

FIRST-ORDER METHODS IN OPTIMIZATION



MOS-SIAM Series on Optimization

This series is published jointly by the Mathematical Optimization Society and the Society for Industrial and Applied Mathematics. It includes research monographs, books on applications, textbooks at all levels, and tutorials. Besides being of high scientific quality, books in the series must advance the understanding and practice of optimization. They must also be written clearly and at an appropriate level for the intended audience.

Editor-in-Chief

Katya Scheinberg
Lehigh University

Editorial Board

Santanu S. Dey, *Georgia Institute of Technology*
 Maryam Fazel, *University of Washington*
 Andrea Lodi, *University of Bologna*
 Arkadi Nemirovski, *Georgia Institute of Technology*
 Stefan Ulbrich, *Technische Universität Darmstadt*
 Luis Nunes Vicente, *University of Coimbra*
 David Williamson, *Cornell University*
 Stephen J. Wright, *University of Wisconsin*

Series Volumes

- Beck, Amir, *First-Order Methods in Optimization*
 Terlaky, Tamás, Anjos, Miguel F., and Ahmed, Shabbir, editors, *Advances and Trends in Optimization with Engineering Applications*
 Todd, Michael J., *Minimum-Volume Ellipsoids: Theory and Algorithms*
 Bienstock, Daniel, *Electrical Transmission System Cascades and Vulnerability: An Operations Research Viewpoint*
 Koch, Thorsten, Hiller, Benjamin, Pfetsch, Marc E., and Schewe, Lars, editors, *Evaluating Gas Network Capacities*
 Corberán, Ángel, and Laporte, Gilbert, *Arc Routing: Problems, Methods, and Applications*
 Toth, Paolo, and Vigo, Daniele, *Vehicle Routing: Problems, Methods, and Applications, Second Edition*
 Beck, Amir, *Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB*
 Attouch, Hedy, Buttazzo, Giuseppe, and Michaille, Gérard, *Variational Analysis in Sobolev and BV Spaces: Applications to PDEs and Optimization, Second Edition*
 Shapiro, Alexander, Dentcheva, Darinka, and Ruszczyński, Andrzej, *Lectures on Stochastic Programming: Modeling and Theory, Second Edition*
 Locatelli, Marco and Schoen, Fabio, *Global Optimization: Theory, Algorithms, and Applications*
 De Loera, Jesús A., Hemmecke, Raymond, and Köppe, Matthias, *Algebraic and Geometric Ideas in the Theory of Discrete Optimization*
 Blekherman, Grigoriy, Parrilo, Pablo A., and Thomas, Rekha R., editors, *Semidefinite Optimization and Convex Algebraic Geometry*
 Delfour, M. C., *Introduction to Optimization and Semidifferential Calculus*
 Ulbrich, Michael, *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*
 Biegler, Lorenz T., *Nonlinear Programming: Concepts, Algorithms, and Applications to Chemical Processes*
 Shapiro, Alexander, Dentcheva, Darinka, and Ruszczyński, Andrzej, *Lectures on Stochastic Programming: Modeling and Theory*
 Conn, Andrew R., Scheinberg, Katya, and Vicente, Luis N., *Introduction to Derivative-Free Optimization*
 Ferris, Michael C., Mangasarian, Olvi L., and Wright, Stephen J., *Linear Programming with MATLAB*
 Attouch, Hedy, Buttazzo, Giuseppe, and Michaille, Gérard, *Variational Analysis in Sobolev and BV Spaces: Applications to PDEs and Optimization*
 Wallace, Stein W. and Ziema, William T., editors, *Applications of Stochastic Programming*
 Grötschel, Martin, editor, *The Sharpest Cut: The Impact of Manfred Padberg and His Work*
 Renegar, James, *A Mathematical View of Interior-Point Methods in Convex Optimization*
 Ben-Tal, Aharon and Nemirovski, Arkadi, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*
 Conn, Andrew R., Gould, Nicholas I. M., and Toint, Phillippe L., *Trust-Region Methods*

FIRST-ORDER METHODS IN OPTIMIZATION

Amir Beck

Tel-Aviv University
Tel-Aviv
Israel



Society for Industrial and Applied Mathematics
Philadelphia



Mathematical
Optimization Society

Mathematical Optimization Society
Philadelphia

Copyright © 2017 by the Society for Industrial and Applied Mathematics and the Mathematical Optimization Society

10 9 8 7 6 5 4 3 2 1

All rights reserved. Printed in the United States of America. No part of this book may be reproduced, stored, or transmitted in any manner without the written permission of the publisher. For information, write to the Society for Industrial and Applied Mathematics, 3600 Market Street, 6th Floor, Philadelphia, PA 19104-2688 USA.

Trademarked names may be used in this book without the inclusion of a trademark symbol. These names are used in an editorial context only; no infringement of trademark is intended.

<i>Publisher</i>	Kivmars Bowling
<i>Acquisitions Editor</i>	Paula Callaghan
<i>Developmental Editor</i>	Gina Rinelli Harris
<i>Managing Editor</i>	Kelly Thomas
<i>Production Editor</i>	Louis R. Primus
<i>Copy Editor</i>	Bruce Owens
<i>Production Manager</i>	Donna Witzleben
<i>Production Coordinator</i>	Cally Shrader
<i>Compositor</i>	Cheryl Hufnagle
<i>Graphic Designer</i>	Lois Sellers

Library of Congress Cataloging-in-Publication Data

Please visit www.siam.org/books/mo25 to view the CIP data.



For

My wife, Nili

My daughters, Noy and Vered

My parents, Nili and Itzhak



Contents

Preface	xi
1 Vector Spaces	1
1.1 Definition	1
1.2 Dimension	2
1.3 Norms	2
1.4 Inner Products	2
1.5 Affine Sets and Convex Sets	3
1.6 Euclidean Spaces	3
1.7 The Space \mathbb{R}^n	4
1.8 The Space $\mathbb{R}^{m \times n}$	6
1.9 Cartesian Product of Vector Spaces	7
1.10 Linear Transformations	8
1.11 The Dual Space	9
1.12 The Bidual Space	10
1.13 Adjoint Transformations	11
1.14 Norms of Linear Transformations	12
2 Extended Real-Valued Functions	13
2.1 Extended Real-Valued Functions and Closedness	13
2.2 Closedness versus Continuity	18
2.3 Convex Functions	21
2.4 Support Functions	26
3 Subgradients	35
3.1 Definitions and First Examples	35
3.2 Properties of the Subdifferential Set	39
3.3 Directional Derivatives	44
3.4 Computing Subgradients	53
3.5 The Value Function	67
3.6 Lipschitz Continuity and Boundedness of Subgradients	71
3.7 Optimality Conditions	72
3.8 Summary of Weak and Strong Subgradient Calculus Results	84
4 Conjugate Functions	87
4.1 Definition and Basic Properties	87
4.2 The Biconjugate	89

4.3	Conjugate Calculus Rules	91
4.4	Examples	93
4.5	Infimal Convolution and Conjugacy	102
4.6	Subdifferentials of Conjugate Functions	104
5	Smoothness and Strong Convexity	107
5.1	<i>L</i> -Smooth Functions	107
5.2	Strong Convexity	117
5.3	Smoothness and Strong Convexity Correspondence	123
6	The Proximal Operator	129
6.1	Definition, Existence, and Uniqueness	129
6.2	First Set of Examples of Proximal Mappings	131
6.3	Prox Calculus Rules	135
6.4	Prox of Indicators—Orthogonal Projections	146
6.5	The Second Prox Theorem	157
6.6	Moreau Decomposition	160
6.7	The Moreau Envelope	163
6.8	Miscellaneous Prox Computations	170
6.9	Summary of Prox Computations	177
7	Spectral Functions	179
7.1	Symmetric Functions	179
7.2	Symmetric Spectral Functions over \mathbb{S}^n	182
7.3	Symmetric Spectral Functions over $\mathbb{R}^{m \times n}$	188
8	Primal and Dual Projected Subgradient Methods	195
8.1	From Gradient Descent to Subgradient Descent	195
8.2	The Projected Subgradient Method	201
8.3	The Stochastic Projected Subgradient Method	221
8.4	The Incremental Projected Subgradient Method	229
8.5	The Dual Projected Subgradient Method	232
9	Mirror Descent	247
9.1	From Projected Subgradient to Mirror Descent	247
9.2	Convergence Analysis	252
9.3	Mirror Descent for the Composite Model	260
10	The Proximal Gradient Method	269
10.1	The Composite Model	269
10.2	The Proximal Gradient Method	270
10.3	Analysis of the Proximal Gradient Method—The Nonconvex Case	272
10.4	Analysis of the Proximal Gradient Method—The Convex Case .	281
10.5	The Proximal Point Method	288
10.6	Convergence of the Proximal Gradient Method—The Strongly Convex Case	288
10.7	The Fast Proximal Gradient Method—FISTA	290
10.8	Smoothing	304

10.9	Non-Euclidean Proximal Gradient Methods	315
11	The Block Proximal Gradient Method	331
11.1	Decomposition Methods	331
11.2	Model and Assumptions	332
11.3	The Toolbox	333
11.4	The Cyclic Block Proximal Gradient Method	338
11.5	The Randomized Block Proximal Gradient Method	347
12	Dual-Based Proximal Gradient Methods	353
12.1	The Primal and Dual Models	353
12.2	The Dual Proximal Gradient Method	355
12.3	Fast Dual Proximal Gradient	358
12.4	Examples I	360
12.5	The Dual Block Proximal Gradient Method	369
12.6	Examples II	374
13	The Generalized Conditional Gradient Method	379
13.1	The Frank–Wolfe/Conditional Gradient Method	379
13.2	The Generalized Conditional Gradient Method	380
13.3	The Strongly Convex Case	391
13.4	The Randomized Generalized Block Conditional Gradient Method	400
14	Alternating Minimization	405
14.1	The Method	405
14.2	Coordinate-wise Minima	407
14.3	The Composite Model	411
14.4	Convergence in the Convex Case	413
14.5	Rate of Convergence in the Convex Case	415
15	ADMM	423
15.1	The Augmented Lagrangian Method	423
15.2	Alternating Direction Method of Multipliers (ADMM)	425
15.3	Convergence Analysis of AD-PMM	427
15.4	Minimizing $f_1(\mathbf{x}) + f_2(\mathbf{A}\mathbf{x})$	432
A	Strong Duality and Optimality Conditions	439
B	Tables	443
C	Symbols and Notation	451
D	Bibliographic Notes	457
	Bibliography	463
	Index	473

Preface

This book, as the title suggests, is about first-order methods, namely, methods that exploit information on values and gradients/subgradients (but not Hessians) of the functions comprising the model under consideration. First-order methods go back to 1847 with the work of Cauchy on the steepest descent method. With the increase in the amount of applications that can be modeled as large- or even huge-scale optimization problems, there has been a revived interest in using simple methods that require low iteration cost as well as low memory storage.

The primary goal of the book is to provide in a self-contained manner a comprehensive study of the main first-order methods that are frequently used in solving large-scale problems. This is done by gathering and reorganizing in a unified manner many results that are currently scattered throughout the literature. Special emphasis is placed on rates of convergence and complexity analysis. Although the name of the book is “first-order methods in optimization,” two disclaimers are in order. First, we will actually also consider methods that exploit additional operations at each iteration such as prox evaluations, linear oracles, exact minimization w.r.t. blocks of variables, and more, so perhaps a more suitable name would have been “simple methods in optimization.” Second, in order to be truly self-contained, the first part of the book (Chapters 1–7) is actually purely theoretical and contains essential topics that are crucial for the developments in the algorithmic part (Chapters 8–15).

The book is intended for students and researchers with a background in advanced calculus and linear algebra, as well as prior knowledge in the fundamentals of optimization (some convex analysis, optimality conditions, and duality). A MATLAB toolbox implementing many of the algorithms described in the book was developed by the author and Nili Guttmann-Beck and can be found at www.siam.org/books/mo25.

The outline of the book is as follows. Chapter 1 reviews important facts about vector spaces. Although the material is quite fundamental, it is advisable not to skip this chapter since many of the conventions regarding the underlying spaces used in the book are explained. Chapter 2 focuses on extended real-valued functions with a special emphasis on properties such as convexity, closedness, and continuity. Chapter 3 covers the topic of subgradients starting from basic definitions, continuing with directional derivatives, differentiability, and subdifferentiability and ending with calculus rules. Optimality conditions are derived for convex problems (Fermat’s optimality condition), but also for the nonconvex composite model, which will be discussed extensively throughout the book. Conjugate functions are the subject of Chapter 4, which covers several issues, such as Fenchel’s

inequality, the biconjugate, calculus rules, conjugate subgradient theorem, relations with the infimal convolution, and Fenchel's duality theorem. Chapter 5 covers two different but closely related subjects: smoothness and strong convexity—several characterizations of each of these concepts are given, and their relation via the conjugate correspondence theorem is established. The proximal operator is discussed in Chapter 6, which includes a large amount of prox computations as well as calculus rules. The basic properties of the proximal mapping (first and second prox theorems and Moreau decomposition) are proved, and the Moreau envelope concludes the theoretical part of the chapter. The first part of the book ends with Chapter 7, which contains a study of symmetric spectral functions. The second, algorithmic part of the book starts with Chapter 8 with primal and dual projected subgradient methods. Several stepsize rules are discussed, and complexity results for both the convex and the strongly convex cases are established. The chapter also includes discussions on the stochastic as well as the incremental projected subgradient methods. The non-Euclidean version of the projected subgradient method, a.k.a. the mirror descent method, is discussed in Chapter 9. Chapter 10 is concerned with the proximal gradient method as well as its many variants and extensions. The chapter also studies several theoretical results concerning the so-called gradient mapping, which plays an important part in the convergence analysis of proximal gradient-based methods. The extension of the proximal gradient method to the *block* proximal gradient method is discussed in Chapter 11, while Chapter 12 considers the dual proximal gradient method and contains a result on a primal-dual relation that allows one to transfer rate of convergence results from the dual problem to the primal problem. The generalized conditional gradient method is the topic of Chapter 13, which contains the basic rate of convergence results of the method, as well as its block version, and discusses the effect of strong convexity assumptions on the model. The alternating minimization method is the subject of Chapter 14, where its convergence (as well as divergence) in many settings is established and illustrated. The book concludes with a discussion on the ADMM method in Chapter 15.

My deepest thanks to Marc Teboulle, whose fundamental works in first-order methods form the basis of many of the results in the book. Marc introduced me to the world of optimization, and he is a constant source and inspiration and admiration. I would like to thank Luba Tetruashvili for reading the book and for her helpful remarks. It has been a pleasure to work with the extremely devoted and efficient SIAM staff. Finally, I would like to acknowledge the support of the Israel Science Foundation for supporting me while writing this book.

Chapter 1

Vector Spaces

This chapter reviews several important facts about different aspects of vectors spaces that will be used throughout the book. More comprehensive and detailed accounts of these subjects can be found in advanced linear algebra books.

