\\ &= \rho_F(\widetilde{u} + t(u - \widetilde{u})) - \rho_F(\widetilde{u})\\ &= \rho_F\big(tu + (1-t)\widetilde{u}\big) - \rho_F(\widetilde{u})\\ &\le t\rho_F(u) + (1-t)\rho_F(\widetilde{u}) - \rho_F(\widetilde{u})\\ &= t\big(\rho_F(u) - \rho_F(\widetilde{u})\big)\end{aligned}$$

by applying (3.22) with $x := \bar{x} - t(u - \widetilde{u})$. It follows that

$$\langle -v, u - \widetilde{u}\rangle \le \rho_F(u) - \rho_F(\widetilde{u}) \;\text{ for all }\; u \in \mathbb{R}^n.$$

Thus $-v \in \partial \rho_F(\widetilde{u})$, and so $v \in -\partial \rho_F(\bar{\omega} - \bar{x})$, which proves the inclusion "$\subset$" in (3.20).

To verify the opposite inclusion, write

$$\rho_F(x) = \inf\big\{t \ge 0 \;\big|\; x \in tF\big\} = \inf\big\{t \ge 0 \;\big|\; (-x + tF) \cap O \ne \emptyset\big\} = \mathcal{T}_O^F(-x)$$

with $O := \{0\}$. Since $\bar{x} \notin \Omega - F_\infty$, for any $\bar{\omega} \in \Omega$ we have $\bar{x} - \bar{\omega} \notin O - F_\infty = -F_\infty$. Furthermore, it follows from $-v \in \partial \rho_F(\bar{\omega} - \bar{x})$ by applying Theorem 3.38 that $v \in S^*$. Hence it remains to justify the inclusion

$$\big[-\partial \rho_F(\bar{\omega} - \bar{x})\big] \cap N(\bar{\omega}; \Omega) \subset N(\bar{x}; \Omega_r) \tag{3.23}$$

and then to apply Theorem 3.38. To proceed with the proof of (3.23), pick any vector $x \in \Omega_r$ and a positive number ϵ arbitrarily small. Then there are $t < r + \epsilon$, $f \in F$, and $\omega \in \Omega$ such that $\omega = x + tf$, which yields

$$\begin{aligned}\langle v, x - \bar{x}\rangle &= \langle v, \omega - tf - \bar{x}\rangle \\ &= t\langle -v, f\rangle + \langle v, \omega - \bar{\omega}\rangle + \langle v, \bar{\omega} - \bar{x}\rangle \\ &\le t + \langle v, \omega - \bar{\omega}\rangle + \langle v, \bar{\omega} - \bar{x}\rangle \\ &\le \mathcal{T}_{\Omega}^{F}(\bar{x}) + \epsilon + \langle v, \omega - \bar{\omega}\rangle + \langle v, \bar{\omega} - \bar{x}\rangle.\end{aligned}$$

Observe that $\langle v, \omega - \bar{\omega}\rangle \le 0$ by $v \in N(\bar{\omega}; \Omega)$. Moreover,

$$\langle v, \bar{\omega} - \bar{x}\rangle = \langle -v, 0 - (\bar{\omega} - \bar{x})\rangle \le \rho_F(0) - \rho_F(\bar{\omega} - \bar{x}) = -\mathcal{T}_{\Omega}^{F}(\bar{x})$$

since $-v \in \partial\rho_F(\bar{\omega} - \bar{x})$. It follows that $\langle v, x - \bar{x}\rangle \le \epsilon$ for all $x \in \Omega_r$. Hence $v \in N(\bar{x}; \Omega_r)$ because $\epsilon > 0$ was chosen arbitrarily. This completes the proof of the theorem. □

3.9 NASH EQUILIBRIUM

This section gives a brief introduction to noncooperative game theory and presents a simple proof of the existence of Nash equilibrium as a consequence of convexity and Brouwer's fixed point theorem. For simplicity, we only consider two-person games while more general situations can be treated similarly.

Definition 3.40 *Let Ω_1 and Ω_2 be nonempty subsets of $\mathbb{R}^n$ and $\mathbb{R}^p$, respectively. A* NONCOOPERATIVE GAME *in the case of two players I and II consists of two strategy sets Ω_i and two real-valued functions $u_i : \Omega_1 \times \Omega_2 \to \mathbb{R}$ for $i = 1, 2$ called the* PAYOFF FUNCTIONS. *Then we refer to this game as $\{\Omega_i, u_i\}$ for $i = 1, 2$.*

The key notion of *Nash equilibrium* was introduced by John Forbes Nash, Jr., in 1950 who proved the *existence* of such an equilibrium and was awarded the Nobel Memorial Prize in Economics in 1994.

Definition 3.41 *Given a noncooperative two-person game $\{\Omega_i, u_i\}$, $i = 1, 2$, a* NASH EQUILIBRIUM *is an element $(\bar{x}_1, \bar{x}_2) \in \Omega_1 \times \Omega_2$ satisfying the conditions*

$$\begin{aligned}u_1(x_1, \bar{x}_2) &\le u_1(\bar{x}_1, \bar{x}_2) \ \textit{ for all } \ x_1 \in \Omega_1, \\ u_2(\bar{x}_1, x_2) &\le u_2(\bar{x}_1, \bar{x}_2) \ \textit{ for all } \ x_2 \in \Omega_2.\end{aligned}$$

The conditions above mean that $(\bar{x}_1, \bar{x}_2)$ is a pair of best strategies that both players agree with in the sense that $\bar{x}_1$ is a best response of Player I when Player II chooses the strategy $\bar{x}_2$, and $\bar{x}_2$ is a best response of Player II when Player I chooses the strategy $\bar{x}_1$. Let us now consider two examples to illustrate this concept.

Example 3.42 In a two-person game suppose that each player can only choose either strategy A or B. If both players choose strategy A, they both get 4 points. If Player I chooses strategy A and Player II chooses strategy B, then Player I gets 1 point and Player II gets 3 points. If Player I chooses strategy B and Player II chooses strategy A, then Player I gets 3 points and Player II gets 1 point. If both choose strategy B, each player gets 2 points. The payoff function of each player is represented in the *payoff matrix* below.

		Player II: A	Player II: B
Player I	A	4, 4	1, 3
	B	3, 1	2, 2

In this example, Nash equilibrium occurs when both players choose strategy A, and it also occurs when both players choose strategy B. Let us consider, for instance, the case where both players choose strategy B. In this case, given that Player II chooses strategy B, Player I also wants to keep strategy B because a change of strategy would lead to a reduction of his/her payoff from 2 to 1. Similarly, given that Player I chooses strategy B, Player II wants to keep strategy B because a change of strategy would also lead to a reduction of his/her payoff from 2 to 1.

Example 3.43 Let us consider another simple two-person game called *matching pennies*. Suppose that Player I and Player II each has a penny. Each player must secretly turn the penny to heads or tails and then reveal his/her choices simultaneously. Player I wins the game and gets Player II's penny if both coins show the same face (heads-heads or tails-tails). In the other case where the coins show different faces (heads-tails or tails-heads), Player II wins the game and gets Player I's penny. The payoff function of each player is represented in the payoff matrix below.

		Player II: Heads	Player II: Tails
Player I	Heads	1, −1	−1, 1
	Tails	−1, 1	1, −1

In this game it is not hard to see that no Nash equilibrium exists.

Denote by A the matrix that represents the payoff function of Player I, and denote by B the matrix that represents the payoff function of Player II:

$$A = \begin{array}{|c|c|}\hline 1 & -1 \\ \hline -1 & 1 \\ \hline\end{array} \quad \text{and } B = \begin{array}{|c|c|}\hline -1 & 1 \\ \hline 1 & -1 \\ \hline\end{array}$$

Now suppose that Player I is playing with a mind reader who knows Player I's choice of faces. If Player I decides to turn heads, then Player II knows about it and chooses to turn tails and wins the game. In the case where Player I decides to turn tails, Player II chooses to turn heads and again wins the game. Thus to have a fair game, he/she decides to randomize his/her strategy by, e.g., tossing the coin instead of putting the coin down. We describe the new game as follows.

Player I		Player II: Heads, q_1	Player II: Tails, q_2
	Heads, p_1	$1, -1$	$-1,\ 1$
	Tails, p_2	$-1,\ 1$	$1, -1$

In this new game, Player I uses a coin randomly with probability of coming up heads p_1 and probability of coming up tails p_2, where $p_1 + p_2 = 1$. Similarly, Player II uses another coin randomly with probability of coming up heads q_1 and probability of coming up tails q_2, where $q_1 + q_2 = 1$. The new strategies are called *mixed strategies* while the original ones are called *pure strategies*.

Suppose that Player II uses mixed strategy $\{q_1, q_2\}$. Then Player I's expected payoff for playing the pure strategy heads is

$$u_H(q_1, q_2) = q_1 - q_2 = 2q_1 - 1.$$

Similarly, Player I's expected payoff for playing the pure strategy tails is

$$u_T(q_1, q_2) = -q_1 + q_2 = -2q_1 + 1.$$

Thus Player I's expected payoff for playing mixed strategy $\{p_1, p_2\}$ is

$$u_1(p, q) = p_1 u_H(q_1, q_2) + p_2 u_T(q_1, q_2) = p_1(q_1 - q_2) + p_2(-q_1 + q_2) = p^T A q,$$

where $p = [p_1, p_2]^T$ and $q = [q_1, q_2]^T$.

By the same arguments, if Player I chooses mixed strategy $\{p_1, p_2\}$, then Player II's expected payoff for playing mixed strategy $\{q_1, q_2\}$ is

$$u_2(p, q) = p^T B q = -u_1(p, q).$$

For this new game, an element $(\bar{p}, \bar{q}) \in \Delta \times \Delta$ is a Nash equilibrium if

$$\begin{aligned} u_1(p, \bar{q}) &\le u_1(\bar{p}, \bar{q}) \ \text{ for all } \ p \in \Delta, \\ u_2(\bar{p}, q) &\le u_2(\bar{p}, \bar{q}) \ \text{ for all } \ q \in \Delta, \end{aligned}$$

where $\Delta := \Big\{(p_1, p_2) \ \Big|\ p_1 \ge 0,\ p_2 \ge 0,\ p_1 + p_2 = 1\Big\}$ is a nonempty, compact, convex set.

Nash [23, 24] proved the *existence* of his equilibrium in the class of *mixed strategies* in a more general setting. His proof was based on Brouwer's fixed point theorem with rather involved arguments. Now we present, following Rockafellar [28], a much simpler proof of the Nash equilibrium theorem that also uses Brouwer's fixed point theorem while applying in addition just elementary tools of convex analysis and optimization.

Theorem 3.44 Consider a two-person game $\{\Omega_i, u_i\}$, where $\Omega_1 \subset \mathbb{R}^n$ and $\Omega_2 \subset \mathbb{R}^p$ are nonempty, compact, convex sets. Let the payoff functions $u_i : \Omega_1 \times \Omega_2 \to \mathbb{R}$ be given by

$$u_1(p,q) := p^T A q, \quad u_2(p,q) := p^T B q,$$

where A and B are $n \times p$ matrices. Then this game admits a Nash equilibrium.

Proof. It follows from the definition that an element $(\bar{p}, \bar{q}) \in \Omega_1 \times \Omega_2$ is a Nash equilibrium of the game under consideration if and only if

$$\begin{aligned} -u_1(p,\bar{q}) &\geq -u_1(\bar{p},\bar{q}) \ \text{ for all } \ p \in \Omega_1, \\ -u_2(\bar{p},q) &\geq -u_2(\bar{p},\bar{q}) \ \text{ for all } \ q \in \Omega_2. \end{aligned}$$

It is a simple exercise to prove (see also an elementary convex optimization result of Corollary 4.15 in Chapter 4) that this holds if and only if we have the normal cone inclusions

$$\nabla_p u_1(\bar{p},\bar{q}) \in N(\bar{p};\Omega_1), \quad \nabla_q u_2(\bar{p},\bar{q}) \in N(\bar{q};\Omega_2). \tag{3.24}$$

By using the structures of u_i and defining $\Omega := \Omega_1 \times \Omega_2$, conditions (3.24) can be expressed in one of the following two equivalent ways:

$$A\bar{q} \in N(\bar{p};\Omega_1) \ \text{ and } \ B^T\bar{p} \in N(\bar{q};\Omega_2);$$

$$(A\bar{q}, B^T\bar{p}) \in N(\bar{p};\Omega_1) \times N(\bar{q};\Omega_2) = N((\bar{p},\bar{q});\Omega_1 \times \Omega_2) = N((\bar{p},\bar{q});\Omega).$$

By Proposition 1.78 on the Euclidean projection, this is equivalent also to

$$(\bar{p},\bar{q}) = \Pi((\bar{p},\bar{q}) + (A\bar{q}, B^T\bar{p});\Omega). \tag{3.25}$$

Defining now the mapping $F\colon \Omega \to \Omega$ by

$$F(p,q) := \Pi\big((p,q) + (Aq, B^T p);\Omega\big) \ \text{ with } \ \Omega = \Omega_1 \times \Omega_2$$

and employing another elementary projection property from Proposition 1.79 allow us to conclude that the mapping F is continuous and the set Ω is nonempty, compact, and convex. By the classical Brouwer fixed point theorem (see, e.g., [33]), the mapping F has a *fixed point* $(\bar{p},\bar{q}) \in \Omega$, which satisfies (3.25), and thus it is a Nash equilibrium of the game. $\square$

3.10 EXERCISES FOR CHAPTER 3

Exercise 3.1 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function finite around $\bar{x}$. Show that f is Gâteaux differentiable at $\bar{x}$ with $f'_G(\bar{x}) = v$ if and only if the directional derivative $f'(\bar{x}; d)$ exists and $f'(\bar{x}; d) = \langle v, d \rangle$ for all $d \in \mathbb{R}^n$.

Exercise 3.2 Let Ω be a nonempty, closed, convex subset of $\mathbb{R}^n$.

(i) Prove that the function $f(x) := [d(x; \Omega)]^2$ is differentiable on $\mathbb{R}^n$ and

$$\nabla f(\bar{x}) = 2[\bar{x} - \Pi(\bar{x}; \Omega)]$$

for every $\bar{x} \in \mathbb{R}^n$.

(ii) Prove that the function $g(x) := d(x; \Omega)$ is differentiable on $\mathbb{R}^n \setminus \Omega$.

Exercise 3.3 Let F be a nonempty, compact, convex subset of $\mathbb{R}^n$ and let $A : \mathbb{R}^n \to \mathbb{R}^p$ be a linear mapping. Define

$$f(x) := \sup_{u \in F} \big(\langle Ax, u \rangle - \frac{1}{2} \|u\|^2 \big).$$

Prove that f is convex and differentiable.

Exercise 3.4 Give an example of a function $f : \mathbb{R}^n \to \mathbb{R}$, which is Gâteaux differentiable but not Fréchet differentiable. Can such a function be convex?

Exercise 3.5 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function and let $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$. Show that if $\dfrac{\partial f}{\partial x_i}$ at $\bar{x}$ exists for all $i = 1, \ldots, n$, then f is (Fréchet) differentiable at $\bar{x}$.

Exercise 3.6 Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable on $\mathbb{R}^n$. Show that f is strictly convex if and only if it satisfies the condition

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle < f(x) - f(\bar{x}) \ \text{ for all } \ x \in \mathbb{R}^n \ \text{ and } \ \bar{x} \in \mathbb{R}^n, \ x \neq \bar{x}.$$

Exercise 3.7 Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable (C^2) on $\mathbb{R}^n$. Prove that if the Hessian $\nabla^2 f(\bar{x})$ is positive definite for all $\bar{x} \in \mathbb{R}^n$, then f is strictly convex. Give an example showing that the converse is not true in general.

Exercise 3.8 A convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is called *strongly convex* with modulus $c > 0$ if for all $x_1, x_2 \in \mathbb{R}^n$ and $\lambda \in (0, 1)$ we have

$$f\big(\lambda x_1 + (1-\lambda)x_2\big) \le \lambda f(x_1) + (1-\lambda) f(x_2) - \frac{1}{2} c \lambda (1-\lambda) \|x_1 - x_2\|^2.$$

Prove that the following are equivalent for any C^2 convex function $f : \mathbb{R}^n \to \mathbb{R}$:

(i) f strongly convex.

(ii) $f - \frac{c}{2}\|\cdot\|^2$ is convex.

(iii) $\langle \nabla^2 f(x)d, d\rangle \geq c\|d\|^2$ for all $x, d \in \mathbb{R}^n$.

Exercise 3.9 A set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is said to be *monotone* if

$$\langle v_2 - v_1, x_2 - x_1\rangle \geq 0 \text{ whenever } v_i \in F(x_i),\ i = 1, 2.$$

F is called *strictly monotone* if

$$\langle v_2 - v_1, x_2 - x_1\rangle > 0 \text{ whenever } v_i \in F(x_i),\ x_1 \neq x_2,\ i = 1, 2.$$

It is called *strongly monotone* with modulus $\ell > 0$ if

$$\langle v_2 - v_1, x_2 - x_1\rangle \geq \ell\|x_1 - x_2\|^2 \text{ whenever } v_i \in F(x_i),\ i = 1, 2.$$

(i) Prove that for any convex function $f : \mathbb{R}^n \to \mathbb{R}$ the mapping $F(x) := \partial f(x)$ is monotone.

(ii) Furthermore, for any convex function $f : \mathbb{R}^n \to \mathbb{R}$ the subdifferential mapping $\partial f(x)$ is strictly monotone if and only if f is strictly convex, and it is strongly monotone if and only if f is strongly convex.

Exercise 3.10 Prove Corollary 3.7.

Exercise 3.11 Let A be an $p \times n$ matrix and let $b \in \mathbb{R}^p$. Use the Farkas lemma to show that the system $Ax = b, x \geq 0$ has a feasible solution $x \in \mathbb{R}^n$ if and only if the system $A^T y \leq 0, b^T y > 0$ has no feasible solution $y \in \mathbb{R}^p$.

Exercise 3.12 Deduce Carathéodory's theorem from Radon's theorem and its proof.

Exercise 3.13 Use Carathéodory's theorem to prove Helly's theorem.

Exercise 3.14 Let C be a subspace of $\mathbb{R}^n$. Show that the polar of C is represented by

$$C^\circ = C^\perp := \big\{v \in \mathbb{R}^n \;\big|\; \langle v, c\rangle = 0 \text{ for all } c \in C\big\}.$$

Exercise 3.15 Let C be a nonempty, convex cone of $\mathbb{R}^n$. Show that

$$C^\circ = \big\{v \in \mathbb{R}^n \;\big|\; \langle v, c\rangle \leq 1 \text{ for all } c \in C\big\}.$$

Exercise 3.16 Let Ω_1 and Ω_2 be convex sets with $\operatorname{int}\Omega_1 \cap \Omega_2 \neq \emptyset$. Show that

$$T(\bar{x}; \Omega_1 \cap \Omega_2) = T(\bar{x}; \Omega_1) \cap T(\bar{x}; \Omega_2) \text{ for any } \bar{x} \in \Omega_1 \cap \Omega_2.$$

Exercise 3.17 Let A be $p \times n$ matrix, let

$$\Omega := \{x \in \mathbb{R}^n \mid Ax = 0,\ x \geq 0\},$$

and let $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_n) \in \Omega$. Verify the tangent and normal cone formulas:

$$T(\bar{x};\Omega) = \{u = (u_1, \ldots, u_n) \in \mathbb{R}^n \mid Au = 0,\ u_i \geq 0 \text{ for any } i \text{ with } \bar{x}_i = 0\},$$

$$N(\bar{x};\Omega) = \{v \in \mathbb{R}^n \mid v = A^* y - z,\ y \in \mathbb{R}^p,\ z \in \mathbb{R}^n,\ z_i \geq 0,\ \langle z, \bar{x}\rangle = 0\}.$$

Exercise 3.18 Give an example of a closed, convex set $\Omega \subset \mathbb{R}^2$ with a point $\bar{x} \in \Omega$ such that the cone $\mathbb{R}^+(\Omega - \bar{x})$ is not closed.

Exercise 3.19 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function such that there is $\ell \geq 0$ satisfying $\partial f(x) \subset \ell I\!B$. Show that f is ℓ–Lipschitz continuous on $\mathbb{R}^n$.

Exercise 3.20 Let $f : \mathbb{R} \to \mathbb{R}$ be convex. Prove that f is nondecreasing if and only if $\partial f(x) \subset [0, \infty)$ for all $x \in \mathbb{R}$.

Exercise 3.21 Find the horizon cones of the following sets:

(i) $F := \{(x, y) \in \mathbb{R}^2 \mid y \geq x^2\}$.
(ii) $F := \{(x, y) \in \mathbb{R}^2 \mid y \geq |x|\}$.
(iii) $F := \{(x, y) \in \mathbb{R}^2 \mid x > 0,\ y \geq 1/x\}$.
(iv) $F := \{(x, y) \in \mathbb{R}^2 \mid x > 0,\ y \geq \sqrt{x^2 + 1}\}$.

Exercise 3.22 Let $v_i \in \mathbb{R}^n$ and $b_i > 0$ for $i = 1, \ldots, m$. Consider the set

$$F := \{x \in \mathbb{R}^n \mid \langle v_i, x\rangle \leq b_i \text{ for all } i = 1, \ldots, m\}. \tag{3.26}$$

Prove the following representation of the horizon cone and Minkowski gauge for F:

$$F_\infty = \{x \in \mathbb{R}^n \mid \langle v_i, x\rangle \leq 0 \text{ for all } i = 1, \ldots, m\},$$

$$\rho_F(u) = \max\Big\{0, \max_{i \in \{1,\ldots,m\}} \frac{\langle v_i, u\rangle}{b_i}\Big\}.$$

Exercise 3.23 Let A, B be nonempty, closed, convex subsets of $\mathbb{R}^n$.

(i) Show that $(A \times B)_\infty = A_\infty \times B_\infty$.
(ii) Find conditions ensuring that $(A + B)_\infty = A_\infty + B_\infty$.

Exercise 3.24 Let $F \neq \emptyset$ be a closed, bounded, convex set and let $\Omega \neq \emptyset$ be a closed set in $\mathbb{R}^n$. Prove that $\mathcal{T}^F_\Omega(x) = 0$ if and only if $x \in \Omega$. This means that the condition $0 \in F$ in Theorem 3.27(i) is not required in this case.

Exercise 3.25 Let $F \neq \emptyset$ be a closed, bounded, convex set and let $\Omega \neq \emptyset$ be a closed, convex set. Consider the sets C^* and S^* defined in Theorem 3.37 and Theorem 3.38, respectively. Prove the following:

(i) If $\bar{x} \in \Omega$, then
$$\partial\mathcal{T}^F_\Omega(\bar{x}) = N(\bar{x};\Omega) \cap C^*.$$

(ii) If $\bar{x} \notin \Omega$, then
$$\partial\mathcal{T}^F_\Omega(\bar{x}) = N(\bar{x};\Omega_r) \cap S^*,\ r := \mathcal{T}^F_\Omega(\bar{x}) > 0.$$

(iii) If $\bar{x} \notin \Omega$, then
$$\partial\mathcal{T}^F_\Omega(\bar{x}) = \big[-\partial\rho_F(\bar{\omega}-\bar{x})\big] \cap N(\bar{\omega};\Omega)$$

for any $\bar{\omega} \in \Pi_F(\bar{x};\Omega)$

Exercise 3.26 Give a direct proof of characterizations (3.24) of Nash equilibrium.

Exercise 3.27 In Example 3.43 with mixed strategies, find $(\bar{p}, \bar{q})$ that yields a Nash equilibrium in this game.

CHAPTER 4

Applications to Optimization and Location Problems

In this chapter we give some applications of the results of convex analysis from Chapters 1–3 to selected problems of convex optimization and facility location. Applications to optimization problems are mainly classical, being presented however in a simplified and unified way. On the other hand, applications to location problems are mostly based on recent journal publications and have not been systematically discussed from the viewpoint of convex analysis in monographs and textbooks.

4.1 LOWER SEMICONTINUITY AND EXISTENCE OF MINIMIZERS

We first discuss the notion of lower semicontinuity for extended-real-valued functions, which plays a crucial role in justifying the existence of absolute minimizers and related topics. Note that most of the results presented below do not require the convexity of f.

Definition 4.1 *An extended-real-valued function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is* LOWER SEMICONTINUOUS *(l.s.c.)* AT $\bar{x} \in \mathbb{R}^n$ *if for every $\alpha \in \mathbb{R}$ with $f(\bar{x}) > \alpha$ there is $\delta > 0$ such that*

$$f(x) > \alpha \ \text{ for all } \ x \in I\!B(\bar{x}; \delta). \tag{4.1}$$

We simply say that f is LOWER SEMICONTINUOUS *if it is l.s.c. at every point of $\mathbb{R}^n$.*

The following examples illustrate the validity or violation of lower semicontinuity.

Example 4.2 **(i)** Define the function $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) := \begin{cases} x^2 & \text{if } x \neq 0, \\ -1 & \text{if } x = 0. \end{cases}$$

It is easy to see that it is l.s.c. at $\bar{x} = 0$ and in fact on the whole real line. On the other hand, the replacement of $f(0) = -1$ by $f(0) = 1$ destroys the lower semicontinuity at $\bar{x} = 0$.

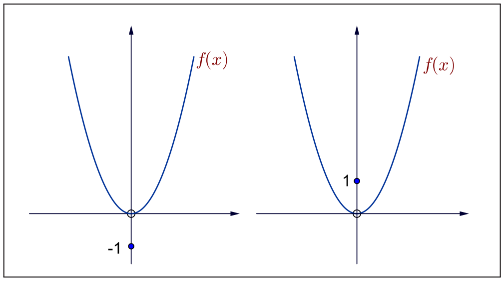

Figure 4.1: Which function is l.s.c?

(ii) Consider the indicator function of the set $[-1, 1] \subset \mathbb{R}$ given by

$$f(x) := \begin{cases} 0 & \text{if } |x| \le 1, \\ \infty & \text{otherwise.} \end{cases}$$

Then f is a lower semicontinuous function on $\mathbb{R}$. However, the small modification

$$g(x) := \begin{cases} 0 & \text{if } |x| < 1, \\ \infty & \text{otherwise} \end{cases}$$

violates the lower semicontinuity on $\mathbb{R}$. In fact, g is not l.s.c. at $x = \pm 1$ while it is lower semicontinuous at any other point of the real line.

The next four propositions present useful characterizations of lower semicontinuity.

Proposition 4.3 *Let $f: \mathbb{R}^n \to \overline{\mathbb{R}}$ and let $\bar{x} \in \operatorname{dom} f$. Then f is l.s.c. at $\bar{x}$ if and only if for any $\epsilon > 0$ there is $\delta > 0$ such that*

$$f(\bar{x}) - \epsilon < f(x) \quad \textit{whenever} \quad x \in I\!B(\bar{x}; \delta). \tag{4.2}$$

Proof. Suppose that f is l.s.c. at $\bar{x}$. Since $\lambda := f(\bar{x}) - \epsilon < f(\bar{x})$, there is $\delta > 0$ with

$$f(\bar{x}) - \epsilon = \lambda < f(x) \quad \text{for all} \quad x \in I\!B(\bar{x}; \delta).$$

Conversely, suppose that for any $\epsilon > 0$ there is $\delta > 0$ such that (4.2) holds. Fix any $\lambda < f(\bar{x})$ and choose $\epsilon > 0$ for which $\lambda < f(\bar{x}) - \epsilon$. Then

$$\lambda < f(\bar{x}) - \epsilon < f(x) \quad \text{whenever} \quad x \in I\!B(\bar{x}; \delta)$$

for some $\delta > 0$, which completes the proof. $\square$

Proposition 4.4 *The function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is l.s.c. if and only if for any number $\lambda \in \mathbb{R}$ the level set $\{x \in \mathbb{R}^n \mid f(x) \le \lambda\}$ is closed.*

Proof. Suppose that f is l.s.c. For any fix $\lambda \in \mathbb{R}$, denote

$$\Omega := \{x \in \mathbb{R}^n \mid f(x) \le \lambda\}.$$

If $\bar{x} \notin \Omega$, we have $f(\bar{x}) > \lambda$. By definition, find $\delta > 0$ such that

$$f(x) > \lambda \text{ for all } x \in I\!B(\bar{x};\delta),$$

which shows that $I\!B(\bar{x};\delta) \subset \Omega^c$, the complement of Ω. Thus Ω^c is open, and so Ω is closed.