1.1 Definition

A *vector space* \mathbb{E} over \mathbb{R} (or a “real vector space”) is a set of elements called *vectors* such that the following holds.

- (A) For any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{E}$, there corresponds a vector $\mathbf{x} + \mathbf{y}$, called the *sum* of \mathbf{x} and \mathbf{y} , satisfying the following properties:
 - 1. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{E}$.
 - 2. $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$ for any $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{E}$.
 - 3. There exists in \mathbb{E} a unique vector $\mathbf{0}$ (called the *zeros vector*) such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for any \mathbf{x} .
 - 4. For any $\mathbf{x} \in \mathbb{E}$, there exists a vector $-\mathbf{x} \in \mathbb{E}$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$.
- (B) For any real number (also called *scalar*) $\alpha \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{E}$, there corresponds a vector $\alpha\mathbf{x}$ called the *scalar multiplication* of α and \mathbf{x} satisfying the following properties:
 - 1. $\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$ for any $\alpha, \beta \in \mathbb{R}, \mathbf{x} \in \mathbb{E}$.
 - 2. $1\mathbf{x} = \mathbf{x}$ for any $\mathbf{x} \in \mathbb{E}$.
- (C) The two operations (summation, scalar multiplication) satisfy the following properties:
 - 1. $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$ for any $\alpha \in \mathbb{R}, \mathbf{x}, \mathbf{y} \in \mathbb{E}$.
 - 2. $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$ for any $\alpha, \beta \in \mathbb{R}, \mathbf{x} \in \mathbb{E}$.

1.2 Dimension

A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ in a vector space \mathbb{E} is called *linearly independent* or just *independent* if the linear system

$$\sum_{i=1}^n \alpha_i \mathbf{v}_i = \mathbf{0}$$

implies that $\alpha_1 = \alpha_2 = \dots = \alpha_n = 0$. In other words, there does not exist a nontrivial linear combination of vectors that is equal to the zeros vector. A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is said to *span* \mathbb{E} if for any $\mathbf{x} \in \mathbb{E}$, there exist $\beta_1, \beta_2, \dots, \beta_n \in \mathbb{R}$ such that

$$\mathbf{x} = \sum_{i=1}^n \beta_i \mathbf{v}_i.$$

A *basis* of a vector space \mathbb{E} is an independent set of vectors that spans \mathbb{E} . It is well known that the number of vectors in all the bases of a vector space \mathbb{E} is the same; this number is called the *dimension* of the space and is denoted by $\dim(\mathbb{E})$. In this book we will discuss only vector spaces with a finite dimension, namely, *finite-dimensional vector spaces*.

1.3 Norms

A *norm* $\|\cdot\|$ on a vector space \mathbb{E} is a function $\|\cdot\| : \mathbb{E} \rightarrow \mathbb{R}$ satisfying the following properties:

1. **(nonnegativity)** $\|\mathbf{x}\| \geq 0$ for any $\mathbf{x} \in \mathbb{E}$ and $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$.
2. **(positive homogeneity)** $\|\lambda \mathbf{x}\| = |\lambda| \cdot \|\mathbf{x}\|$ for any $\mathbf{x} \in \mathbb{E}$ and $\lambda \in \mathbb{R}$.
3. **(triangle inequality)** $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{E}$.

We will sometimes denote the norm of a space \mathbb{E} by $\|\cdot\|_{\mathbb{E}}$ to emphasize the identity of the space and to distinguish it from other norms. The *open ball* with center $\mathbf{c} \in \mathbb{E}$ and radius $r > 0$ is denoted by $B(\mathbf{c}, r)$ and defined by

$$B(\mathbf{c}, r) = \{\mathbf{x} \in \mathbb{E} : \|\mathbf{x} - \mathbf{c}\| < r\}.$$

The *closed ball* with center $\mathbf{c} \in \mathbb{E}$ and radius $r > 0$ is denoted by $B[\mathbf{c}, r]$ and defined by

$$B[\mathbf{c}, r] = \{\mathbf{x} \in \mathbb{E} : \|\mathbf{x} - \mathbf{c}\| \leq r\}.$$

We will sometimes use the notation $B_{\|\cdot\|}[\mathbf{c}, r]$ or $B_{\|\cdot\|}(\mathbf{c}, r)$ to identify the specific norm that is being used.

1.4 Inner Products

An *inner product* of a real vector space \mathbb{E} is a function that associates to each pair of vectors \mathbf{x}, \mathbf{y} a real number, which is denoted by $\langle \mathbf{x}, \mathbf{y} \rangle$ and satisfies the following properties:

1. **(commutativity)** $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{E}$.

2. **(linearity)** $\langle \alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2, \mathbf{y} \rangle = \alpha_1 \langle \mathbf{x}_1, \mathbf{y} \rangle + \alpha_2 \langle \mathbf{x}_2, \mathbf{y} \rangle$ for any $\alpha_1, \alpha_2 \in \mathbb{R}$ and $\mathbf{x}_1, \mathbf{x}_2, \mathbf{y} \in \mathbb{E}$.
3. **(positive definiteness)** $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ for any $\mathbf{x} \in \mathbb{E}$ and $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ if and only if $\mathbf{x} = \mathbf{0}$.

A vector space endowed with an inner product is also called an *inner product space*. At this point we would like to make the following important note:

Underlying Spaces: In this book the underlying vector spaces, usually denoted by \mathbb{V} or \mathbb{E} , are always finite dimensional real inner product spaces with endowed inner product $\langle \cdot, \cdot \rangle$ and endowed norm $\| \cdot \|$.

1.5 Affine Sets and Convex Sets

Given a real vector space \mathbb{E} , a set $S \subseteq \mathbb{E}$ is called *affine* if for any $\mathbf{x}, \mathbf{y} \in S$ and $\lambda \in \mathbb{R}$, the inclusion $\lambda \mathbf{x} + (1 - \lambda) \mathbf{y} \in S$ holds. For a set $S \subseteq \mathbb{E}$, the *affine hull* of S , denoted by $\text{aff}(S)$, is the intersection of all affine sets containing S . Clearly, $\text{aff}(S)$ is by itself an affine set, and it is the smallest affine set containing S (w.r.t. inclusion). A *hyperplane* is a subset of \mathbb{E} given by

$$H_{\mathbf{a}, b} = \{\mathbf{x} \in \mathbb{E} : \langle \mathbf{a}, \mathbf{x} \rangle = b\},$$

where $\mathbf{a} \in \mathbb{E}$ and $b \in \mathbb{R}$. It is an easy exercise to show that hyperplanes are affine sets.

A set $C \subseteq \mathbb{E}$ is called *convex* if for any $\mathbf{x}, \mathbf{y} \in C$ and $\lambda \in [0, 1]$ it holds that $\lambda \mathbf{x} + (1 - \lambda) \mathbf{y} \in C$. Evidently, affine sets are always convex. Open and closed balls are always convex regardless of the choice of norm. For given $\mathbf{x}, \mathbf{y} \in \mathbb{E}$, the *closed line segment* between \mathbf{x} and \mathbf{y} is a subset of \mathbb{E} denoted by $[\mathbf{x}, \mathbf{y}]$ and defined as

$$[\mathbf{x}, \mathbf{y}] = \{\alpha \mathbf{x} + (1 - \alpha) \mathbf{y} : \alpha \in [0, 1]\}.$$

The *open line segment* (\mathbf{x}, \mathbf{y}) is similarly defined as

$$(\mathbf{x}, \mathbf{y}) = \{\alpha \mathbf{x} + (1 - \alpha) \mathbf{y} : \alpha \in (0, 1)\}$$

when $\mathbf{x} \neq \mathbf{y}$ and is the empty set \emptyset when $\mathbf{x} = \mathbf{y}$. Closed and open line segments are convex sets. Another example of convex sets are *half-spaces*, which are sets of the form

$$H_{\mathbf{a}, b}^- = \{\mathbf{x} \in \mathbb{E} : \langle \mathbf{a}, \mathbf{x} \rangle \leq b\},$$

where $\mathbf{a} \in \mathbb{E}$ and $b \in \mathbb{R}$.

1.6 Euclidean Spaces

A finite dimensional real vector space equipped with an inner product $\langle \cdot, \cdot \rangle$ is called a *Euclidean space* if it is endowed with the norm $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$, which is referred to as the *Euclidean norm*.

1.7 The Space \mathbb{R}^n

The vector space \mathbb{R}^n (n being a positive integer) is the set of n -dimensional column vectors with real components endowed with the component-wise addition operator,

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix},$$

and the scalar-vector product,

$$\lambda \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{pmatrix},$$

where in the above $x_1, x_2, \dots, x_n, \lambda$ are real numbers. We will denote the standard basis of \mathbb{R}^n by $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$, where \mathbf{e}_i is the n -length column vector whose i th component is one while all the others are zeros. The column vectors of all ones and all zeros will be denoted by \mathbf{e} and $\mathbf{0}$, respectively, where the length of the vectors will be clear from the context.

By far the most used inner product in \mathbb{R}^n is the *dot product* defined by

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

Inner Product in \mathbb{R}^n : In this book, unless otherwise stated, the endowed inner product in \mathbb{R}^n is the dot product.

Of course, the dot product is not the only possible inner product that can be defined over \mathbb{R}^n . Another useful option is the \mathbf{Q} -inner product, which is defined as

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{Q}} = \mathbf{x}^T \mathbf{Q} \mathbf{y},$$

where \mathbf{Q} is a positive definite $n \times n$ matrix. Obviously, the \mathbf{Q} -inner product amounts to the dot product when $\mathbf{Q} = \mathbf{I}$. If \mathbb{R}^n is endowed with the dot product, then the associated Euclidean norm is the l_2 -norm

$$\|\mathbf{x}\|_2 = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\sum_{i=1}^n x_i^2}.$$

If \mathbb{R}^n is endowed with the \mathbf{Q} -inner product, then the associated Euclidean norm is the \mathbf{Q} -norm

$$\|\mathbf{x}\|_{\mathbf{Q}} = \sqrt{\mathbf{x}^T \mathbf{Q} \mathbf{x}}.$$

For a given $p \geq 1$, the l_p -norm on \mathbb{R}^n is given by the formula

$$\|\mathbf{x}\|_p = \sqrt[p]{\sum_{i=1}^n |x_i|^p}.$$

The l_∞ -norm on \mathbb{R}^n is defined by

$$\|\mathbf{x}\|_\infty = \max_{i=1,2,\dots,n} |x_i|.$$

1.7.1 Subsets of \mathbb{R}^n

The *nonnegative orthant* is the subset of \mathbb{R}^n consisting of all vectors in \mathbb{R}^n with nonnegative components and is denoted by \mathbb{R}_+^n :

$$\mathbb{R}_+^n = \{(x_1, x_2, \dots, x_n)^T : x_1, x_2, \dots, x_n \geq 0\}.$$

Similarly, the *positive orthant* consists of all the vectors in \mathbb{R}^n with positive components and is denoted by \mathbb{R}_{++}^n :

$$\mathbb{R}_{++}^n = \{(x_1, x_2, \dots, x_n)^T : x_1, x_2, \dots, x_n > 0\}.$$

The *unit simplex*, denoted by Δ_n , is the subset of \mathbb{R}^n comprising all nonnegative vectors whose components sum up to one:

$$\Delta_n = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \geq \mathbf{0}, \mathbf{e}^T \mathbf{x} = 1\}.$$

Given two vectors $\ell, \mathbf{u} \in \mathbb{R}^n$ that satisfy $\ell \leq \mathbf{u}$, the *box* with lower bounds ℓ and upper bounds \mathbf{u} is denoted by $\text{Box}[\ell, \mathbf{u}]$ and defined as

$$\text{Box}[\ell, \mathbf{u}] = \{\mathbf{x} \in \mathbb{R}^n : \ell \leq \mathbf{x} \leq \mathbf{u}\}.$$

Thus, for example, $\text{Box}[-\mathbf{e}, \mathbf{e}] = [-1, 1]^n$.

1.7.2 Operations on Vectors in \mathbb{R}^n

There are several operations on vectors in \mathbb{R}^n that will be frequently used in the book. For a given vector $\mathbf{x} \in \mathbb{R}^n$, the vector $[\mathbf{x}]_+$ is the *nonnegative part* of \mathbf{x} defined by $[\mathbf{x}]_+ = (\max\{x_i, 0\})_{i=1}^n$. For a given $\mathbf{x} \in \mathbb{R}^n$, the vector $|\mathbf{x}|$ is the vector of component-wise absolute values $(|x_i|)_{i=1}^n$, and the vector $\text{sgn}(\mathbf{x})$ is defined as

$$\text{sgn}(\mathbf{x})_i = \begin{cases} 1, & x_i \geq 0, \\ -1, & x_i < 0. \end{cases}$$

For two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, their *Hadamard product*, denoted by $\mathbf{a} \odot \mathbf{b}$, is the vector comprising the component-wise products: $\mathbf{a} \odot \mathbf{b} = (a_i b_i)_{i=1}^n$.

1.8 The Space $\mathbb{R}^{m \times n}$

The set of all real-valued $m \times n$ matrices is denoted by $\mathbb{R}^{m \times n}$. This is a vector space with the component-wise addition as the summation operation and the component-wise scalar multiplication as the “scalar-vector multiplication” operation. The *dot product* in $\mathbb{R}^{m \times n}$ is defined by

$$\langle \mathbf{A}, \mathbf{B} \rangle = \text{Tr}(\mathbf{A}^T \mathbf{B}) = \sum_{i=1}^m \sum_{j=1}^n A_{ij} B_{ij}, \quad \mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}.$$

The space $\mathbb{R}^{m \times n}$ is sometimes associated with the space \mathbb{R}^{mn} in the sense that each matrix in $\mathbb{R}^{m \times n}$ corresponds to the mn -length vector constructed by stacking the columns of the matrix. Unless otherwise stated, we will assume that the inner product in $\mathbb{R}^{m \times n}$ is the dot product.

Inner Product in $\mathbb{R}^{m \times n}$: In this book, unless otherwise stated, the endowed inner product in $\mathbb{R}^{m \times n}$ is the dot product.

1.8.1 Subsets of $\mathbb{R}^{n \times n}$

The set of all $n \times n$ symmetric matrices is denoted by \mathbb{S}^n :

$$\mathbb{S}^n = \{\mathbf{A} \in \mathbb{R}^{n \times n} : \mathbf{A} = \mathbf{A}^T\}.$$

Note that \mathbb{S}^n is also a vector space with the same summation and scalar multiplication operations as in $\mathbb{R}^{n \times n}$. The inner product in \mathbb{S}^n , unless otherwise stated, is the dot product.