To verify the converse implication, suppose that the set $\{x \in \mathbb{R}^n \mid f(x) \le \lambda\}$ is closed for any $\lambda \in \mathbb{R}$. Fix $\bar{x} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$ with $f(\bar{x}) > \lambda$ and then let

$$\Theta := \{x \in \mathbb{R}^n \mid f(x) \le \lambda\}.$$

Since Θ is closed and $\bar{x} \notin \Theta$, there is $\delta > 0$ such that $I\!B(\bar{x};\delta) \subset \Theta^c$. This implies

$$f(x) > \lambda \text{ for all } x \in I\!B(\bar{x};\delta)$$

and thus completes the proof of the proposition. □

Proposition 4.5 *The function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is l.s.c. if and only if its epigraph is closed.*

Proof. To verify the "only if" part, fix $(\bar{x}, \lambda) \notin \operatorname{epi} f$ meaning $\lambda < f(\bar{x})$. For any $\epsilon > 0$ with $f(\bar{x}) > \lambda + \epsilon > \lambda$, there is $\delta > 0$ such that $f(x) > \lambda + \epsilon$ whenever $x \in I\!B(\bar{x};\delta)$. Thus

$$I\!B(\bar{x};\delta) \times (\lambda - \epsilon, \lambda + \epsilon) \subset (\operatorname{epi} f)^c,$$

and hence the set epi f is closed.

Conversely, suppose that epi f is closed. Employing Proposition 4.4, it suffices to verify that for any $\lambda \in \mathbb{R}$ the set $\Omega := \{x \in \mathbb{R}^n \mid f(x) \le \lambda\}$ is closed. To see this, fix any sequence $\{x_k\} \subset \Omega$ that converges to $\bar{x}$. Since $f(x_k) \le \lambda$, we have $(x_k, \lambda) \in \operatorname{epi} f$ for all k. It follows from $(x_k, \lambda) \to (\bar{x}, \lambda)$ that $(\bar{x}, \lambda) \in \operatorname{epi} f$, and so $f(\bar{x}) \le \lambda$. Thus $\bar{x} \in \Omega$, which justifies the closedness of the set Ω. □

Proposition 4.6 *The function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is l.s.c. at $\bar{x}$ if and only if for any sequence $x_k \to \bar{x}$ we have* $\liminf_{k\to\infty} f(x_k) \ge f(\bar{x})$.

Proof. If f is l.s.c. at $\bar{x}$, take a sequence of $x_k \to \bar{x}$. For any $\lambda < f(\bar{x})$, there exists $\delta > 0$ such that (4.1) holds. Then $x_k \in I\!B(\bar{x};\delta)$, and so $\lambda < f(x_k)$ for large k. It follows that $\lambda \le \liminf_{k\to\infty} f(x_k)$, and hence $f(\bar{x}) \le \liminf_{k\to\infty} f(x_k)$ by the arbitrary choice of $\lambda < f(\bar{x})$.

To verify the converse, suppose by contradiction that f is not l.s.c. at $\bar{x}$. Then there is $\lambda < f(\bar{x})$ such that for every $\delta > 0$ we have $x_\delta \in I\!B(\bar{x};\delta)$ with $\lambda \ge f(x_\delta)$. Applying this to $\delta_k := \dfrac{1}{k}$ gives us a sequence $\{x_k\} \subset \mathbb{R}^n$ converging to $\bar{x}$ with $\lambda \ge f(x_k)$. It yields

$$f(\bar{x}) > \lambda \ge \liminf_{k\to\infty} f(x_k),$$

which is a contradiction that completes the proof. □

Now we are ready to obtain a significant result on the lower semicontinuity of the minimal value function (3.9), which is useful in what follows.

Theorem 4.7 In the setting of Theorem 3.27 with $0 \in F$ we have that the minimal time function $\mathcal{T}^F_\Omega$ is lower semicontinuous.

Proof. Pick $\bar{x} \in \mathbb{R}^n$ and fix an arbitrary sequence $\{x_k\}$ converging to $\bar{x}$ as $k \to \infty$. Based on Proposition 4.6, we check that

$$\liminf_{k\to\infty} \mathcal{T}^F_\Omega(x_k) \ge \mathcal{T}^F_\Omega(\bar{x}).$$

The above inequality obviously holds if $\liminf_{k\to\infty} \mathcal{T}^F_\Omega(x_k) = \infty$; thus it remains to consider the case of $\liminf_{k\to\infty} \mathcal{T}^F_\Omega(x_k) = \gamma \in [0,\infty)$. Since it suffices to show that $\gamma \ge \mathcal{T}^F_\Omega(\bar{x})$, we assume without loss of generality that $\mathcal{T}^F_\Omega(x_k) \to \gamma$. By the definition of the minimal time function, we find a sequence $\{t_k\} \subset [0,\infty)$ such that

$$\mathcal{T}^F_\Omega(x_k) \le t_k < \mathcal{T}^F_\Omega(x_k) + \frac{1}{k} \quad \text{and} \quad (x_k + t_k F) \cap \Omega \ne \emptyset \quad \text{as} \quad k \in \mathbb{N}.$$

For each $k \in \mathbb{N}$, fix $f_k \in F$ and $\omega_k \in \Omega$ with $x_k + t_k f_k = \omega_k$ and suppose without loss of generality that $\omega_k \to \omega \in \Omega$. Consider the two cases: $\gamma > 0$ and $\gamma = 0$. If $\gamma > 0$, then

$$f_k \to \frac{\omega - \bar{x}}{\gamma} \in F \quad \text{as} \quad k \to \infty,$$

which implies that $(\bar{x} + \gamma F) \cap \Omega \ne \emptyset$, and hence $\gamma \ge \mathcal{T}^F_\Omega(\bar{x})$. Consider the other case where $\gamma = 0$. In this case the sequence $\{t_k\}$ converges to 0, and so $t_k f_k \to \omega - \bar{x}$. It follows from Proposition 3.24 that $\omega - \bar{x} \in F_\infty$, which yields $\bar{x} \in \Omega - F_\infty$. Employing Theorem 3.27 tells us that $\mathcal{T}^F_\Omega(\bar{x}) = 0$ and also $\gamma \ge \mathcal{T}^F_\Omega(\bar{x})$. This verifies the lower semicontinuity of $\mathcal{T}^F_\Omega$ in setting of Theorem 3.27. □

The following two propositions indicate two cases of preserving lower semicontinuity under important operations on functions.

Proposition 4.8 *Let $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$, $i = 1, \ldots, m$, be l.s.c. at some point $\bar{x}$. Then the sum $\sum_{i=1}^m f_i$ is l.s.c. at this point.*

Proof. It suffices to verify it for $m = 2$. Taking $x_k \to \bar{x}$, we get

$$\liminf_{k\to\infty} f_i(x_k) \geq f(\bar{x}) \quad \text{for} \quad i = 1, 2,$$

which immediately implies that

$$\liminf_{k\to\infty}\big[f_1(x_k) + f_2(x_k)\big] \geq \liminf_{k\to\infty} f_1(x_k) + \liminf_{k\to\infty} f_2(x_k) \geq f_1(\bar{x}) + f_2(\bar{x}),$$

and thus $f_1 + f_2$ is l.s.c. at $\bar{x}$ by Proposition 4.6. $\square$

Proposition 4.9 *Let I be any nonempty index set and let $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ be l.s.c. for all $i \in I$. Then the supremum function $f(x) := \sup_{i\in I} f_i(x)$ is also lower semicontinuous.*

Proof. For any number $\lambda \in \mathbb{R}$, we have directly from the definition that

$$\big\{x \in \mathbb{R}^n \bigm| f(x) \leq \lambda\big\} = \big\{x \in \mathbb{R}^n \bigm| \sup_{i\in I} f_i(x) \leq \lambda\big\} = \bigcap_{i\in I}\big\{x \in \mathbb{R}^n \bigm| f_i(x) \leq \lambda\big\}.$$

Thus all the level sets of f are closed as intersections of closed sets. This insures the lower semicontinuity of f by Proposition 4.4. $\square$

The next result provides two interrelated *unilateral* versions of the classical Weierstrass existence theorem presented in Theorem 1.18. The crucial difference is that now we assume the *lower semicontinuity* versus continuity while ensuring the existence of only *minima*, not together with maxima as in Theorem 1.18.

Theorem 4.10 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a l.s.c. function. The following hold:

(i) The function f attains its absolute minimum on any nonempty, compact subset $\Omega \subset \mathbb{R}^n$ that intersects its domain.

(ii) Assume that $\inf\{f(x) \mid x \in \mathbb{R}^n\} < \infty$ and that there is a real number $\lambda > \inf\{f(x) \mid x \in \mathbb{R}^n\}$ for which the level set $\{x \in \mathbb{R}^n \mid f(x) < \lambda\}$ of f is bounded. Then f attains its absolute minimum on $\mathbb{R}^n$ at some point $\bar{x} \in \operatorname{dom} f$.

Proof. To verify **(i)**, show first that f is bounded below on Ω. Suppose on the contrary that for every $k \in \mathbb{N}$ there is $x_k \in \Omega$ such that $f(x_k) \le -k$. Since Ω is compact, we get without loss of generality that $x_k \to \bar{x} \in \Omega$. It follows from the lower semicontinuity that

$$f(\bar{x}) \le \liminf_{k\to\infty} f(x_k) = -\infty,$$

which is a contradiction showing that f is bounded below on Ω. To justify now assertion **(i)**, observe that $\gamma := \inf_{x\in\Omega} f(x) < \infty$. Take $\{x_k\} \subset \Omega$ such that $f(x_k) \to \gamma$. By the compactness of Ω, suppose that $x_k \to \bar{x} \in \Omega$ as $k \to \infty$. Then

$$f(\bar{x}) \le \liminf_{k\to\infty} f(x_k) = \gamma,$$

which shows that $f(\bar{x}) = \gamma$, and so $\bar{x}$ realizes the absolute minimum of f on Ω.

It remains to prove assertion **(ii)**. It is not hard to show that f is bounded below, so $\gamma := \inf_{x\in\mathbb{R}^n} f(x)$ is a real number. Let $\{x_k\} \subset \mathbb{R}^n$ for which $f(x_k) \to \gamma$ as $k \to \infty$ and let $k_0 \in \mathbb{N}$ be such that $f(x_k) < \lambda$ for all $k \ge k_0$. Since the level set $\{x \in \mathbb{R}^n \mid f(x) < \lambda\}$ is bounded, we select a subsequence of $\{x_k\}$ (no relabeling) that converges to some $\bar{x}$. Then

$$f(\bar{x}) \le \liminf_{k\to\infty} f(x_k) \le \gamma,$$

which shows that $\bar{x} \in \operatorname{dom} f$ is a minimum point of f on $\mathbb{R}^n$. $\square$

In contrast to all the results given above in this section, the next theorem requires *convexity*. It provides effective characterizations of the level set boundedness imposed in Theorem 4.10(ii). Note that condition **(iv)** of the theorem below is known as *coercivity*.

Theorem 4.11 Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex and let $\inf\{f(x) \mid x \in \mathbb{R}^n\} < \infty$. The following assertions are equivalent:

(i) There exists a real number $\beta > \inf\{f(x) \mid x \in \mathbb{R}^n\}$ for which the level set $\mathcal{L}_\beta := \{x \in \mathbb{R}^n \mid f(x) < \beta\}$ is bounded.
(ii) All the level sets $\{x \in \mathbb{R}^n \mid f(x) < \lambda\}$ of f are bounded.
(iii) $\lim_{\|x\|\to\infty} f(x) = \infty$.
(iv) $\liminf_{\|x\|\to\infty} \dfrac{f(x)}{\|x\|} > 0$.

Proof. Let $\gamma := \inf\{f(x) \mid x \in \mathbb{R}^n\}$, which is a real number or $-\infty$. We first verify **(i)**$\Longrightarrow$**(ii)**. Suppose by contradiction that there is $\lambda \in \mathbb{R}$ such that the level set $\mathcal{L}_\lambda$ is not bounded. Then there exists a sequence $\{x_k\} \subset \mathcal{L}_\lambda$ with $\|x_k\| \to \infty$. Fix $\alpha \in (\gamma, \beta)$, $\bar{x} \in \mathcal{L}_\alpha$ and define the convex combination

$$y_k := \bar{x} + \frac{1}{\sqrt{\|x_k\|}}\Big(x_k - \bar{x}\Big) = \frac{1}{\sqrt{\|x_k\|}}x_k + \Big(1 - \frac{1}{\sqrt{\|x_k\|}}\Big)\bar{x}.$$

By the convexity of f, we have the relationships

$$f(y_k) \le \frac{1}{\sqrt{\|x_k\|}} f(x_k) + \Big(1 - \frac{1}{\sqrt{\|x_k\|}}\Big) f(\bar{x}) < \frac{\lambda}{\sqrt{\|x_k\|}} + \Big(1 - \frac{1}{\sqrt{\|x_k\|}}\Big)\alpha \to \alpha < \beta,$$

and hence $y_k \in \mathcal{L}_\beta$ for k sufficiently large, which ensures the boundedness of $\{y_k\}$. It follows from the triangle inequality that

$$\|y_k\| \ge \Big\| \frac{x_k}{\sqrt{\|x_k\|}} \Big\| - \Big(1 - \frac{1}{\sqrt{\|x_k\|}}\Big)\|\bar{x}\| \ge \sqrt{\|x_k\|} - \Big(1 - \frac{1}{\sqrt{\|x_k\|}}\Big)\|\bar{x}\| \to \infty,$$

which is a contradiction. Thus **(ii)** holds.

Next we show that **(ii)**$\Longrightarrow$**(iii)**. Indeed, the violation of **(iii)** yields the existence of a constant $M > 0$ and a sequence $\{x_k\}$ such that $\|x_k\| \to \infty$ as $k \to \infty$ and $f(x_k) < M$ for all k. But this obviously contradicts the boundedness of the level set $\{x \in \mathbb{R}^n \mid f(x) < M\}$.

Finally, we justify **(iii)**$\Longrightarrow$**(i)** leaving the proof of **(iii)**$\Longleftrightarrow$**(iv)** as an exercise. Suppose that **(iii)** is satisfied and for $M > \gamma$, find α such that $f(x) \ge M$ whenever $\|x\| > \alpha$. Then

$$\mathcal{L}_M := \big\{x \in \mathbb{R}^n \;\big|\; f(x) < M\big\} \subset I\!B(0;\alpha),$$

which shows that this set is bounded and thus verifies **(i)**. □

4.2 OPTIMALITY CONDITIONS

In this section we first derive necessary and sufficient optimality conditions for optimal solutions to the following *set-constrained* convex optimization problem:

$$\text{minimize} \;\; f(x) \;\; \text{subject to} \;\; x \in \Omega, \tag{4.3}$$

where Ω is a *nonempty, convex set* and $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a *convex function*.

Definition 4.12 *We say that $\bar{x} \in \Omega$ is a* LOCAL OPTIMAL SOLUTION *to* (4.3) *if $f(\bar{x}) < \infty$ and there is $\gamma > 0$ such that*

$$f(x) \ge f(\bar{x}) \;\; \text{for all} \;\; x \in I\!B(\bar{x};\gamma) \cap \Omega. \tag{4.4}$$

If $f(x) \ge f(\bar{x})$ for all $x \in \Omega$, then $\bar{x} \in \Omega$ is a GLOBAL OPTIMAL SOLUTION *to* (4.3).

The next result shows, in particular, that there is *no difference* between local and global solutions for problems of convex optimization.

Proposition 4.13 *The following are equivalent in the convex setting of* (4.3):

(i) *$\bar{x}$ is a local optimal solution to the constrained problem* (4.3).
(ii) *$\bar{x}$ is a local optimal solution to the unconstrained problem*

$$\textit{minimize} \;\; g(x) := f(x) + \delta(x;\Omega) \;\; \textit{on} \;\; \mathbb{R}^n. \tag{4.5}$$

(iii) $\bar{x}$ *is a global optimal solution to the unconstrained problem* (4.5).
(iv) $\bar{x}$ *is a global optimal solution to the constrained problem* (4.3).

Proof. To verify **(i)**$\Longrightarrow$**(ii)**, select $\gamma > 0$ such that (4.4) is satisfied. Since $g(x) = f(x)$ for $x \in \Omega \cap I\!B(\bar{x}; \gamma)$ and since $g(x) = \infty$ for $x \in I\!B(\bar{x}; \gamma)$ and $x \notin \Omega$, we have

$$g(x) \geq g(\bar{x}) \quad \text{for all} \quad x \in I\!B(\bar{x}; \gamma),$$

which means **(ii)**. Implication **(ii)**$\Longrightarrow$**(iii)** is a consequence of Proposition 2.33. The verification of **(iii)**$\Longrightarrow$**(iv)** is similar to that of **(i)**$\Longrightarrow$**(ii)**, and **(vi)**$\Longrightarrow$**(i)** is obvious. □

In what follows we use the term "optimal solutions" for problems of convex optimization, without mentioning global or local. Observe that the objective function g of the unconstrained problem (4.5) associated with the constrained one (4.3) is always *extended-real-valued* when $\Omega \neq \mathbb{R}^n$. This reflects the "infinite penalty" for the constraint violation.

The next theorem presents basic *necessary and sufficient* optimality conditions for optimal solutions to the constrained problem (4.3) with a general (convex) cost function.

Theorem 4.14 Consider the convex problem (4.3) with $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ such that $f(\bar{x}) < \infty$ for some point $\bar{x} \in \Omega$ under the qualification condition

$$\partial^\infty f(\bar{x}) \cap \big[-N(\bar{x}; \Omega) \big] = \{0\}, \tag{4.6}$$

which automatically holds if f is continuous at $\bar{x}$. Then the following are equivalent:

(i) $\bar{x}$ is an optimal solution to (4.3).
(ii) $0 \in \partial f(\bar{x}) + N(\bar{x}; \Omega)$, i.e., $\partial f(\bar{x}) \cap [-N(\bar{x}; \Omega)] \neq \emptyset$.
(iii) $f'(\bar{x}; d) \geq 0$ for all $d \in T(\bar{x}; \Omega)$.

Proof. Observe first that $\partial^\infty f(\bar{x}) = \{0\}$ and thus (4.6) holds if f is continuous at $\bar{x}$. This follows from Theorem 2.29. To verify now **(i)**$\Longrightarrow$**(ii)**, we use the unconstrained form (4.5) of problem (4.3) due to Proposition 4.13. Then Proposition 2.35 provides the optimal solution characterization $0 \in \partial g(\bar{x})$. Applying the subdifferential sum rule from Theorem 2.44 under the qualification condition (4.6) gives us the optimality condition in **(ii)**.

To show that **(ii)**$\Longrightarrow$**(iii)**, we have by the tangent-normal duality (3.4) the equivalence

$$d \in T(\bar{x}; \Omega) \iff \big[\langle v, d \rangle \leq 0 \quad \text{for all} \quad v \in N(\bar{x}; \Omega) \big].$$

Fix an element $w \in \partial f(\bar{x})$ with $-w \in N(\bar{x}; \Omega)$. Then $\langle w, d \rangle \geq 0$ for all $d \in T(\bar{x}; \Omega)$ and we arrive at $f'(\bar{x}; d) \geq \langle w, d \rangle \geq 0$ by Theorem 2.84, achieving **(iii)** in this way.

It remains to justify implication **(iii)**$\Longrightarrow$**(i)**. It follows from Lemma 2.83 that

$$f(\bar{x} + d) - f(\bar{x}) \geq f'(\bar{x}; d) \geq 0 \quad \text{for all} \quad d \in T(\bar{x}; \Omega).$$

By definition of the tangent cone, we have $T(\bar{x};\Omega) = \text{cl}\big[\text{cone}(\Omega - \bar{x})\big]$, which tells us that $d := w - \bar{x} \in T(\bar{x};\Omega)$ if $w \in \Omega$. This yields

$$f(w) - f(\bar{x}) = f(w - \bar{x} + \bar{x}) - f(\bar{x}) = f(\bar{x} + d) - f(\bar{x}) \geq 0,$$

which verifies **(i)** and completes the proof of the theorem. □

Corollary 4.15 *Assume that in the framework of Theorem 4.14 the cost function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is differentiable at $\bar{x} \in \text{dom}\, f$. Then the following are equivalent:*

(i) *$\bar{x}$ is an optimal solution to problem (4.3).*
(ii) $-\nabla f(\bar{x}) \in N(\bar{x};\Omega)$*, which reduces to* $\nabla f(\bar{x}) = 0$ *if* $\bar{x} \in \text{int}\,\Omega$.

Proof. Since the differentiability of the convex function f at $\bar{x}$ implies its continuity and so Lipschitz continuity around this point by Theorem 2.29. Since $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$ in this case by Theorem 3.3, we deduce this corollary directly from Theorem 4.14. □

Next we consider the following convex optimization problem containing functional *inequality constraints* given by finite *convex* functions, along with the set/geometric constraint as in (4.3). For simplicity, suppose that f is also finite on $\mathbb{R}^n$. Problem (P) is:

$$\begin{aligned}
&\text{minimize} \quad f(x)\\
&\text{subject to } g_i(x) \leq 0 \;\text{ for }\; i = 1,\ldots,m,\\
&x \in \Omega.
\end{aligned}$$

Define the *Lagrange function* for (P) involving only the cost and functional constraints by

$$L(x,\lambda) := f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

with $\lambda_i \in \mathbb{R}$ and the *active index set* given by $I(x) := \{i \in \{1,\ldots,m\} \mid g_i(x) = 0\}$.

Theorem 4.16 Let $\bar{x}$ be an optimal solution to (P). Then there are multipliers $\lambda_0 \geq 0$ and $\lambda = (\lambda_1,\ldots,\lambda_m) \in \mathbb{R}^m$, not equal to zero simultaneously, such that $\lambda_i \geq 0$ as $i = 0,\ldots,m$,

$$0 \in \lambda_0 \partial f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \partial g_i(\bar{x}) + N(\bar{x};\Omega),$$

and $\lambda_i g_i(\bar{x}) = 0$ for all $i = 1,\ldots,m$.

Proof. Consider the finite convex function

$$\varphi(x) := \max\left\{ f(x) - f(\bar{x}),\ g_i(x) \;\middle|\; i = 1, \ldots, m \right\}$$

and observe that $\bar{x}$ solves the following problem with no functional constraints:

$$\text{minimize } \varphi(x) \text{ subject to } x \in \Omega.$$

Since φ is finite and hence locally Lipschitz continuous, we get from Theorem 4.14 that $0 \in \partial\varphi(\bar{x}) + N(\bar{x}; \Omega)$. The subdifferential maximum rule from Proposition 2.54 yields

$$0 \in \operatorname{co}\Big[\partial f(\bar{x}) \cup \big[\bigcup_{i \in I(\bar{x})} \partial g_i(\bar{x})\big]\Big] + N(\bar{x}; \Omega).$$

This gives us $\lambda_0 \geq 0$ and $\lambda_i \geq 0$ for $i \in I(\bar{x})$ such that $\lambda_0 + \sum_{i \in I(\bar{x})} \lambda_i = 1$ and

$$0 \in \lambda_0 \partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \lambda_i \partial g_i(\bar{x}) + N(\bar{x}; \Omega). \tag{4.7}$$

Letting $\lambda_i := 0$ for $i \notin I(\bar{x})$, we get $\lambda_i g_i(\bar{x}) = 0$ for all $i = 1, \ldots, m$ with $(\lambda_0, \lambda) \neq 0$. □

Many applications including numerical algorithms to solve problem (P) require effective *constraint qualification conditions* that ensure $\lambda_0 \neq 0$ in Theorem 4.16. The following one is classical in convex optimization.

Definition 4.17 SLATER'S CONSTRAINT QUALIFICATION (CQ) *holds in* (P) *if there exists* $u \in \Omega$ *such that* $g_i(u) < 0$ *for all* $i = 1, \ldots, m$.

The next theorem provides necessary and sufficient conditions for convex optimization in the *qualified*, *normal*, or *Karush-Kuhn-Tucker (KKT) form*.

Theorem 4.18 Suppose that Slater's CQ holds in (P) and let $\bar{x}$ be a feasible solution. Then $\bar{x}$ is an optimal solution to (P) if and only if there exist nonnegative Lagrange multipliers $(\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^m$ such that

$$0 \in \partial f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \partial g_i(\bar{x}) + N(\bar{x}; \Omega) \tag{4.8}$$

and $\lambda_i g_i(\bar{x}) = 0$ for all $i = 1, \ldots, m$.

Proof. To verify the necessity part, we only need to show that $\lambda_0 \neq 0$ in (4.7). Suppose on the contrary that $\lambda_0 = 0$ and find $(\lambda_1, \ldots, \lambda_m) \neq 0$, $v_i \in \partial g_i(\bar{x})$, and $v \in N(\bar{x}; \Omega)$ satisfying

$$0 = \sum_{i=1}^{m} \lambda_i v_i + v.$$

This readily implies that

$$0 = \sum_{i=1}^{m} \lambda_i \langle v_i, x - \bar{x} \rangle + \langle v, x - \bar{x} \rangle \le \sum_{i=1}^{m} \lambda_i \big(g_i(x) - g_i(\bar{x})\big) = \sum_{i=1}^{m} \lambda_i g_i(x) \;\text{ for all }\; x \in \Omega,$$

which contradicts since the Slater condition implies $\sum_{i=1}^{m} \lambda_i g_i(u) < 0$.

To check the sufficiency of (4.8), take $v_0 \in \partial f(\bar{x})$, $v_i \in \partial g_i(\bar{x})$, and $v \in N(\bar{x}; \Omega)$ such that $0 = v_0 + \sum_{i=1}^{m} \lambda_i v_i + v$ with $\lambda_i g_i(\bar{x}) = 0$ and $\lambda_i \ge 0$ for all $i = 1, \ldots, m$. Then the optimality of $\bar{x}$ in (P) follows directly from the definitions of the subdifferential and normal cone. Indeed, for any $x \in \Omega$ with $g_i(x) \le 0$ for $i = 1, \ldots, m$ one has

$$\begin{aligned}
0 = \langle \lambda_i v_i + v, x - \bar{x} \rangle &= \sum_{i=1}^{m} \lambda_i \langle v_i, x - \bar{x} \rangle + \langle v, x - \bar{x} \rangle \\
&\le \sum_{i=1}^{m} \lambda_i [g_i(x) - g_i(\bar{x})] + f(x) - f(\bar{x}) \\
&= \sum_{i=1}^{m} \lambda_i g_i(x) + f(x) - f(\bar{x}) \le f(x) - f(\bar{x}),
\end{aligned}$$

which completes the proof. □

Finally in this section, we consider the convex optimization problem (Q) with inequality and linear *equality* constraints

$$\begin{aligned}
&\text{minimize } \; f(x) \\
&\text{subject to } \; g_i(x) \le 0 \;\text{ for }\; i = 1, \ldots, m, \\
&Ax = b,
\end{aligned}$$

where A is an $p \times n$ matrix and $b \in \mathbb{R}^p$.

Corollary 4.19 *Suppose that Slater's CQ holds for problem (Q) with the set $\Omega := \{x \in \mathbb{R}^n \mid Ax - b = 0\}$ in Definition 4.17 and let $\bar{x}$ be a feasible solution. Then $\bar{x}$ is an optimal solution to (Q) if and only if there exist nonnegative multipliers $\lambda_1, \ldots, \lambda_m$ such that*

$$0 \in \partial f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \partial g_i(\bar{x}) + \operatorname{im} A^*$$

and $\lambda_i g_i(\bar{x}) = 0$ for all $i = 1, \ldots, m$.

Proof. The result follows from Theorem 4.16 and Proposition 2.12, where the normal cone to the set Ω is calculated via the image of the adjoint operator A^*. □

The concluding result presents the classical version of the Lagrange multiplier rule for convex problems with differentiable data.

Corollary 4.20 *Let f and g_i for $i = 1, \ldots, m$ be differentiable at a feasible solution $\bar{x}$ and let $\{\nabla g_i(\bar{x}) \mid i \in I(\bar{x})\}$ be linearly independent. Then $\bar{x}$ is an optimal solution to problem (P) with $\Omega = \mathbb{R}^n$ if and only if there are nonnegative multipliers $\lambda_1, \ldots, \lambda_m$ such that*

$$0 \in \nabla f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\bar{x}) = L_x(\bar{x}, \lambda)$$

and $\lambda_i g_i(\bar{x}) = 0$ for all $i = 1, \ldots, m$.

Proof. In this case $N(\bar{x}; \Omega) = \{0\}$ and it follows from (4.7) that $\lambda_0 = 0$ contradicts the linear independence of the gradient elements $\{\nabla g_i(\bar{x}) \mid i \in I(\bar{x})\}$. □

4.3 SUBGRADIENT METHODS IN CONVEX OPTIMIZATION

This section is devoted to algorithmic aspects of convex optimization and presents two versions of the classical subgradient method: one for unconstrained and the other for constrained optimization problems. First we study the *unconstrained problem*:

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad x \in \mathbb{R}^n, \tag{4.9}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function. Let $\{\alpha_k\}$ as $k \in \mathbb{N}$ be a sequence of positive numbers. The subgradient algorithm generated by $\{\alpha_k\}$ is defined as follows. Given a starting point $x_1 \in \mathbb{R}^n$, consider the *iterative procedure*:

$$x_{k+1} := x_k - \alpha_k v_k \quad \text{with some} \quad v_k \in \partial f(x_k), \quad k \in \mathbb{N}. \tag{4.10}$$

In this section we assume that problem (4.9) *admits an optimal solution* and that f is *convex and Lipschitz continuous* on $\mathbb{R}^n$. Our main goal is to find conditions that ensure the *convergence of the algorithm*. We split the proof of the main result into several propositions, which are of their own interest. The first one gives us an important subgradient property of convex Lipschitzian functions.