The set of all $n \times n$ positive semidefinite matrices is denoted by \mathbb{S}_+^n :

$$\mathbb{S}_+^n = \{\mathbf{A} \in \mathbb{R}^{n \times n} : \mathbf{A} \succeq \mathbf{0}\}.$$

The set of all $n \times n$ positive definite matrices is denoted by \mathbb{S}_{++}^n :

$$\mathbb{S}_{++}^n = \{\mathbf{A} \in \mathbb{R}^{n \times n} : \mathbf{A} \succ \mathbf{0}\}.$$

Obviously, the inclusion $\mathbb{S}_{++}^n \subseteq \mathbb{S}_+^n \subseteq \mathbb{S}^n$ holds. Similarly, \mathbb{S}_-^n is the set of all $n \times n$ negative semidefinite matrices, and \mathbb{S}_{--}^n is the set of all $n \times n$ negative definite matrices:

$$\begin{aligned} \mathbb{S}_-^n &= \{\mathbf{A} \in \mathbb{R}^{n \times n} : \mathbf{A} \preceq \mathbf{0}\}, \\ \mathbb{S}_{--}^n &= \{\mathbf{A} \in \mathbb{R}^{n \times n} : \mathbf{A} \prec \mathbf{0}\}. \end{aligned}$$

The set of all $n \times n$ orthogonal matrices is denoted by \mathbb{O}^n :

$$\mathbb{O}^n = \{\mathbf{A} \in \mathbb{R}^{n \times n} : \mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A} = \mathbf{I}\}.$$

1.8.2 Norms in $\mathbb{R}^{m \times n}$

If $\mathbb{R}^{m \times n}$ is endowed with the dot product, then the corresponding Euclidean norm is the *Frobenius norm* defined by

$$\|\mathbf{A}\|_F = \sqrt{\text{Tr}(\mathbf{A}^T \mathbf{A})} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2}, \quad \mathbf{A} \in \mathbb{R}^{m \times n}.$$

Many examples of matrix norms are generated by using the concept of induced norms, which we now describe. Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and two norms $\|\cdot\|_a$ and $\|\cdot\|_b$ on \mathbb{R}^n and \mathbb{R}^m , respectively, the *induced matrix norm* $\|\mathbf{A}\|_{a,b}$ is defined by

$$\|\mathbf{A}\|_{a,b} = \max_{\mathbf{x}} \{\|\mathbf{Ax}\|_b : \|\mathbf{x}\|_a \leq 1\}.$$

It can be easily shown that the above definition implies that for any $\mathbf{x} \in \mathbb{R}^n$, the inequality

$$\|\mathbf{Ax}\|_b \leq \|\mathbf{A}\|_{a,b} \|\mathbf{x}\|_a$$

holds. We refer to the matrix norm $\|\cdot\|_{a,b}$ as the (a,b) -norm. When $a = b$, we will simply refer to it as an a -norm and omit one of the subscripts in its notation, that is, use the notation $\|\cdot\|_a$ instead of $\|\cdot\|_{a,a}$.

Example 1.1 (spectral norm). If $\|\cdot\|_a = \|\cdot\|_b = \|\cdot\|_2$, then the induced norm of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the maximum singular value of \mathbf{A} :

$$\|\mathbf{A}\|_2 = \|\mathbf{A}\|_{2,2} = \sqrt{\lambda_{\max}(\mathbf{A}^T \mathbf{A})} \equiv \sigma_{\max}(\mathbf{A}). \quad \blacksquare$$

Example 1.2 (1-norm). When $\|\cdot\|_a = \|\cdot\|_b = \|\cdot\|_1$, the induced matrix norm of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is given by

$$\|\mathbf{A}\|_1 = \max_{j=1,2,\dots,n} \sum_{i=1}^m |A_{i,j}|.$$

This norm is also called the *maximum absolute column sum norm*. \blacksquare

Example 1.3 (∞ -norm). When $\|\cdot\|_a = \|\cdot\|_b = \|\cdot\|_\infty$, the induced matrix norm of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is given by

$$\|\mathbf{A}\|_\infty = \max_{i=1,2,\dots,m} \sum_{j=1}^n |A_{i,j}|.$$

This norm is also called the *maximum absolute row sum norm*. \blacksquare

1.9 Cartesian Product of Vector Spaces

Given m vector spaces $\mathbb{E}_1, \mathbb{E}_2, \dots, \mathbb{E}_m$ equipped with inner products $\langle \cdot, \cdot \rangle_{\mathbb{E}_i}$, their Cartesian product $\mathbb{E}_1 \times \mathbb{E}_2 \times \dots \times \mathbb{E}_m$ is the vector space of all m -tuples $(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m)$ equipped with the component-wise addition between vectors:

$$(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m) + (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m) = (\mathbf{v}_1 + \mathbf{w}_1, \mathbf{v}_2 + \mathbf{w}_2, \dots, \mathbf{v}_m + \mathbf{w}_m)$$

and the scalar-vector multiplication operation given by

$$\alpha(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m) = (\alpha\mathbf{v}_1, \alpha\mathbf{v}_2, \dots, \alpha\mathbf{v}_m).$$

The inner product in the Cartesian product space is defined as

$$\langle (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m), (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m) \rangle_{\mathbb{E}_1 \times \mathbb{E}_2 \times \dots \times \mathbb{E}_m} = \sum_{i=1}^m \langle \mathbf{v}_i, \mathbf{w}_i \rangle_{\mathbb{E}_i}. \quad (1.1)$$

The space $\mathbb{R} \times \mathbb{R}$, for example, consists of all two-dimensional row vectors, so in that respect it is different than \mathbb{R}^2 , which comprises all two-dimensional *column* vectors. However, with only a slight abuse of notation, we will occasionally refer to $\mathbb{R} \times \mathbb{R}$ as \mathbb{R}^2 .

Suppose that $\mathbb{E}_1, \mathbb{E}_2, \dots, \mathbb{E}_m$ are vector spaces with endowed norms $\|\cdot\|_{\mathbb{E}_1}, \|\cdot\|_{\mathbb{E}_2}, \dots, \|\cdot\|_{\mathbb{E}_m}$, respectively. There are many ways to define a norm on the Cartesian product space $\mathbb{E}_1 \times \mathbb{E}_2 \times \dots \times \mathbb{E}_m$. For example, for any $p \geq 1$, we can define the composite l_p -norm as

$$\|(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m)\| = \sqrt[p]{\sum_{i=1}^m \|\mathbf{u}_i\|_{\mathbb{E}_i}^p}.$$

Another norm is a composite weighted l_2 -norm:

$$\|(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m)\| = \sqrt{\sum_{i=1}^m \omega_i \|\mathbf{u}_i\|_{\mathbb{E}_i}^2},$$

where $\omega_1, \omega_2, \dots, \omega_m$ are given positive real numbers.

We will use the convention that if $\mathbb{E}_1, \mathbb{E}_2, \dots, \mathbb{E}_m$ are Euclidean spaces, then $\mathbb{E}_1 \times \mathbb{E}_2 \times \dots \times \mathbb{E}_m$ is also a Euclidean space, and consequently, by the definition (1.1) of the inner product in product spaces,

$$\|(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m)\|_{\mathbb{E}_1 \times \mathbb{E}_2 \times \dots \times \mathbb{E}_m} = \sqrt{\sum_{i=1}^m \|\mathbf{u}_i\|_{\mathbb{E}_i}^2}.$$

1.10 Linear Transformations

Given two vector spaces \mathbb{E} and \mathbb{V} , a function $\mathcal{A} : \mathbb{E} \rightarrow \mathbb{V}$ is called a *linear transformation* if the following property holds for any $\mathbf{x}, \mathbf{y} \in \mathbb{E}$ and $\alpha, \beta \in \mathbb{R}$:

$$\mathcal{A}(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha\mathcal{A}(\mathbf{x}) + \beta\mathcal{A}(\mathbf{y}).$$

All linear transformations from \mathbb{R}^n to \mathbb{R}^m have the form

$$\mathcal{A}(\mathbf{x}) = \mathbf{Ax}$$

for some matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$. All linear transformations from $\mathbb{R}^{m \times n}$ to \mathbb{R}^k have the form

$$\mathcal{A}(\mathbf{X}) = \begin{pmatrix} \text{Tr}(\mathbf{A}_1^T \mathbf{X}) \\ \text{Tr}(\mathbf{A}_2^T \mathbf{X}) \\ \vdots \\ \text{Tr}(\mathbf{A}_k^T \mathbf{X}) \end{pmatrix}$$

for some $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_k \in \mathbb{R}^{m \times n}$. The *identity transformation*, denoted by \mathcal{I} , is defined by the relation $\mathcal{I}(\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{E}$.

1.11 The Dual Space

A *linear functional* on a vector space \mathbb{E} is a linear transformation from \mathbb{E} to \mathbb{R} . Given a vector space \mathbb{E} , the set of all linear functionals on \mathbb{E} is called the *dual space* and is denoted by \mathbb{E}^* . For inner product spaces, it is known that given a linear functional $f \in \mathbb{E}^*$, there always exists $\mathbf{v} \in \mathbb{E}$ such that

$$f(\mathbf{x}) = \langle \mathbf{v}, \mathbf{x} \rangle. \quad (1.2)$$

For the sake of simplicity of notation, we will represent the linear functional f given in (1.2) by the vector \mathbf{v} . This correspondence between linear functionals and elements in \mathbb{E} leads us to consider the elements in \mathbb{E}^* as exactly the same as those in \mathbb{E} . The inner product in \mathbb{E}^* is the same as the inner product in \mathbb{E} . Essentially, the only difference between \mathbb{E} and \mathbb{E}^* will be in the choice of norms of each of the spaces. Suppose that \mathbb{E} is endowed with a norm $\|\cdot\|$. Then the norm of the dual space, called the *dual norm*, is given by

$$\|\mathbf{y}\|_* \equiv \max_{\mathbf{x}} \{\langle \mathbf{y}, \mathbf{x} \rangle : \|\mathbf{x}\| \leq 1\}, \quad \mathbf{y} \in \mathbb{E}^*. \quad (1.3)$$

It is not difficult to show that the dual norm is indeed a norm. A useful property is that the maximum in (1.3) can be taken over the unit sphere rather than over the unit ball, meaning that the following formula is valid:

$$\|\mathbf{y}\|_* = \max_{\mathbf{x}} \{\langle \mathbf{y}, \mathbf{x} \rangle : \|\mathbf{x}\| = 1\}, \quad \mathbf{y} \in \mathbb{E}^*.$$

The definition of the dual norm readily implies the following generalized version of the Cauchy–Schwarz inequality.

Lemma 1.4 (generalized Cauchy–Schwarz inequality). *Let \mathbb{E} be an inner product vector space endowed with a norm $\|\cdot\|$. Then*

$$|\langle \mathbf{y}, \mathbf{x} \rangle| \leq \|\mathbf{y}\|_* \|\mathbf{x}\| \text{ for any } \mathbf{y} \in \mathbb{E}^*, \mathbf{x} \in \mathbb{E}. \quad (1.4)$$

Proof. If $\mathbf{x} = \mathbf{0}$, the inequality is trivially satisfied. Otherwise, take $\tilde{\mathbf{x}} = \frac{\mathbf{x}}{\|\mathbf{x}\|}$. Obviously, $\|\tilde{\mathbf{x}}\| = 1$, and hence, by the definition of the dual norm, we have

$$\|\mathbf{y}\|_* \geq \langle \mathbf{y}, \tilde{\mathbf{x}} \rangle = \frac{1}{\|\mathbf{x}\|} \langle \mathbf{y}, \mathbf{x} \rangle,$$

showing that $|\langle \mathbf{y}, \mathbf{x} \rangle| \leq \|\mathbf{y}\|_* \|\mathbf{x}\|$. Plugging $-\mathbf{x}$ instead of \mathbf{x} in the latter inequality, we obtain that $|\langle \mathbf{y}, \mathbf{x} \rangle| \geq -\|\mathbf{y}\|_* \|\mathbf{x}\|$, thus showing the validity of inequality (1.4). \square

Another important result is that Euclidean norms are self-dual, meaning that $\|\cdot\| = \|\cdot\|_*$. Here of course we use our convention that the elements in the dual space \mathbb{E}^* are the same as the elements in \mathbb{E} . We can thus write, in only a slight abuse of notation,¹ that for any Euclidean space \mathbb{E} , $\mathbb{E} = \mathbb{E}^*$.

¹Disregarding the fact that the members of \mathbb{E}^* are actually linear functionals on \mathbb{E} .

Example 1.5 (l_p -norms). Consider the space \mathbb{R}^n endowed with the l_p -norm. When $p > 1$, the dual norm is the l_q -norm, where $q > 1$ is the number satisfying $\frac{1}{p} + \frac{1}{q} = 1$. When $p = 1$, the dual norm is the l_∞ -norm, and vice versa—the dual norm of the l_∞ -norm is the l_1 -norm. ■

Example 1.6 (Q-norms). Consider the space \mathbb{R}^n endowed with the \mathbf{Q} -norm, where $\mathbf{Q} \in \mathbb{S}_{++}^n$. The dual norm of $\|\cdot\|_{\mathbf{Q}}$ is $\|\cdot\|_{\mathbf{Q}^{-1}}$, meaning

$$\|\mathbf{x}\|_{\mathbf{Q}^{-1}} = \sqrt{\mathbf{x}^T \mathbf{Q}^{-1} \mathbf{x}}.$$

As an example, consider the case where \mathbf{Q} is diagonal: $\mathbf{Q} = \text{diag}(w_1, w_2, \dots, w_n)$ with $w_1, w_2, \dots, w_n > 0$. The \mathbf{Q} -norm in this case takes the form

$$\|\mathbf{x}\| = \sqrt{\sum_{i=1}^n w_i x_i^2},$$

and its dual norm is

$$\|\mathbf{x}\|_* = \sqrt{\sum_{i=1}^n \frac{1}{w_i} x_i^2}. \quad \blacksquare$$

Example 1.7 (dual norm of Cartesian products of spaces). Consider the space $\mathbb{E} = \mathbb{E}_1 \times \mathbb{E}_2 \times \cdots \times \mathbb{E}_m$, where $\mathbb{E}_1, \mathbb{E}_2, \dots, \mathbb{E}_m$ are inner product vectors spaces with norms $\|\cdot\|_{\mathbb{E}_1}, \|\cdot\|_{\mathbb{E}_2}, \dots, \|\cdot\|_{\mathbb{E}_m}$, respectively. Recall that we assume that the inner product in the product space is given by

$$\langle (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m), (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m) \rangle = \sum_{i=1}^m \langle \mathbf{v}_i, \mathbf{w}_i \rangle.$$

The dual space to $\mathbb{E}_1 \times \mathbb{E}_2 \times \cdots \times \mathbb{E}_m$ is the product space $\mathbb{E}_1^* \times \mathbb{E}_2^* \times \cdots \times \mathbb{E}_m^*$ with endowed norm defined as usual in dual spaces. For example, suppose that the norm on the product space is the composite weighted l_2 -norm:

$$\|(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m)\| = \sqrt{\sum_{i=1}^m \omega_i \|\mathbf{u}_i\|_{\mathbb{E}_i}^2}, \quad \mathbf{u}_i \in \mathbb{E}_i, i = 1, 2, \dots, p,$$

where $\omega_1, \omega_2, \dots, \omega_m > 0$ are given positive weights. Then it is simple to show that the dual norm in this case is given by

$$\|(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m)\|_* = \sqrt{\sum_{i=1}^m \frac{1}{\omega_i} \|\mathbf{v}_i\|_{\mathbb{E}_i^*}^2}, \quad \mathbf{v}_i \in \mathbb{E}_i^*, i = 1, 2, \dots, p.$$

where $\|\cdot\|_{\mathbb{E}_i^*}$ is the dual norm to $\|\cdot\|_{\mathbb{E}_i}$, namely, the norm of the dual space \mathbb{E}_i^* . ■

1.12 The Bidual Space

Given a vector space \mathbb{E} , the dual space \mathbb{E}^* is also a vector space, and we can also consider its dual space, namely, \mathbb{E}^{**} . This is the so-called *bidual* space. In the

setting of finite dimensional spaces, the bidual space is the same as the original space (under our convention that the elements in the dual space are the same as the elements in the original space), and the corresponding norm (bidual norm) is the same as the original norm.