Proposition 4.21 *Let $\ell \geq 0$ be a Lipschitz constant of f on $\mathbb{R}^n$. For any $x \in \mathbb{R}^n$, we have the subgradient estimate*

$$\|v\| \leq \ell \quad \textit{whenever} \quad v \in \partial f(x).$$

Proof. Fix $x \in \mathbb{R}^n$ and take any subgradient $v \in \partial f(x)$. It follows from the definition that

$$\langle v, u - x\rangle \le f(u) - f(x) \text{ for all } u \in \mathbb{R}^n .$$

Combining this with the Lipschitz property of f gives us

$$\langle v, u - x\rangle \le \ell \|u - x\|, \quad u \in \mathbb{R}^n,$$

which yields in turn that

$$\langle v, y + x - x\rangle \le \ell \|y + x - x\|, \text{ i.e. } \langle v, y\rangle \le \ell \|y\|, \quad y \in \mathbb{R}^n .$$

Using $y := v$, we arrive at the conclusion $\|v\| \le \ell$. □

Proposition 4.22 *Let $\ell \ge 0$ be a Lipschitz constant of f on $\mathbb{R}^n$. For the sequence $\{x_k\}$ generated by algorithm* (4.10), *we have*

$$\|x_{k+1} - x\|^2 \le \|x_k - x\|^2 - 2\alpha_k\big(f(x_k) - f(x)\big) + \alpha_k^2\ell^2 \text{ for all } x \in \mathbb{R}^n,\ k \in \mathbb{N}. \tag{4.11}$$

Proof. Fix $x \in \mathbb{R}^n$, $k \in \mathbb{N}$ and get from (4.10) that

$$\begin{aligned}\|x_{k+1} - x\|^2 &= \|x_k - \alpha_k v_k - x\|^2 = \|x_k - x - \alpha_k v_k\|^2\\ &= \|x_k - x\|^2 - 2\alpha_k\langle v_k, x_k - x\rangle + \alpha_k^2 \|v_k\|^2 .\end{aligned}$$

Since $v_k \in \partial f(x_k)$ and $\|v_k\| \le \ell$ by Proposition 4.21, it follows that

$$\langle v_k, x - x_k\rangle \le f(x) - f(x_k).$$

This gives us (4.11) and completes the proof of the proposition. □

The next proposition provides an estimate of closeness of iterations to the optimal value of (4.9) for an arbitrary sequence $\{\alpha_k\}$ in the algorithm (4.10). Denote

$$\overline{V} := \min\{f(x)\ \big|\ x \in \mathbb{R}^n\} \text{ and } V_k := \min\{f(x_1), \ldots, f(x_k)\}.$$

Proposition 4.23 *Let $\ell \ge 0$ be a Lipschitz constant of f on $\mathbb{R}^n$ and let S be the set of optimal solutions to* (4.9). *For all $k \in \mathbb{N}$, we have*

$$0 \le V_k - \overline{V} \le \frac{d(x_1; S)^2 + \ell^2 \sum_{i=1}^{k} \alpha_i^2}{2\sum_{i=1}^{k} \alpha_i}. \tag{4.12}$$

Proof. Substituting any $\bar{x} \in S$ into estimate (4.11) gives us

$$\|x_{k+1} - \bar{x}\|^2 \le \|x_k - \bar{x}\|^2 - 2\alpha_k\big(f(x_k) - f(\bar{x})\big) + \alpha_k^2\ell^2.$$

Since $V_k \le f(x_i)$ for all $i = 1, \ldots, k$ and $\overline{V} = f(\bar{x})$, we get

$$\begin{aligned}\|x_{k+1} - \bar{x}\|^2 &\le \|x_1 - \bar{x}\|^2 - 2\sum_{i=1}^{k}\alpha_i\big(f(x_i) - f(\bar{x})\big) + \ell^2\sum_{i=1}^{k}\alpha_i^2\\ &\le \|x_1 - \bar{x}\|^2 - 2\sum_{i=1}^{k}\alpha_i(V_k - \overline{V}) + \ell^2\sum_{i=1}^{k}\alpha_i^2.\end{aligned}$$

It follows from $\|x_{k+1} - \bar{x}\|^2 \ge 0$ that

$$2\sum_{i=1}^{k}\alpha_i\left(V_k - \overline{V}\right) \le \ \|x_1 - \bar{x}\|^2 + \ell^2\sum_{i=1}^{k}\alpha_i^2.$$

Since $V_k \ge \overline{V}$ for every k, we deduce the estimate

$$0 \le V_k - \overline{V} \le \frac{\|x_1 - \bar{x}\|^2 + \ell^2\sum_{i=1}^{k}\alpha_i^2}{2\sum_{i=1}^{k}\alpha_i}$$

and hence arrive at (4.12) since $\bar{x} \in S$ was chosen arbitrarily. $\square$

The corollary below shows the number of steps in order to achieve an appropriate error estimate for the optimal value in the case of *constant step sizes*.

Corollary 4.24 *Suppose that $\alpha_k = \epsilon$ for all $k \in \mathbb{N}$ in the setting of Proposition 4.23. Then there exists a positive constant C such that*

$$0 \le V_k - \overline{V} < \ell^2\epsilon \ \ \textit{whenever} \ \ k > C/\epsilon^2.$$

Proof. From (4.12) we readily get

$$0 \le V_k - \overline{V} \le \frac{a + k\ell^2\epsilon^2}{2k\epsilon} = \frac{a}{2k\epsilon} + \frac{\ell^2\epsilon}{2} \ \ \text{with} \ \ a := d(x_1; S)^2.$$

Since $a/(2k\epsilon) + \ell^2\epsilon/2 < \ell^2\epsilon$ as $k > a/(\ell^2\epsilon^2)$, the conclusion holds with $C := a/\ell^2$. $\square$

As a direct consequence of Proposition 4.23, we get our first convergence result.

Proposition 4.25 *In the setting of Proposition 4.23 suppose in addition that*

$$\sum_{k=1}^{\infty}\alpha_k = \infty \ \ \textit{and} \ \ \lim_{k\to\infty}\alpha_k = 0. \tag{4.13}$$

Then we have the convergence $V_k \to \overline{V}$ as $k \to \infty$.

Proof. Given any $\epsilon > 0$, choose $K \in \mathbb{N}$ such that $\alpha_i \ell^2 < \epsilon$ for all $i \geq K$. For any $k > K$, using (4.12) ensures the estimates

$$\begin{aligned} 0 \leq V_k - \overline{V} &\leq \frac{d(x_1;S)^2 + \ell^2 \sum\limits_{i=1}^{K} \alpha_i^2}{2\sum\limits_{i=1}^{k} \alpha_i} + \frac{\ell^2 \sum\limits_{i=K}^{k} \alpha_i^2}{2\sum\limits_{i=1}^{k} \alpha_i} \\ &\leq \frac{d(x_1;S)^2 + \ell^2 \sum\limits_{i=1}^{K} \alpha_i^2}{2\sum\limits_{i=1}^{k} \alpha_i} + \frac{\epsilon \sum\limits_{i=K}^{k} \alpha_i}{2\sum\limits_{i=1}^{k} \alpha_i} \\ &\leq \frac{d(x_1;S)^2 + \ell^2 \sum\limits_{i=1}^{K} \alpha_i^2}{2\sum\limits_{i=1}^{k} \alpha_i} + \epsilon/2. \end{aligned}$$

The assumption $\sum_{k=1}^{\infty} \alpha_k = \infty$ yields $\lim_{k\to\infty} \sum_{i=1}^{k} \alpha_i = \infty$, and hence

$$0 \leq \liminf_{k\to\infty}(V_k - \overline{V}) \leq \limsup_{k\to\infty}(V_k - \overline{V}) \leq \epsilon/2.$$

Since $\epsilon > 0$ is chosen arbitrarily, one has $\lim_{k\to\infty}(V_k - \overline{V}) = 0$ or $V_k \to \overline{V}$ as $k \to \infty$. $\square$

The next result gives us important information about the *value convergence* of the subgradient algorithm (4.10).

Proposition 4.26 *In the setting of Proposition 4.25 we have*

$$\liminf_{k\to\infty} f(x_k) = \overline{V}.$$

Proof. Since $\overline{V}$ is the optimal value in (4.9), we get $f(x_k) \geq \overline{V}$ for any $k \in \mathbb{N}$. This yields

$$\liminf_{k\to\infty} f(x_k) \geq \overline{V}.$$

To verify the opposite inequality, suppose on the contrary that $\liminf_{k\to\infty} f(x_k) > \overline{V}$ and then find a small number $\epsilon > 0$ such that

$$\liminf_{k\to\infty} f(x_k) - 2\epsilon > \overline{V} = \min_{x\in\mathbb{R}^n} f(x) = f(\bar{x}),$$

where $\bar{x}$ is an optimal solution to problem (4.9). Thus

$$\liminf_{k\to\infty} f(x_k) - \epsilon > f(\bar{x}) + \epsilon.$$

In the case where $\liminf_{k\to\infty} f(x_k) < \infty$, by the definition of lim inf, there is $k_0 \in \mathbb{N}$ such that for every $k \geq k_0$ we have that

$$f(x_k) > \liminf_{k\to\infty} f(x_k) - \epsilon,$$

and hence $f(x_k) - f(\bar{x}) > \epsilon$ for every $k \geq k_0$. Note that this conclusion also holds if $\liminf_{k\to\infty} f(x_k) = \infty$. It follows from (4.11) with $x = \bar{x}$ that

$$\|x_{k+1} - \bar{x}\|^2 \leq \|x_k - \bar{x}\|^2 - 2\alpha_k\big(f(x_k) - f(\bar{x})\big) + \alpha_k^2\ell^2 \quad \text{for every} \quad k \geq k_0. \tag{4.14}$$

Let k_1 be such that $\alpha_k\ell^2 < \epsilon$ for all $k \geq k_1$ and let $\bar{k}_1 := \max\{k_1, k_0\}$. For any $k > \bar{k}_1$, it easily follows from estimate (4.14) that

$$\|x_{k+1} - \bar{x}\|^2 \leq \|x_k - \bar{x}\|^2 - \alpha_k\epsilon \leq \ldots \leq \left\|x_{\bar{k}_1} - \bar{x}\right\|^2 - \epsilon \sum_{j=\bar{k}_1}^{k} \alpha_j,$$

which contradicts the assumption $\sum_{k=1}^{\infty} \alpha_k = \infty$ in (4.13) and so completes the proof. $\square$

Now we are ready to justify the convergence of the subgradient algorithm (4.10).

Theorem 4.27 Suppose that the generating sequence $\{\alpha_k\}$ in (4.10) satisfies

$$\sum_{k=1}^{\infty} \alpha_k = \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \alpha_k^2 < \infty. \tag{4.15}$$

Then the sequence $\{V_k\}$ converges to the optimal value $\overline{V}$ and the sequence of iterations $\{x_k\}$ in (4.10) converges to some optimal solution $\bar{x}$ of (4.9).

Proof. Since (4.15) implies (4.13), the sequence $\{V_k\}$ converges to $\overline{V}$ by Proposition 4.25. It remains to prove that $\{x_k\}$ converges to some optimal solution $\bar{x}$ of problem (4.9). Let $\ell \geq 0$ be a Lipschitz constant of f on $\mathbb{R}^n$. Since the set S of optimal solutions to (4.9) is assumed to be nonempty, we pick any $\widetilde{x} \in S$ and substitute $x = \widetilde{x}$ in estimate (4.11) of Proposition 4.22. It gives us

$$\|x_{k+1} - \widetilde{x}\|^2 \leq \|x_k - \widetilde{x}\|^2 - 2\alpha_k\big(f(x_k) - \overline{V}\big) + \alpha_k^2\ell^2 \quad \text{for all} \quad k \in \mathbb{N}\,. \tag{4.16}$$

Since $f(x_k) - \overline{V} \geq 0$, inequality (4.16) yields

$$\|x_{k+1} - \widetilde{x}\|^2 \leq \|x_k - \widetilde{x}\|^2 + \alpha_k^2\ell^2.$$

Then we get by induction that

$$\|x_{k+1} - \widetilde{x}\|^2 \leq \|x_1 - \widetilde{x}\|^2 + \ell^2 \sum_{i=1}^{k} \alpha_i^2. \tag{4.17}$$

The right-hand side of (4.17) is finite by our assumption that $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$, and hence

$$\sup_{k \in \mathbb{N}} \|x_{k+1} - \widetilde{x}\| < \infty.$$

This implies that the sequence $\{x_k\}$ is bounded. Furthermore, Proposition 4.26 tells us that

$$\liminf_{k \to \infty} f(x_k) = \overline{V}.$$

Let $\{x_{k_j}\}$ be a subsequence of $\{x_k\}$ such that

$$\lim_{j \to \infty} f(x_{k_j}) = \overline{V}. \tag{4.18}$$

Then $\{x_{k_j}\}$ is also bounded and contains a limiting point $\bar{x}$. Without loss of generality assume that $x_{k_j} \to \bar{x}$ as $j \to \infty$. We directly deduce from (4.18) and the continuity of f that $\bar{x}$ is an optimal solution to (4.9). Fix any $j \in \mathbb{N}$ and any $k \geq k_j$. Thus we can rewrite (4.16) with $\widetilde{x} = \bar{x}$ and see that

$$\|x_{k+1} - \bar{x}\|^2 \leq \|x_k - \bar{x}\|^2 + \alpha_k^2 \ell^2 \leq \ldots \leq \left\|x_{k_j} - \bar{x}\right\|^2 + \ell^2 \sum_{i=k_j}^{k} \alpha_i^2.$$

Taking first the limit as $k \to \infty$ and then the limit as $j \to \infty$ in the resulting inequality above gives us the estimate

$$\limsup_{k \to \infty} \|x_{k+1} - \bar{x}\|^2 \leq \lim_{j \to \infty} \left\|x_{k_j} - \bar{x}\right\|^2 + \ell^2 \lim_{j \to \infty} \sum_{i=k_j}^{\infty} \alpha_i^2.$$

Since $x_{k_j} \to \bar{x}$ and $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$, this implies that

$$\limsup_{k \to \infty} \|x_{k+1} - \bar{x}\|^2 = 0,$$

and thus justifies the claimed convergence of $x_k \to \bar{x}$ as $k \to \infty$. $\square$

Finally in this section, we present a version of the subgradient algorithm for convex problems of *constrained optimization* given by

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad x \in \Omega, \tag{4.19}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function that is Lipschitz continuous on $\mathbb{R}^n$ with some constant $\ell \geq 0$, while—in contrast to (4.9)—we have now the constraint on x given by a nonempty, closed, convex set $\Omega \subset \mathbb{R}^n$. The following counterpart of the subgradient algorithm (4.10) for constrained problems is known as the *projected subgradient method*.

Given a sequence of positive numbers $\{\alpha_k\}$ and given a starting point $x_1 \in \Omega$, the sequence of iterations $\{x_k\}$ is constructed by

$$x_{k+1} := \Pi(x_k - \alpha_k v_k; \Omega) \text{ with some } v_k \in \partial f(x_k), \quad k \in \mathbb{N}, \tag{4.20}$$

where $\Pi(x; \Omega)$ stands for the (unique) Euclidean projection of x onto Ω.

The next proposition, which is a counterpart of Proposition 4.22 for algorithm (4.20), plays a major role in justifying convergence results for the projected subgradient method.

Proposition 4.28 *In the general setting of* (4.20) *we have the estimate*

$$\|x_{k+1} - x\|^2 \le \|x_k - x\|^2 - 2\alpha_k\big(f(x_k) - f(x)\big) + \alpha_k^2 \ell^2 \textit{ for all } x \in \Omega,\ k \in \mathbb{N}\,.$$

Proof. Define $z_{k+1} := x_k - \alpha_k v_k$ and fix any $x \in \Omega$. It follows from Proposition 4.22 that

$$\|z_{k+1} - x\|^2 \le \|x_k - x\|^2 - 2\alpha_k\big(f(x_k) - f(x)\big) + \alpha_k^2 \ell^2 \text{ for all } k \in \mathbb{N}\,. \tag{4.21}$$

Now by projecting z_{k+1} onto Ω and using Proposition 1.79, we get

$$\|x_{k+1} - x\| = \|\Pi(z_{k+1}; \Omega) - \Pi(x; \Omega)\| \le \|z_{k+1} - x\|\,.$$

Combining this with estimate (4.21) gives us

$$\|x_{k+1} - x\|^2 \le \|x_k - x\|^2 - 2\alpha_k\big(f(x_k) - f(x)\big) + \alpha_k^2 \ell^2 \text{ for all } k \in \mathbb{N},$$

which thus completes the proof of the proposition. □

Based on this proposition and proceeding in the same way as for the subgradient algorithm (4.10) above, we can derive similar convergence results for the projected subgradient algorithm (4.20) under the assumptions (4.13) on the generating sequence $\{\alpha_k\}$.

4.4 THE FERMAT-TORRICELLI PROBLEM

In 1643 Pierre de Fermat proposed to Evangelista Torricelli the following optimization problem: Given three points in the plane, find another point such that the sum of its distances to the given points is minimal. This problem was solved by Torricelli and was named the *Fermat-Torricelli problem*. A more general version of the Fermat-Torricelli problem asks for a point that minimizes the sum of the distances to a finite number of given points in $\mathbb{R}^n$. This is one of the main problems in location science, which has been studied by many researchers. The first numerical algorithm for solving the general Fermat-Torricelli problem was introduced by Endre Weiszfeld in 1937. Assumptions that guarantee the convergence of the Weiszfeld algorithm were given by Harold Kuhn in 1972. Kuhn also pointed out an example in which the Weiszfeld algorithm fails to converge. Several new algorithms have been introduced recently to improve the Weiszfeld algorithm. The goal of this section is to revisit the Fermat-Torricelli problem from both theoretical and numerical viewpoints by using techniques of convex analysis and optimization.

Given points $a_i \in \mathbb{R}^n$ as $i = 1, \ldots, m$, define the (nonsmooth) convex function

$$\varphi(x) := \sum_{i=1}^{m} \|x - a_i\|. \tag{4.22}$$

Then the mathematical description of the *Fermat-Torricelli problem* is

$$\text{minimize } \varphi(x) \text{ over all } x \in \mathbb{R}^n. \tag{4.23}$$

Proposition 4.29 *The Fermat-Torricelli problem* (4.23) *has at least one solution.*

Proof. The conclusion follows directly from Theorem 4.10. □

Proposition 4.30 *Suppose that a_i as $i = 1, \ldots, m$ are not collinear, i.e., they do not belong to the same line. Then the function φ in* (4.22) *is strictly convex, and hence the Fermat-Torricelli problem* (4.23) *has a unique solution.*

Proof. Since each function $\varphi_i(x) := \|x - a_i\|$ as $i = 1, \ldots, m$ is obviously convex, their sum $\varphi = \sum_{i=1}^{m} \varphi_i$ is convex as well, i.e., for any $x, y \in \mathbb{R}^n$ and $\lambda \in (0, 1)$ we have

$$\varphi\big(\lambda x + (1-\lambda)y\big) \le \lambda\varphi(x) + (1-\lambda)\varphi(y). \tag{4.24}$$

Supposing by contradiction that φ is not strictly convex, find $\bar{x}, \bar{y} \in \mathbb{R}^n$ with $\bar{x} \neq \bar{y}$ and $\lambda \in (0, 1)$ such that (4.24) holds as equality. It follows that

$$\varphi_i\big(\lambda\bar{x} + (1-\lambda)\bar{y}\big) = \lambda\varphi_i(\bar{x}) + (1-\lambda)\varphi_i(\bar{y}) \text{ for all } i = 1, \ldots, m,$$

which ensures therefore that

$$\|\lambda(\bar{x} - a_i) + (1-\lambda)(\bar{y} - a_i)\| = \|\lambda(\bar{x} - a_i)\| + \|(1-\lambda)(\bar{y} - a_i)\|, \quad i = 1, \ldots, m.$$

If $\bar{x} \neq a_i$ and $\bar{y} \neq a_i$, then there exists $t_i > 0$ such that $t_i\lambda(\bar{x} - a_i) = (1-\lambda)(\bar{y} - a_i)$, and hence

$$\bar{x} - a_i = \gamma_i(\bar{y} - a_i) \text{ with } \gamma_i := \frac{1-\lambda}{t_i\lambda}.$$

Since $\bar{x} \neq \bar{y}$, we have $\gamma_i \neq 1$. Thus

$$a_i = \frac{1}{1-\gamma_i}\bar{x} - \frac{\gamma_i}{1-\gamma_i}\bar{y} \in \mathcal{L}(\bar{x}, \bar{y}),$$

where $\mathcal{L}(\bar{x}, \bar{y})$ signifies the line connecting $\bar{x}$ and $\bar{y}$. Both cases where $\bar{x} = a_i$ and $\bar{y} = a_i$ give us $a_i \in \mathcal{L}(\bar{x}, \bar{y})$. Hence $a_i \in \mathcal{L}(\bar{x}, \bar{y})$ for $i = 1, \ldots, m$, a contradiction. □

We proceed further with solving the Fermat-Torricelli problem for three points in $\mathbb{R}^n$. Let us first present two elementary lemmas.

Lemma 4.31 *Let a_1 and a_2 be any two unit vectors in $\mathbb{R}^n$. Then we have $-a_1 - a_2 \in I\!B$ if and only if $\langle a_1, a_2\rangle \leq -\dfrac{1}{2}$.*

Proof. If $-a_1 - a_2 \in I\!B$, then $a_1 + a_2 \in I\!B$ and

$$\|a_1 + a_2\| \leq 1.$$

By squaring both sides, we get the inequality

$$\|a_1 + a_2\|^2 \leq 1,$$

which subsequently implies that

$$\|a_1\|^2 + 2\langle a_1, a_2\rangle + \|a_2\|^2 = 2 + 2\langle a_1, a_2\rangle \leq 1.$$

Therefore, $\langle a_1, a_2\rangle \leq -\dfrac{1}{2}$. The proof of the converse implication is similar. □

Lemma 4.32 *Let $a_1, a_2, a_3 \in \mathbb{R}^n$ with $\|a_1\| = \|a_2\| = \|a_3\| = 1$. Then $a_1 + a_2 + a_3 = 0$ if and only if we have the equalities*

$$\langle a_1, a_2\rangle = \langle a_2, a_3\rangle = \langle a_3, a_1\rangle = -\frac{1}{2}. \tag{4.25}$$

Proof. It follows from the condition $a_1 + a_2 + a_3 = 0$ that

$$\|a_1\|^2 + \langle a_1, a_2\rangle + \langle a_1, a_3\rangle = 0, \tag{4.26}$$

$$\langle a_2, a_1\rangle + \|a_2\|^2 + \langle a_2, a_3\rangle = 0, \tag{4.27}$$

$$\langle a_3, a_1\rangle + \langle a_3, a_2\rangle + \|a_3\|^2 = 0. \tag{4.28}$$

Subtracting (4.28) from (4.26) gives us $\langle a_1, a_2\rangle = \langle a_2, a_3\rangle$ and similarly $\langle a_2, a_3\rangle = \langle a_3, a_1\rangle$. Substituting these equalities into (4.26), (4.27), and (4.28) ensures the validity of (4.25).

Conversely, assume that (4.25) holds. Then we have

$$\|a_1 + a_2 + a_3\|^2 = \sum_{i=1}^{3} \|a_i\|^2 + \sum_{i,j=1, i\neq j}^{3} \langle a_i, a_j\rangle = 0,$$

which readily yields the equality $a_1 + a_2 + a_3 = 0$ and thus completes the proof. □

Now we use subdifferentiation of convex functions to solve the Fermat-Torricelli problem for three points in $\mathbb{R}^n$. Given two nonzero vectors $u, v \in \mathbb{R}^n$, define

$$\cos(u, v) := \frac{\langle u, v \rangle}{\|u\| \cdot \|v\|}$$

and also consider, when $\bar{x} \neq a_i$, the three vectors

$$v_i = \frac{\bar{x} - a_i}{\|\bar{x} - a_i\|}, \; i = 1, 2, 3.$$

Geometrically, each vector v_i is a unit one pointing in the direction from the vertex a_i to $\bar{x}$. Observe that the classical (three-point) Fermat-Torricelli problem always has a unique solution even if the given points belong to the same line. Indeed, in the latter case the middle point solves the problem.

The next theorem fully describes optimal solutions of the Fermat-Torricelli problem for three points in two distinct cases.

Theorem 4.33 Consider the Fermat-Torricelli problem (4.23) for the three points $a_1, a_2, a_3 \in \mathbb{R}^n$. The following assertions hold:

(i) Let $\bar{x} \notin \{a_1, a_2, a_3\}$. Then $\bar{x}$ is the solution of the problem if and only if

$$\cos(v_1, v_2) = \cos(v_2, v_3) = \cos(v_3, v_1) = -1/2.$$

(ii) Let $\bar{x} \in \{a_1, a_2, a_3\}$, say $\bar{x} = a_1$. Then $\bar{x}$ is the solution of the problem if and only if

$$\cos(v_2, v_3) \leq -1/2.$$

Proof. In case **(i)** function (4.22) is differentiable at $\bar{x}$. Then $\bar{x}$ solves (4.23) if and only if

$$\nabla\varphi(\bar{x}) = v_1 + v_2 + v_3 = 0.$$

By Lemma 4.32 this holds if and only if

$$\langle v_i, v_j \rangle = \cos(v_i, v_j) = -1/2 \text{ for any } i, j \in \big\{1, 2, 3\big\} \text{ with } i \neq j.$$

To verify **(ii)**, employ the subdifferential Fermat rule (2.35) and sum rule from Corollary 2.45, which tell us that $\bar{x} = a_1$ solves the Fermat-Torricelli problem if and only if

$$0 \in \partial\varphi(a_1) = v_2 + v_3 + I\!B.$$

Since v_2 and v_3 are unit vectors, it follows from Lemma 4.31 that

$$\langle v_2, v_3 \rangle = \cos(v_2, v_3) \leq -1/2,$$

which completes the proof of the theorem. $\square$

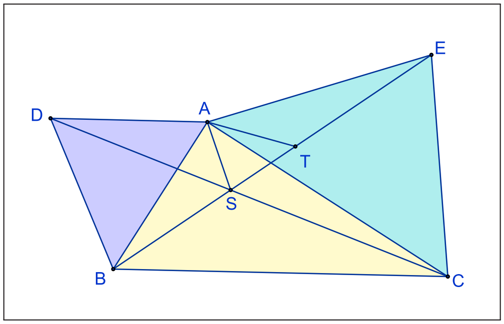

Figure 4.2: Construction of the Torricelli point.

Example 4.34 This example illustrates geometrically the solution construction for the Fermat-Torricelli problem in the plane. Consider (4.23) with the three given points A, B, and C on $\mathbb{R}^2$ as shown in Figure 4.2. If one of the angles of the triangle ABC is greater than or equal to $120°$, then the corresponding vertex is the solution to (4.23) by Theorem 4.33(ii). Consider the case where none of the angles of the triangle is greater than or equal to $120°$. Construct the two equilateral triangles ABD and ACE and let S be the intersection of DC and BE as in the figure. Two quadrilaterals $ADBC$ and $ABCE$ are convex, and hence S lies inside the triangle ABC. It is clear that two triangles DAC and BAE are congruent. A rotation of $60°$ about A maps the triangle DAC to the triangle BAE. The rotation maps CD to BE, so $\angle DSB = 60°$. Let T be the image of S though this rotation. Then T belongs to BE. It follows that $\angle AST = \angle ASE = 60°$. Moreover, $\angle DSA = 60°$, and hence $\angle BSA = 120°$. It is now clear that $\angle ASC = 120°$ and $\angle BSC = 120°$. By Theorem 4.33(i) the point S is the solution of the problem under consideration.

Next we present the *Weiszfeld algorithm* [32] to solve numerically the Fermat-Torricelli problem in the general case of (4.23) for m points in $\mathbb{R}^n$ that are *not collinear*. This is our *standing assumption* in the rest of this section. To proceed, observe that the gradient of the function φ from (4.22) is calculated by

$$\nabla\varphi(x) = \sum_{i=1}^{m} \frac{x - a_i}{\|x - a_i\|} \quad \text{for} \quad x \notin \{a_1, \ldots, a_m\}.$$

Solving the gradient equation $\nabla\varphi(x) = 0$ gives us

$$x = \frac{\sum_{i=1}^{m} \dfrac{a_i}{\|x - a_i\|}}{\sum_{i=1}^{m} \dfrac{1}{\|x - a_i\|}} =: f(x). \tag{4.29}$$

We define $f(x)$ on the whole space $\mathbb{R}^n$ by letting $f(x) := x$ for $x \in \{a_1, \ldots, a_m\}$.