1.13 Adjoint Transformations

Given two inner product vector spaces \mathbb{E}, \mathbb{V} and a linear transformation \mathcal{A} from \mathbb{V} to \mathbb{E} , the *adjoint transformation*, denoted by \mathcal{A}^T , is a transformation from \mathbb{E}^* to \mathbb{V}^* defined by the relation

$$\langle \mathbf{y}, \mathcal{A}(\mathbf{x}) \rangle = \langle \mathcal{A}^T(\mathbf{y}), \mathbf{x} \rangle$$

for any $\mathbf{x} \in \mathbb{V}, \mathbf{y} \in \mathbb{E}^*$. When $\mathbb{V} = \mathbb{R}^n, \mathbb{E} = \mathbb{R}^m$ (endowed with the dot product), and $\mathcal{A}(\mathbf{x}) = \mathbf{Ax}$ for some matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, then the adjoint transformation is given by $\mathcal{A}^T(\mathbf{x}) = \mathbf{A}^T \mathbf{x}$.

Example 1.8 (adjoint of a transformation from $\mathbb{R}^{m \times n}$ to \mathbb{R}^k). Consider now a linear transformation from the space $\mathbb{R}^{m \times n}$ to \mathbb{R}^k . As was already mentioned in Section 1.10, such a transformation has the form

$$\mathcal{A}(\mathbf{X}) = \begin{pmatrix} \text{Tr}(\mathbf{A}_1^T \mathbf{X}) \\ \text{Tr}(\mathbf{A}_2^T \mathbf{X}) \\ \vdots \\ \text{Tr}(\mathbf{A}_k^T \mathbf{X}) \end{pmatrix},$$

where $\mathbf{A}_i \in \mathbb{R}^{m \times n}$ are given matrices. The adjoint transformation \mathcal{A}^T will be a transformation from \mathbb{R}^k to $\mathbb{R}^{m \times n}$. To find it, let us write the defining relation of the adjoint operator:

$$\langle \mathbf{y}, \mathcal{A}(\mathbf{X}) \rangle = \langle \mathcal{A}^T(\mathbf{y}), \mathbf{X} \rangle \text{ for all } \mathbf{X} \in \mathbb{R}^{m \times n}, \mathbf{y} \in \mathbb{R}^k,$$

which is the same as (recall that unless otherwise stated, the inner products in $\mathbb{R}^{m \times n}$ and \mathbb{R}^k are the dot products)

$$\sum_{i=1}^k y_i \text{Tr}(\mathbf{A}_i^T \mathbf{X}) = \langle \mathcal{A}^T(\mathbf{y}), \mathbf{X} \rangle \text{ for all } \mathbf{X} \in \mathbb{R}^{m \times n}, \mathbf{y} \in \mathbb{R}^k,$$

that is,

$$\text{Tr} \left(\left[\sum_{i=1}^k y_i \mathbf{A}_i \right]^T \mathbf{X} \right) = \langle \mathcal{A}^T(\mathbf{y}), \mathbf{X} \rangle \text{ for all } \mathbf{X} \in \mathbb{R}^{m \times n}, \mathbf{y} \in \mathbb{R}^k.$$

Obviously, the above relation implies that the adjoint transformation is given by

$$\mathcal{A}^T(\mathbf{y}) = \sum_{i=1}^k y_i \mathbf{A}_i. \quad \blacksquare$$

The adjoint of the adjoint transformation is the original transformation: $(\mathcal{A}^T)^T = \mathcal{A}$. It also holds that whenever \mathcal{A} is an invertible transformation,

$$(\mathcal{A}^T)^{-1} = (\mathcal{A}^{-1})^T.$$

1.14 Norms of Linear Transformations

Let $\mathcal{A} : \mathbb{E} \rightarrow \mathbb{V}$ be a linear transformation from a vector space \mathbb{E} to a vector space \mathbb{V} . Assume that \mathbb{E} and \mathbb{V} are endowed with the norms $\|\cdot\|_{\mathbb{E}}$ and $\|\cdot\|_{\mathbb{V}}$, respectively. The norm of the linear transformation is defined by

$$\|\mathcal{A}\| \equiv \max\{\|\mathcal{A}(\mathbf{x})\|_{\mathbb{V}} : \|\mathbf{x}\|_{\mathbb{E}} \leq 1\}.$$

It is not difficult to show that $\|\mathcal{A}\| = \|\mathcal{A}^T\|$. There is a close connection between the notion of induced norms discussed in Section 1.8.2 and norms of linear transformations. Specifically, suppose that \mathcal{A} is a linear transformation from \mathbb{R}^n to \mathbb{R}^m given by

$$\mathcal{A}(\mathbf{x}) = \mathbf{Ax}, \tag{1.5}$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$, and assume that \mathbb{R}^n and \mathbb{R}^m are endowed with the norms $\|\cdot\|_a$ and $\|\cdot\|_b$, respectively. Then $\|\mathcal{A}\| = \|\mathbf{A}\|_{a,b}$, meaning that the induced norm of a matrix is actually the norm of the corresponding linear transformation given by the relation (1.5).

Chapter 2

Extended Real-Valued Functions

Underlying Space: Recall that in this book, the underlying spaces (denoted usually by \mathbb{E} or \mathbb{V}) are finite-dimensional inner product vector spaces with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$.

2.1 Extended Real-Valued Functions and Closedness

An *extended real-valued function* is a function defined over the entire underlying space that can take any real value, as well as the infinite values $-\infty$ and ∞ . Since infinite values are allowed, we also define the appropriate arithmetic operations with $-\infty$ and ∞ as follows:

$$\begin{aligned} a + \infty &= \infty + a = \infty \quad (-\infty < a < \infty), \\ a - \infty &= -\infty + a = -\infty \quad (-\infty < a < \infty), \\ a \cdot \infty &= \infty \cdot a = \infty \quad (0 < a < \infty), \\ a \cdot (-\infty) &= (-\infty) \cdot a = -\infty \quad (0 < a < \infty), \\ a \cdot \infty &= \infty \cdot a = -\infty \quad (-\infty < a < 0), \\ a \cdot (-\infty) &= (-\infty) \cdot a = \infty \quad (-\infty < a < 0), \\ 0 \cdot \infty &= \infty \cdot 0 = 0 \cdot (-\infty) = (-\infty) \cdot 0 = 0. \end{aligned}$$

In a sense, the only “unnatural” rule is the last one, since the expression “ $0 \cdot \infty$ ” is considered to be undefined in some branches of mathematics, but in the context of extended real-valued functions, defining it as zero is the “correct” choice in the sense of consistency. We will also use the following natural order between finite and infinite numbers:

$$\begin{aligned} \infty &> a \quad (-\infty \leq a < \infty), \\ -\infty &< a \quad (-\infty < a \leq \infty). \end{aligned}$$

For an extended real-valued function $f : \mathbb{E} \rightarrow [-\infty, \infty]$, the *effective domain* or just the *domain* is the set

$$\text{dom}(f) = \{\mathbf{x} \in \mathbb{E} : f(\mathbf{x}) < \infty\}.$$

The notation “ $f : \mathbb{E} \rightarrow [-\infty, \infty]$ ” means that f is (potentially) extended real-valued (even if not explicitly stated). The notation “ $f : \mathbb{E} \rightarrow (-\infty, \infty]$ ” means that f is extended real-valued and does not attain the value $-\infty$.

The simplest examples of extended real-valued functions are indicators.

Example 2.1 (indicator functions). For any subset $C \subseteq \mathbb{E}$, the *indicator function* of C is defined to be the extended real-valued function given by

$$\delta_C(\mathbf{x}) = \begin{cases} 0, & \mathbf{x} \in C, \\ \infty, & \mathbf{x} \notin C. \end{cases} \quad \blacksquare$$

We obviously have

$$\text{dom}(\delta_C) = C.$$

The *epigraph* of an extended real-valued function $f : \mathbb{E} \rightarrow [-\infty, \infty]$ is defined by

$$\text{epi}(f) = \{(\mathbf{x}, y) : f(\mathbf{x}) \leq y, \mathbf{x} \in \mathbb{E}, y \in \mathbb{R}\}.$$

The epigraph is a subset of $\mathbb{E} \times \mathbb{R}$. Note that if $(\mathbf{x}, y) \in \text{epi}(f)$, then obviously $\mathbf{x} \in \text{dom}(f)$. A function $f : \mathbb{E} \rightarrow [-\infty, \infty]$ is called *proper* if it does not attain the value $-\infty$ and there exists at least one $\mathbf{x} \in \mathbb{E}$ such that $f(\mathbf{x}) < \infty$, meaning that $\text{dom}(f)$ is nonempty. The notion of closedness will play an important role in much of the analysis in this book.

Definition 2.2 (closed functions). A function $f : \mathbb{E} \rightarrow [-\infty, \infty]$ is **closed** if its epigraph is closed.

The indicator function δ_C is closed if and only if its underlying set C is closed.

Proposition 2.3 (closedness of indicators of closed sets). The indicator function δ_C is closed if and only if C is a closed set.

Proof. The epigraph of δ_C is given by

$$\text{epi}(\delta_C) = \{(\mathbf{x}, y) \in \mathbb{E} \times \mathbb{R} : \delta_C(\mathbf{x}) \leq y\} = C \times \mathbb{R}_+,$$

which is evidently closed if and only if C is closed. \square

We thus obtained in particular that the domain of a closed indicator function is necessarily a closed set. However, in general, we note that the domain of a closed function might not be closed. A classical example for this observation is given below.

Example 2.4. Consider the function $f : \mathbb{R} \rightarrow [-\infty, \infty]$ given by

$$f(x) = \begin{cases} \frac{1}{x}, & x > 0, \\ \infty, & \text{else.} \end{cases}$$

The domain of the function, which is the open interval $(0, \infty)$, is obviously not closed, but the function is closed since its epigraph

$$\text{epi}(f) = \{(x, y) : xy \geq 1, x > 0\}$$

is a closed set; see Figure 2.1. ■

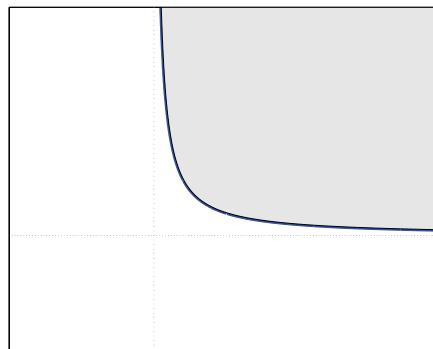


Figure 2.1. The epigraph of the function $f(x) = \frac{1}{x}$ for $x > 0$ and ∞ otherwise.

A property that will be later shown to be equivalent to closedness is *lower semicontinuity*.

Definition 2.5 (lower semicontinuity). A function $f : \mathbb{E} \rightarrow [-\infty, \infty]$ is called **lower semicontinuous at $\mathbf{x} \in \mathbb{E}$** if

$$f(\mathbf{x}) \leq \liminf_{n \rightarrow \infty} f(\mathbf{x}_n)$$

for any sequence $\{\mathbf{x}_n\}_{n \geq 1} \subseteq \mathbb{E}$ for which $\mathbf{x}_n \rightarrow \mathbf{x}$ as $n \rightarrow \infty$. A function $f : \mathbb{E} \rightarrow [-\infty, \infty]$ is called **lower semicontinuous** if it is lower semicontinuous at each point in \mathbb{E} .

For any $\alpha \in \mathbb{R}$, the α -level set of a function $f : \mathbb{E} \rightarrow [-\infty, \infty]$ is the set

$$\text{Lev}(f, \alpha) = \{\mathbf{x} \in \mathbb{E} : f(\mathbf{x}) \leq \alpha\}.$$

The following theorem shows that closedness and lower semicontinuity are equivalent properties, and they are both equivalent to the property that all the level sets of the function are closed.

Theorem 2.6 (equivalence of closedness, lower semicontinuity, and closedness of level sets). *Let $f : \mathbb{E} \rightarrow [-\infty, \infty]$. Then the following three claims are equivalent:*

- (i) *f is lower semicontinuous.*
- (ii) *f is closed.*
- (iii) *For any $\alpha \in \mathbb{R}$, the level set*

$$\text{Lev}(f, \alpha) = \{\mathbf{x} \in \mathbb{E} : f(\mathbf{x}) \leq \alpha\}$$

is closed.

Proof. (i \Rightarrow ii) Suppose that f is lower semicontinuous. We will show that $\text{epi}(f)$ is closed. For that, take $\{(\mathbf{x}_n, y_n)\}_{n \geq 1} \subseteq \text{epi}(f)$ such that $(\mathbf{x}_n, y_n) \rightarrow (\mathbf{x}^*, y^*)$ as $n \rightarrow \infty$. Then for any $n \geq 1$,

$$f(\mathbf{x}_n) \leq y_n.$$

Therefore, by the lower semicontinuity of f at \mathbf{x}^* , we have

$$f(\mathbf{x}^*) \leq \liminf_{n \rightarrow \infty} f(\mathbf{x}_n) \leq \liminf_{n \rightarrow \infty} y_n = y^*,$$

showing that $(\mathbf{x}^*, y^*) \in \text{epi}(f)$ and hence that f is closed.

(ii \Rightarrow iii) Suppose that f is closed, namely, that $\text{epi}(f)$ is closed. Let $\alpha \in \mathbb{R}$. We will show that $\text{Lev}(f, \alpha)$ is closed. If $\text{Lev}(f, \alpha) = \emptyset$, we are done. Otherwise, take a sequence $\{\mathbf{x}_n\}_{n \geq 1} \subseteq \text{Lev}(f, \alpha)$ that converges to $\bar{\mathbf{x}}$. Obviously $(\mathbf{x}_n, \alpha) \in \text{epi}(f)$ for any n and $(\mathbf{x}_n, \alpha) \rightarrow (\bar{\mathbf{x}}, \alpha)$ as $n \rightarrow \infty$. By the closedness of $\text{epi}(f)$, it follows that $(\bar{\mathbf{x}}, \alpha) \in \text{epi}(f)$, establishing the fact that $\bar{\mathbf{x}} \in \text{Lev}(f, \alpha)$.

(iii \Rightarrow i) Suppose that all the level sets of f are closed. We will show that it is lower semicontinuous. Assume by contradiction that f is not lower semicontinuous, meaning that there exists $\mathbf{x}^* \in \mathbb{E}$ and $\{\mathbf{x}_n\}_{n \geq 1} \subseteq \mathbb{E}$ such that $\mathbf{x}_n \rightarrow \mathbf{x}^*$ and $\liminf_{n \rightarrow \infty} f(\mathbf{x}_n) < f(\mathbf{x}^*)$. Take α that satisfies

$$\liminf_{n \rightarrow \infty} f(\mathbf{x}_n) < \alpha < f(\mathbf{x}^*). \quad (2.1)$$

Then there exists a subsequence $\{\mathbf{x}_{n_k}\}_{k \geq 1}$ such that $f(\mathbf{x}_{n_k}) \leq \alpha$ for all $k \geq 1$. By the closedness of the level set $\text{Lev}(f, \alpha)$ and the fact that $\mathbf{x}_{n_k} \rightarrow \mathbf{x}^*$ as $k \rightarrow \infty$, it follows that $f(\mathbf{x}^*) \leq \alpha$, which is a contradiction to (2.1), showing that (iii) implies (i). \square

The next result shows that closedness of functions is preserved under affine change of variables, summation, multiplication by a nonnegative number, and maximization. Before stating the theorem, we note that in this book we will not use the inf/sup notation but rather use only the min/max notation, where the usage of this notation does not imply that the maximum or minimum is actually attained.

Theorem 2.7 (operations preserving closedness).

- (a) Let $\mathcal{A} : \mathbb{E} \rightarrow \mathbb{V}$ be a linear transformation from \mathbb{E} to \mathbb{V} and $\mathbf{b} \in \mathbb{V}$ and let $f : \mathbb{V} \rightarrow [-\infty, \infty]$ be an extended real-valued closed function. Then the function $g : \mathbb{E} \rightarrow [-\infty, \infty]$ given by

$$g(\mathbf{x}) = f(\mathcal{A}(\mathbf{x}) + \mathbf{b})$$

is closed.

- (b) Let $f_1, f_2, \dots, f_m : \mathbb{E} \rightarrow (-\infty, \infty]$ be extended real-valued closed functions and let $\alpha_1, \alpha_2, \dots, \alpha_m \in \mathbb{R}_+$. Then the function $f = \sum_{i=1}^m \alpha_i f_i$ is closed.

- (c) Let $f_i : \mathbb{E} \rightarrow (-\infty, \infty]$, $i \in I$ be extended real-valued closed functions, where I is a given index set. Then the function

$$f(\mathbf{x}) = \max_{i \in I} f_i(\mathbf{x})$$

is closed.

Proof. (a) To show that g is closed, take a sequence $\{(\mathbf{x}_n, y_n)\}_{n \geq 1} \subseteq \text{epi}(g)$ such that $(\mathbf{x}_n, y_n) \rightarrow (\mathbf{x}^*, y^*)$ as $n \rightarrow \infty$, where $\mathbf{x}^* \in \mathbb{E}$ and $y^* \in \mathbb{R}$. The relation $\{(\mathbf{x}_n, y_n)\}_{n \geq 1} \subseteq \text{epi}(g)$ can be written equivalently as

$$f(\mathcal{A}(\mathbf{x}_n) + \mathbf{b}) \leq y_n \text{ for all } n \geq 1.$$

Therefore, $(\mathcal{A}(\mathbf{x}_n) + \mathbf{b}, y_n) \in \text{epi}(f)$. Hence, since f is closed and $\mathcal{A}(\mathbf{x}_n) + \mathbf{b} \rightarrow \mathcal{A}(\mathbf{x}^*) + \mathbf{b}$, $y_n \rightarrow y^*$ as $n \rightarrow \infty$ (by the continuity of linear transformations), it follows that $(\mathcal{A}(\mathbf{x}^*) + \mathbf{b}, y^*) \in \text{epi}(f)$, meaning that

$$f(\mathcal{A}(\mathbf{x}^*) + \mathbf{b}) \leq y^*,$$

which is the same as the relation $(\mathbf{x}^*, y^*) \in \text{epi}(g)$. We have shown that $\text{epi}(g)$ is closed or, equivalently, that g is closed.

(b) We will prove that f is lower semicontinuous, which by Theorem 2.6 is equivalent to the closedness of f . Let $\{\mathbf{x}_n\}_{n \geq 1}$ be a sequence converging to \mathbf{x}^* . Then by the lower semicontinuity of f_i , for any $i = 1, 2, \dots, m$,

$$f_i(\mathbf{x}^*) \leq \liminf_{n \rightarrow \infty} f_i(\mathbf{x}_n).$$

Multiplying the above inequality by α_i and summing for $i = 1, 2, \dots, m$ gives

$$\left(\sum_{i=1}^m \alpha_i f_i \right) (\mathbf{x}^*) \leq \sum_{i=1}^m \liminf_{n \rightarrow \infty} \alpha_i f_i(\mathbf{x}_n) \leq \liminf_{n \rightarrow \infty} \left(\sum_{i=1}^m \alpha_i f_i \right) (\mathbf{x}_n),$$

where in the last inequality we used the fact that for any two sequences of real numbers $\{a_n\}_{n \geq 1}$ and $\{b_n\}_{n \geq 1}$, it holds that

$$\liminf_{n \rightarrow \infty} a_n + \liminf_{n \rightarrow \infty} b_n \leq \liminf_{n \rightarrow \infty} (a_n + b_n).$$

A simple induction argument shows that this property holds for an arbitrary number of sequences. We have thus established the lower semicontinuity and hence closedness of $\sum_{i=1}^m \alpha_i f_i$.

(c) Since f_i is closed for any $i \in I$, it follows that $\text{epi}(f_i)$ is closed for any i , and hence $\text{epi}(f) = \bigcap_{i \in I} \text{epi}(f_i)$ is closed as an intersection of closed sets, implying that f is closed. \square

2.2 Closedness versus Continuity

A relation between continuity and closedness is described in the following theorem stating that if an extended real-valued function is continuous over its domain,² which is assumed to be closed, then it is closed.

Theorem 2.8. *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be an extended real-valued function that is continuous over its domain and suppose that $\text{dom}(f)$ is closed. Then f is closed.*

Proof. To show that $\text{epi}(f)$ is closed (which is the same as saying that f is closed), take a sequence $\{(\mathbf{x}_n, y_n)\}_{n \geq 1} \subseteq \text{epi}(f)$ for which $(\mathbf{x}_n, y_n) \rightarrow (\mathbf{x}^*, y^*)$ as $n \rightarrow \infty$ for some $\mathbf{x}^* \in \mathbb{E}$ and $y^* \in \mathbb{R}$. Since $\{\mathbf{x}_n\}_{n \geq 1} \subseteq \text{dom}(f)$, $\mathbf{x}_n \rightarrow \mathbf{x}^*$ and $\text{dom}(f)$ is closed, it follows that $\mathbf{x}^* \in \text{dom}(f)$. By the definition of the epigraph, we have for all $n \geq 1$,

$$f(\mathbf{x}_n) \leq y_n. \quad (2.2)$$

Since f is continuous over $\text{dom}(f)$, and in particular at \mathbf{x}^* , it follows by taking n to ∞ in (2.2) that

$$f(\mathbf{x}^*) \leq y^*,$$

showing that $(\mathbf{x}^*, y^*) \in \text{epi}(f)$, thus establishing the closedness of $\text{epi}(f)$. \square

In particular, any real-valued continuous function over \mathbb{E} is closed.

Corollary 2.9. *Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be continuous. Then f is closed.*

The above results demonstrate that there is a connection between continuity and closedness. However, these two notions are different, as the following example illustrates.

Example 2.10. Consider the function $f_\alpha : \mathbb{R} \rightarrow (-\infty, \infty]$ given by

$$f_\alpha(x) = \begin{cases} \alpha, & x = 0, \\ x, & 0 < x \leq 1, \\ \infty, & \text{else.} \end{cases}$$

²A function $g : \mathbb{E} \rightarrow (-\infty, \infty]$ is *continuous over its domain* if for any sequence $\{\mathbf{x}_n\}_{n \geq 1} \subseteq \text{dom}(g)$ satisfying $\mathbf{x}_n \rightarrow \mathbf{x}^*$ as $n \rightarrow \infty$ for some $\mathbf{x}^* \in \text{dom}(g)$, it holds that $g(\mathbf{x}_n) \rightarrow g(\mathbf{x}^*)$ as $n \rightarrow \infty$.

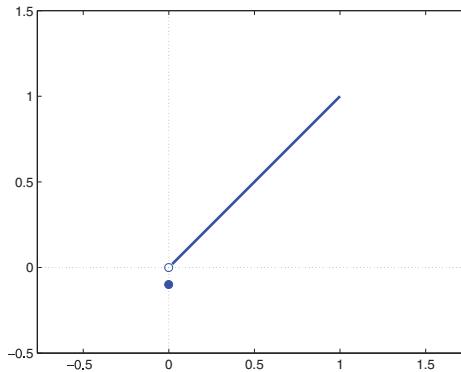


Figure 2.2. An example of a closed function, which is not continuous over its domain.

This function is closed if and only if $\alpha \leq 0$, and it is continuous over its domain if and only if $\alpha = 0$. Thus, the function $f_{-0,1}$, plotted in Figure 2.2, is closed but not continuous over its domain. ■

Example 2.11 (l_0 -norm). Consider the l_0 -norm function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ given by

$$f(\mathbf{x}) = \|\mathbf{x}\|_0 \equiv \#\{i : x_i \neq 0\}.$$

That is, $\|\mathbf{x}\|_0$ is the number of nonzero elements in \mathbf{x} . Note the l_0 -norm is actually not a norm. It does not satisfy the homogeneity property. Nevertheless, this terminology is widely used in the literature, and we will therefore adopt it. Although f is obviously not continuous, it is closed. To show this, note that

$$f(\mathbf{x}) = \sum_{i=1}^n I(x_i),$$

where $I : \mathbb{R} \rightarrow \{0, 1\}$ is given by

$$I(y) = \begin{cases} 0, & y = 0, \\ 1, & y \neq 0. \end{cases}$$

The function I is closed since its level sets, which are given by

$$\text{Lev}(I, \alpha) = \begin{cases} \emptyset, & \alpha < 0, \\ \{0\}, & \alpha \in [0, 1), \\ \mathbb{R}, & \alpha \geq 1, \end{cases}$$

are closed sets. Therefore, f , as a sum of closed functions, is closed (Theorem 2.7(b)). ■

It is well known that a continuous function over a nonempty compact³ set attains a minimum. This is the well-known Weierstrass theorem. We will now show that this property also holds for closed functions.

Theorem 2.12 (Weierstrass theorem for closed functions). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper closed function and assume that C is a compact set satisfying $C \cap \text{dom}(f) \neq \emptyset$. Then*

- (a) *f is bounded below over C .*
- (b) *f attains its minimal value over C .*

Proof. (a) Suppose by contradiction that f is not bounded below over C . Then there exists a sequence $\{\mathbf{x}_n\}_{n \geq 1} \subseteq C$ such that

$$\lim_{n \rightarrow \infty} f(\mathbf{x}_n) = -\infty. \quad (2.3)$$

By the Bolzano–Weierstrass theorem, since C is compact, there exists a subsequence $\{\mathbf{x}_{n_k}\}_{k \geq 1}$ that converges to a point $\bar{\mathbf{x}} \in C$. By Theorem 2.6, f is lower semicontinuous, and hence

$$f(\bar{\mathbf{x}}) \leq \liminf_{k \rightarrow \infty} f(\mathbf{x}_{n_k}),$$

which is a contradiction to (2.3).

(b) Denote by f_{opt} the minimal value of f over C . Then there exists a sequence $\{\mathbf{x}_n\}_{n \geq 1}$ for which $f(\mathbf{x}_n) \rightarrow f_{\text{opt}}$ as $n \rightarrow \infty$. As before, take a subsequence $\{\mathbf{x}_{n_k}\}_{k \geq 1}$ that converges to some point $\bar{\mathbf{x}} \in C$. By the lower semicontinuity of f , it follows that

$$f(\bar{\mathbf{x}}) \leq \lim_{k \rightarrow \infty} f(\mathbf{x}_{n_k}) = f_{\text{opt}},$$

showing that $\bar{\mathbf{x}}$ is a minimizer of f over C . \square

When the set C in the premise of Theorem 2.12 is not compact, the Weierstrass theorem does not guarantee the attainment of a minimizer, but attainment of a minimizer can be shown when the compactness of C is replaced by closedness if the function has a property called *coerciveness*.

Definition 2.13 (coerciveness). *A proper function $f : \mathbb{E} \rightarrow (-\infty, \infty]$ is called coercive if*

$$\lim_{\|\mathbf{x}\| \rightarrow \infty} f(\mathbf{x}) = \infty.$$

An important property of closed coercive functions is that they possess a minimizer on any closed set that has a nonempty intersection with the domain of the function.

Theorem 2.14 (attainment under coerciveness). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper closed and coercive function and let $S \subseteq \mathbb{E}$ be a nonempty closed set satisfying $S \cap \text{dom}(f) \neq \emptyset$. Then f attains its minimal value over S .*

³A set is called *compact* if it is closed and bounded.

Proof. Let \mathbf{x}_0 be an arbitrary point in $S \cap \text{dom}(f)$. By the coerciveness of f , there exists an $M > 0$ such that

$$f(\mathbf{x}) > f(\mathbf{x}_0) \text{ for any } \mathbf{x} \text{ satisfying } \|\mathbf{x}\| > M. \quad (2.4)$$

Since any minimizer \mathbf{x}^* of f over S satisfies $f(\mathbf{x}^*) \leq f(\mathbf{x}_0)$, it follows from (2.4) that the set of minimizers of f over S is the same as the set of minimizers of f over $S \cap B_{\|\cdot\|}[\mathbf{0}, M]$, which is compact (both sets are closed, and $B_{\|\cdot\|}[\mathbf{0}, M]$ is bounded) and nonempty (as it contains \mathbf{x}_0). Therefore, by the Weierstrass theorem for closed functions (Theorem 2.12), there exists a minimizer of f over $S \cap B[\mathbf{0}, M]$ and hence also over S . \square

2.3 Convex Functions

2.3.1 Definition and Basic Properties

Like closedness, the definition of convexity for extended real-valued functions can be written in terms of the epigraph.

Definition 2.15 (convex functions). An extended real-valued function $f : \mathbb{E} \rightarrow [-\infty, \infty]$ is called **convex** if $\text{epi}(f)$ is a convex set.