Weiszfeld algorithm consists of choosing an arbitrary starting point $x_1 \in \mathbb{R}^n$, taking f is from (4.29), and defining recurrently

$$x_{k+1} := f(x_k) \text{ for } k \in \mathbb{N}, \tag{4.30}$$

where $f(x)$ is called the *algorithm mapping*. In the rest of this section we present the precise conditions and the proof of convergence for Weiszfeld algorithm in the way suggested by Kuhn [10] who corrected some errors from [32]. In contrast to [10], we employ below subdifferential rules of convex analysis, which allow us to simplify the proof of convergence.

The next proposition ensures decreasing the function value of φ after each iteration, i.e., the *descent property* of the algorithm.

Proposition 4.35 *If $f(x) \neq x$, then $\varphi(f(x)) < \varphi(x)$.*

Proof. The assumption of $f(x) \neq x$ tells us that x is not a vertex from $\{a_1, \ldots, a_m\}$. Observe also that the point $f(x)$ is a unique minimizer of the strictly convex function

$$g(z) := \sum_{i=1}^{m} \frac{\|z - a_i\|^2}{\|x - a_i\|}$$

and that $g(f(x)) < g(x) = \varphi(x)$ since $f(x) \neq x$. We have furthermore that

$$\begin{aligned} g\big(f(x)\big) &= \sum_{i=1}^{m} \frac{\|f(x) - a_i\|^2}{\|x - a_i\|} \\ &= \sum_{i=1}^{m} \frac{(\|x - a_i\| + \|f(x) - a_i\| - \|x - a_i\|)^2}{\|x - a_i\|} \\ &= \varphi(x) + 2\big(\varphi(f(x)) - \varphi(x)\big) + \sum_{i=1}^{m} \frac{(\|f(x) - a_i\| - \|x - a_i\|)^2}{\|x - a_i\|}. \end{aligned}$$

This clearly implies the inequality

$$2\varphi\big(f(x)\big) + \sum_{i=1}^{m} \frac{(\|f(x) - a_i\| - \|x - a_i\|)^2}{\|x - a_i\|} < 2\varphi(x)$$

and thus completes the proof of the proposition. □

The following two propositions show how the algorithm mapping f behaves near a vertex that solves the Fermat-Torricelli problem. First we present a necessary and sufficient condition ensuring the optimality of a vertex in (4.23). Define

$$R_j := \sum_{i=1, i \neq j}^{m} \frac{a_i - a_j}{\|a_i - a_j\|}, \quad j = 1, \ldots, m.$$

Proposition 4.36 *The vertex a_j solves problem (4.23) if and only if $\|R_j\| \le 1$.*

Proof. Employing the Fermat rule from Proposition 2.35 and the subdifferential sum rule from Corollary 2.45 shows that the vertex a_j solves problem (4.23) if and only if

$$0 \in \partial\varphi(a_j) = -R_j + I\!B(0,1),$$

which is clearly equivalent to the condition $\|R_j\| \le 1$. □

Proposition 4.36 from convex analysis allows us to essentially simplify the proof of the following result established by Kuhn. We use the notation:

$$f^s(x) := f\big(f^{s-1}(x)\big) \text{ for } s \in \mathbb{N} \text{ with } f^0(x) := x, \quad x \in \mathbb{R}^n.$$

Proposition 4.37 *Suppose that the vertex a_j does not solve (4.23). Then there is $\delta > 0$ such that the condition $0 < \|x - a_j\| \le \delta$ implies the existence of $s \in \mathbb{N}$ with*

$$\|f^s(x) - a_j\| > \delta \text{ and } \|f^{s-1}(x) - a_j\| \le \delta. \tag{4.31}$$

Proof. For any $x \notin \{a_1, \ldots, a_m\}$, we have from (4.29) that

$$f(x) - a_j = \frac{\sum_{i=1, i\neq j}^m \dfrac{a_i - a_j}{\|x - a_i\|}}{\sum_{i=1}^m \dfrac{1}{\|x - a_i\|}} \text{ and so}$$

$$\lim_{x\to a_j} \frac{f(x) - a_j}{\|x - a_j\|} = \lim_{x\to a_j} \frac{\sum_{i=1, i\neq j}^m \dfrac{a_i - a_j}{\|x - a_i\|}}{1 + \sum_{i=1, i\neq j}^m \dfrac{\|x - a_j\|}{\|x - a_i\|}} = R_j, \quad j = 1, \ldots, m.$$

It follows from Proposition 4.36 that

$$\lim_{x\to a_j} \frac{\|f(x) - a_j\|}{\|x - a_j\|} = \|R_j\| > 1, \quad j = 1, \ldots, m, \tag{4.32}$$

which allows us to select $\epsilon > 0$, $\delta > 0$ such that

$$\frac{\|f(x) - a_j\|}{\|x - a_j\|} > 1 + \epsilon \text{ whenever } 0 < \|x - a_j\| \le \delta.$$

Thus for every $s \in \mathbb{N}$ the validity of $0 < \|f^{q-1}(x) - a_j\| \le \delta$ with any $q = 1, \ldots, s$ yields

$$\|f^s(x) - a_j\| \ge (1+\epsilon)\|f^{s-1}(x) - a_j\| \ge \ldots \ge (1+\epsilon)^s\|x - a_j\|.$$

Since $(1+\epsilon)^s\|x - a_j\| \to \infty$ as $s \to \infty$, it gives us (4.31) and completes the proof. □

We finally present the main result on the convergence of the Weiszfeld algorithm.

Theorem 4.38 Let $\{x_k\}$ be the sequence generated by Weiszfeld algorithm (4.30) and let $x_k \notin \{a_1, \ldots, a_m\}$ for $k \in \mathbb{N}$. Then $\{x_k\}$ converges to the unique solution $\bar{x}$ of (4.23).

Proof. If $x_k = x_{k+1}$ for some $k = k_0$, then $\{x_k\}$ is a constant sequence for $k \ge k_0$, and hence it converges to x_{k_0}. Since $f(x_{k_0}) = x_{k_0}$ and x_{k_0} is not a vertex, x_{k_0} solves the problem. Thus we can assume that $x_{k+1} \ne x_k$ for every k. It follows from Proposition 4.35 that the sequence $\{\varphi(x_k)\}$ of nonnegative numbers is decreasing, so it converges. This gives us

$$\lim_{k\to\infty} \big(\varphi(x_k) - \varphi(x_{k+1})\big) = 0. \tag{4.33}$$

By the construction of $\{x_k\}$, we have that each x_k belongs to the compact set $\operatorname{co}\{a_1, \ldots, a_m\}$. This allows us to select from $\{x_k\}$ a subsequence $\{x_{k_\nu}\}$ converging to some point $\bar{z}$ as $\nu \to \infty$. We have to show that $\bar{z} = \bar{x}$, i.e., it is the unique solution to (4.23). Indeed, we get from (4.33) that

$$\lim_{\nu\to\infty} \big(\varphi(x_{k_\nu}) - \varphi\big(f(x_{k_\nu})\big)\big) = 0,$$

which implies by the continuity of the functions φ and f that $\varphi(\bar{z}) = \varphi(f(\bar{z}))$, i.e., $f(\bar{z}) = \bar{z}$ and thus $\bar{z}$ solves (4.23) if it is not a vertex.

It remains to consider the case where $\bar{z}$ is a vertex, say a_1. Suppose by contradiction that $\bar{z} = a_1 \ne \bar{x}$ and select $\delta > 0$ sufficiently small so that (4.31) holds for some $s \in \mathbb{N}$ and that $I\!B(a_1; \delta)$ does not contain $\bar{x}$ and a_i for $i = 2, \ldots, m$. Since $x_{k_\nu} \to \bar{z} = a_1$, we can assume without loss of generality that this sequence is entirely contained in $I\!B(a_1; \delta)$. For $x = x_{k_1}$ select l_1 so that $x_{l_1} \in I\!B(a_1; \delta)$ and $f(x_{l_1}) \notin I\!B(a_1; \delta)$. Choosing now $k_\nu > l_1$ and applying Proposition 4.37 give us $l_2 > l_1$ with $x_{l_2} \in I\!B(a_1; \delta)$ and $f(x_{l_2}) \notin I\!B(a_1; \delta)$. Repeating this procedure, find a subsequence $\{x_{l_\nu}\}$ such that $x_{l_\nu} \in I\!B(a_1; \delta)$ while $f(x_{l_\nu})$ is not in this ball. Extracting another subsequence if needed, we can assume that $\{x_{l_\nu}\}$ converges to some point $\bar{q}$ satisfying $f(\bar{q}) = \bar{q}$. If $\bar{q}$ is not a vertex, then it solves (4.23) by the above, which is a contradiction because the solution $\bar{x}$ is not in $I\!B(a_1; \delta)$. Thus $\bar{q}$ is a vertex, which must be a_1 since the other vertexes are also not in the ball. Then we have

$$\lim_{\nu\to\infty} \frac{\|f(x_{l_\nu}) - a_1\|}{\|x_{l_\nu} - a_1\|} = \infty,$$

which contradicts (4.32) in Proposition 4.37 and completes the proof of the theorem. □

4.5 A GENERALIZED FERMAT-TORRICELLI PROBLEM

There are several generalized versions of the Fermat-Torricelli problem known in the literature under different names; see, e.g., our papers [15, 16, 17] and the references therein. Note that the constrained version of the problem below is also called a "generalized Heron problem" after Heron from Alexandria who formulated its simplest version in the plane in the 1st century AD. Following [15, 16], we consider here a generalized Fermat-Torricelli problem defined via the *minimal time functions* (3.9) for finitely many closed, convex target sets in $\mathbb{R}^n$ with a constraint set of the same structure. In [6, 11, 29] and their references, the reader can find some particular versions of this problem and related material on location problems. We also refer the reader to [15, 17] to *nonconvex* problems of this type.

Let the *target sets* Ω_i for $i = 1, \ldots, m$ and the *constraint set* Ω_0 be nonempty, closed, convex subsets of $\mathbb{R}^n$. Consider the minimal time functions $\mathcal{T}_F(x; \Omega_i) = \mathcal{T}^F_{\Omega_i}(x)$ from (3.9) for Ω_i, $i = 1, \ldots, m$, with the *constant dynamics* defined by a closed, bounded, convex set $F \subset \mathbb{R}^n$ that contains the origin as an interior point. Then the *generalized Fermat-Torricelli problem* is formulated as follows:

$$\text{minimize } \mathcal{H}(x) := \sum_{i=1}^{m} \mathcal{T}_F(x; \Omega_i) \text{ subject to } x \in \Omega_0. \tag{4.34}$$

Our first goal in this section is to establish the *existence* and *uniqueness* of solutions to the optimization problem (4.34). We split the proof into several propositions of their own interest. Let us start with the following property of the Minkowski gauge from (3.11).

Proposition 4.39 *Assume that F is a closed, bounded, convex subset of $\mathbb{R}^n$ such that $0 \in \operatorname{int} F$. Then $\rho_F(x) = 1$ if and only if $x \in \operatorname{bd} F$.*

Proof. Suppose that $\rho_F(x) = 1$. Then there exists a sequence of real numbers $\{t_k\}$ that converges to $t = 1$ and such that $x \in t_k F$ for every k. Since F is closed, it yields $x \in F$. To complete the proof of the implication, we show that $x \notin \operatorname{int} F$. By contradiction, assume that $x \in \operatorname{int} F$ and choose $t > 0$ so small that $x + tx \in F$. Then $x \in (1+t)^{-1} F$. This tells us $\rho_F(x) \le (1+t)^{-1} < 1$, a contradiction.

Now let $x \in \operatorname{bd} F$. Since F is closed, we have $x \in F$, and hence $\rho_F(x) \le 1$. Suppose on the contrary that $\gamma := \rho_F(x) < 1$. Then there exists a sequence of real numbers $\{t_k\}$ that converges to γ with $x \in t_k F$ for every k. Then $x \in \gamma F = \gamma F + (1-\gamma)0 \subset F$. By Proposition 3.32, the Minkowski gauge ρ_F is continuous, so the set $\{u \in \mathbb{R}^n \mid \rho_F(u) < 1\}$ is an open subset of F containing x. Thus $x \in \operatorname{int} F$, which completes the proof. $\square$

Recall further that a set $Q \subset \mathbb{R}^n$ is said to be *strictly convex* if for any $x, y \in Q$ with $x \neq y$ and for any $t \in (0, 1)$ we have $tx + (1-t)y \in \operatorname{int} Q$.

Proposition 4.40 *Assume that F is a closed, bounded, strictly convex subset of $\mathbb{R}^n$ and such that $0 \in \operatorname{int} F$. Then*

$$\rho_F(x+y) = \rho_F(x) + \rho_F(y) \ \textit{with} \ x, y \neq 0 \tag{4.35}$$

if and only if $x = \lambda y$ for some $\lambda > 0$.

Proof. Denote $\alpha := \rho_F(x)$, $\beta := \rho_F(y)$ and observe that $\alpha, \beta > 0$ since F is bounded. If the linearity property (4.35) holds, then

$$\rho_F\Big(\frac{x+y}{\alpha+\beta}\Big) = 1,$$

which implies in turn that

$$\rho_F\Big(\frac{x}{\alpha}\frac{\alpha}{\alpha+\beta} + \frac{y}{\beta}\frac{\beta}{\alpha+\beta}\Big) = 1.$$

This allows us to conclude that

$$\frac{x}{\alpha}\frac{\alpha}{\alpha+\beta} + \frac{y}{\beta}\frac{\beta}{\alpha+\beta} \in \operatorname{bd} F.$$

Since $\dfrac{x}{\alpha} \in F$, $\dfrac{y}{\beta} \in F$, and $\dfrac{\alpha}{\alpha+\beta} \in (0,1)$, it follows from the strict convexity of F that

$$\frac{x}{\alpha} = \frac{y}{\beta}, \ \text{i.e.,} \ x = \lambda y \ \text{ with } \ \lambda = \frac{\rho_F(x)}{\rho_F(y)}.$$

The opposite implication in the proposition is obvious. □

Proposition 4.41 *Let F be as in Proposition 4.40 and let Ω be a nonempty, closed, convex subset of $\mathbb{R}^n$. For any $x \in \mathbb{R}^n$, the set $\Pi_F(x;\Omega)$ is a singleton.*

Proof. We only need to consider the case where $x \notin \Omega$. It is clear that $\Pi_F(x;\Omega) \neq \emptyset$. Suppose by contradiction that there are $u_1, u_2 \in \Pi_F(x;\Omega)$ with $u_1 \neq u_2$. Then

$$\mathcal{T}_F(x;\Omega) = \rho_F(u_1 - x) = \rho_F(u_2 - x) = r > 0,$$

which implies that $\dfrac{u_1 - x}{r} \in \operatorname{bd} F$ and $\dfrac{u_2 - x}{r} \in \operatorname{bd} F$. Since F is strictly convex, we have

$$\frac{1}{2}\frac{u_1 - x}{r} + \frac{1}{2}\frac{u_2 - x}{r} = \frac{1}{r}\Big(\frac{u_1+u_2}{2} - x\Big) \in \operatorname{int} F,$$

and so we get $\rho_F(u - x) < r = \mathcal{T}_F(x;\Omega)$ with $u := \dfrac{u_1+u_2}{2} \in \Omega$. This contradiction completes the proof of the proposition. □

In what follows we identify the generalized projection $\Pi_F(x;\Omega)$ with its unique element when both sets F and Ω are strictly convex.

Proposition 4.42 *Let F be a closed, bounded, convex subset of $\mathbb{R}^n$ such that $0 \in \operatorname{int} F$ and let Ω be a nonempty, closed set $\Omega \subset \mathbb{R}^n$ with $\bar{x} \notin \Omega$. If $\bar{\omega} \in \Pi_F(\bar{x};\Omega)$, then $\bar{\omega} \in \operatorname{bd}\Omega$.*

Proof. Suppose that $\bar{\omega} \in \operatorname{int}\Omega$. Then there is $\epsilon > 0$ such that $I\!B(\bar{\omega};\epsilon) \subset \Omega$. Define

$$z := \bar{\omega} + \epsilon\frac{\bar{x}-\bar{\omega}}{\|\bar{x}-\bar{\omega}\|} \in I\!B(\bar{\omega};\epsilon).$$

and observe from the construction of the Minkowski gauge (3.11) that

$$\rho_F(z-\bar{x}) = \rho_F\Big(\bar{\omega} + \epsilon\frac{\bar{x}-\bar{\omega}}{\|\bar{x}-\bar{\omega}\|} - \bar{x}\Big) = \Big(1 - \frac{\epsilon}{\|\bar{x}-\bar{\omega}\|}\Big)\rho_F(\bar{\omega}-\bar{x}) < \rho_F(\bar{\omega}-\bar{x}) = \mathcal{T}_F(\bar{x};\Omega).$$

This shows that $\bar{\omega}$ must be a boundary point of Ω. $\square$

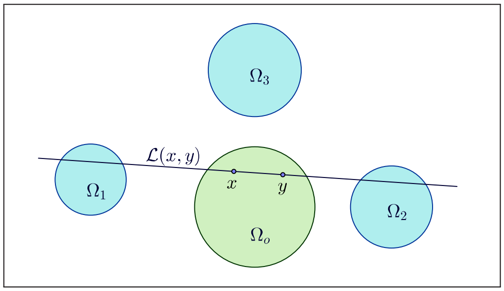

Figure 4.3: Three sets that satisfy assumptions of Proposition 4.43.

Proposition 4.43 *In the setting of problem (4.34) suppose that the sets Ω_i for $i = 1,\ldots,m$ are strictly convex. If for any $x, y \in \Omega_0$ with $x \neq y$ there is $i_0 \in \{1,\ldots,m\}$ such that*

$$\mathcal{L}(x,y) \cap \Omega_{i_0} = \emptyset$$

for the line $\mathcal{L}(x,y)$ connecting x and y, then the function

$$\mathcal{H}(x) = \sum_{i=1}^{m} \mathcal{T}_F(x;\Omega_i)$$

defined in (4.34) is strictly convex on the constraint set Ω_0.

Proof. Suppose by contradiction that there are $x, y \in \Omega_0$, $x \neq y$, and $t \in (0,1)$ with

$$\mathcal{H}\big(tx + (1-t)y\big) = t\mathcal{H}(x) + (1-t)\mathcal{H}(y).$$

Using the convexity of each function $\mathcal{T}_F(x;\Omega_i)$ for $i = 1, \ldots, m$, we have

$$\mathcal{T}_F\big(tx + (1-t)y;\Omega_i\big) = t\mathcal{T}_F(x;\Omega_i) + (1-t)\mathcal{T}_F(y;\Omega_i), \quad i = 1, \ldots, m. \tag{4.36}$$

Suppose further that $\mathcal{L}(x,y) \cap \Omega_{i_0} = \emptyset$ for some $i_0 \in \{1, \ldots, m\}$ and define

$$u := \Pi_F(x;\Omega_{i_0}) \text{ and } v := \Pi_F(y;\Omega_{i_0}).$$

Then it follows from (4.36), Theorem 3.33, and the convexity of the Minkowski gauge that

$$\begin{aligned}
t\rho_F(u-x) + (1-t)\rho_F(v-y) &= t\mathcal{T}_F(x;\Omega_{i_0}) + (1-t)\mathcal{T}_F(y;\Omega_{i_0})\\
&= \mathcal{T}_F\big(tx + (1-t)y;\Omega_{i_0}\big)\\
&\le \rho_F\big(tu + (1-t)v - (tx + (1-t)y)\big)\\
&\le t\rho_F(u-x) + (1-t)\rho_F(v-y),
\end{aligned}$$

which shows that $tu + (1-t)v = \Pi_F(tx + (1-t)y;\Omega_{i_0})$. Thus $u = v$, since otherwise we have $tu + (1-t)v \in \operatorname{int}\Omega_{i_0}$, a contradiction. This gives us $u = v = \Pi_F(tx + (1-t)y;\Omega_{i_0})$. Applying (4.36) again tells us that

$$\rho_F\big(u - (tx + (1-t)y\big) = \rho_F\big(t(u-x)\big) + \rho_F\big((1-t)(u-y)\big),$$

so $u - x, u - y \neq 0$ due to $x, y \notin \Omega_{i_0}$. By Proposition 4.40, we find $\lambda > 0$ such that $t(u-x) = \lambda(1-t)(u-y)$, which yields

$$u - x = \gamma(u-y) \text{ with } \gamma = \frac{\lambda(1-t)}{t} \neq 1.$$

This justifies the inclusion

$$u = \frac{1}{1-\gamma}x - \frac{\gamma}{1-\gamma}y \in \mathcal{L}(x,y),$$

which is a contradiction that completes the proof of the proposition. □

Now we are ready to establish the existence and uniqueness theorem for (4.34).

Theorem 4.44 In the setting of Proposition 4.43 suppose further that at least one of the sets among Ω_i for $i = 0, \ldots, m$ is bounded. Then the generalized Fermat-Torricelli problem (4.34) has a unique solution.

Proof. By the strict convexity of $\mathcal{H}$ from Proposition 4.43, it suffices to show that problem (4.34) admits an optimal solution. If Ω_0 is bounded, then it is compact and the conclusion follows from the (Weierstrass) Theorem 4.10 since the function $\mathcal{H}$ in (4.34) Lipschitz continuous by Corollary 3.34. Consider now the case where one of the sets Ω_i for $i = 1, \ldots, m$ is bounded; say it is Ω_1. In this case we get that for any $\alpha \in \mathbb{R}$ the set

$$\mathcal{L}_\alpha = \big\{x \in \Omega_0 \;\big|\; \mathcal{H}(x) \le \alpha\big\} \subset \big\{x \in \Omega_0 \;\big|\; \mathcal{T}_F(x; \Omega_1) \le \alpha\big\} \subset \Omega_0 \cap (\Omega_1 - \alpha F)$$

is compact. This ensures the existence of a minimizer for (4.34) by Theorem 4.10. □

Next we completely solve by means of convex analysis and optimization the following simplified version of problem (4.34):

$$\text{minimize } \; \mathcal{H}(x) = \sum_{i=1}^{3} d(x; \Omega_i), \quad x \in \mathbb{R}^2, \tag{4.37}$$

where the *target sets* $\Omega_i := I\!B(c_i; r_i)$ for $i = 1, 2, 3$ are the *closed balls* of $\mathbb{R}^2$ whose centers c_i are supposed to be *distinct* from each other to avoid triviality.

We first study properties of this problem that enable us to derive verifiable conditions in terms of the initial data (c_i, r_i) ensuring the *uniqueness* of solutions to (4.37).

Proposition 4.45 *The solution set S of (4.37) is a nonempty, compact, convex set in $\mathbb{R}^2$.*

Proof. The existence of solutions (4.37) follows from the proof of Theorem 4.44. It is also easy to see that S is closed and bounded. The convexity of S can be deduced from the convexity of the distance functions $d(x; \Omega_i)$. □

To proceed, for any $u \in \mathbb{R}^2$ define the index subset

$$\mathcal{I}(u) := \big\{i \in \{1, 2, 3\} \;\big|\; u \in \Omega_i\big\}, \tag{4.38}$$

which is widely used in what follows.

Proposition 4.46 *Suppose that $\bar{x}$ does not belong to any of the sets Ω_i, $i = 1, 2, 3$, i.e., $\mathcal{I}(\bar{x}) = \emptyset$. Then $\bar{x}$ is an optimal solution to (4.37) if and only if $\bar{x}$ solves the classical Fermat-Torricelli problem (4.23) generated by the centers of the balls $a_i := c_i$ for $i = 1, 2, 3$ in (4.22).*

Proof. Assume that $\bar{x}$ is a solution to (4.37) with $\bar{x} \notin \Omega_i$ for $i = 1, 2, 3$. Choose $\delta > 0$ such that $I\!B(\bar{x}; \delta) \cap \Omega_i = \emptyset$ as $i = 1, 2, 3$ and get following for every $x \in I\!B(\bar{x}; \delta)$:

$$\sum_{i=1}^{3} \|\bar{x} - c_i\| - \sum_{i=1}^{3} r_i = \inf_{x \in \mathbb{R}^2} \mathcal{H}(x) = \mathcal{H}(\bar{x}) \le \mathcal{H}(x) = \sum_{i=1}^{3} \|x - c_i\| - \sum_{i=1}^{3} r_i.$$

This shows that $\bar{x}$ is a local minimum of the problem

$$\text{minimize } \mathcal{F}(x) := \sum_{i=1}^{3} \|x - c_i\|, \quad x \in \mathbb{R}^2, \tag{4.39}$$

so it is a global absolute minimum of this problem due to the convexity of $\mathcal{F}$. Thus $\bar{x}$ solves the classical Fermat-Torricelli problem (4.23) generated by c_1, c_2, c_3. The converse implication is also straightforward. □

The next result gives a condition for the uniqueness of solutions to (4.37) in terms of the index subset defined in (4.38).

Proposition 4.47 *Let $\bar{x}$ be a solution to (4.37) such that $\mathcal{I}(\bar{x}) = \emptyset$ in (4.38). Then $\bar{x}$ is the unique solution to (4.37), i.e., $S = \{\bar{x}\}$.*

Proof. Suppose by contradiction that there is $u \in S$ such that $u \neq \bar{x}$. Since $\mathcal{I}(\bar{x}) = \emptyset$, we have $\bar{x} \notin \Omega_i$ for $i = 1, 2, 3$. Then Proposition 4.46 tells us that $\bar{x}$ solves the classical Fermat-Torricelli problem generated by c_1, c_2, c_3. It follows from the convexity of S in Proposition 4.45 that $[\bar{x}, u] \subset S$. Thus we find $v \neq \bar{x}$ such that $v \in S$ and $v \notin \Omega_i$ for $i = 1, 2, 3$. It shows that v also solves the classical Fermat-Torricelli problem (4.39). This contradicts the uniqueness result for (4.39) and hence justifies the claim of the proposition. □

The following proposition verifies a visible geometric property of ellipsoids.

Proposition 4.48 *Given $\alpha > 0$ and $a, b \in \mathbb{R}^2$ with $a \neq b$, consider the set*

$$E := \left\{x \in \mathbb{R}^2 \;\middle|\; \|x - a\| + \|x - b\| = \alpha\right\}$$

and suppose that $\|a - b\| < \alpha$. For $x, y \in E$ with $x \neq y$, we have

$$\left\|\frac{x+y}{2} - a\right\| + \left\|\frac{x+y}{2} - b\right\| < \alpha,$$

which ensures, in particular, that $\dfrac{x+y}{2} \notin E$.

Proof. It follows from the triangle inequality that

$$\begin{aligned}\left\|\frac{x+y}{2} - a\right\| + \left\|\frac{x+y}{2} - b\right\| &= \frac{1}{2}\Big(\|x - a + y - a\| + \|x - b + y - b\|\Big)\\ &\leq \frac{1}{2}\Big(\|x - a\| + \|y - a\| + \|x - b\| + \|y - b\|\Big) = \alpha.\end{aligned}$$

Suppose by contradiction that

$$\left\|\frac{x+y}{2}-a\right\|+\left\|\frac{x+y}{2}-b\right\|=\alpha, \tag{4.40}$$

which means that $\frac{x+y}{2}\in E$. Since our norm is (always) Euclidean, it gives

$$x-a=\beta(y-a) \text{ and } x-b=\gamma(y-b)$$

for some numbers $\beta,\gamma\in(0,\infty)\setminus\{1\}$ since $x\neq y$. This implies in turn that

$$a=\frac{1}{1-\beta}x-\frac{\beta}{1-\beta}y\in\mathcal{L}(x,y) \text{ and } b=\frac{1}{1-\gamma}x-\frac{\gamma}{1-\gamma}y\in\mathcal{L}(x,y).$$

Since $\alpha>\|a-b\|$, we see that $x,y\in\mathcal{L}(a,b)\setminus[a,b]$. To check now that $\frac{x+y}{2}\in[a,b]$, impose without loss of generality the following order of points: x, $\frac{x+y}{2}$, a,b, and y. Then

$$\left\|\frac{x+y}{2}-a\right\|+\left\|\frac{x+y}{2}-b\right\|<\|x-a\|+\|x-b\|=\alpha,$$

which contradicts (4.40). Since $\frac{x+y}{2}\in[a,b]$, it yields

$$\left\|\frac{x+y}{2}-a\right\|+\left\|\frac{x+y}{2}-b\right\|=\|a-b\|<\alpha,$$

which is a contradiction. Thus $\frac{x+y}{2}\notin E$ and the proof is complete. □

The next result ensures the uniqueness of solutions to (4.37) via the index set (4.38) differently from Proposition 4.47.