It is not difficult to show that a proper extended real-valued function $f : \mathbb{E} \rightarrow (-\infty, \infty]$ is convex if and only if $\text{dom}(f)$ is convex and the restriction of f to $\text{dom}(f)$ is convex over $\text{dom}(f)$ in the sense of convexity of real-valued functions over convex domains. Using this observation, we conclude that a proper extended real-valued function f is convex if and only if

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}) \text{ for all } \mathbf{x}, \mathbf{y} \in \mathbb{E}, \lambda \in [0, 1], \quad (2.5)$$

or, equivalently, if and only if $\text{dom}(f)$ is convex and (2.5) is satisfied for any $\mathbf{x}, \mathbf{y} \in \text{dom}(f)$ and $\lambda \in [0, 1]$. Inequality (2.5) is a special case of *Jensen's inequality*, stating that for any $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \in \mathbb{E}$ and $\boldsymbol{\lambda} \in \Delta_k$, the following inequality holds:

$$f\left(\sum_{i=1}^k \lambda_i \mathbf{x}_i\right) \leq \sum_{i=1}^k \lambda_i f(\mathbf{x}_i).$$

There are several operations that preserve convexity of extended real-valued convex functions. Some of them are summarized in Theorem 2.16 below. The proof can be easily deduced by combining two facts: (i) the same properties are known to hold for real-valued convex functions defined on a given convex domain, and (ii) the observation that a proper extended real-valued function is convex if and only if its domain is convex and its restriction to its domain is a real-valued convex function.

Theorem 2.16 (operations preserving convexity).

- (a) Let $\mathcal{A} : \mathbb{E} \rightarrow \mathbb{V}$ be a linear transformation from \mathbb{E} to \mathbb{V} (two underlying vector spaces) and $\mathbf{b} \in \mathbb{V}$, and let $f : \mathbb{V} \rightarrow (-\infty, \infty]$ be an extended real-valued

convex function. Then the extended real-valued function $g : \mathbb{E} \rightarrow (-\infty, \infty]$ given by

$$g(\mathbf{x}) = f(\mathcal{A}(\mathbf{x}) + \mathbf{b})$$

is convex.

- (b) Let $f_1, f_2, \dots, f_m : \mathbb{E} \rightarrow (-\infty, \infty]$ be extended real-valued convex functions, and let $\alpha_1, \alpha_2, \dots, \alpha_m \in \mathbb{R}_+$. Then the function $\sum_{i=1}^m \alpha_i f_i$ is convex.
- (c) Let $f_i : \mathbb{E} \rightarrow (-\infty, \infty]$, $i \in I$, be extended real-valued convex functions, where I is a given index set. Then the function

$$f(\mathbf{x}) = \max_{i \in I} f_i(\mathbf{x})$$

is convex.

Given a nonempty set $C \subseteq \mathbb{E}$, the *distance function to C* is defined by

$$d_C(\mathbf{x}) = \min_{\mathbf{y} \in C} \|\mathbf{x} - \mathbf{y}\|.$$

The next example shows that for Euclidean spaces, the function $\frac{1}{2} (\|\mathbf{x}\|^2 - d_C^2(\mathbf{x}))$ is always convex, regardless of whether C is convex or not.

Example 2.17.⁴ Suppose that the underlying space \mathbb{E} is Euclidean (meaning that $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$). Let $C \subseteq \mathbb{E}$ be a nonempty set, and consider the function

$$\varphi_C(\mathbf{x}) = \frac{1}{2} (\|\mathbf{x}\|^2 - d_C^2(\mathbf{x})).$$

To show that φ_C is convex, note that

$$d_C^2(\mathbf{x}) = \min_{\mathbf{y} \in C} \|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 - \max_{\mathbf{y} \in C} [2\langle \mathbf{y}, \mathbf{x} \rangle - \|\mathbf{y}\|^2].$$

Hence,

$$\varphi_C(\mathbf{x}) = \max_{\mathbf{y} \in C} \left[\langle \mathbf{y}, \mathbf{x} \rangle - \frac{1}{2} \|\mathbf{y}\|^2 \right]. \quad (2.6)$$

Therefore, since φ_C is a maximization of affine—and hence convex—functions, by Theorem 2.16(c), it is necessarily convex. ■

Another operation that preserves convexity is partial minimization of jointly convex functions.

Theorem 2.18 (convexity under partial minimization). *Let $f : \mathbb{E} \times \mathbb{V} \rightarrow (-\infty, \infty]$ be a convex function satisfying the following property:*

$$\text{for any } \mathbf{x} \in \mathbb{E} \text{ there exists } \mathbf{y} \in \mathbb{V} \text{ for which } f(\mathbf{x}, \mathbf{y}) < \infty. \quad (2.7)$$

⁴Example 2.17 is from Hiriart-Urruty and Lemaréchal [67, Example 2.1.4].

Let⁵ $g : \mathbb{E} \rightarrow [-\infty, \infty)$ be defined by

$$g(\mathbf{x}) \equiv \min_{\mathbf{y} \in \mathbb{E}} f(\mathbf{x}, \mathbf{y}).$$

Then g is convex.

Proof. Let $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{E}$ and $\lambda \in [0, 1]$. To show the convexity of g , we will prove that

$$g(\lambda\mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2) \leq \lambda g(\mathbf{x}_1) + (1 - \lambda)g(\mathbf{x}_2). \quad (2.8)$$

The inequality is obvious if $\lambda = 0$ or 1 . We will therefore assume that $\lambda \in (0, 1)$. The proof is split into two cases.

Case I: Here we assume that $g(\mathbf{x}_1), g(\mathbf{x}_2) > -\infty$. Take $\varepsilon > 0$. Then there exist $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{V}$ such that

$$f(\mathbf{x}_1, \mathbf{y}_1) \leq g(\mathbf{x}_1) + \varepsilon, \quad (2.9)$$

$$f(\mathbf{x}_2, \mathbf{y}_2) \leq g(\mathbf{x}_2) + \varepsilon. \quad (2.10)$$

By the convexity of f , we have

$$\begin{aligned} f(\lambda\mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2, \lambda\mathbf{y}_1 + (1 - \lambda)\mathbf{y}_2) &\leq \lambda f(\mathbf{x}_1, \mathbf{y}_1) + (1 - \lambda)f(\mathbf{x}_2, \mathbf{y}_2) \\ &\stackrel{(2.9),(2.10)}{\leq} \lambda(g(\mathbf{x}_1) + \varepsilon) + (1 - \lambda)(g(\mathbf{x}_2) + \varepsilon) \\ &= \lambda g(\mathbf{x}_1) + (1 - \lambda)g(\mathbf{x}_2) + \varepsilon. \end{aligned}$$

Therefore, by the definition of g , we can conclude that

$$g(\lambda\mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2) \leq \lambda g(\mathbf{x}_1) + (1 - \lambda)g(\mathbf{x}_2) + \varepsilon.$$

Since the above inequality holds for any $\varepsilon > 0$, it follows that (2.8) holds.

Case II: Assume that at least one of the values $g(\mathbf{x}_1), g(\mathbf{x}_2)$ is equal $-\infty$. We will assume without loss of generality that $g(\mathbf{x}_1) = -\infty$. In this case, (2.8) is equivalent to saying that $g(\lambda\mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2) = -\infty$. Take any $M \in \mathbb{R}$. Then since $g(\mathbf{x}_1) = -\infty$, it follows that there exists $\mathbf{y}_1 \in \mathbb{V}$ for which

$$f(\mathbf{x}_1, \mathbf{y}_1) \leq M.$$

By property (2.7), there exists $\mathbf{y}_2 \in \mathbb{V}$ for which $f(\mathbf{x}_2, \mathbf{y}_2) < \infty$. Using the convexity of f , we obtain that

$$\begin{aligned} f(\lambda\mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2, \lambda\mathbf{y}_1 + (1 - \lambda)\mathbf{y}_2) &\leq \lambda f(\mathbf{x}_1, \mathbf{y}_1) + (1 - \lambda)f(\mathbf{x}_2, \mathbf{y}_2) \\ &\leq \lambda M + (1 - \lambda)f(\mathbf{x}_2, \mathbf{y}_2), \end{aligned}$$

which by the definition of g implies the inequality

$$g(\lambda\mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2) \leq \lambda M + (1 - \lambda)f(\mathbf{x}_2, \mathbf{y}_2).$$

Since the latter inequality holds for any $M \in \mathbb{R}$ and since $f(\mathbf{x}_2, \mathbf{y}_2) < \infty$, it follows that $g(\lambda\mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2) = -\infty$, proving the result for the second case. \square

⁵The fact that g does not attain the value ∞ is a direct consequence of property (2.7).

2.3.2 The Infimal Convolution

Let $h_1, h_2 : \mathbb{E} \rightarrow (-\infty, \infty]$ be two proper functions. The *infimal convolution* of h_1, h_2 is defined by the following formula:

$$(h_1 \square h_2)(\mathbf{x}) \equiv \min_{\mathbf{u} \in \mathbb{E}} \{h_1(\mathbf{u}) + h_2(\mathbf{x} - \mathbf{u})\}.$$

A direct consequence of Theorem 2.18 is the following result stating that the infimal convolution of a proper convex function and a real-valued convex function is always convex.

Theorem 2.19 (convexity of the infimal convolution). *Let $h_1 : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper convex function and let $h_2 : \mathbb{E} \rightarrow \mathbb{R}$ be a real-valued convex function. Then $h_1 \square h_2$ is convex.*

Proof. Define $f(\mathbf{x}, \mathbf{y}) \equiv h_1(\mathbf{y}) + h_2(\mathbf{x} - \mathbf{y})$. The convexity of h_1 and h_2 implies that f is convex. In addition, property (2.7) holds since for any $\mathbf{x} \in \mathbb{E}$, we can pick any $\mathbf{y} \in \text{dom}(h_1)$ and obtain that $f(\mathbf{x}, \mathbf{y}) = h_1(\mathbf{y}) + h_2(\mathbf{x} - \mathbf{y}) < \infty$. Thus, by Theorem 2.18, the function $h_1 \square h_2$, as a partial minimization function of $f(\cdot, \cdot)$ w.r.t. the second argument is a convex function. \square

Example 2.20 (convexity of the distance function). Let $C \subseteq \mathbb{E}$ be a nonempty convex set. The distance function can be written as the following infimal convolution:

$$d_C(\mathbf{x}) = \min_{\mathbf{y}} \{\|\mathbf{x} - \mathbf{y}\| : \mathbf{y} \in C\} = \min_{\mathbf{y} \in \mathbb{E}} \{\delta_C(\mathbf{y}) + \|\mathbf{x} - \mathbf{y}\|\} = (\delta_C \square h_1)(\mathbf{x}),$$

where $h_1(\cdot) = \|\cdot\|$. Since δ_C is proper and convex and h_1 is real-valued convex, it follows by Theorem 2.19 that d_C is convex. \blacksquare

2.3.3 Continuity of Convex Functions

It is well known that convex functions are continuous at points in the interior of their domain. This is explicitly recalled in the next result, which actually states a stronger property of convex functions—local Lipschitz continuity.

Theorem 2.21 (local Lipschitz continuity of convex functions [10, Theorem 7.36]). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be convex. Let $\mathbf{x}_0 \in \text{int}(\text{dom}(f))$. Then there exist $\varepsilon > 0$ and $L > 0$ such that $B[\mathbf{x}_0, \varepsilon] \subseteq C$ and*

$$|f(\mathbf{x}) - f(\mathbf{x}_0)| \leq L\|\mathbf{x} - \mathbf{x}_0\| \tag{2.11}$$

for all $\mathbf{x} \in B[\mathbf{x}_0, \varepsilon]$.

Convex functions are not necessarily continuous at boundary points. Continuity is not guaranteed even when the function at hand is closed and convex (cf. Example 2.32). However, for univariate functions we will now show that closed and convex functions are continuous.

Theorem 2.22 (continuity of closed convex univariate functions). *Let $f : \mathbb{R} \rightarrow (-\infty, \infty]$ be a proper closed and convex function. Then f is continuous over $\text{dom}(f)$.*

Proof. Since f is convex, its domain is some interval $I = \text{dom}(f)$. If $\text{int}(I) = \emptyset$, then I is a singleton, and consequently the continuity of f over I is obvious. Assume then that $\text{int}(I) \neq \emptyset$. The fact that f is continuous over $\text{int}(I)$ follows from Theorem 2.21. We only need to show the continuity of f at the endpoints of I (if it exists). For that, we can assume without loss of generality that the interval I has a left endpoint a , and we will prove the right continuity of f at a . We begin by showing that $\lim_{t \rightarrow a^+} f(t)$ exists. Let $c > a$ be an arbitrary scalar in I and define the function

$$g(t) \equiv \frac{f(c-t) - f(c)}{t}.$$

Obviously, g is defined on $(0, c-a]$. We will show that g is nondecreasing and upper bounded over $(0, c-a]$. For that, take $0 < t \leq s \leq c-a$. Then

$$c-t = \left(1 - \frac{t}{s}\right)c + \frac{t}{s}(c-s),$$

and hence, by the convexity of f ,

$$f(c-t) \leq \left(1 - \frac{t}{s}\right)f(c) + \frac{t}{s}f(c-s),$$

which after some rearrangement of terms can be seen to be equivalent to

$$\frac{f(c-t) - f(c)}{t} \leq \frac{f(c-s) - f(c)}{s}.$$

Thus,

$$g(t) \leq g(s) \text{ for any } 0 < t \leq s \leq c-a. \quad (2.12)$$

Namely, g is nondecreasing over $(0, c-a]$. To show the upper boundedness, just plug $s = c-a$ into (2.12) and obtain that

$$g(t) \leq g(c-a) \text{ for any } t \in (0, c-a]. \quad (2.13)$$

We can thus conclude that $\lim_{t \rightarrow (c-a)^-} g(t)$ exists and is equal to some real number ℓ . Hence,

$$f(c-t) = f(c) + tg(t) \rightarrow f(c) + (c-a)\ell,$$

as $t \rightarrow (c-a)^-$, and consequently $\lim_{t \rightarrow a^+} f(t)$ exists and is equal to $f(c) + (c-a)\ell$. Using (2.13), we obtain that for any $t \in (0, c-a]$,

$$f(c-t) = f(c) + tg(t) \leq f(c) + (c-a)g(c-a) = f(c) + (c-a)\frac{f(a) - f(c)}{c-a} = f(a),$$

implying the inequality $\lim_{t \rightarrow a^+} f(t) \leq f(a)$. On the other hand, since f is closed, it is also lower semicontinuous (Theorem 2.6), and thus $\lim_{t \rightarrow a^+} f(t) \geq f(a)$. Consequently, $\lim_{t \rightarrow a^+} f(t) = f(a)$, proving the right continuity of f at a . \square

2.4 Support Functions

Let $C \subseteq \mathbb{E}$ be a nonempty set. Then the *support function* of C is the function $\sigma_C : \mathbb{E}^* \rightarrow (-\infty, \infty]$ given by

$$\sigma_C(\mathbf{y}) = \max_{\mathbf{x} \in C} \langle \mathbf{y}, \mathbf{x} \rangle.$$

For a fixed \mathbf{x} , the linear function $\mathbf{y} \mapsto \langle \mathbf{y}, \mathbf{x} \rangle$ is obviously closed and convex. Therefore, by Theorems 2.7(c) and 2.16(c), the support function, as a maximum of closed and convex functions, is always closed and convex, regardless of whether C is closed and/or convex. We summarize this property in the next lemma.