Proposition 4.49 *Let S be the solution set for problem (4.37) and let $\bar{x}\in S$ satisfy the conditions: $\mathcal{I}(\bar{x})=\{i\}$ and $\bar{x}\notin[c_j,c_k]$, where i,j,k are distinct indices in $\{1,2,3\}$. Then S is a singleton, namely $S=\{\bar{x}\}$.*

Proof. Suppose for definiteness that $\mathcal{I}(\bar{x})=\{1\}$. Then we have $\bar{x}\in\Omega_1$, $\bar{x}\notin\Omega_2$, $\bar{x}\notin\Omega_3$, and $\bar{x}\notin[c_2,c_3]$. Denote

$$\alpha:=\|\bar{x}-c_2\|+\|\bar{x}-c_3\|=\inf_{x\in\mathbb{R}^2}\mathcal{H}(x)+r_2+r_3$$

and observe that $\alpha>\|c_2-c_3\|$. Consider the ellipse

$$E:=\left\{x\in\mathbb{R}^2\,\middle|\,\|x-c_2\|+\|x-c_3\|=\alpha\right\}$$

for which we get $\bar{x} \in E \cap \Omega_1$. Arguing by contradiction, suppose that S is not a singleton, i.e., there is $u \in S$ with $u \neq \bar{x}$. The convexity of S implies that $[\bar{x}, u] \subset S$. We can choose $v \in [\bar{x}, u]$ sufficiently close to but different from $\bar{x}$ so that $[\bar{x}, v] \cap \Omega_2 = \emptyset$ and $[\bar{x}, v] \cap \Omega_3 = \emptyset$. Let us first show that $v \notin \Omega_1$. Assuming the contrary gives us $d(v; \Omega_1) = 0$, and hence

$$\mathcal{H}(v) = \inf_{x \in \mathbb{R}^2} \mathcal{H}(x) = d(v; \Omega_2) + d(v; \Omega_3) = \|v - c_2\| - r_2 + \|v - c_3\| - r_3,$$

which implies that $v \in E$. It follows from the convexity of S and Ω_1 that $\dfrac{\bar{x}+v}{2} \in S \cap \Omega_1$. We have furthermore that $\dfrac{\bar{x}+v}{2} \notin \Omega_2$ and $\dfrac{\bar{x}+v}{2} \notin \Omega_3$, which tell us that $\dfrac{\bar{x}+v}{2} \in E$. But this cannot happen due to Proposition 4.48. Therefore, $v \in S$ and $\mathcal{I}(v) = \emptyset$, which contradict the result from Proposition 4.47 and thus justify that S is a singleton. □

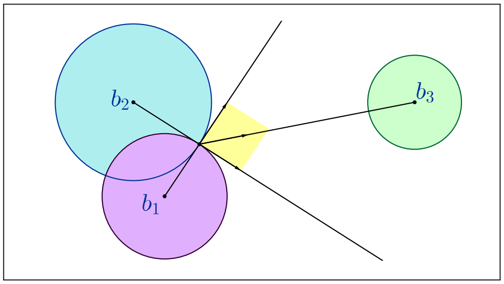

Figure 4.4: A Fermat-Torricelli problem with $|I(\bar{x})| = 2$.

Now we verify the expected boundary property of optimal solutions to (4.37).

Proposition 4.50 *Let $\bar{x} \in S$ and let $\mathcal{I}(\bar{x}) = \{i, j\} \subset \{1, 2, 3\}$. Then $\bar{x} \in \operatorname{bd}(\Omega_i \cap \Omega_j)$.*

Proof. Assume for definiteness $\mathcal{I}(\bar{x}) = \{1, 2\}$, i.e., $\bar{x} \in \Omega_1 \cap \Omega_2$ and $\bar{x} \notin \Omega_3$. Arguing by contradiction, suppose that $\bar{x} \in \operatorname{int}(\Omega_1 \cap \Omega_2)$, and therefore there exists $\delta > 0$ such that $I\!B(\bar{x}; \delta) \subset \Omega_1 \cap \Omega_2$. Denote $p := \Pi(\bar{x}; \Omega_3)$, $\gamma := \|\bar{x} - p\| > 0$ and take $q \in [\bar{x}, p] \cap I\!B(\bar{x}, \delta)$ satisfying the condition $\|q - p\| < \|\bar{x} - p\| = d(\bar{x}; \Omega_3)$. Then we get

$$\mathcal{H}(q) = \sum_{i=1}^{3} d(q; \Omega_i) = d(q; \Omega_3) \leq \|q - p\| < \|\bar{x} - p\| = d(\bar{x}; \Omega_3) = \sum_{i=1}^{3} d(\bar{x}; \Omega_i) = \mathcal{H}(\bar{x}),$$

which contradicts the fact that $\bar{x} \in S$ and thus completes the proof. □

The following result gives us verifiable necessary and sufficient conditions for the solution uniqueness in (4.37).

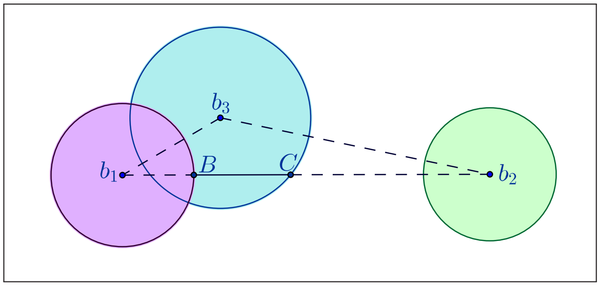

Figure 4.5: A generalized Fermat-Torricelli problem with infinite solutions.

Theorem 4.51 Let $\cap_{i=1}^{3}\Omega_i = \emptyset$. Then problem (4.37) has more than one solutions if and only if there is a set $[c_i, c_j] \cap \Omega_k$ for distinct indices $i, j, k \in \{1, 2, 3\}$ that contains a point $u \in \mathbb{R}^2$ belonging to the interior of Ω_k and such that the index set $\mathcal{I}(u)$ is a singleton.

Proof. Let $u \in [c_1, c_2] \cap \Omega_3$, $u \in \operatorname{int}\Omega_3$ satisfy the conditions in the "if" part of the theorem. Then $u \notin \Omega_1$, $u \notin \Omega_2$ and we can choose $v \neq u$ with $v \in [c_1, c_2]$, $v \notin \Omega_1$, $v \notin \Omega_2$, and $v \in \operatorname{int}\Omega_3$. This shows that the set $\mathcal{I}(v)$ is a singleton. Let us verify that in this case u is a solution to (4.37). To proceed, we observe first that

$$\mathcal{H}(u) = d(u; \Omega_1) + d(u; \Omega_2) + d(u; \Omega_3) = \|c_1 - c_2\| - r_1 - r_2.$$

Choose $\delta > 0$ such that $I\!B(u; \delta) \cap \Omega_1 = \emptyset$, $I\!B(u; \delta) \cap \Omega_2 = \emptyset$, and $I\!B(u; \delta) \subset \Omega_3$. For every $x \in I\!B(u; \delta)$, we get

$$\mathcal{H}(x) = d(x; \Omega_2) + d(x; \Omega_3) = \|x - c_1\| + \|x - c_2\| - r_1 - r_2 \geq \|c_1 - c_2\| - r_1 - r_2 = \mathcal{H}(u),$$

which implies that u is a local (and hence global) minimizer of $\mathcal{H}$. Similar arguments show that $v \in S$, so the solution set S is not a singleton.

To justify the "only if" part, suppose that S is not a singleton and pick two distinct elements $\bar{x}, \bar{y} \in S$, which yields $[\bar{x}, \bar{y}] \subset S$ by Proposition 4.45. If there is a solution to (4.37) not belonging to Ω_i for every $i \in \{1, 2, 3\}$, then S reduces to a singleton due to Proposition 4.47, a contradiction. Thus we can assume by the above that Ω_1 contains infinitely many solutions. Then its interior $\operatorname{int}\Omega_1$ also contains infinitely many solutions by the obvious strict convexity of the ball Ω_1. If there is such a solution u with $\mathcal{I}(u) = \{1\}$, then $u \in \operatorname{int}\Omega_1$ and $u \notin \Omega_2$ as well as $u \notin \Omega_3$. This implies that $u \in [c_2, c_3]$, since the opposite ensures the solution uniqueness by Proposition 4.49. Consider the case where the set $\mathcal{I}(u)$ contains two elements for every solution belonging to $\operatorname{int}\Omega_1$. Then there are infinitely many solutions belonging to the intersection of these two sets, which is strictly convex in this case. Hence there is a solution to (4.37) that belongs to the interior of this intersection, which contradicts Proposition 4.50 and thus completes the proof of the theorem. □

Let us now consider yet another particular case of problem (4.34), where the *constant dynamics* F is given by the *closed unit ball*. It is formulated as

$$\text{minimize } \mathcal{D}(x) := \sum_{i=1}^{m} d(x;\Omega_i) \text{ subject to } x \in \Omega_0. \tag{4.41}$$

For $x \in \mathbb{R}^n$, define the index subsets of $\{1,\ldots,m\}$ by

$$I(x) := \big\{ i \in \{1,\ldots,m\} \,\big|\, x \in \Omega_i \big\} \text{ and } J(x) := \big\{ i \in \{1,\ldots,m\} \,\big|\, x \notin \Omega_i \big\}. \tag{4.42}$$

It is obvious that $I(x) \cup J(x) = \{1,\ldots,m\}$ and $I(x) \cap J(x) = \emptyset$ for all $x \in \mathbb{R}^n$. The next theorem fully characterizes optimal solutions to (4.41) via projections to Ω_i.

Theorem 4.52 The following assertions hold for problem (4.41):

(i) Let $\bar{x} \in \Omega_0$ be a solution to (4.41). For $a_i(\bar{x}) \equiv A_i(\bar{x})$ as $i \in J(\bar{x})$, we have

$$-\sum_{i \in J(\bar{x})} a_i(\bar{x}) \in \sum_{i \in I(\bar{x})} A_i(\bar{x}) + N(\bar{x};\Omega_0), \tag{4.43}$$

where each set $A_i(\bar{x})$, $i = 1,\ldots,m$, is calculated by

$$A_i(\bar{x}) = \begin{cases} \dfrac{\bar{x} - \Pi(\bar{x};\Omega_i)}{d(\bar{x};\Omega_i)} & \text{for } \bar{x} \notin \Omega_i, \\[2ex] N(\bar{x};\Omega_i) \cap I\!B & \text{for } \bar{x} \in \Omega_i. \end{cases} \tag{4.44}$$

(ii) Conversely, if $\bar{x} \in \Omega_0$ satisfies (4.43), then $\bar{x}$ is a solution to (4.41).

Proof. To verify **(i)**, fix an optimal solution $\bar{x}$ to problem (4.41) and equivalently describe it as an optimal solution to the *unconstrained* optimization problem:

$$\text{minimize } \mathcal{D}(x) + \delta(x;\Omega), \quad x \in \mathbb{R}^n. \tag{4.45}$$

Applying the subdifferential Fermat rule to (4.45), we characterize $\bar{x}$ by

$$0 \in \partial\Big(\sum_{i=1}^{n} d(\cdot;\Omega_i) + \delta(\cdot;\Omega)\Big)(\bar{x}). \tag{4.46}$$

Since all of the functions $d(\cdot;\Omega_i)$ for $i = 1,\ldots,n$ are convex and continuous, we employ the subdifferential sum rule of Theorem 2.15 to (4.46) and arrive at

$$\begin{aligned}0 \in &\sum_{i\in J(\bar{x})} \partial d(\bar{x};\Omega_i) + \sum_{i\in I(\bar{x})} d(\bar{x};\Omega_i) + N(\bar{x};\Omega_0)\\ &= \sum_{i\in J(\bar{x})} A_i(\bar{x}) + \sum_{i\in I(\bar{x})} A_i(\bar{x}) + N(\bar{x};\Omega_0)\\ &= \sum_{i\in J(\bar{x})} a_i(\bar{x}) + \sum_{i\in I(\bar{x})} A_i(\bar{x}) + N(\bar{x};\Omega_0),\end{aligned}$$

which justifies the claimed inclusion

$$-\sum_{i\in J(\bar{x})} a_i(\bar{x}) \in \sum_{i\in I(\bar{x})} A_i(\bar{x}) + N(\bar{x};\Omega_0).$$

The converse assertion **(ii)** is verified similarly by taking into account the "if and only if" property of the proof of **(i)**. □

Let us specifically examine problem (4.41) in the case where the intersections of the constraint set Ω_0 with the target sets Ω_i are empty.

Corollary 4.53 *Consider problem* (4.41) *with* $\Omega_0 \cap \Omega_i = \emptyset$ *for* $i = 1,\ldots,m$. *Given* $\bar{x} \in \Omega_0$, *define the elements*

$$a_i(\bar{x}) := \frac{\bar{x} - \Pi(\bar{x};\Omega_i)}{d(\bar{x};\Omega_i)} \neq 0, \quad i = 1,\ldots,m. \tag{4.47}$$

Then $\bar{x} \in \Omega_0$ *is an optimal solution to* (4.41) *if and only if*

$$-\sum_{i=1}^{m} a_i(\bar{x}) \in N(\bar{x};\Omega_0). \tag{4.48}$$

Proof. In this case we have that $\bar{x} \notin \Omega_i$ for all $i = 1,\ldots,m$, which means that $I(\bar{x}) = \emptyset$ and $J(\bar{x}) = \{1,\ldots,m\}$. Thus our statement is a direct consequence of Theorem 4.52. □

For more detailed solution of (4.41), consider its special case of $m = 3$.

Theorem 4.54 Let $m = 3$ in the framework of Theorem 4.52, where the target sets Ω_1, Ω_2, and Ω_3 are pairwise disjoint, and where $\Omega_0 = \mathbb{R}^n$. The following hold for the optimal solution $\bar{x}$ to (4.41) with the sets $a_i(\bar{x})$ defined by (4.47):

(i) If $\bar{x}$ belongs to one of the sets Ω_i, say Ω_1, then we have

$$\langle a_2, a_3\rangle \le -1/2 \text{ and } -a_2 - a_3 \in N(\bar{x};\Omega_1).$$

(ii) If $\bar{x}$ does not belong to any of the three sets Ω_1, Ω_2, and Ω_3, then

$$\langle a_i, a_j \rangle = -1/2 \text{ for } i \neq j \text{ as } i, j \in \{1, 2, 3\}.$$

Conversely, if $\bar{x}$ satisfies either **(i)** or **(ii)**, then it is an optimal solution to (4.41).

Proof. To verify **(i)**, we have from Theorem 4.52 that $\bar{x} \in \Omega_1$ solves (4.41) if and only if

$$\begin{aligned} 0 &\in \partial d(\bar{x}; \Omega_1) + \partial d(\bar{x}; \Omega_2) + \partial d(\bar{x}; \Omega_3) \\ &= \partial d(\bar{x}; \Omega_1) + a_2 + a_3 = N(\bar{x}; \Omega_1) \cap I\!B + a_2 + a_3. \end{aligned}$$

Thus $-a_2 - a_3 \in N(\bar{x}; \Omega_1) \cap I\!B$, which is equivalent to the validity of $-a_2 - a_3 \in I\!B$ and $-a_2 - a_3 \in N(\bar{x}; \Omega_1)$. By Lemma 4.31, this is also equivalent to

$$\langle a_2, a_3 \rangle \leq -1/2 \text{ and } -a_2 - a_3 \in N(\bar{x}; \Omega_1),$$

which gives **(i)**. To justify **(ii)**, we similarly conclude that in the case where $\bar{x} \notin \Omega_i$ for $i = 1, 2, 3$, this element solves (4.41) if and only if

$$0 \in \partial d(\bar{x}; \Omega_1) + \partial d(\bar{x}; \Omega_2) + \partial d(\bar{x}; \Omega_3) = a_1 + a_2 + a_3.$$

Then Lemma 4.32 tells us that the latter is equivalent to

$$\langle a_i, a_j \rangle = -1/2 \text{ for } i \neq j \text{ as } i, j \in \{1, 2, 3\}.$$

Conversely, the validity of either **(i)** or **(ii)** and the convexity of Ω_i ensure that

$$0 \in \partial d(\bar{x}; \Omega_1) + \partial d(\bar{x}; \Omega_2) + \partial d(\bar{x}; \Omega_3) = \partial \mathcal{D}(\bar{x}),$$

and thus $\bar{x}$ is an optimal solution to the problem under consideration. □

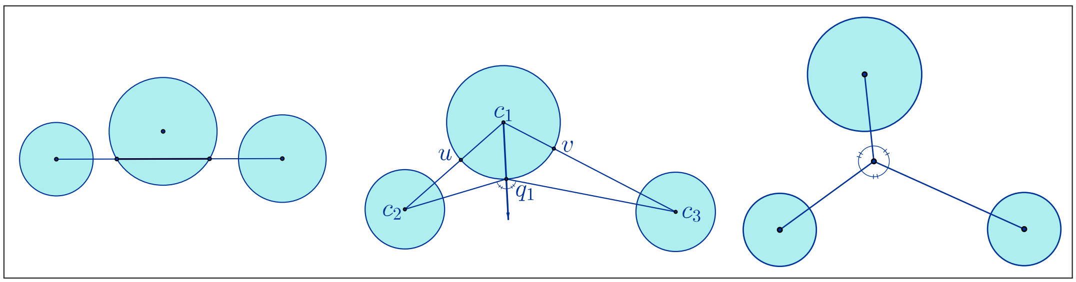

Figure 4.6: Generalized Fermat-Torricelli problem for three disks.

Example 4.55 Let the sets Ω_1, Ω_2, and Ω_3 in problem (4.37) be closed balls in $\mathbb{R}^2$ of radius $r = 1$ centered at the points $(0, 2)$, $(-2, 0)$, and $(2, 0)$, respectively. We can easily see that the point

$(0, 1) \in \Omega_1$ satisfies all the conditions in Theorem 4.54(i), and hence it is an optimal solution (in fact the unique one) to this problem.

More generally, consider problem (4.37) in $\mathbb{R}^2$ generated by three pairwise disjoint disks denoted by Ω_i, $i = 1, 2, 3$. Let c_1, c_2, and c_3 be the centers of the disks. Assume first that either the line segment $[c_2, c_3] \in \mathbb{R}^2$ intersects Ω_1, or $[c_1, c_3]$ intersects Ω_2, or $[c_1, c_2]$ intersects Ω_3. It is not hard to check that any point of the intersections (say, of the sets Ω_1 and $[c_2, c_3]$ for definiteness) is an optimal solution to the problem under consideration, since it satisfies the necessary and sufficient optimality conditions of Theorem 4.54(i). Indeed, if $\bar{x}$ is such a point, then a_2 and a_3 are unit vectors with $\langle a_2, a_3 \rangle = -1$ and $-a_2 - a_3 = 0 \in N(\bar{x}; \Omega_1)$. If the above intersection assumptions are violated, we define the three points q_1, q_2, and q_3 as follows. Let u and v be intersection points of $[c_1, c_2]$ and $[c_1, c_3]$ with the boundary of the disk centered at c_1, respectively. Then we see that there is a unique point q_1 on the minor curve generated by u and v such that the measures of angle $c_1 q_1 c_2$ and $c_1 q_1 c_3$ are equal. The points q_2 and q_3 are defined similarly. Theorem 4.54 yields that whenever the angle $c_2 q_1 c_3$, or $c_1 q_2 c_3$, or $c_2 q_3 c_1$ equals or exceeds $120°$ (say, the angle $c_2 q_1 c_3$ does), then the point $\bar{x} := q_1$ is an optimal solution to the problem under consideration. Indeed, in this case a_2 and a_3 from (4.47) are unit vectors with $\langle a_2, a_3 \rangle \le -1/2$ and $-a_2 - a_3 \in N(\bar{x}; \Omega_1)$ because the vector $-a_2 - a_3$ is *orthogonal* to Ω_1.

If none of these angles equals or exceeds $120°$, there is a point q not belonging to Ω_i as $i = 1, 2, 3$ such that the angles $c_1 q c_2 = c_2 q c_3 = c_3 q c_1$ are of $120°$ and that q is an optimal solution to the problem under consideration. Observe that in this case the point q is also a unique optimal solution to the classical Fermat-Torricelli problem determined by the points c_1, c_2, and c_3; see Figure 4.6.

Finally in this section, we discuss the application of the *subgradient algorithm* from Section 4.3 to the numerical solution of the generalized Fermat-Torricelli problem (4.34).

Theorem 4.56 Consider the generalized Fermat-Torricelli problem (4.34) and assume that $S \neq \emptyset$ for its solution set. Picking a sequence $\{\alpha_k\}$ of positive numbers and a starting point $x_1 \in \Omega_0$, form the iterative sequence

$$x_{k+1} := \Pi\Big(x_k - \alpha_k \sum_{i=1}^{m} v_{ik}; \Omega_0\Big), \quad k = 1, 2, \ldots, \tag{4.49}$$

with an arbitrary choice of subgradient elements

$$v_{ik} \in -\partial\rho_F(\omega_{ik} - x_k) \cap N(\omega_{ik}; \Omega_i) \text{ for some } \omega_{ik} \in \Pi_F(x_k; \Omega_i) \text{ if } x_k \notin \Omega_i \tag{4.50}$$

and $v_{ik} := 0$ otherwise, where $\Pi_F(x; \Omega)$ stands for the generalized projection operator (3.21), and where ρ_F is the Minkowski gauge (3.11). Define the value sequence

$$V_k := \min\big\{\mathcal{H}(x_j) \;\big|\; j = 1, \ldots, k\big\}.$$

(i) If the sequence $\{\alpha_k\}$ satisfies conditions (4.13), then $\{V_k\}$ converges to the optimal value $\overline{V}$ of problem (4.34). Furthermore, we have the estimate

$$0 \le V_k - \overline{V} \le \frac{d(x_1;S)^2 + \ell^2 \sum_{k=1}^{\infty} \alpha_k^2}{2\sum_{i=1}^{k} \alpha_k},$$

where $\ell \ge 0$ is a Lipschitz constant of cost function $\mathcal{H}(\cdot)$ on $\mathbb{R}^n$.
(ii) If the sequence $\{\alpha_k\}$ satisfies conditions (4.15), then $\{V_k\}$ converges to the optimal value $\overline{V}$ and $\{x_k\}$ in (4.49) converges to an optimal solution $\bar{x} \in S$ of problem (4.34).

Proof. This is a direct application of the projected subgradient algorithm described in Proposition 4.28 to the constrained optimization problem (4.34) by taking into account the results for the subgradient method in convex optimization presented in Section 4.3. □

Let us specify the above algorithm for the case of the unit ball $F = I\!B$ in (4.34).

Corollary 4.57 *Let $F = I\!B \subset \mathbb{R}^n$ in (4.34), let $\{\alpha_k\}$ be a sequence of positive numbers, and let $x_1 \in \Omega_0$ be a starting point. Consider algorithm (4.49) with $\partial\rho_F(\cdot)$ calculated by*

$$\partial\rho_F(\omega_{ki} - x_k) = \begin{cases} \dfrac{\Pi(x_k;\Omega_i) - x_k}{d(x_k;\Omega_i)} & \text{if } x_k \notin \Omega_i, \\[2ex] 0, & \text{if } x_k \in \Omega_i. \end{cases}$$

Then all the conclusions of Theorem 4.56 hold for this specification.

Proof. Immediately follows from Theorem 4.56 with $\rho_F(x) = \|x\|$. □

Next we illustrate applications of the subgradient algorithm of Theorem 4.56 to solving the generalized Fermat-Torricelli problem (4.34) formulated via the minimal time function with *non-ball dynamics*. Note that to implement the subgradient algorithm (4.49), it is sufficient to calculate just one subgradient from the set on the right-hand side of (4.50) for each target set. In the following example we consider for definiteness the dynamics F given by the square $[-1,1] \times [-1,1]$ in the plane. In this case the corresponding Minkowski gauge (3.11) is given by the formula

$$\rho_F(x_1,x_2) = \max\big\{|x_1|,|x_2|\big\}, \quad (x_1,x_2) \in \mathbb{R}^2. \tag{4.51}$$

Example 4.58 Consider the unconstrained generalized Fermat-Torricelli problem (4.34) with the dynamics $F = [-1,1] \times [-1,1]$ and the square targets Ω_i of *right position* (with sides parallel to the axes) centered at (a_i, b_i) with radii r_i (half of the side) as $i = 1,\ldots,m$. Given a sequence

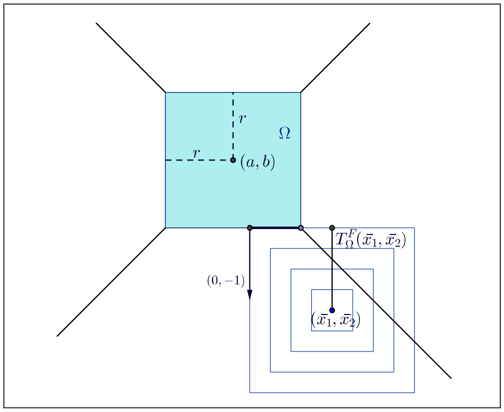

Figure 4.7: Minimal time function with square dynamic

of positive numbers $\{\alpha_k\}$ and a starting point x_1, construct the iterations $x_k = (x_{1k}, x_{2k})$ by algorithm (4.49). By Theorem 3.39, subgradients v_{ik} can be chosen by

$$v_{ik} = \begin{cases} (1,0) & \text{if } |x_{2k} - b_i| \le x_{1k} - a_i \text{ and } x_{1k} > a_i + r_i, \\ (-1,0) & \text{if } |x_{2k} - b_i| \le a_i - x_{1k} \text{ and } x_{1k} < a_i - r_i, \\ (0,1) & \text{if } |x_{1k} - a_i| \le x_{2k} - b_i \text{ and } x_{2k} > b_i + r_i, \\ (0,-1) & \text{if } |x_{1k} - a_i| \le b_i - x_{2k} \text{ and } x_{2k} < b_i - r_i, \\ (0,0) & \text{otherwise.} \end{cases}$$

4.6 A GENERALIZED SYLVESTER PROBLEM

In [31] James Joseph Sylvester proposed the *smallest enclosing circle problem*: Given a finite number of points in the plane, find the smallest circle that encloses all of the points. After more than a century, location problems of this type remain active as they are mathematically beautiful and have meaningful real-world applications.

This section is devoted to the study by means of convex analysis of the following generalized version of the Sylvester problem formulated in [18]; see also [19, 20] for further developments. Let $F \subset \mathbb{R}^n$ be a closed, bounded, convex set containing the origin as an interior point. Define the *generalized ball* centered at $x \in \mathbb{R}^n$ with radius $r > 0$ by

$$G_F(x; r) := x + rF.$$

It is obvious that for $F = I\!B$ the generalized ball $G_F(x;r)$ reduces to the closed ball of radius r centered at x. Given now finitely many nonempty, closed, convex sets Ω_i for $i = 0, \ldots, m$, the *generalized Sylvester problem* is formulated as follows: Find a point $x \in \Omega_0$ and the smallest number $r \geq 0$ such that

$$G_F(x;r) \cap \Omega_i \neq \emptyset \text{ for all } i = 1, \ldots, m.$$

Observe that when $\Omega_0 = \mathbb{R}^2$ and all the sets Ω_i for $i = 1, \ldots, m$ are singletons in $\mathbb{R}^2$, the generalized Sylvester problem becomes the classical Sylvester enclosing circle problem. The sets Ω_i for $i = 1, \ldots, m$ are often called the *target sets*.

To study the generalized Sylvester problem, we introduce the cost function

$$\mathcal{T}(x) := \max\left\{\mathcal{T}_F(x;\Omega_i) \;\middle|\; i = 1, \ldots, m\right\}$$

via the minimal time function (3.9) and consider the following optimization problem:

$$\text{minimize } \mathcal{T}(x) \text{ subject to } x \in \Omega_0. \tag{4.52}$$

We can show that the generalized ball $G_F(x;r)$ with $\bar{x} \in \Omega_0$ is an optimal solution of the generalized Sylvester problem if and only if $\bar{x}$ is an optimal solution of the optimization problem (4.52) with $r = \mathcal{T}(\bar{x})$.

Our first result gives sufficient conditions for the *existence* of solutions to (4.52).

Proposition 4.59 *The generalized Sylvester problem (4.52) admits an optimal solution if the set $\Omega_0 \cap \big[\bigcap_{i=1}^m (\Omega_i - \alpha F)\big]$ is bounded for all $\alpha \geq 0$. This holds, in particular, if there exists an index $i_0 \in \{0, \ldots, m\}$ such that the set Ω_{i_0} is compact.*

Proof. For any $\alpha \geq 0$, consider the level set

$$\mathcal{L}_\alpha := \left\{x \in \Omega_0 \;\middle|\; \mathcal{T}(x) \leq \alpha\right\}$$

and then show that in fact we have

$$\mathcal{L}_\alpha = \Omega_0 \cap [\bigcap_{i=1}^m (\Omega_i - \alpha F)]. \tag{4.53}$$

Indeed, it holds for any nonempty, closed, convex set Q that $\mathcal{T}_F(x;Q) \leq \alpha$ if and only if $(x + \alpha F) \cap Q \neq \emptyset$, which is equivalent to $x \in Q - \alpha F$. If $x \in \mathcal{L}_\alpha$, then $x \in \Omega_0$ and $\mathcal{T}_F(x;\Omega_i) \leq \alpha$ for $i = 1, \ldots, m$. Thus $x \in \Omega_i - \alpha F$ for $i = 1, \ldots, m$, which verifies the inclusion "$\subset$" in (4.53). The proof of the opposite inclusion is also straightforward. It is clear furthermore that the boundedness of the set $\Omega_0 \cap \big[\bigcap_{i=1}^m (\Omega_i - \alpha F)\big]$ for all $\alpha \geq 0$ ensures that the level sets for the cost function in (4.52) are bounded. Thus it follows from Theorem 4.11 that problem (4.52) has an optimal solution. The last statement is obvious. $\square$

Now we continue with the study of the *uniqueness* of optimal solutions to (4.52) splitting the proof into several propositions.

Proposition 4.60 *Consider a convex function $g(x) \geq 0$ on a nonempty, convex set $\Omega \subset \mathbb{R}^n$. Then the function defined by $h(x) := (g(x))^2$ is strictly convex on Ω if and only if g is not constant on any line segment $[a, b] \subset \Omega$ with $a \neq b$.*

Proof. Suppose that h is strictly convex on Ω and suppose on the contrary that g is constant on a line segment $[a, b]$ with $a \neq b$, then h is also constant on this line segment, so it cannot be strictly convex. Conversely, suppose g is not constant on any segment $[a, b] \subset \Omega$ with $a \neq b$. Observe that h is a convex function on Ω since it is a composition of a quadratic function, which is a nondecreasing convex function on $[0, \infty)$, and a convex function whose range is a subset of $[0, \infty)$. Let us show that it is strictly convex on Ω. Suppose by contradiction that there are $t \in (0, 1)$ and $x, y \in \Omega$, $x \neq y$, with

$$h(z) = th(x) + (1-t)h(y) \text{ with } z := tx + (1-t)y.$$

Then $(g(z))^2 = t(g(x))^2 + (1-t)(g(y))^2$. Since $g(z) \leq tg(x) + (1-t)g(y)$, one has

$$\big(g(z)\big)^2 \leq t^2\big(g(x)\big)^2 + (1-t)^2\big(g(y)\big)^2 + 2t(1-t)g(x)g(y),$$

which readily implies that

$$t\big(g(x)\big)^2 + (1-t)\big(g(y)\big)^2 \leq t^2\big(g(x)\big)^2 + (1-t)^2\big(g(y)\big)^2 + 2t(1-t)g(x)g(y).$$

Thus $(g(x) - g(y))^2 \leq 0$, so $g(x) = g(y)$. This shows that $h(x) = h(z) = h(y)$ for the point $z \in (x, y)$ defined above. We arrive at a contradiction by showing that h is constant on the line segment $[z, y]$. Indeed, fix any $u \in (z, y)$ and observe that

$$h(u) \leq \nu h(z) + (1-\nu)h(y) = h(z) \text{ for some } \nu \in (0, 1).$$

On the other hand, since z lies between x and u, we have

$$h(z) \leq \mu h(x) + (1-\mu)h(u) \leq \mu h(z) + (1-\mu)h(z) = h(z) \text{ for some } \mu \in (0, 1).$$

It gives us $h(z) = \mu h(x) + (1-\mu)h(u) = \mu h(z) + (1-\mu)h(u)$, and hence $h(u) = h(z)$. This contradicts the assumption that $g(u) = \sqrt{h(u)}$ is not constant on any line segment $[a, b]$ with $a \neq b$ and thus completes the proof of the proposition. □

Proposition 4.61 *Let both sets F and Ω in $\mathbb{R}^n$ be nonempty, closed, strictly convex, where F contains the origin as an interior point. Then the convex function $\mathcal{T}_F(\cdot; \Omega)$ is not constant on any line segment $[a, b] \subset \mathbb{R}^n$ such that $a \neq b$ and $[a, b] \cap \Omega = \emptyset$.*

Proof. To prove the statement of the proposition, suppose on the contrary that there is $[a,b] \subset \mathbb{R}^n$ with $a \neq b$ such that $[a,b] \cap \Omega = \emptyset$ and $\mathcal{T}_F(x;\Omega) = r$ for all $x \in [a,b]$. Choosing $u \in \Pi_F(a;\Omega)$ and $v \in \Pi_F(b;\Omega)$, we get $\rho_F(u-a) = \rho_F(v-b) = r$. Let us verify that $u \neq v$. Indeed, the opposite gives us $\rho_F(u-a) = \rho_F(u-b) = r$, which implies by $r > 0$ that

$$\rho_F\Big(\frac{u-a}{r}\Big) = \rho_F\Big(\frac{u-b}{r}\Big) = 1,$$

and thus $\dfrac{u-a}{r} \in \operatorname{bd} F$ and $\dfrac{u-b}{r} \in \operatorname{bd} F$. Since F is strictly convex, it yields

$$\frac{1}{2}\Big(\frac{u-a}{r}\Big) + \frac{1}{2}\Big(\frac{u-b}{r}\Big) = \frac{1}{r}\Big(u - \frac{a+b}{2}\Big) \in \operatorname{int} F.$$

Hence we arrive at a contradiction by having

$$\mathcal{T}_F\Big(\frac{a+b}{2};\Omega\Big) \le \rho_F\Big(u - \frac{a+b}{2}\Big) < r.$$

To finish the proof of the proposition, fix $t \in (0,1)$ and get

$$\begin{aligned} r = \mathcal{T}_F\big(ta + (1-t)b;\Omega\big) &\le \rho_F\big(tu + (1-t)v - (ta + (1-t)b)\big) \\ &\le t\rho_F(u-a) + (1-t)\rho_F(v-b) = r, \end{aligned}$$

which implies that $tu + (1-t)v \in \Pi_F(ta + (1-t)b;\Omega)$. Thus $tu + (1-t)v \in \operatorname{bd}\Omega$, which contradicts the strict convexity of Ω and completes the proof. $\square$

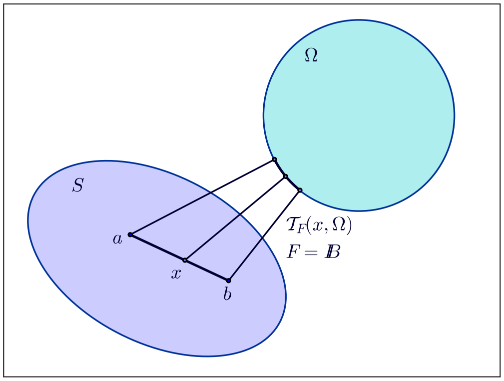

Figure 4.8: Minimal time function for strictly convex sets.

Proposition 4.62 *Let $\Omega \subset \mathbb{R}^n$ be a nonempty, closed, convex set and let $h_i\colon \Omega \to \mathbb{R}$ be nonnegative, continuous, convex functions for $i = 1,\dots,m$. Define*

$$\phi(x) := \max\big\{h_1(x),\dots,h_m(x)\big\}$$

and suppose that $\phi(x) = r > 0$ *on* $[a, b]$ *for some line segment* $[a, b] \subset \Omega$ *with* $a \neq b$. *Then there exists a line segment* $[\alpha, \beta] \subset [a, b]$ *with* $\alpha \neq \beta$ *and an index* $i_0 \in \{1, \ldots, m\}$ *such that we have* $h_{i_0}(x) = r$ *for all* $x \in [\alpha, \beta]$.

Proof. There is nothing to prove for $m = 1$. When $m = 2$, consider the function

$$\phi(x) := \max\big\{h_1(x), h_2(x)\big\}.$$

Then the conclusion is obvious if $h_1(x) = r$ for all $x \in [a, b]$. Otherwise we find $x_0 \in [a, b]$ with $h_1(x_0) < r$ and a subinterval $[\alpha, \beta] \subset [a, b]$, $\alpha \neq \beta$, such that

$$h_1(x) < r \quad \text{for all} \quad x \in [\alpha, \beta].$$

Thus $h_2(x) = r$ on this subinterval. To proceed further by induction, we assume that the conclusion holds for a positive integer $m \geq 2$ and show that the conclusion holds for $m + 1$ functions. Denote

$$\phi(x) := \max\big\{h_1(x), \ldots, h_m(x), h_{m+1}(x)\big\}.$$

Then we have $\phi(x) = \max\{h_1(x), k_1(x)\}$ with $k_1(x) := \max\{h_2(x), \ldots, h_{m+1}(x)\}$, and the conclusion follows from the case of two functions and the induction hypothesis. □

Now we are ready to study the *uniqueness* of an optimal solution to the generalized Sylvester problem (4.52).

Theorem 4.63 In the setting of problem (4.52) suppose further that the sets F and Ω_i for $i = 1, \ldots, m$ are strictly convex. Then problem (4.52) has at most one optimal solution if and only if $\cap_{i=0}^m \Omega_i$ contains at most one element.

Proof. Define $\mathcal{K}(x) := (\mathcal{T}(x))^2$ and consider the optimization problem

$$\text{minimize} \;\; \mathcal{K}(x) \;\; \text{subject to} \;\; x \in \Omega_0. \tag{4.54}$$

It is obvious that $\bar{x} \in \Omega_0$ is an optimal solution to (4.52) if and only if it is an optimal solution to problem (4.54). We are going to prove that if $\cap_{i=0}^m \Omega_i$ contains at most one element, then $\mathcal{K}$ is strictly convex on Ω_0, and hence problem (4.52) has at most one optimal solution. Indeed, suppose that $\mathcal{K}$ is not strictly convex on Ω_0. By Proposition 4.60, we find a line segment $[a, b] \subset \Omega_0$ with $a \neq b$ and a number $r \geq 0$ so that

$$\mathcal{T}(x) = \max\big\{\mathcal{T}_F(x; \Omega_i) \;\big|\; i = 1, \ldots, m\big\} = r \quad \text{for all} \quad x \in [a, b].$$

It is clear that $r > 0$, since otherwise $[a, b] \subset \bigcap_{i=0}^m \Omega_i$, which contradicts the assumption that this set contains at most one element. Proposition 4.62 tells us that there exist a line segment $[\alpha, \beta] \subset [a, b]$ with $\alpha \neq \beta$ and an index $i_0 \in \{1, \ldots, m\}$ such that

$$\mathcal{T}_F(x; \Omega_{i_0}) = r \quad \text{for all} \quad x \in [\alpha, \beta].$$

Since $r > 0$, we have $x \notin \Omega_{i_0}$ for all $x \in [\alpha, \beta]$, which contradicts Proposition 4.61.

Conversely, let problem (4.52) have at most one solution. Suppose now that the set $\cap_{i=0}^m \Omega_i$ contains more than one element. It is clear that $\bigcap_{i=0}^m \Omega_i$ is the solution set of (4.52) with the optimal value equal to zero. Thus this problem has more than one solution, which is a contradiction that completes the proof of the theorem. □

The following corollary gives a sufficient condition for the uniqueness of optimal solutions to the generalized Sylvester problem (4.52).

Corollary 4.64 *In the setting of* (4.52) *suppose that the sets F and Ω_i for $i = 1, \ldots, m$ are strictly convex. If $\bigcap_{i=0}^m \Omega_i$ contains at most one element and if one of the sets among Ω_i for $i = 0, \ldots, m$ is bounded, then problem* (4.52) *has a unique optimal solution.*

Proof. Since $\bigcap_{i=0}^m \Omega_i$ contains at most one element, problem (4.52) has at most one optimal solution by Theorem 4.63. Thus it has exactly one solution because the solution set is nonempty by Proposition 4.59. □

Next we consider a special unconstrained case of (4.52) given by

$$\text{minimize } \mathcal{M}(x) := \max\left\{d(x; \Omega_i) \;\middle|\; i = 1, \ldots, m\right\} \text{ subject to } x \in \mathbb{R}^n, \tag{4.55}$$

for nonempty, closed, convex sets Ω_i under the assumption that $\cap_{i=1}^n \Omega_i = \emptyset$. Problem (4.55) corresponds to (4.52) with $F = I\!B \subset \mathbb{R}^n$. Denote the set of *active indices* for (4.55) by

$$I(x) := \left\{i \in \{1, \ldots, m\} \;\middle|\; \mathcal{M}(x) = d(x; \Omega_i)\right\}, \quad x \in \mathbb{R}^n.$$

Further, for each $x \notin \Omega_i$, define the singleton

$$a_i(x) := \frac{x - \Pi(x; \Omega_i)}{d(x; \Omega_i)} \tag{4.56}$$

where the projection $\Pi(x; \Omega_i)$ is unique. We have $\mathcal{M}(x) > 0$ as $i \in I(x)$ due to the assumption $\bigcap_{i=1}^m \Omega_i = \emptyset$. This ensures that $x \notin \Omega_i$ automatically for any $i \in I(x)$.

Let us now present *necessary and sufficient optimality conditions* for (4.55) based on the general results of convex analysis and optimization developed above.

Theorem 4.65 The element $\bar{x} \in \mathbb{R}^n$ is an optimal solution to problem (4.55) under the assumptions made if and only if we have the inclusion

$$\bar{x} \in \operatorname{co}\left\{\Pi(\bar{x}; \Omega_i) \;\middle|\; i \in I(\bar{x})\right\} \tag{4.57}$$

In particular, for the case where $\Omega_i = \{a_i\}$, $i = 1, \ldots, m$, problem (4.55) has a unique solution and $\bar{x}$ is the problem generated by a_i if and only if

$$\bar{x} \in \operatorname{co}\left\{a_i \;\middle|\; i \in I(\bar{x})\right\}. \tag{4.58}$$

Proof. It follows from Proposition 2.35 and Proposition 2.54 that the condition

$$0 \in \text{co} \bigcup_{i \in I(\bar{x})} \partial d(\bar{x}; \Omega_i)$$

is necessary and sufficient for optimality of $\bar{x}$ in (4.55). Employing now the distance function subdifferentiation from Theorem 2.39 gives us the minimum characterization

$$0 \in \text{co}\big\{a_i(\bar{x}) \bigm| i \in I(\bar{x})\big\}, \tag{4.59}$$

via the singletons $a_i(\bar{x})$ calculated in (4.56). It is easy to observe that (4.59) holds if and only if there are $\lambda_i \geq 0$ for $i \in I(\bar{x})$ such that $\sum_{i \in I(\bar{x})} \lambda_i = 1$ and

$$0 = \sum_{i \in I(\bar{x})} \lambda_i \frac{\bar{x} - \bar{\omega}_i}{\mathcal{M}(\bar{x})} \text{ with } \omega_i := \Pi(\bar{x}; \Omega_i).$$

This equation is clearly equivalent to

$$0 = \sum_{i \in I(\bar{x})} \lambda_i (\bar{x} - \bar{\omega}_i), \text{ or } \bar{x} = \sum_{i \in I(\bar{x})} \lambda_i \bar{\omega}_i,$$

which in turn is equivalent to inclusion (4.57). Finally, for $\Omega_i = \{a_i\}$ as $i = 1, \ldots, m$, the latter inclusion obviously reduces to condition (4.58). In this case the problem has a unique optimal solution by Corollary 4.64. □

To formulate the following result, we say that *the ball* $I\!B(\bar{x}; r)$ *touches the set* Ω_i if the intersection set $\Omega_i \cap I\!B(\bar{x}; r)$ is a singleton.

Corollary 4.66 *Consider problem* (4.55) *under the assumptions above. Then any smallest intersecting ball touches at least two target sets among* Ω_i, $i = 1, \ldots, m$.

Proof. Let us first verify that there are at least two indices in $I(\bar{x})$ provided that $\bar{x}$ is a solution to (4.55). Suppose the contrary, i.e., $I(\bar{x}) = \{i_0\}$. Then it follows from (4.57) that

$$\bar{x} \in \Pi(\bar{x}; \Omega_{i_0}) \text{ and } d(\bar{x}; \Omega_{i_0}) = \mathcal{M}(\bar{x}) > 0,$$

a contradiction. Next we show that $I\!B(\bar{x}; r)$ with $r := \mathcal{M}(\bar{x})$ touches Ω_i when $i \in I(\bar{x})$. Indeed, if on the contrary there are $u, v \in \Omega_i$ such that

$$u, v \in I\!B(\bar{x}; r) \cap \Omega_i \text{ with } u \neq v,$$

Then, by taking into account that $d(\bar{x}; \Omega_i) = r$, we get

$$\|u - \bar{x}\| \leq r = d(\bar{x}; \Omega_i) \text{ and } \|v - \bar{x}\| \leq r = d(\bar{x}; \Omega_i),$$

so $u, v \in \Pi(\bar{x}; \Omega_i)$. This contradicts the fact that the projector $\Pi(\bar{x}; \Omega_i)$ is a singleton and thus completes the proof of the corollary. □

Now we illustrate the results obtained by solving the following two examples.

Example 4.67 Let $\bar{x}$ solve the smallest enclosing ball problem generated by $a_i \in \mathbb{R}^2, i = 1, 2, 3$. Corollary 4.66 tells us that there are either two or three active indices. If $I(\bar{x})$ consists of two indices, say $I(\bar{x}) = \{2, 3\}$, then $\bar{x} \in \text{co}\{a_2, a_3\}$ and $\|\bar{x} - a_2\| = \|\bar{x} - a_3\| > \|\bar{x} - a_1\|$ by Theorem 4.65. In this case we have $\bar{x} = \dfrac{a_2 + a_3}{2}$ and $\langle a_2 - a_1, a_3 - a_1 \rangle < 0$. If conversely $\langle a_2 - a_1, a_3 - a_1 \rangle < 0$, then the angle of the triangle formed by a_1, a_2, and a_3 at the vertex a_1 is *obtuse*; we include here the case where all the three points a_i are on a straight line. It is easy to see that $I\Big(\dfrac{a_2 + a_3}{2}\Big) = \{2, 3\}$ and that the point $\bar{x} = \dfrac{a_2 + a_3}{2}$ satisfies the assumption of Theorem 4.65 thus solving the problem. Observe that in this case the solution of (4.55) is the *midpoint* of the side opposite to the obtuse vertex.

If none of the angles of the triangle formed by a_1, a_2, a_3 is obtuse, then the active index set $I(\bar{x})$ consists of three elements. In this case $\bar{x}$ is the unique point satisfying

$$\bar{x} \in \text{co}\big\{a_1, a_2, a_3\big\} \text{ and } \|\bar{x} - a_1\| = \|\bar{x} - a_2\| = \|\bar{x} - a_3\|.$$

In the other words, $\bar{x}$ is the *center of the circumscribing circle of the triangle*.

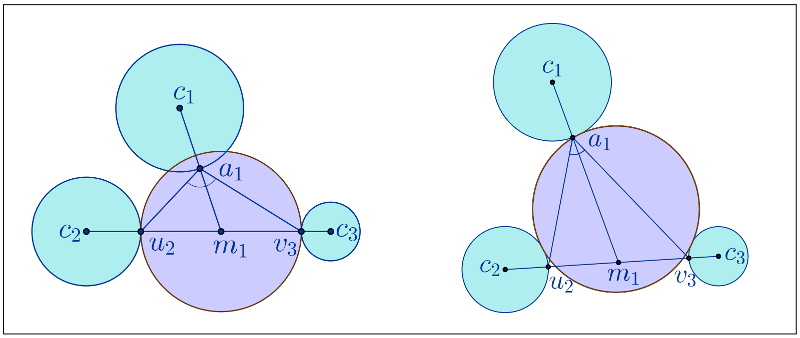

Figure 4.9: Smallest intersecting ball problem for three disks.

Example 4.68 Let $\Omega_i = I\!B(c_i, r_i)$ with $r_i > 0$ and $i = 1, 2, 3$ be *disjoint disks* in $\mathbb{R}^2$, and let $\bar{x}$ be the unique solution of the corresponding problem (4.55). Consider first the case where one of the line segments connecting two of the centers intersects the other disks, e.g., the line segment connecting c_2 and c_3 intersects Ω_1. Denote $u_2 := \overline{c_2 c_3} \cap \text{bd}\,\Omega_2$ and $u_3 := \overline{c_2 c_3} \cap \text{bd}\,\Omega_3$, and let $\bar{x}$ be the midpoint of $\overline{u_2 u_3}$. Then $I(\bar{x}) = \{2, 3\}$ and we can apply Theorem 4.65 to see that $\bar{x}$ is the solution of the problem.

It remains to consider the case where any line segment connecting two centers of the disks does not intersect the remaining disk. In this case denote

$$\begin{aligned}
&u_1 := \overline{c_1c_2} \cap \operatorname{bd}\Omega_1, \quad v_1 := \overline{c_1c_3} \cap \operatorname{bd}\Omega_1,\\
&u_2 := \overline{c_2c_3} \cap \operatorname{bd}\Omega_2, \quad v_2 := \overline{c_2c_1} \cap \operatorname{bd}\Omega_2,\\
&u_3 := \overline{c_3c_1} \cap \operatorname{bd}\Omega_3, \quad v_3 := \overline{c_3c_2} \cap \operatorname{bd}\Omega_3,\\
&a_1 := \overline{c_1m_1} \cap \operatorname{bd}\Omega_1,\ a_2 = \overline{c_2m_2} \cap \operatorname{bd}\Omega_2,\ a_3 := \overline{c_3m_3} \cap \operatorname{bd}\Omega_3,
\end{aligned}$$

where m_1 is the midpoint of $\overline{u_2v_3}$, m_2 is the midpoint of $\overline{u_3v_1}$, and m_3 is the midpoint of $\overline{u_1v_2}$. Suppose that one of the angles $\widehat{u_2a_1v_3}$, $\widehat{u_3a_2v_1}$, $\widehat{u_1a_3v_2}$ is greater than or equal to 90°; e.g., $\widehat{u_2a_1v_3}$ is greater than 90°. Then $I(m_1) = \{2, 3\}$, and thus $\bar{x} = m_1$ is the unique solution of this problem by Theorem 4.65. If now we suppose that all of the aforementioned angles are less than or equal to 90°, then $I(\bar{x}) = 3$ and the smallest disk we are looking for is the unique disk that touches three other disks. The construction of this disk is the celebrated *problem of Apollonius*.

Finally in this section, we present and illustrate the *subgradient algorithm* to solve numerically the generalized Sylvester problem (4.52).

Theorem 4.69 In the setting of problem (4.52) suppose that the solution set S of the problem is nonempty. Fix $x_1 \in \Omega_0$ and define the sequences of iterates by

$$x_{k+1} := \Pi\big(x_k - \alpha_k v_k; \Omega_0\big),\ k \in \mathbb{N},$$

where $\{\alpha_k\}$ is a sequence of positive numbers, and where

$$v_k \in \begin{cases} \{0\} & \text{if } x_k \in \Omega_i,\\ \big[-\partial\rho_F(\omega_k - x_k)\big] \cap N(\omega_k; \Omega_i) & \text{if } x_k \notin \Omega_i, \end{cases}$$

where $\omega_k \in \Pi_F(x_k; \Omega_i)$ for some $i \in I(x_k)$. Define the value sequence

$$V_k := \min\big\{\mathcal{T}(x_j)\ \big|\ j = 1, \ldots, k\big\}.$$

(i) If the sequence $\{\alpha_k\}$ satisfies conditions (4.13), then $\{V_k\}$ converges to the optimal value $\overline{V}$ of the problem. Furthermore, we have the estimate

$$0 \le V_k - \overline{V} \le \frac{d(x_1; S)^2 + \ell^2 \sum_{k=1}^{\infty} \alpha_k^2}{2\sum_{i=1}^{k} \alpha_k},$$

where $\ell \ge 0$ is a Lipschitz constant of cost function $\mathcal{T}(\cdot)$ on $\mathbb{R}^n$.
(ii) If the sequence $\{\alpha_k\}$ satisfies conditions (4.15), then $\{V_k\}$ converges to the optimal value $\overline{V}$ and $\{x_k\}$ in (4.49) converges to an optimal solution $\bar{x} \in S$ of the problem.

Proof. Follows from the results of Section 4.3 applied to problem (4.52). □

The next examples illustrate the application of the subgradient algorithm of Theorem 4.69 in particular situations.

Example 4.70 Consider problem (4.55) in $\mathbb{R}^2$ with m target sets given by the squares

$$\Omega_i = S(\omega_i; r_i) := \big[\omega_{1i} - r_i, \omega_{1i} + r_i\big] \times \big[\omega_{2i} - r_i, \omega_{2i} + r_i\big], \quad i = 1, \ldots, m.$$

Denote the vertices of the ith square by $v_{1i} := (\omega_{1i} + r_i, \omega_{2i} + r_i)$, $v_{2i} := (\omega_{1i} - r_i, \omega_{2i} + r_i)$, $v_{3i} := (\omega_{1i} - r_i, \omega_{2i} - r_i)$, $v_{4i} := (\omega_{1i} + r_i, \omega_{2i} - r_i)$, and let $x_k := (x_{1k}, x_{2k})$. Fix an index $i \in I(x_k)$ and then choose the elements v_k from Theorem 4.69 by

$$v_k = \begin{cases} \dfrac{x_k - v_{1i}}{\|x_k - v_{1i}\|} & \text{if } x_{1k} - \omega_{1i} > r_i \text{ and } x_{2k} - \omega_{2i} > r_i, \\ \dfrac{x_k - v_{2i}}{\|x_k - v_{2i}\|} & \text{if } x_{1k} - \omega_{1i} < -r_i \text{ and } x_{2k} - \omega_{2i} > r_i, \\ \dfrac{x_k - v_{3i}}{\|x_k - v_{3i}\|} & \text{if } x_{1k} - \omega_{1i} < -r_i \text{ and } x_{2k} - \omega_{2i} < -r_i, \\ \dfrac{x_k - v_{4i}}{\|x_k - v_{4i}\|} & \text{if } x_{1k} - \omega_{1i} > r_i \text{ and } x_{2k} - \omega_{2i} < -r_i, \\ (0,1) & \text{if } |x_{1k} - \omega_{1i}| \le r_i \text{ and } x_{2k} - \omega_{2i} > r_i, \\ (0,-1) & \text{if } |x_{1k} - \omega_{1i}| \le r_i \text{ and } x_{2k} - \omega_{2i} < -r_i, \\ (1,0) & \text{if } x_{1k} - \omega_{1i} > r_i \text{ and } |x_{2k} - \omega_{2i}| \le r_i, \\ (-1,0) & \text{if } x_{1k} - \omega_{1i} < -r_i \text{ and } |x_{2k} - \omega_{2i}| \le r_i. \end{cases}$$

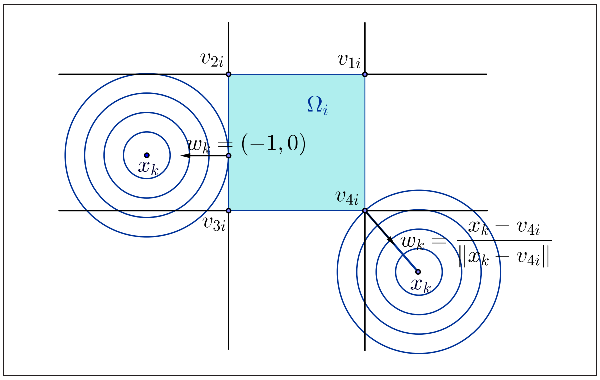

Figure 4.10: Euclidean projection to a square.

Now we consider the case of a non-Euclidean norm in the minimal time function.

Example 4.71 Consider problem (4.52) with $F := \{(x_1, x_2) \in \mathbb{R}^2 \mid |x_1| + |x_2| \le 1\}$ and a nonempty, closed, convex target set Ω. The generalized ball $G_F(\bar{x}; t)$ centered at $\bar{x} = (\bar{x}_1, \bar{x}_2)$ with radius $t > 0$ has the following *diamond shape*:

$$G_F(\bar{x}; t) = \{(x_1, x_2) \in \mathbb{R}^2 \mid |x_1 - \bar{x}_1| + |x_2 - \bar{x}_2| \le t\}.$$

The distance from $\bar{x}$ to Ω and the corresponding projection are given by

$$\mathcal{T}_F(\bar{x}; \Omega) = \min\{t \ge 0 \mid G_F(\bar{x}; t) \cap \Omega \ne \emptyset\}, \tag{4.60}$$

$$\Pi_F(\bar{x}; \Omega) = G_F(\bar{x}; t) \cap \Omega \quad \text{with} \quad t = \mathcal{T}_F(\bar{x}; \Omega).$$

Let Ω_i be the squares $S(\omega_i; r_i)$, $i = 1, \ldots, m$, given in Example 4.70. Then problem (4.55) can be interpreted as follows: Find a set of the diamond shape in $\mathbb{R}^2$ that intersects all m given squares. Using the same notation as in Example 4.70, we can see that the elements v_k in Theorem 4.69 are chosen by

$$v_k = \begin{cases} (1,1) & \text{if } x_{1k} - \omega_{1i} > r_i \text{ and } x_{2k} - \omega_{2i} > r_i, \\ (-1,1) & \text{if } x_{1k} - \omega_{1i} < -r_i \text{ and } x_{2k} - \omega_{2i} > r_i, \\ (-1,-1) & \text{if } x_{1k} - \omega_{1i} < -r_i \text{ and } x_{2k} - \omega_{2i} < -r_i, \\ (1,-1) & \text{if } x_{1k} - \omega_{1i} > r_i \text{ and } x_{2k} - \omega_{2i} < -r_i, \\ (0,1) & \text{if } |x_{1k} - \omega_{1i}| \le r_i \text{ and } x_{2k} - \omega_{2i} > r_i, \\ (0,-1) & \text{if } |x_{1k} - \omega_{1i}| \le r_i \text{ and } x_{2k} - \omega_{2i} < -r_i, \\ (1,0) & \text{if } x_{1k} - \omega_{1i} > r_i \text{ and } |x_{2k} - \omega_{2i}| \le r_i, \\ (-1,0) & \text{if } x_{1k} - \omega_{1i} < -r_i \text{ and } |x_{2k} - \omega_{2i}| \le r_i. \end{cases}$$

4.7 EXERCISES FOR CHAPTER 4

Exercise 4.1 Given a lower semicontinuous function $f: \mathbb{R}^n \to \overline{\mathbb{R}}$, show that its multiplication $(\alpha f) \mapsto \alpha f(x)$ by any scalar $\alpha > 0$ is lower semicontinuous as well.