Lemma 2.23 (closedness and convexity of support functions). *Let $C \subseteq \mathbb{E}$ be a nonempty set. Then σ_C is a closed and convex function.*

In most of our discussions on support functions in this chapter, the fact that σ_C operates on the dual space \mathbb{E}^* instead of \mathbb{E} will have no importance—recall that we use the convention that the elements of \mathbb{E}^* and \mathbb{E} are the same. However, when norms will be involved, naturally, the dual norm will have to be used (cf. Example 2.31).

Additional properties of support functions that follow directly by definition are given in Lemma 2.24 below. Note that for two sets A, B that reside in the same space, the sum $A + B$ stands for the *Minkowski sum* given by

$$A + B = \{\mathbf{a} + \mathbf{b} : \mathbf{a} \in A, \mathbf{b} \in B\}.$$

Also, for a scalar $\alpha \in \mathbb{R}$ and a set $A \subseteq \mathbb{E}$, the set αA is

$$\alpha A = \{\alpha \mathbf{a} : \mathbf{a} \in A\}.$$

Lemma 2.24.

(a) **(positive homogeneity)** For any nonempty set $C \subseteq \mathbb{E}$, $\mathbf{y} \in \mathbb{E}^*$ and $\alpha \geq 0$,

$$\sigma_C(\alpha \mathbf{y}) = \alpha \sigma_C(\mathbf{y}).$$

(b) **(subadditivity)** For any nonempty set $C \subseteq \mathbb{E}$ and $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{E}^*$,

$$\sigma_C(\mathbf{y}_1 + \mathbf{y}_2) \leq \sigma_C(\mathbf{y}_1) + \sigma_C(\mathbf{y}_2).$$

(c) For any nonempty set $C \subseteq \mathbb{E}$, $\mathbf{y} \in \mathbb{E}^*$ and $\alpha \geq 0$,

$$\sigma_{\alpha C}(\mathbf{y}) = \alpha \sigma_C(\mathbf{y}).$$

(d) For any two nonempty sets $A, B \subseteq \mathbb{E}$ and $\mathbf{y} \in \mathbb{E}^*$,

$$\sigma_{A+B}(\mathbf{y}) = \sigma_A(\mathbf{y}) + \sigma_B(\mathbf{y}).$$

Proof. (a) $\sigma_C(\alpha\mathbf{y}) = \max_{\mathbf{x} \in C} \langle \alpha\mathbf{y}, \mathbf{x} \rangle = \alpha \max_{\mathbf{x} \in C} \langle \mathbf{y}, \mathbf{x} \rangle = \alpha\sigma_C(\mathbf{y})$.

(b)

$$\begin{aligned}\sigma_C(\mathbf{y}_1 + \mathbf{y}_2) &= \max_{\mathbf{x} \in C} \langle \mathbf{y}_1 + \mathbf{y}_2, \mathbf{x} \rangle = \max_{\mathbf{x} \in C} [\langle \mathbf{y}_1, \mathbf{x} \rangle + \langle \mathbf{y}_2, \mathbf{x} \rangle] \\ &\leq \max_{\mathbf{x} \in C} \langle \mathbf{y}_1, \mathbf{x} \rangle + \max_{\mathbf{x} \in C} \langle \mathbf{y}_2, \mathbf{x} \rangle = \sigma_C(\mathbf{y}_1) + \sigma_C(\mathbf{y}_2).\end{aligned}$$

(c)

$$\sigma_{\alpha C}(\mathbf{y}) = \max_{\mathbf{x} \in \alpha C} \langle \mathbf{y}, \mathbf{x} \rangle = \max_{\mathbf{x}_1 \in C} \langle \mathbf{y}, \alpha \mathbf{x}_1 \rangle = \alpha \max_{\mathbf{x}_1 \in C} \langle \mathbf{y}, \mathbf{x}_1 \rangle = \alpha\sigma_C(\mathbf{y}).$$

(d)

$$\begin{aligned}\sigma_{A+B}(\mathbf{y}) &= \max_{\mathbf{x} \in A+B} \langle \mathbf{y}, \mathbf{x} \rangle = \max_{\mathbf{x}_1 \in A, \mathbf{x}_2 \in B} \langle \mathbf{y}, \mathbf{x}_1 + \mathbf{x}_2 \rangle \\ &= \max_{\mathbf{x}_1 \in A, \mathbf{x}_2 \in B} [\langle \mathbf{y}, \mathbf{x}_1 \rangle + \langle \mathbf{y}, \mathbf{x}_2 \rangle] = \max_{\mathbf{x}_1 \in A} \langle \mathbf{y}, \mathbf{x}_1 \rangle + \max_{\mathbf{x}_2 \in B} \langle \mathbf{y}, \mathbf{x}_2 \rangle \\ &= \sigma_A(\mathbf{y}) + \sigma_B(\mathbf{y}). \quad \blacksquare\end{aligned}$$

Following are some basic examples of support functions.

Example 2.25 (support functions of finite sets). Suppose that

$$C = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m\},$$

where $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m \in \mathbb{E}$. Then

$$\sigma_C(\mathbf{y}) = \max\{\langle \mathbf{b}_1, \mathbf{y} \rangle, \langle \mathbf{b}_2, \mathbf{y} \rangle, \dots, \langle \mathbf{b}_m, \mathbf{y} \rangle\}.$$

■

Recall that $S \subseteq \mathbb{E}$ is called a *cone* if it satisfies the following property: for any $\mathbf{x} \in S$ and $\lambda \geq 0$, the inclusion $\lambda\mathbf{x} \in S$ holds.

Example 2.26 (support functions of cones). Let $K \subseteq \mathbb{E}$ be a cone. Define the polar cone of K as

$$K^\circ = \{\mathbf{y} \in \mathbb{E}^*: \langle \mathbf{y}, \mathbf{x} \rangle \leq 0 \text{ for all } \mathbf{x} \in K\}.$$

We will show that

$$\sigma_K(\mathbf{y}) = \delta_{K^\circ}(\mathbf{y}). \quad (2.14)$$

Indeed, if $\mathbf{y} \in K^\circ$, then $\langle \mathbf{y}, \mathbf{x} \rangle \leq 0$ for all $\mathbf{x} \in K$ and for $\mathbf{x} = \mathbf{0}$, $\langle \mathbf{y}, \mathbf{x} \rangle = 0$. Therefore,

$$\sigma_K(\mathbf{y}) = \max_{\mathbf{x} \in K} \langle \mathbf{y}, \mathbf{x} \rangle = 0.$$

If $\mathbf{y} \notin K^\circ$, then there exists $\tilde{\mathbf{x}} \in K$ such that $\langle \mathbf{y}, \tilde{\mathbf{x}} \rangle > 0$. Since $\lambda\tilde{\mathbf{x}} \in K$ for all $\lambda \geq 0$, it follows that

$$\sigma_K(\mathbf{y}) \geq \langle \mathbf{y}, \lambda\tilde{\mathbf{x}} \rangle = \lambda\langle \mathbf{y}, \tilde{\mathbf{x}} \rangle \text{ for all } \lambda \geq 0.$$

Taking $\lambda \rightarrow \infty$, we obtain that $\sigma_K(\mathbf{y}) = \infty$ for $\mathbf{y} \notin K^\circ$, and hence formula (2.14) is proven. ■

Example 2.27 (support function of the nonnegative orthant). Consider the space $\mathbb{E} = \mathbb{R}^n$. As a special case of Example 2.26, since $(\mathbb{R}_+^n)^\circ = \mathbb{R}_-^n$, it follows that

$$\sigma_{\mathbb{R}_+^n}(\mathbf{y}) = \delta_{\mathbb{R}_-^n}(\mathbf{y}).$$

■

The next example uses Farkas's lemma,⁶ which we now recall.

Lemma 2.28 (Farkas's lemma—second formulation). *Let $\mathbf{c} \in \mathbb{R}^n$ and $\mathbf{A} \in \mathbb{R}^{m \times n}$. Then the following two claims are equivalent:*

- A. *The implication $\mathbf{Ax} \leq \mathbf{0} \Rightarrow \mathbf{c}^T \mathbf{x} \leq 0$ holds true.*
- B. *There exists $\mathbf{y} \in \mathbb{R}_+^m$ such that $\mathbf{A}^T \mathbf{y} = \mathbf{c}$.*

Example 2.29 (support functions of convex polyhedral cones). Let the underlying space be $\mathbb{E} = \mathbb{R}^n$ and let $\mathbf{A} \in \mathbb{R}^{m \times n}$. Define the set

$$S = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} \leq \mathbf{0}\}.$$

Since S is a cone, we can use Example 2.26 to conclude that

$$\sigma_S(\mathbf{y}) = \delta_{S^\circ}(\mathbf{y}).$$

Note that $\mathbf{y} \in S^\circ$ if and only if

$$\langle \mathbf{y}, \mathbf{x} \rangle \leq 0 \text{ for any } \mathbf{x} \text{ satisfying } \mathbf{Ax} \leq \mathbf{0}. \quad (2.15)$$

By Farkas's lemma (Lemma 2.28), (2.15) is equivalent to the statement

there exists $\boldsymbol{\lambda} \in \mathbb{R}_+^m$ such that $\mathbf{A}^T \boldsymbol{\lambda} = \mathbf{y}$.

Hence,

$$S^\circ = \{\mathbf{A}^T \boldsymbol{\lambda} : \boldsymbol{\lambda} \in \mathbb{R}_+^m\}.$$

To conclude,

$$\sigma_S(\mathbf{y}) = \delta_{\{\mathbf{A}^T \boldsymbol{\lambda} : \boldsymbol{\lambda} \in \mathbb{R}_+^m\}}(\mathbf{y}).$$

■

Example 2.30 (support functions of affine sets). Let the underlying space be $\mathbb{E} = \mathbb{R}^n$ and let $\mathbf{B} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$. Define the affine set

$$C = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Bx} = \mathbf{b}\}.$$

⁶The lemma and its proof can be found, for example, in [10, Lemma 10.3].

We assume that C is nonempty, namely, that there exists $\mathbf{x}_0 \in \mathbb{R}^n$ for which $\mathbf{Bx}_0 = \mathbf{b}$. The support function is obviously given by

$$\sigma_C(\mathbf{y}) = \max_{\mathbf{x}} \{\langle \mathbf{y}, \mathbf{x} \rangle : \mathbf{Bx} = \mathbf{b}\}.$$

Making the change of variables $\mathbf{x} = \mathbf{z} + \mathbf{x}_0$, we obtain that the support function can be rewritten as

$$\begin{aligned} \sigma_C(\mathbf{y}) &= \max_{\mathbf{z}} \{\langle \mathbf{y}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{x}_0 \rangle : \mathbf{Bz} = \mathbf{0}\} \\ &= \langle \mathbf{y}, \mathbf{x}_0 \rangle + \max_{\mathbf{z}} \{\langle \mathbf{y}, \mathbf{z} \rangle : \mathbf{Bz} = \mathbf{0}\} \\ &= \langle \mathbf{y}, \mathbf{x}_0 \rangle + \sigma_{\tilde{C}}(\mathbf{y}), \end{aligned} \quad (2.16)$$

where $\tilde{C} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Bx} = \mathbf{0}\}$. The set \tilde{C} is a convex polyhedral cone that can be written as

$$\tilde{C} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} \leq \mathbf{0}\},$$

where $\mathbf{A} = \begin{pmatrix} \mathbf{B} \\ -\mathbf{B} \end{pmatrix}$. By Example 2.29, it follows that

$$\sigma_{\tilde{C}} = \delta_{\tilde{C}^\circ}, \quad (2.17)$$

where \tilde{C}° is the polar cone of \tilde{C} , which is given by

$$\tilde{C}^\circ = \{\mathbf{B}^T \boldsymbol{\lambda}_1 - \mathbf{B}^T \boldsymbol{\lambda}_2 : \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in \mathbb{R}_+^m\}.$$

We will show that

$$\tilde{C}^\circ = \text{Range}(\mathbf{B}^T). \quad (2.18)$$

Indeed, if $\mathbf{v} \in \tilde{C}^\circ$, then there exists $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in \mathbb{R}_+^m$ for which $\mathbf{v} = \mathbf{B}^T \boldsymbol{\lambda}_1 - \mathbf{B}^T \boldsymbol{\lambda}_2 = \mathbf{B}^T (\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2) \in \text{Range}(\mathbf{B}^T)$. In the other direction, if $\mathbf{v} \in \text{Range}(\mathbf{B}^T)$, then there exists $\boldsymbol{\lambda} \in \mathbb{R}^m$ for which $\mathbf{v} = \mathbf{B}^T \boldsymbol{\lambda}$. Defining $\boldsymbol{\lambda}_1 = [\boldsymbol{\lambda}]_+, \boldsymbol{\lambda}_2 = [-\boldsymbol{\lambda}]_+$, we obtain that $\boldsymbol{\lambda} = \boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2$ with $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in \mathbb{R}_+^m$, and hence

$$\mathbf{v} = \mathbf{B}^T \boldsymbol{\lambda} = \mathbf{B}^T (\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2) = \mathbf{B}^T \boldsymbol{\lambda}_1 - \mathbf{B}^T \boldsymbol{\lambda}_2 \in \tilde{C}^\circ.$$

Combining (2.16), (2.17), and (2.18), we finally conclude that

$$\sigma_C(\mathbf{y}) = \langle \mathbf{y}, \mathbf{x}_0 \rangle + \delta_{\text{Range}(\mathbf{B}^T)}(\mathbf{y}).$$

Example 2.31 (support functions of unit balls). Suppose that \mathbb{E} is the underlying space endowed with a norm $\|\cdot\|$. Consider the unit ball given by

$$B_{\|\cdot\|}[\mathbf{0}, 1] = \{\mathbf{x} \in \mathbb{E} : \|\mathbf{x}\| \leq 1\}.$$

By the definition of the dual norm, we have for any $\mathbf{y} \in \mathbb{E}^*$

■

$$\sigma_{B_{\|\cdot\|}[\mathbf{0},1]}(\mathbf{y}) = \max_{\|\mathbf{x}\| \leq 1} \langle \mathbf{y}, \mathbf{x} \rangle = \|\mathbf{y}\|_*.$$

Thus, for example, for the space \mathbb{R}^n we have

$$\begin{aligned}\sigma_{B_{\|\cdot\|_p}[\mathbf{0},1]}(\mathbf{y}) &= \|\mathbf{y}\|_q \quad \left(1 \leq p \leq \infty, \frac{1}{p} + \frac{1}{q} = 1\right), \\ \sigma_{B_{\|\cdot\|_{\mathbf{Q}}}[\mathbf{0},1]}(\mathbf{y}) &= \|\mathbf{y}\|_{\mathbf{Q}^{-1}} \quad (\mathbf{Q} \in \mathbb{S}_{++}^n).\end{aligned}$$

In the first formula we use the convention that $p = 1/\infty$ corresponds to $q = \infty/1$. \blacksquare

The next example is also an example of a closed and convex function that is not continuous (recall that such an example does not exist for one-dimensional functions; see Theorem 2.22).