Exercise 4.2 Given a function $f: \mathbb{R}^n \to \overline{\mathbb{R}}$, show that its lower semicontinuity is equivalent to openess, for every $\lambda \in \mathbb{R}$, of the set

$$\mathcal{U}_\lambda := \{x \in \mathbb{R}^n \mid f(x) > \lambda\}.$$

Exercise 4.3 Give an example of two lower semicontinuous real-valued functions whose product is not lower semicontinuous.

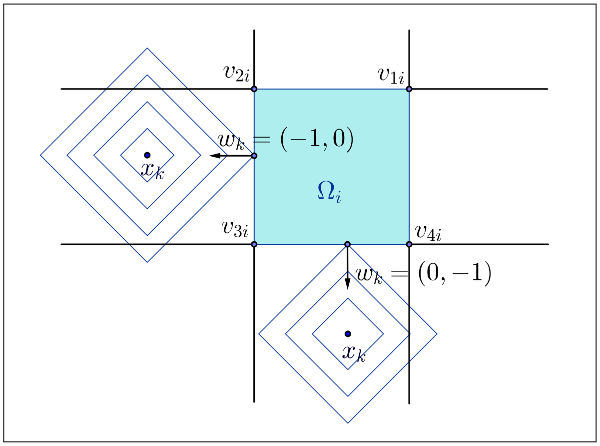

Figure 4.11: Minimal time function with diamond-shape dynamic.

Exercise 4.4 Let $f : \mathbb{R}^n \to \mathbb{R}$. Prove the following assertions:

(i) f is upper semicontinuous (u.s.c.) at $\bar{x}$ in the sense of Definition 2.89 if and only if $-f$ is lower semicontinuous at this point.
(ii) f is u.s.c. at $\bar{x}$ if and only if $\limsup_{k\to\infty} f(x_k) \le f(\bar{x})$ for every $x_k \to \bar{x}$.
(iii) f is u.s.c. at $\bar{x}$ if and only if we have $\limsup_{x\to\bar{x}} f(x) \le f(\bar{x})$.
(iv) f is u.s.c. on $\mathbb{R}^n$ if and only if the upper level set $\{x \in \mathbb{R}^n \mid f(x) \ge \lambda\}$ is closed for any real number λ.
(v) f is u.s.c. on $\mathbb{R}^n$ if and only if its hypergraphical set (or simply HYPERGRAPH) hypo $f := \{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R} \mid \lambda \le f(x)\}$ is closed.

Exercise 4.5 Show that a function $f : \mathbb{R}^n \to \mathbb{R}$ is continuous at $\bar{x}$ if and only if it is both lower semicontinuous and upper semicontinuous at this point.

Exercise 4.6 Prove the equivalence **(iii)**$\Longleftrightarrow$**(iv)** in Theorem 4.11.

Exercise 4.7 Consider the real-valued function

$$g(x) := \sum_{i=1}^{m} \|x - a_i\|^2, \ x \in \mathbb{R}^n,$$

where $a_i \in \mathbb{R}^n$ for $i = 1, \ldots, m$. Show that g is a convex function and has an absolute minimum at the mean point $\bar{x} = \dfrac{a_1 + a_2 + \ldots + a_m}{m}$.

Exercise 4.8 Consider the problem:

$$\text{minimize } \|x\|^2 \text{ subject to } Ax = b,$$

where A is an $p \times n$ matrix with rank $A = p$, and where $b \in \mathbb{R}^p$. Find the unique optimal solution of this problem.

Exercise 4.9 Give a direct proof of Corollary 4.15 without using Theorem 4.14.

Exercise 4.10 Given $a_i \in \mathbb{R}^n$ for $i = 1, \ldots, m$, write MATLAB programs to implement the subgradient algorithm for minimizing the function f given by:

(i) $f(x) := \sum_{i=1}^m \|x - a_i\|, \ x \in \mathbb{R}^n$.
(ii) $f(x) := \sum_{i=1}^m \|x - a_i\|_1, \ x \in \mathbb{R}^n$.
(iii) $f(x) := \sum_{i=1}^m \|x - a_i\|_\infty, \ x \in \mathbb{R}^n$.

Exercise 4.11 Write a MATLAB program to implement the projected subgradient algorithm for solving the problem:

$$\text{minimize } \|x\|_1 \text{ subject to } Ax = b,$$

where A is an $p \times n$ matrix with rank $A = p$, and where $b \in \mathbb{R}^p$.

Exercise 4.12 Consider the optimization problem (4.9) and the iterative procedure (4.10) under the standing assumptions imposed in Section 4.3. Suppose that the optimal value $\overline{V}$ is known and that the sequence $\{\alpha_k\}$ is given by

$$\alpha_k := \begin{cases} 0 & \text{if } 0 \in \partial f(x_k), \\ \dfrac{f(x_k) - \overline{V}}{\|v_k\|^2} & \text{otherwise,} \end{cases} \tag{4.61}$$

where $v_k \in \partial f(x_k)$. Prove that:

(i) $f(x_k) \to \overline{V}$ as $k \to \infty$.
(ii) $0 \le V_k - \overline{V} \le \dfrac{\ell d(x_1; S)}{\sqrt{k}}$ for $k \in \mathbb{N}$.

Exercise 4.13 Let Ω_i for $i = 1, \ldots, m$ be nonempty, closed, convex subsets of $\mathbb{R}^n$ having a nonempty intersection. Use the iterative procedure (4.10) for the function

$$f(x) := \max\left\{d(x; \Omega_i) \;\middle|\; i = 1, \ldots, m\right\}$$

with the sequence $\{\alpha_k\}$ given in (4.61) to derive an algorithm for finding $\bar{x} \in \bigcap_{i=1}^m \Omega_i$.

Exercise 4.14 Derive the result of Corollary 4.20 from condition **(ii)** of Theorem 4.14.

Exercise 4.15 Formulate and prove a counterpart of Theorem 4.27 for the projected subgradient algorithm (4.20) to solve the constrained optimization problem (4.19).

Exercise 4.16 Prove the converse implication in Lemma 4.31.

Exercise 4.17 Write a MATLAB program to implement the Weiszfeld algorithm.

Exercise 4.18 Let $a_i \in \mathbb{R}$ for $i = 1, \ldots, m$. Solve the problem

$$\text{minimize } \sum_{i=1}^{m} |x - a_i|, \quad x \in \mathbb{R}.$$

Exercise 4.19 Give a simplified proof of Theorem 4.51.

Exercise 4.20 Prove assertion **(ii)** of Theorem 4.52.

Exercise 4.21 Give a complete proof of Theorem 4.56.

Exercise 4.22 Let $\Omega_i := [a_i, b_i] \in \mathbb{R}$ for $i = 1, \ldots, m$ be m disjoint intervals. Solve the following problem:

$$\text{minimize } f(x) := \sum_{i=1}^{m} d(x; \Omega_i), \quad x \in \mathbb{R}.$$

Exercise 4.23 Write a MATLAB program to implement the subgradient algorithm for solving the problem:

$$\text{minimize } f(x), \quad x \in \mathbb{R}^n,$$

where $f(x) := \sum_{i=1}^{m} d(x; \Omega_i)$ with $\Omega_i := I\!B(c_i; r_i)$ for $i = 1, \ldots, m$.

Exercise 4.24 Complete the solution of the problem in Example 4.58. Write a MATLAB program to implement the algorithm.

Exercise 4.25 Give a simplified proof of Theorem 4.63 in the case where F is the closed unit ball of $\mathbb{R}^n$ and the target sets are closed Euclidean balls of $\mathbb{R}^n$.

Exercise 4.26 Give a simplified proof of Theorem 4.63 in the case where F is the closed unit ball of $\mathbb{R}^n$.

Exercise 4.27 Write a MATLAB program to implement the subgradient algorithm for solving the problem:

$$\text{minimize } f(x), \quad x \in \mathbb{R}^n,$$

where $f(x) := \max\{d(x; \Omega_i) \mid i = 1, \ldots, m\}$ with $\Omega_i := I\!B(c_i; r_i)$ for $i = 1, \ldots, m$. Note that if $\bar{x}$ is an optimal solution to this problem with the optimal value r, then $I\!B(\bar{x}; r)$ is the smallest ball that intersects Ω_i for $i = 1, \ldots, m$.

Exercise 4.28 Give a complete proof of Theorem 4.69.

Exercise 4.29 Complete the solution of the problem in Example 4.70. Write a MATLAB program to implement the algorithm.

Exercise 4.30 Complete the solution of the problem in Example 4.71. Write a MATLAB program to implement the algorithm.

Solutions and Hints for Selected Exercises

This final part of the book contains hints for selected exercises. Rather detailed solutions are given for several problems.

EXERCISES FOR CHAPTER 1

Exercise 1.1. Suppose that Ω_1 is bounded. Fix a sequence $\{x_k\} \subset \Omega_1 - \Omega_2$ converging to $\bar{x}$. For every $k \in \mathbb{N}$, x_k is represented as $x_k = y_k - z_k$ with $y_k \in \Omega_1$, $z_k \in \Omega_2$. Since Ω_1 is closed and bounded, $\{y_k\}$ has a subsequence $\{y_{k_\ell}\}$ converging to $\bar{y} \in \Omega_1$. Then $\{z_{k_\ell}\}$ converges to $\bar{y} - \bar{x} \in \Omega_2$. Thus $\bar{x} \in \Omega_1 - \Omega_2$ and Ω is closed. Let $\Omega_1 := \{(x, y) \in \mathbb{R}^2 \mid y \geq 1/x,\ x > 0\}$ and $\Omega_2 := \{(x, y) \in \mathbb{R}^2 \mid y \leq -1/x,\ x > 0\}$. Then Ω_1 and Ω_2 are nonempty, closed, convex sets while $\Omega_1 - \Omega_2 = \{(x, y) \in \mathbb{R}^2 \mid y > 0\}$ is not closed.

Exercise 1.4. **(i)** Use Exercise 1.3(ii).
(ii) We only consider the case where $m = 2$ since the general case is similar. From the definition of cones, both sets Ω_1 and Ω_2 contain 0, and so $\Omega_1 \cup \Omega_2 \subset \Omega_1 + \Omega_2$. Then $\operatorname{co}\{\Omega_1 \cup \Omega_2\} \subset \Omega_1 + \Omega_2$. To verify the opposite inclusion, fix any $x = w_1 + w_2 \in \Omega_1 + \Omega_2$ with $w_1 \in \Omega_1$ and $w_2 \in \Omega_2$ and then write $x = [(2w_1 + 2w_2)/2] \in \operatorname{co}\{\Omega_1 \cup \Omega_2\}$.

Exercise 1.7. Compute the first or second derivative of each function. Then apply either Theorem 1.45 or Corollary 1.46.

Exercise 1.8. **(i)** Use $f(x, y) := xy,\ (x, y) \in \mathbb{R}^2$. **(ii)** Take $f_1(x) := x$ and $f_2(x) := -x$.

Exercise 1.9. Let $f_1(x) := x$ and $f_2(x) := x^2$, $x \in \mathbb{R}$.

Exercise 1.12. Use Proposition 1.39. Note that the function $f(x) := x^q$ is convex on $[0, \infty)$.

Exercise 1.16. Define $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^p$ by $F(x) := K$, $x \in \mathbb{R}^n$. Then gph F is a convex subset of $\mathbb{R}^n \times \mathbb{R}^p$. The conclusion follows from Proposition 1.50.

Exercise 1.18. **(i)** $\operatorname{gph} F = C$. **(ii)** Consider the function

$$f(x) := \begin{cases} -\sqrt{1-x^2} & \text{if } |x| \le 1, \\ \infty & \text{otherwise.} \end{cases}$$

(iii) Apply Proposition 1.50.

Exercise 1.20. **(i)** Use the definition. **(ii)** Take $f(x) := \sqrt{|x|}$.

Exercise 1.22. Suppose that $\lambda v + \sum_{i=1}^m \lambda_i v_i = 0$ with $\lambda + \sum_{i=1}^m \lambda_i = 0$. If $\lambda \neq 0$, then $-\sum_{i=1}^m \dfrac{\lambda_i}{\lambda} = 1$, and so

$$v = -\sum_{i=1}^m \frac{\lambda_i}{\lambda} v_i \in \operatorname{aff}\{v_1, \ldots, v_m\},$$

which is a contradiction. Thus $\lambda = 0$, and hence $\lambda_i = 0$ for $i = 1, \ldots, m$ because the elements $v_1, \ldots, v_m$ are affinely independent.

Exercise 1.23. Since $\Delta_m \subset \Omega \subset \operatorname{aff}\Omega$, we have $\operatorname{aff}\Delta_m \subset \operatorname{aff}\Omega$. To prove the opposite inclusion, it remains to show that $\Omega \subset \operatorname{aff}\Delta_m$. By contradiction, suppose that there are $v \in \Omega$ such that $v \notin \operatorname{aff}\Delta_m$. It follows from Exercise 1.22 that the vectors $v, v_1, \ldots, v_m$ are affinely independent and all of them belong to Ω. This contradicts the condition $\dim \Omega = m$. The second equality can also be proved similarly.

Exercise 1.24. **(i)** We obviously have $\operatorname{aff}\Omega \subset \operatorname{aff}\overline{\Omega}$. Observe that the set $\operatorname{aff}\Omega$ is closed and contains Ω since $\operatorname{aff}\Omega = w_0 + L$, where $w_0 \in \Omega$ and L is the linear subspace parallel to $\operatorname{aff}\Omega$, which is a closed set. Thus $\overline{\Omega} \subset \operatorname{aff}\Omega$ and the conclusion follows.
(ii) Choose $\delta > 0$ so that $I\!B(\bar{x};\delta) \cap \operatorname{aff}\Omega \subset \Omega$. Since $\bar{x} + t(\bar{x} - x) = (1+t)\bar{x} + (-t)x$ is an affine combination of some elements in $\operatorname{aff}\Omega = \operatorname{aff}\overline{\Omega}$, we have $\bar{x} + t(\bar{x} - x) \in \operatorname{aff}\Omega$. Thus $u := \bar{x} + t(\bar{x} - x) \in \Omega$ when t is small.
(iii) It follows from **(i)** that $\operatorname{ri}\Omega \subset \operatorname{ri}\overline{\Omega}$. Fix any $x \in \operatorname{ri}\overline{\Omega}$ and $\bar{x} \in \operatorname{ri}\Omega$. By **(ii)** we have $z := x + t(x - \bar{x}) \in \overline{\Omega}$ when $t > 0$ is small, and so $x = z/(1+t) + t\bar{x}/(1+t) \in (z, \bar{x}) \subset \operatorname{ri}\Omega$.

Exercise 1.26. If $\overline{\Omega}_1 = \overline{\Omega}_2$, then $\operatorname{ri}\overline{\Omega}_1 = \operatorname{ri}\overline{\Omega}_2$, and so $\operatorname{ri}\Omega_1 = \operatorname{ri}\Omega_2$ by Exercise 1.24(iii).

Exercise 1.27. **(i)** Fix $y \in B(\operatorname{ri}\Omega)$ and find $x \in \operatorname{ri}\Omega$ such that $y = B(x)$. Then pick a vector $\bar{y} \in \operatorname{ri} B(\Omega) \subset B(\Omega)$ and take $\bar{x} \in \Omega$ with $\bar{y} = B(\bar{x})$. If $x = \bar{x}$, then $y = \bar{y} \in \operatorname{ri} B(\Omega)$. Consider the case where $x \neq \bar{x}$. Following the solution of Exercise 1.24(iii), find $\widetilde{x} \in \Omega$ such that $x \in (\widetilde{x}, \bar{x})$ and let $\widetilde{y} := B(\widetilde{x}) \in B(\Omega)$. Thus $y = B(x) \in (B(\widetilde{x}),\ B(\bar{x})) = (\widetilde{y}, \bar{y})$, and so $y \in \operatorname{ri} B(\Omega)$. To complete the proof, it remains to show that $\overline{B(\Omega)} = \overline{B(\operatorname{ri}\Omega)}$ and then use Exercise 1.26 to obtain the inclusion

$$\text{ri}\, B(\Omega) = \text{ri}\, B(\text{ri}\,\Omega) \subset B(\text{ri}\,\Omega).$$

Since $\overline{B(\text{ri}\,\Omega)} \subset \overline{B(\Omega)}$, the continuity of B and Proposition 1.73 imply that

$$B(\Omega) \subset B(\overline{\Omega}) = B(\overline{\text{ri}\,\Omega}) \subset \overline{B(\text{ri}\,\Omega)}.$$

Thus we have $\overline{B(\Omega)} \subset \overline{B(\text{ri}\,\Omega)}$ and complete the proof.
(ii) Consider $B(x,y) := x - y$ defined on $\mathbb{R}^n \times \mathbb{R}^n$.

Exercise 1.28. (i) Let us show that

$$\text{aff}(\text{epi}\, f) = \text{aff}(\text{dom}\, f) \times \mathbb{R}\,. \tag{4.62}$$

Fix any $(x,\gamma) \in \text{aff}(\text{epi}\, f)$. Then there exist $\lambda_i \in \mathbb{R}$ and $(x_i, \gamma_i) \in \text{epi}\, f$ for $i = 1,\ldots,m$ with $\sum_{i=1}^m \lambda_i = 1$ such that

$$(x,\gamma) = \sum_{i=1}^m \lambda_i (x_i, \gamma_i).$$

Since $x_i \in \text{dom}\, f$ for $i = 1,\ldots,m$, we have $x = \sum_{i=1}^m \lambda_i x_i \in \text{aff}(\text{dom}\, f)$, which yields

$$\text{aff}(\text{epi}\, f) \subset \text{aff}(\text{dom}\, f) \times \mathbb{R}\,.$$

To verify the opposite inclusion, fix any $(x,\gamma) \in \text{aff}(\text{dom}\, f) \times \mathbb{R}$ and find $x_i \in \text{dom}\, f$ and $\lambda_i \in \mathbb{R}$ for $i = 1,\ldots,m$ with $\sum_{i=1}^m \lambda_i = 1$ and $x = \sum_{i=1}^m \lambda_i x_i$. Define further $\alpha_i := f(x_i) \in \mathbb{R}$ and $\alpha := \sum_{i=1}^m \lambda_i \alpha_i$ for which $(x_i, \alpha_i) \in \text{epi}\, f$ and $(x_i, \alpha_i + 1) \in \text{epi}\, f$. Thus

$$\sum_{i=1}^m \lambda_i (x_i, \alpha_i) = (x, \alpha) \in \text{aff}(\text{epi}\, f),$$
$$\sum_{i=1}^m \lambda_i (x_i, \alpha_i + 1) = (x, \alpha + 1) \in \text{aff}(\text{epi}\, f).$$

Letting $\lambda := \alpha - \gamma + 1$ gives us

$$(x,\gamma) = \lambda(x,\alpha) + (1-\lambda)(x, \alpha+1) \in \text{aff}(\text{epi}\, f),$$

which justifies the opposite inclusion in (4.62).

(ii) See the proof of [9, Proposition 1.1.9].

Exercise 1.30. (i) $\Pi(x;\Omega) = x + \dfrac{b - \langle a, x\rangle}{\|a\|^2} a.$
(ii) $\Pi(x;\Omega) = x + A^T(AA^T)^{-1}(b - Ax).$
(iii) $\Pi(x;\Omega) = \max\{x, 0\}$ (componentwise maximum).

EXERCISES FOR CHAPTER 2

Exercise 2.1. **(i)** Apply Proposition 2.1 and the fact that $tx \in \Omega$ for all $t \geq 0$.
(ii) Use **(i)** and observe that $-x \in \Omega$ whenever $x \in \Omega$.

Exercise 2.2. Observe that the sets $\operatorname{ri}\Omega_i$ for $i = 1, 2$ are nonempty and convex. By Theorem 2.5, there is $0 \neq v \in \mathbb{R}^n$ such that

$$\langle v, x\rangle \leq \langle v, y\rangle \ \text{ for all } \ x \in \operatorname{ri}\Omega_1,\ \ y \in \operatorname{ri}\Omega_2.$$

Fix $x_0 \in \operatorname{ri}\Omega_1$ and $y_0 \in \operatorname{ri}\Omega_2$. It follows from Theorem 1.72(ii) that for any $x \in \Omega_1$, $y \in \Omega_2$, and $t \in (0, 1)$ we have that $x + t(x_0 - x) \in \operatorname{ri}\Omega_1$ and $y + t(y_0 - y) \in \operatorname{ri}\Omega_2$. Then apply the inequality above for these elements and let $t \to 0^+$.

Exercise 2.3. We split the solution into four steps.

Step 1: If L is a subspace of $\mathbb{R}^n$ that contains a nonempty, convex set Ω with $\bar{x} \notin \overline{\Omega}$ and $\bar{x} \in L$, then there is $0 \neq v \in L$ such that

$$\sup\big\{\langle v, x\rangle \ \big|\ x \in \Omega\big\} < \langle v, \bar{x}\rangle. \tag{4.63}$$

Proof. By Proposition 2.1, find $w \in \mathbb{R}^n$ such that

$$\sup\big\{\langle w, x\rangle \ \big|\ x \in \Omega\big\} < \langle w, \bar{x}\rangle. \tag{4.64}$$

Represent $\mathbb{R}^n$ as $L \oplus L^{\perp}$, where $L^{\perp} := \{u \in \mathbb{R}^n \mid \langle u, x\rangle = 0 \text{ for all } \ x \in L\}$. Then $w = u + v$ with $u \in L^{\perp}$ and $v \in L$. Substituting w into (4.64) and taking into account that $\langle u, x\rangle = 0$ for all $x \in \Omega$ and that $\langle u, \bar{x}\rangle = 0$, we get (4.63). Note that (4.63) yields $v \neq 0$.

Step 2: Let $\Omega \subset \mathbb{R}^n$ be a nonempty, convex set such that $0 \notin \operatorname{ri}\Omega$. Then the sets Ω and $\{0\}$ can be separated properly.

Proof. In the case where $0 \notin \overline{\Omega}$, we can separate Ω and $\{0\}$ strictly, so the conclusion is obvious. Suppose now that $0 \in \overline{\Omega} \setminus \operatorname{ri}\Omega$. Let $L := \operatorname{aff}\Omega = \operatorname{span}\Omega$, where the latter holds due to $0 \in \operatorname{aff}\Omega$, which follows from the fact that $\overline{\Omega} \subset L$ since L is closed. Repeating the proof of Lemma 2.4, find a sequence $\{x_k\} \subset L$ with $x_k \to 0$ and $x_k \notin \overline{\Omega}$ for every k. It follows from Step 1 that there is $\{v_k\} \subset L$ with $v_k \neq 0$ and

$$\sup\big\{\langle v_k, x\rangle \ \big|\ x \in \Omega\big\} < \langle v_k, x_k\rangle.$$

Dividing both sides by $\|v_k\|$, we suppose without loss of generality that $\|v_k\| = 1$ and that $v_k \to v \in L$ with $\|v\| = 1$. Then

$$\sup\big\{\langle v, x\rangle \ \big|\ x \in \Omega\big\} \leq 0.$$

To finish this proof, it remains to show that there exists $x \in \Omega$ with $\langle v, x\rangle < 0$. By contradiction, suppose that $\langle v, x\rangle \geq 0$ for all $x \in \Omega$. Then $\langle v, x\rangle = 0$ on Ω. Since $v \in L = \operatorname{aff}\Omega$, we can represent v as $v = \sum_{i=1}^m \lambda_i w_i$, where $\sum_{i=1}^m \lambda_i = 1$ and $w_i \in \Omega$ for $i = 1, \ldots, m$. This implies that

$\|v\|^2 = \langle v, v\rangle = \sum_{i=1}^m \lambda_i \langle v, w_i\rangle = 0$, which is a contradiction.

Step 3: Let $\Omega \subset \mathbb{R}^n$ be a nonempty, convex set. If the sets Ω and $\{0\}$ can be separated properly, then we have $0 \notin \operatorname{ri} \Omega$.

Proof. By definition, find $0 \neq v \in \mathbb{R}^n$ such that

$$\sup \big\{\langle v, x\rangle \ \big|\ x \in \Omega\big\} \le 0$$

and then take $\bar{x} \in \Omega$ with $\langle v, \bar{x}\rangle < 0$. Arguing by contradiction, suppose that $0 \in \operatorname{ri} \Omega$. Then Exercise 1.24(ii) gives us $t > 0$ such that $0 + t(0 - \bar{x}) = -t\bar{x} \in \Omega$. Thus $\langle v, -t\bar{x}\rangle \le 0$, which yields $\langle v, \bar{x}\rangle \ge 0$, a contradiction.

Step 4: Justifying the statement of Exercise 2.3.

Proof. Define $\Omega := \Omega_1 - \Omega_2$. By Exercise 1.27(ii) we have $\operatorname{ri} \Omega_1 \cap \operatorname{ri} \Omega_2 = \emptyset$ if and only if

$$0 \notin \operatorname{ri}(\Omega_1 - \Omega_2) = \operatorname{ri} \Omega_1 - \operatorname{ri} \Omega_2.$$

Then the conclusion follows from Step 2 and Step 3.

Exercise 2.4. Observe that $\Omega_1 \cap [\Omega_2 - (0, \lambda)] = \emptyset$ for any $\lambda > 0$.

Exercise 2.5. Employing Theorem 2.5 and arguments similar to those in Exercise 2.2, we can show that there is $0 \neq v \in \mathbb{R}^n$ such that

$$\langle v, x\rangle \le \langle v, y\rangle \ \text{ for all } \ x \in \Omega_1,\ y \in \Omega_2.$$

By Proposition 2.13, choose $0 \neq v \in N(\bar{x}; \Omega_1) \cap [-N(\bar{x}; \Omega_2)]$ and, using the normal cone definition, arrive at the conclusion with $v_1 := v$, $v_2 := -v$, $\alpha_1 := \langle v_1, \bar{x}\rangle$, and $\alpha_2 := \langle v_2, \bar{x}\rangle$.

Exercise 2.6. Fix $\bar{x} \le 0$ and $v \in N(\bar{x}; \Omega)$. Then

$$\langle v, x - \bar{x}\rangle \le 0 \ \text{ for all } \ x \in \Omega.$$

Using this inequality for $0 \in \Omega$ and $2\bar{x} \in \Omega$, we get $\langle v, \bar{x}\rangle = 0$. Since $\Omega = (-\infty, 0] \times \ldots \times (-\infty, 0]$, it follows from Proposition 2.11 that $v \ge 0$. The opposite inclusion is a direct consequence of the definition.

Exercise 2.8. Suppose that this holds for any k convex sets with $k \le m$ and $m \ge 2$. Consider the sets Ω_i for $i = 1, \ldots, m+1$ satisfying the qualification condition

$$\Big[\sum_{i=1}^{m+1} v_i = 0,\ v_i \in N(\bar{x}; \Omega_i)\Big] \Longrightarrow \big[v_i = 0 \ \text{ for all } \ i = 1, \ldots, m+1\big].$$

Since $0 \in N(\bar{x}; \Omega_{m+1})$, this condition also holds for the first m sets. Thus

$$N\Big(\bar{x}; \bigcap_{i=1}^{m} \Omega_i\Big) = \sum_{i=1}^{m} N(\bar{x}; \Omega_i).$$

Then apply the induction hypothesis to $\Theta_1 := \cap_{i=1}^{m} \Omega_i$ and $\Theta_2 := \Omega_{m+1}$.

Exercise 2.10. Since $v \in \partial f(\bar{x})$, we have

$$v(x - \bar{x}) \le f(x) - f(\bar{x}) \ \text{ for all } \ x \in \mathbb{R}\,.$$

Then pick any $x_1 < \bar{x}$ and use the inequality above.

Exercise 2.12. Define the function

$$g(x) := f(x) + \delta_{\{\bar{x}\}}(x) = \begin{cases} f(\bar{x}) & \text{if } \ x = \bar{x}, \\ \infty & \text{otherwise} \end{cases}$$

via the indicator function of $\{\bar{x}\}$. Then $\operatorname{epi} g = \{\bar{x}\} \times [f(\bar{x}), \infty)$, and hence $N((\bar{x}, g(\bar{x})); \operatorname{epi} g) = \mathbb{R}^n \times (-\infty, 0]$. It follows from Proposition 2.31 that $\partial g(\bar{x}) = \mathbb{R}^n$ and $\partial \delta_{\{\bar{x}\}}(\bar{x}) = \mathbb{R}^n$. Employing the subdifferential sum rule from Corollary 2.45 gives us

$$\mathbb{R}^n = \partial g(\bar{x}) = \partial f(\bar{x}) + \mathbb{R}^n,$$

which ensures that $\partial f(\bar{x}) \neq \emptyset$.

Exercise 2.13. We proceed step by step as in the proof of Theorem 2.15. It follows from Exercise 1.28 that $\operatorname{ri} \Theta_1 = \operatorname{ri} \Omega_1 \times (0, \infty)$ and

$$\operatorname{ri} \Theta_2 = \big\{(x, \lambda) \bigm| x \in \operatorname{ri} \Omega_2,\ \lambda < \langle v, x - \bar{x}\rangle\big\}.$$

It is easy to check, arguing by contradiction, that $\operatorname{ri} \Theta_1 \cap \operatorname{ri} \Theta_2 = \emptyset$. Applying the separation result from Exercise 2.3, we find $0 \neq (w, \gamma) \in \mathbb{R}^n \times \mathbb{R}$ such that

$$\langle w, x\rangle + \lambda_1 \gamma \le \langle w, y\rangle + \lambda_2 \gamma \ \text{ for all } \ (x, \lambda_1) \in \Theta_1,\ (y, \lambda_2) \in \Theta_2.$$

Moreover, there are $(\widetilde{x}, \widetilde{\lambda}_1) \in \Theta_1$ and $(\widetilde{y}, \widetilde{\lambda}_2) \in \Theta_2$ satisfying

$$\langle w, \widetilde{x}\rangle + \widetilde{\lambda}_1 \gamma < \langle w, \widetilde{y}\rangle + \widetilde{\lambda}_2 \gamma.$$

As in the proof of Theorem 2.15, it remains to verify that $\gamma < 0$. By contradiction, assume that $\gamma = 0$ and then get

$$\begin{aligned} &\langle w, x\rangle \le \langle w, y\rangle \ \text{ for all } \ x \in \Omega_1,\ y \in \Omega_2, \\ &\langle w, \widetilde{x}\rangle < \langle w, \widetilde{y}\rangle \ \text{ with } \ \widetilde{x} \in \Omega_1,\ \widetilde{y} \in \Omega_2. \end{aligned}$$

Thus the sets Ω_1 and Ω_2 are properly separated, and hence it follows from Exercise 2.3 that $\text{ri}\,\Omega_1 \cap \text{ri}\,\Omega_2 = \emptyset$. This is a contradiction, which shows that $\gamma < 0$. To complete the proof, we only need to repeat the proof of Case 2 in Theorem 2.15.

Exercise 2.14. Use Exercise 2.13 and proceed as in the proof of Theorem 2.44.

Exercise 2.16. Use the same procedure as in the proof of Theorem 2.51 and the construction of the singular subdifferential to obtain

$$\partial^\infty(f \circ B)(\bar{x}) = A^*\big(\partial^\infty f(\bar{y})\big) = \big\{A^* v \;\big|\; v \in \partial^\infty f(\bar{y})\big\}. \tag{4.65}$$

Alternatively, observe the equality $\text{dom}\,(f \circ B) = B^{-1}(\text{dom}\; f)$. Then apply Proposition 2.23 and Corollary 2.53 to complete the proof.

Exercise 2.17. Use the subdifferential formula from Theorem 2.61 and then the representation of the composition from Proposition 1.54.

Exercise 2.19. For any $v \in \mathbb{R}^n$, we have $\sigma_{\Omega_1}(v) = \|v\|_\infty$ and $\sigma_{\Omega_2}(v) = \|v\|_1$.

Exercise 2.21. Use first the representations

$$(\delta_\Omega)^*(v) = \sup\big\{\langle v, x\rangle - \delta(x; \Omega) \;\big|\; x \in \mathbb{R}^n\big\} = \sup\big\{\langle v, x\rangle \;\big|\; x \in \Omega\big\} = \sigma_\Omega(v), \quad v \in \mathbb{R}^n\,.$$

Applying then the Fenchel conjugate one more time gives us

$$(\sigma_\Omega)^*(x) = \sup\big\{\langle x, v\rangle - \sigma_\Omega(v) \;\big|\; v \in \mathbb{R}^n\big\} = \delta_{\overline{\text{co}}\,\Omega}(x).$$

Exercise 2.23. Since $f : \mathbb{R}^n \to \mathbb{R}$ is convex and positively homogeneous, we have $\partial f(0) \neq \emptyset$ and $f(0) = 0$. This implies in the case where $v \in \Omega = \partial f(0)$ that $\langle v, x\rangle \le f(x)$ for all $x \in \mathbb{R}^n$, and so the equality holds when $x = 0$. Thus we have

$$f^*(v) = \sup\big\{\langle v, x\rangle - f(x) \;\big|\; x \in \mathbb{R}^n\big\} = 0.$$

If $v \notin \Omega = f(0)$, there is $\bar{x} \in \mathbb{R}^n$ satisfying $\langle v, \bar{x}\rangle > f(\bar{x})$. It gives us by homogeneity that

$$f^*(v) = \sup\big\{\langle v, x\rangle - f(x) \;\big|\; x \in \mathbb{R}^n\big\} \ge \sup\big\{\langle v, t\bar{x}\rangle - f(t\bar{x}) \;\big|\; t > 0\big\} = \infty.$$

Exercise 2.26. $f'(-1; d) = -\infty$ and $f'(1; d) = \infty$.

Exercise 2.27. Write

$$\bar{x} + t_2 d = \frac{t_2}{t_1}(\bar{x} + t_1 d) + (1 - \frac{t_2}{t_1})\bar{x} \;\text{ for fixed }\; t_1 < t_2 < 0.$$

Exercise 2.28. **(i)** For $d \in \mathbb{R}^n$, consider the function φ defined in Lemma 2.80. Then show that the inequalities $\lambda < 0 < t$ imply that $\varphi(\lambda) \le \varphi(t)$.
(ii) It follows from **(i)** that $\varphi(t) \le \varphi(1)$ for all $t \in (0,1)$. Similarly, $\varphi(-1) \le \varphi(t)$ for all $t \in (-1,0)$. This easily yields the conclusion.

Exercise 2.31. Fix $\bar{x} \in \mathbb{R}^n$, define $\mathcal{P}(\bar{x};\Omega) := \{w \in \Omega \mid \|\bar{x} - w\| = \mu_\Omega(\bar{x})\}$, and consider the following function

$$f_w(x) := \|x - w\| \text{ for } w \in \Omega \text{ and } x \in \mathbb{R}^n.$$

Then we get from Theorem 2.93 that

$$\partial \mu_\Omega(\bar{x}) = \operatorname{co}\big\{\partial f_w(\bar{x}) \bigm| w \in \mathcal{P}(\bar{x};\Omega)\big\}.$$

In the case where Ω is not a singleton we have $\partial f_w(\bar{x}) = \dfrac{\bar{x} - w}{\mu_\Omega(\bar{x})}$, and so

$$\partial \mu_\Omega(\bar{x}) = \frac{\bar{x} - \operatorname{co}\mathcal{P}(\bar{x};\Omega)}{\mu_\Omega(\bar{x})}.$$

The case where Ω is a singleton is trivial.

Exercise 2.33. **(i)** Fix any $v \in \mathbb{R}^n$ with $\|v\| \le 1$ and any $w \in \Omega$. Then

$$\langle x, v\rangle - \sigma_\Omega(v) \le \langle x, v\rangle - \langle w, v\rangle = \langle x - w, v\rangle \le \|x - w\| \cdot \|v\| \le \|x - w\|.$$

It follows therefore that

$$\sup_{\|v\| \le 1} \big\{\langle x, v\rangle - \sigma_\Omega(v)\big\} \le d(x;\Omega).$$

To verify the opposite inequality, observe first that it holds trivially if $x \in \Omega$. In the case where $x \notin \Omega$, let $w := \Pi(x;\Omega)$ and $u := x - w$; thus $\|u\| = d(x;\Omega)$. Following the proof of Proposition 2.1, we obtain the estimate

$$\langle u, z\rangle \le \langle u, x\rangle - \|u\|^2 \text{ for all } z \in \Omega.$$

Divide both sides by $\|u\|$ and let $v := u/\|u\|$. Then $\sigma_\Omega(v) \le \langle v, x\rangle - d(x;\Omega)$, and so

$$d(x;\Omega) \le \sup_{\|v\| \le 1} \big\{\langle x, v\rangle - \sigma_\Omega(v)\big\}.$$

EXERCISES FOR CHAPTER 3

Exercise 3.1. Suppose that f is Gâteaux differentiable at $\bar{x}$ with $f'_G(\bar{x}) = v$. Then

$$\langle f'_G(\bar{x}), v\rangle = \lim_{t \to 0} \frac{f(\bar{x} + td) - f(\bar{x})}{t} = \lim_{t \to 0^+} \frac{f(\bar{x} + td) - f(\bar{x})}{t} \text{ for all } d \in \mathbb{R}^n.$$

This implies by the definition of the directional derivative that $f'(\bar{x};d) = \langle v, d\rangle$ for all $d \in \mathbb{R}^n$. Conversely, suppose that $f'(\bar{x};d) = \langle v, d\rangle$ for all $d \in \mathbb{R}^n$. Then

$$f'(\bar{x};d) = \langle v, d\rangle = \lim_{t\to 0^+} \frac{f(\bar{x}+td) - f(\bar{x})}{t} \text{ for all } d \in \mathbb{R}^n .$$

Moreover, $f'_-(\bar{x};d) = -f'(\bar{x};-d) = -\langle v, -d\rangle = \langle v, d\rangle$ for all $d \in \mathbb{R}^n$. Thus

$$f'_-(\bar{x};d) = \langle v, d\rangle = \lim_{t\to 0^-} \frac{f(\bar{x}+td) - f(\bar{x})}{t} \text{ for all } d \in \mathbb{R}^n .$$

This implies that f is Gâteaux differentiable at $\bar{x}$ with $f'_G(\bar{x}) = v$.

Exercise 3.2. **(i)** Since the function $\varphi(t) := t^2$ is nondecreasing on $[0,\infty)$, following the proof of Corollary 2.62, we can show that if $g : \mathbb{R}^n \to [0,\infty)$ is a convex function, then $\partial(g^2)(\bar{x}) = 2g(\bar{x})\partial g(\bar{x})$ for every $\bar{x} \in \mathbb{R}^n$, where $g^2(x) := [g(x)]^2$. It follows from Theorem 2.39 that

$$\partial f(\bar{x}) = 2d(\bar{x};\Omega)\partial d(\bar{x};\Omega) = \begin{cases} \{0\} & \text{if } \bar{x} \in \Omega, \\ \{2[\bar{x} - \Pi(\bar{x};\Omega)]\} & \text{otherwise.} \end{cases}$$

Thus in any case $\partial f(\bar{x}) = \{2[\bar{x} - \Pi(\bar{x};\Omega)]\}$ is a singleton, which implies the differentiability of f at $\bar{x}$ by Theorem 3.3 and $\nabla f(\bar{x}) = 2[\bar{x} - \Pi(\bar{x};\Omega)]$.
(ii) Use Theorem 2.39 in the out-of-set case and Theorem 3.3.
Exercise 3.3. For every $u \in F$, the function $f_u(x) := \langle Ax, u\rangle - \frac{1}{2}\|u\|^2$ is convex, and hence f is a convex function by Proposition 1.43. It follows from the equality $\langle Ax, u\rangle = \langle x, A^*u\rangle$ that

$$\nabla f_u(x) = A^*u.$$

Fix $x \in \mathbb{R}^n$ and observe that the function $g_x(u) := \langle Ax, u\rangle - \frac{1}{2}\|u\|^2$ has a unique maximizer on F denoted by $u(x)$. By Theorem 2.93 we have $\partial f(x) = \{A^*u(x)\}$, and thus f is differentiable at x by Theorem 3.3. We refer the reader to [22] for more general results and their applications to optimization.

Exercise 3.5. Consider the set $Y := \{d \in \mathbb{R}^n \mid f'(\bar{x};d) = f'_-(\bar{x};d)\}$ and verify that Y is a subspace containing the standard orthogonal basis $\{e_i \mid i = 1,\ldots,n\}$. Thus $V = \mathbb{R}^n$, which implies the differentiability of f at $\bar{x}$. Indeed, fix any $v \in \partial f(\bar{x})$ and get by Theorem 2.84(iii) that $f'(\bar{x};d) = \langle v, d\rangle$ for all $d \in \mathbb{R}^n$. Hence $\partial f(\bar{x})$ is a singleton, which ensures that f is differentiable at $\bar{x}$ by Theorem 3.3.

Exercise 3.6. Use Corollary 2.37 and follow the proof of Theorem 2.40.

Exercise 3.7. An appropriate example is given by the function $f(x) := x^4$ on $\mathbb{R}$.

Exercise 3.10. Fix any sequence $\{x_k\} \subset \operatorname{co} \Omega$. By the Carathéodory theorem, find $\lambda_{k,i} \geq 0$ and $w_{k,i} \in \Omega$ for $i = 0, \ldots, n$ with

$$x_k = \sum_{i=0}^{n} \lambda_{k,i} w_{k,i}, \quad \sum_{i=0}^{n} \lambda_{k,i} = 1 \text{ for every } k \in \mathbb{N}.$$

Then use the boundedness of $\{\lambda_{k,i}\}_k$ for $i = 0, \ldots, n$ and the compactness of Ω to extract a subsequence of $\{x_k\}$ converging to some element of $\operatorname{co} \Omega$.

Exercise 3.16. By Corollary 2.18 we have $N(\bar{x}; \Omega_1 \cap \Omega_2) = N(\bar{x}; \Omega_1) + N(\bar{x}; \Omega_2)$. Applying Theorem 3.16 tells us that

$$\begin{aligned} T(\bar{x}; \Omega_1 \cap \Omega_2) &= \big[N(\bar{x}; \Omega_1 \cap \Omega_2)\big]^\circ \\ &= \big[N(\bar{x}; \Omega_1) + N(\bar{x}; \Omega_2)\big]^\circ = \big[N(\bar{x}; \Omega_1)\big]^\circ \cap \big[N(\bar{x}; \Omega_2)\big]^\circ \\ &= T(\bar{x}; \Omega_1) \cap T(\bar{x}; \Omega_2). \end{aligned}$$

Exercise 3.18. Consider the set $\Omega := \{(x, y) \in \mathbb{R}^2 \mid y \geq x^2\}$.

Exercise 3.19. Fix any $a, b \in \mathbb{R}^n$ with $a \neq b$. It follows from Theorem 3.20 that there exists $c \in (a, b)$ such that

$$f(b) - f(a) = \langle v, b - a \rangle \text{ for some } c \in \partial f(c).$$

Then $|f(b) - f(a)| \leq \|v\| \cdot \|b - a\| \leq \ell \|b - a\|$.

Exercise 3.21. **(i)** $F_\infty = \{0\} \times \mathbb{R}_+$. **(ii)** $F_\infty = F$. **(iii)** $F_\infty = \mathbb{R}_+ \times \mathbb{R}_+$. **(iv)** $F_\infty = \{(x, y) \mid y \geq |x|\}$.

Exercise 3.23. **(i)** Apply the definition. **(ii)** A condition is $A_\infty \cap (-B_\infty) = \{0\}$.

Exercise 3.24. Suppose that $\mathcal{T}_\Omega^F(x) = 0$. Then there exists a sequence of $t_k \to 0$, $t_k \geq 0$, with $(x + t_k F) \cap \Omega \neq \emptyset$ for every $k \in \mathbb{N}$. Choose $f_k \in F$ and $w_k \in \Omega$ such that $x + t_k f_k = w_k$. Since F is bounded, the sequence $\{w_k\}$ converges to x, and so $x \in \Omega$ since Ω is closed. The converse implication is trivial.

Exercise 3.25. Follow the proof of Theorem 3.37 and Theorem 3.38.

EXERCISES FOR CHAPTER 4

Exercise 4.2. The set $\mathcal{U}_\lambda = \mathcal{L}_\lambda^c$ is closed by Proposition 4.4.

Exercise 4.3. Consider the l.s.c. function $f(x) := 0$ for $x \neq 0$ and $f(0) := -1$. Then the product of $f \cdot f$ is not l.s.c.

Exercise 4.7. Solve the equation $\nabla g(x) = 0$, where $\nabla g(x) = \sum_{i=1}^{m} 2(x - a_i)$.

Exercise 4.8. Let $\Omega := \{x \in \mathbb{R}^n \mid Ax = b\}$. Then $N(x; \Omega) = \{A^T\lambda \mid \lambda \in \mathbb{R}^p\}$ for $x \in \Omega$. Corollary 4.15 tells us that $x \in \Omega$ is an optimal solution to this problem if and only if there exists $\lambda \in \mathbb{R}^p$ such that

$$-2x = A^T\lambda \ \text{ and } \ Ax = b.$$

Solving this system of equations yields $x = A^T(AA^T)^{-1}b$.

Exercise 4.11. The MATLAB function $\operatorname{sign} x$ gives us a subgradient of the function $f(x) := \|x\|_1$ for all $x \in \mathbb{R}$. Considering the set $\Omega := \{x \in \mathbb{R}^n \mid Ax = b\}$, we have $\Pi(x; \Omega) = x + A^T(AA^T)^{-1}(b - Ax)$, which can be found by minimizing the quadratic function $\|x - u\|^2$ over $u \in \Omega$; see Exercise 4.8.

Exercise 4.12. If $0 \in \partial f(x_{k_0})$ for some $k_0 \in \mathbb{N}$, then $\alpha_{k_0} = 0$, and hence $x_k = x_{k_0} \in S$ and $f(x_k) = \overline{V}$ for all $k \geq k_0$. Thus we assume without loss of generality that $0 \notin \partial f(x_k)$ as $k \in \mathbb{N}$. It follows from the proof of Proposition 4.23 that

$$0 \leq \|x_{k+1} - \bar{x}\|^2 \leq \|x_1 - \bar{x}\|^2 - 2\sum_{i=1}^{k} \alpha_i [f(x_i) - \overline{V}] + \sum_{i=1}^{k} \alpha_i^2 \|v_i\|^2$$

for any $\bar{x} \in S$. Substituting here the formula for α_k gives us

$$\begin{aligned} 0 \leq \|x_{k+1} - \bar{x}\|^2 &\leq \|x_1 - \bar{x}\|^2 - 2\sum_{i=1}^{k} \frac{[f(x_i) - \overline{V}]^2}{\|v_i\|^2} + \sum_{i=1}^{k} \frac{[f(x_i) - \overline{V}]^2}{\|v_i\|^2} \\ &= \|x_1 - \bar{x}\|^2 - \sum_{i=1}^{k} \frac{[f(x_i) - \overline{V}]^2}{\|v_i\|^2} \quad \text{and} \end{aligned}$$

$$\sum_{i=1}^{k} \frac{[f(x_i) - \overline{V}]^2}{\|v_i\|^2} \leq \|x_1 - \bar{x}\|^2.$$

Since $\|v_i\| \leq \ell$ for every i, we get

$$\sum_{i=1}^{k} \frac{[f(x_i) - \overline{V}]^2}{\ell^2} \leq \|x_1 - \bar{x}\|^2,$$

which implies in turn that

$$\sum_{i=1}^{k} [f(x_i) - \overline{V}]^2 \leq \ell^2 \|x_1 - \bar{x}\|^2. \tag{4.66}$$

Then the series $\sum_{k=1}^{\infty}[f(x_k)-\overline{V}]^2$ is convergent. Hence $f(x_k)\to\overline{V}$ as $k\to\infty$, which justifies **(i)**. From (4.66) and the fact that $V_k\le f(x_i)$ for $i\in\{1,\ldots,k\}$ it follows that

$$k(V_k-\overline{V})^2\le \ell^2\|x_1-\bar{x}\|^2.$$

Thus $0\le V_k-\overline{V}\le \dfrac{\ell\|x_1-\bar{x}\|}{\sqrt{k}}$. Since $\bar{x}$ was taken arbitrarily in S, we arrive at **(ii)**.

Exercise 4.17. Define the MATLAB function

$$f(x):=\frac{\sum_{i=1}^{m}\dfrac{a_i}{\|x-a_i\|}}{\sum_{i=1}^{m}\dfrac{1}{\|x-a_i\|}},\quad x\notin\{a_1,\ldots,a_m\}$$

and $f(x):=x$ for $x\in\{a_1,\ldots,a_m\}$. Use the FOR loop with a stopping criteria to find $x_{k+1}:=f(x_k)$, where x_0 is a starting point.

Exercise 4.18. Use the subdifferential Fermat rule. Consider the cases where m is odd and where m is even. In the first case the problem has a unique optimal solution. In the second case the problem has infinitely many solutions.

Exercise 4.22. For the interval $\Omega:=[a,b]\subset\mathbb{R}$, the subdifferential formula for the distance function 2.39 gives us

$$\partial d(\bar{x};\Omega)=\begin{cases}\{0\} & \text{if } \bar{x}\in(a,b),\\ [0,1] & \text{if } \bar{x}=b,\\ \{1\} & \text{if } \bar{x}>b,\\ [-1,0] & \text{if } \bar{x}=a,\\ \{-1\} & \text{if } \bar{x}<b.\end{cases}$$

Exercise 4.23. For $\Omega:=I\!B(c;r)\subset\mathbb{R}^n$, a subgradient of the function $d(x;\Omega)$ at $\bar{x}$ is

$$v(\bar{x}):=\begin{cases}0 & \text{if } \bar{x}\in\Omega,\\ \dfrac{\bar{x}-c}{\|\bar{x}-c\|} & \text{if } \bar{x}\notin\Omega.\end{cases}$$

Then a subgradient of the function f can be found by the subdifferential sum rule of Corollary 2.46. Note that the qualification condition (2.29) holds automatically in this case.

Exercise 4.27. Use the subgradient formula in the hint of Exercise 4.23 together with Proposition 2.54. Note that we can easily find an element $i\in I(\bar{x})$, and then any subgradient of the function $f_i(x):=d(x;\Omega_i)$ at $\bar{x}$ belongs to $\partial f(\bar{x})$.

Bibliography

[1] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, 2011. DOI: 10.1007/978-1-4419-9467-7.

[2] D. P. Bertsekas, A. Nedić and A. E. Ozdaglar, *Convex Analysis and Optimization*, Athena Scientific, 2003.

[3] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004. DOI: 10.1017/CBO9780511804441. xi

[4] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization*, Springer, 2000. DOI: 10.1007/978-1-4757-9859-3. xi

[5] J. M. Borwein and Q. J. Zhu, *Techniques of Variational Analysis*, Springer, 2005.

[6] Z. Drezner and H. W. Hamacher (Eds.), *Facility Location: Applications and Theory*, Springer, 2004. 154

[7] G. Colombo and P. R. Wolenski, "Variational analysis for a class of minimal time functions in Hilbert spaces," *J. Convex Anal.*, vol. 11, pp. 335–361, 2004.

[8] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, Vols. I and II, Springer, 1993. DOI: 10.1007/978-3-662-06409-2. xi

[9] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*, Springer, 2001. DOI: 10.1007/978-3-642-56468-0. xi, 185

[10] H. W. Kuhn, "A note on Fermat-Torricelli problem," *Math. Program.*, vol. 54, pp. 98–107, 1973. DOI: 10.1007/BF01584648. 151

[11] Y. S. Kupitz, H. Martini and M. Spirova, "The Fermat-Torricelli problem, Part I: A discrete gradient-method approach," *Optim. Theory Appl.*, vol. 158, pp. 305–327, 2013. DOI: 10.1007/s10957-013-0266-z. 154

[12] S. R. Lay, *Convex Sets and Their Applications*, Wiley & Sons, 1982.

[13] G. G. Magaril-Il'yaev and V. M. Tikhomirov, *Convex Analysis: Theory and Applications*, American Mathematical Society, 2003.

[14] B. S. Mordukhovich, *Variational Analayis and Generalized Differentiation*, Vols. I and II, Springer, 2006. DOI: 10.1007/3-540-31246-3. 40, 41, 49

[15] B. S. Mordukhovich and N. M. Nam, "Applications of variational analysis to a generalized Fermat-Torricelli problem," *Optim. Theory Appl.*, vol. 148, pp. 431–454, 2011. DOI: 10.1007/s10957-010-9761-7. 154

[16] B. S. Mordukhovich, N. M. Nam and J. Salinas, "Solving a generalized Heron problem by means of convex analysis," *Amer. Math. Monthly*, vol. 119, pp. 87–99, 2012. DOI: 10.4169/amer.math.monthly.119.02.087. 154

[17] B. S. Mordukhovich, N. M. Nam and J. Salinas, "Applications of variational analysis to a generalized Heron problem," *Applic. Anal.*, vol. 91, pp. 1915–1942, 2012. DOI: 10.1080/00036811.2011.604849. 154

[18] B. S. Mordukhovich, N. M. Nam and C. Villalobos, "The smallest enclosing ball problem and the smallest intersecting ball problem: Existence and uniqueness of solutions," *Optim. Letters*, vol. 154, pp. 768–791, 2012. DOI: 10.1007/s11590-012-0483-7. 168

[19] N. M. Nam and N. Hoang, "A generalized Sylvester problem and a generalized Fermat-Torricelli problem," *J. Convex Anal.*, vol. 20, pp. 669–687, 2013. 168

[20] N. M. Nam, N. Hoang and N. T. An, "Constructions of solutions to generalized Sylvester and Fermat-Torricelli problems for Euclidean balls," *Optim. Theory Appl.*, in press. DOI: 10.1007/s10957-013-0366-9. 168

[21] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Kluwer Academic Publishers, 2004. DOI: 10.1007/978-1-4419-8853-9. xi

[22] Y. Nesterov, "Smooth minimization of nonsmooth functions," *Math Program.*, vol. 103, pp. 127–152, 2005. DOI: 10.1007/s10107-004-0552-5. 191

[23] J. F. Nash, *Noncooperative Games*, Doctoral thesis, Princeton University, 1950. 122

[24] J. F. Nash, "Equilibrium points in N-person games," *Proc. Nat. Acad. Sci.* vol. 36, pp. 48–49, 1950. DOI: 10.1073/pnas.36.1.48. 122

[25] B. T. Polyak, *An Introduction to Optimization*, Optimization Solftware, 1987.

[26] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970. xi, 54

[27] R. T. Rockafellar and R. J-B. Wets, *Variational Analysis*, Springer, 1998. DOI: 10.1007/978-3-642-02431-3. 49

[28] R. T. Rockafellar, "Applications of convex variational analysis to Nash equilibrium," *Proceedings of 7th International Conference on Nonlinear Analysis and Convex Analysis* (Busan, Korea), pp. 173–183, 2011. 123

[29] A. Ruszczynski, *Nonlinear Optimization*, Princeton University Press, 2006. 154

[30] N. Z. Shor, *Minimization Methods for Nondifferentiable Functions*, Springer, 1985. DOI: 10.1007/978-3-642-82118-9.

[31] J. J. Sylvester, "A question in the geometry of situation," *Quart. J. Math.*, vol. 1, p. 79, 1857. 168

[32] E. Weiszfeld, "Sur le point pour lequel la somme des distances de *n* points donnés est minimum," *Tôhoku Mathematics Journal*, vol. 43, pp. 355–386, 1937. 150, 151

[33] E. Zeidler, *Applied Functional Analysis. Applications to Mathematical Physics*, Springer, 1997. 123

Authors' Biographies

BORIS MORDUKHOVICH

Boris Mordukhovich is Distinguished University Professor of Mathematics at Wayne State University holding also Chair Professorship of Mathematics and Statistics at the King Fahd University of Petroleum and Minerals. He has more than 350 publications including several monographs. Among his best known achievements are the introduction and development of powerful constructions of generalized differentiation and their applications to broad classes of problems in variational analysis, optimization, equilibrium, control, economics, engineering, and other fields. Mordukhovich is a SIAM Fellow, an AMS Fellow, and a recipient of many international awards and honors including Doctor Honoris Causa degrees from six universities over the world. He is named a Highly Cited Researcher in Mathematics by the Institute of Scientific Information (ISI). His research has been supported by continued grants from the National Science Foundations.

Department of Mathematics, Wayne State University, Detroit, MI 48202
boris@math.wayne.edu. Homepage: math.wayne.edu/∼boris

NGUYEN MAU NAM

Nguyen Mau Nam received his B.S. from Hue University, Vietnam, in 1998 and a Ph.D. from Wayne State University in 2007. He is currently Assistant Professor of Mathematics at Portland State University. He has around 30 papers on convex and variational analysis, optimization, and their applications published in or accepted by high-level mathematical journals. His research is supported by a collaboration grant from the Simons Foundation.

Fariborz Maseeh Department of Mathematics and Statistics,
Portland State University, Portland, OR 97201
mau.nam.nguyen@pdx.edu. Homepage: web.pdx.edu/∼mnn3

Index