Example 2.32.⁷ Consider the following set in \mathbb{R}^2 :

$$C = \left\{ (x_1, x_2)^T : x_1 + \frac{x_2^2}{2} \leq 0 \right\}.$$

Then the support function of C is given by

$$\sigma_C(\mathbf{y}) = \max_{x_1, x_2} \left\{ y_1 x_1 + y_2 x_2 : x_1 + \frac{x_2^2}{2} \leq 0 \right\}. \quad (2.19)$$

Obviously, $\sigma_C(\mathbf{0}) = 0$. We will compute the support function at $\mathbf{y} \neq \mathbf{0}$. In this case, it is easy to see that the maximum of problem (2.19) is attained at the boundary of C .⁸ Therefore,

$$\sigma_C(\mathbf{y}) = \max_{x_1, x_2} \left\{ y_1 x_1 + y_2 x_2 : x_1 + \frac{x_2^2}{2} = 0 \right\} = \max_{x_2} \left\{ -\frac{y_1}{2} x_2^2 + y_2 x_2 \right\}.$$

If $y_1 < 0$, then the maximal value is ∞ . If $y_1 = 0$ and $y_2 \neq 0$, then the maximal value is also ∞ . If $y_1 > 0$, the maximum is attained at $x_2 = \frac{y_2}{y_1}$, and the corresponding maximal value is $\frac{y_2^2}{2y_1}$. Thus, the support function is given by

$$\sigma_C(\mathbf{y}) = \begin{cases} \frac{y_2^2}{2y_1}, & y_1 > 0, \\ 0, & y_1 = y_2 = 0, \\ \infty & \text{else.} \end{cases}$$

⁷Example 2.32 is from Rockafellar [108, p. 83].

⁸This fact can be shown by contradiction. If the maximum was attained at an interior point of C , then the gradient of the objective function, meaning \mathbf{y} , would be the zeros vector, which is a contradiction to the assumption that $\mathbf{y} \neq \mathbf{0}$.

By Lemma 2.23, σ_C is closed and convex. However, it is not continuous at $(y_1, y_2) = (0, 0)$. Indeed, taking for any $\alpha > 0$ the path $y_1(t) = \frac{t^2}{2\alpha}, y_2(t) = t(t > 0)$, we obtain that

$$\sigma_C(y_1(t), y_2(t)) = \alpha,$$

and hence the limit of $\sigma_C(y_1(t), y_2(t))$ as $t \rightarrow 0^+$ is α , which combined with the fact that $\sigma_C(0, 0) = 0$ implies the discontinuity of f at $(0, 0)$. The contour lines of σ_C are plotted in Figure 2.3. ■

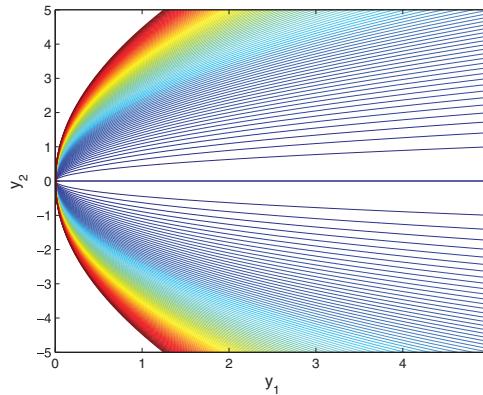


Figure 2.3. Contour lines of the closed, convex, and noncontinuous function from Example 2.32.

An important property of support functions is that they are completely determined by their underlying sets as long as these sets are closed and convex. The proof of this result requires the strict separation theorem,⁹ which is now recalled.

Theorem 2.33 (strict separation theorem). *Let $C \subseteq \mathbb{E}$ be a nonempty closed and convex set, and let $\mathbf{y} \notin C$. Then there exist $\mathbf{p} \in \mathbb{E}^* \setminus \{\mathbf{0}\}$ and $\alpha \in \mathbb{R}$ such that*

$$\langle \mathbf{p}, \mathbf{y} \rangle > \alpha$$

and

$$\langle \mathbf{p}, \mathbf{x} \rangle \leq \alpha \text{ for all } \mathbf{x} \in C.$$

Lemma 2.34. *Let $A, B \subseteq \mathbb{E}$ be nonempty closed and convex sets. Then $A = B$ if and only if $\sigma_A = \sigma_B$.*

Proof. If $A = B$, then obviously $\sigma_A = \sigma_B$. Suppose now that $\sigma_A = \sigma_B$. We will prove that $A = B$. Assume by contradiction that this is not the case, and without loss of generality suppose that there exists $\mathbf{y} \in A$ such that $\mathbf{y} \notin B$. Since $\mathbf{y} \notin B$ and B is nonempty closed and convex, by the strict separation theorem, there exists a

⁹The theorem and its proof can be found, for example, in [10, Theorem 10.1].

hyperplane separating \mathbf{y} from B , meaning that there exists $\mathbf{p} \in \mathbb{E}^* \setminus \{\mathbf{0}\}$ and $\alpha > 0$ such that

$$\langle \mathbf{p}, \mathbf{x} \rangle \leq \alpha < \langle \mathbf{p}, \mathbf{y} \rangle \text{ for any } \mathbf{x} \in B.$$

Taking the maximum over $\mathbf{x} \in B$, we conclude that $\sigma_B(\mathbf{p}) \leq \alpha < \langle \mathbf{p}, \mathbf{y} \rangle \leq \sigma_A(\mathbf{y})$, a contradiction to the assertion that the support functions are the same. \square

A related result states that the support function stays the same under the operations of closure and convex hull of the underlying set.

Lemma 2.35. *Let $A \subseteq \mathbb{E}$ be nonempty. Then*

- (a) $\sigma_A = \sigma_{\text{cl}(A)}$;
- (b) $\sigma_A = \sigma_{\text{conv}(A)}$.

Proof. (a) Since $A \subseteq \text{cl}(A)$,

$$\sigma_A(\mathbf{y}) \leq \sigma_{\text{cl}(A)}(\mathbf{y}) \text{ for any } \mathbf{y} \in \mathbb{E}^*. \quad (2.20)$$

We will show the reverse inequality. Let $\mathbf{y} \in \mathbb{E}^*$. Then by the definition of the support function, there exists a sequence $\{\mathbf{x}^k\}_{k \geq 1} \subseteq \text{cl}(A)$ such that

$$\langle \mathbf{y}, \mathbf{x}^k \rangle \rightarrow \sigma_{\text{cl}(A)}(\mathbf{y}) \text{ as } k \rightarrow \infty. \quad (2.21)$$

By the definition of the closure, it follows that there exists a sequence $\{\mathbf{z}^k\}_{k \geq 1} \subseteq A$ such that $\|\mathbf{z}^k - \mathbf{x}^k\| \leq \frac{1}{k}$ for all k , and hence

$$\mathbf{z}^k - \mathbf{x}^k \rightarrow 0 \text{ as } k \rightarrow \infty. \quad (2.22)$$

Now, since $\mathbf{z}^k \in A$,

$$\sigma_A(\mathbf{y}) \geq \langle \mathbf{y}, \mathbf{z}^k \rangle = \langle \mathbf{y}, \mathbf{x}^k \rangle + \langle \mathbf{y}, \mathbf{z}^k - \mathbf{x}^k \rangle.$$

Taking $k \rightarrow \infty$ and using (2.21), (2.22), we obtain that

$$\sigma_A(\mathbf{y}) \geq \sigma_{\text{cl}(A)}(\mathbf{y}) + 0 = \sigma_{\text{cl}(A)}(\mathbf{y}),$$

which combined with (2.20) yields the desired result $\sigma_A = \sigma_{\text{cl}(A)}$.

(b) Since $A \subseteq \text{conv}(A)$, we have that $\sigma_A(\mathbf{y}) \leq \sigma_{\text{conv}(A)}(\mathbf{y})$ for any $\mathbf{y} \in \mathbb{E}^*$. We will show the reverse inequality. Let $\mathbf{y} \in \mathbb{E}^*$. Then by the definition of the support function, there exists a sequence $\{\mathbf{x}^k\}_{k \geq 1} \subseteq \text{conv}(A)$ such that

$$\langle \mathbf{y}, \mathbf{x}^k \rangle \rightarrow \sigma_{\text{conv}(A)}(\mathbf{y}) \text{ as } k \rightarrow \infty. \quad (2.23)$$

By the definition of the convex hull, it follows that for any k , there exist vectors $\mathbf{z}_1^k, \mathbf{z}_2^k, \dots, \mathbf{z}_{n_k}^k \in A$ (n_k is a positive integer) and $\boldsymbol{\lambda}^k \in \Delta_{n_k}$ such that

$$\mathbf{x}^k = \sum_{i=1}^{n_k} \lambda_i^k \mathbf{z}_i^k.$$

Now,

$$\langle \mathbf{y}, \mathbf{x}^k \rangle = \left\langle \mathbf{y}, \sum_{i=1}^{n_k} \lambda_i^k \mathbf{z}_i^k \right\rangle = \sum_{i=1}^{n_k} \lambda_i^k \langle \mathbf{y}, \mathbf{z}_i^k \rangle \leq \sum_{i=1}^{n_k} \lambda_i^k \sigma_A(\mathbf{y}) = \sigma_A(\mathbf{y}),$$

where the inequality follows by the fact that $\mathbf{z}_i^k \in A$. Taking the limit as $k \rightarrow \infty$ and using (2.23), we obtain that $\sigma_{\text{conv}(A)}(\mathbf{y}) \leq \sigma_A(\mathbf{y})$. \square

Example 2.36 (support of the unit simplex). Suppose that the underlying space is \mathbb{R}^n and consider the unit simplex set $\Delta_n = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{e}^T \mathbf{x} = 1, \mathbf{x} \geq \mathbf{0}\}$. Since the unit simplex can be written as the convex hull of the standard basis of \mathbb{R}^n ,

$$\Delta_n = \text{conv}\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\},$$

it follows by Lemma 2.35(b) that

$$\sigma_{\Delta_n}(\mathbf{y}) = \sigma_{\{\mathbf{e}_1, \dots, \mathbf{e}_n\}}(\mathbf{y}) = \max\{\langle \mathbf{e}_1, \mathbf{y} \rangle, \langle \mathbf{e}_2, \mathbf{y} \rangle, \dots, \langle \mathbf{e}_n, \mathbf{y} \rangle\}.$$

Since we always assume (unless otherwise stated) that \mathbb{R}^n is endowed with the dot product, the support function is

$$\sigma_{\Delta_n}(\mathbf{y}) = \max\{y_1, y_2, \dots, y_n\}.$$

■

The table below summarizes the main support function computations that were considered in this section.

C	$\sigma_C(\mathbf{y})$	Assumptions	Reference
$\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$	$\max_{i=1,2,\dots,n} \langle \mathbf{b}_i, \mathbf{y} \rangle$	$\mathbf{b}_i \in \mathbb{E}$	Example 2.25
K	$\delta_K^\circ(\mathbf{y})$	K —cone	Example 2.26
\mathbb{R}_+^n	$\delta_{\mathbb{R}_+^n}(\mathbf{y})$	$\mathbb{E} = \mathbb{R}^n$	Example 2.27
Δ_n	$\max\{y_1, y_2, \dots, y_n\}$	$\mathbb{E} = \mathbb{R}^n$	Example 2.36
$\{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} \leq \mathbf{0}\}$	$\delta_{\{\mathbf{A}^T \boldsymbol{\lambda} : \boldsymbol{\lambda} \in \mathbb{R}_+^m\}}(\mathbf{y})$	$\mathbb{E} = \mathbb{R}^n, \mathbf{A} \in \mathbb{R}^{m \times n}$	Example 2.29
$\{\mathbf{x} \in \mathbb{R}^n : \mathbf{Bx} = \mathbf{b}\}$	$\langle \mathbf{y}, \mathbf{x}_0 \rangle + \delta_{\text{Range}(\mathbf{B}^T)}(\mathbf{y})$	$\mathbb{E} = \mathbb{R}^n, \mathbf{B} \in \mathbb{R}^{m \times n}, \mathbf{b} \in \mathbb{R}^m, \mathbf{Bx}_0 = \mathbf{b}$	Example 2.30
$B_{\ \cdot\ }[\mathbf{0}, 1]$	$\ \mathbf{y}\ _*$	-	Example 2.31

Chapter 3

Subgradients

3.1 Definitions and First Examples

Definition 3.1 (subgradient). Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper function and let $\mathbf{x} \in \text{dom}(f)$. A vector $\mathbf{g} \in \mathbb{E}^*$ is called a **subgradient** of f at \mathbf{x} if

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle \text{ for all } \mathbf{y} \in \mathbb{E}. \quad (3.1)$$

Recall (see Section 1.11) that we use in this book the convention that the elements of \mathbb{E}^* are exactly the elements of \mathbb{E} , whereas the asterisk just marks the fact that the endowed norm on \mathbb{E}^* is the dual norm $\|\cdot\|_*$ rather than the endowed norm $\|\cdot\|$ on \mathbb{E} .

The inequality (3.1) is also called *the subgradient inequality*. It actually says that each subgradient is associated with an underestimate affine function, which is tangent to the surface of the function at \mathbf{x} . Since the subgradient inequality (3.1) is trivial for $\mathbf{y} \notin \text{dom}(f)$, it is frequently restricted to points in $\text{dom}(f)$ and is thus written as

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle \text{ for all } \mathbf{y} \in \text{dom}(f).$$

Given a point $\mathbf{x} \in \text{dom}(f)$, there might be more than one subgradient of f at \mathbf{x} , and the set of all subgradients is called the *subdifferential*.

Definition 3.2 (subdifferential). The set of all subgradients of f at \mathbf{x} is called the **subdifferential** of f at \mathbf{x} and is denoted by $\partial f(\mathbf{x})$:

$$\partial f(\mathbf{x}) \equiv \{\mathbf{g} \in \mathbb{E}^* : f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle \text{ for all } \mathbf{y} \in \mathbb{E}\}.$$

When $\mathbf{x} \notin \text{dom}(f)$, we define $\partial f(\mathbf{x}) = \emptyset$. Actually, for proper functions, this is a direct consequence of the definition of the subdifferential set since the subgradient inequality (3.1) does not hold for $\mathbf{x} \notin \text{dom}f$ and $\mathbf{y} \in \text{dom}f$.

Example 3.3 (subdifferential of norms at 0). Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}) = \|\mathbf{x}\|$, where $\|\cdot\|$ is the endowed norm on \mathbb{E} . We will show that the subdifferential of f at $\mathbf{x} = \mathbf{0}$ is the dual norm unit ball:

$$\partial f(\mathbf{0}) = B_{\|\cdot\|_*}[\mathbf{0}, 1] = \{\mathbf{g} \in \mathbb{E}^* : \|\mathbf{g}\|_* \leq 1\}. \quad (3.2)$$

To show (3.2), note that $\mathbf{g} \in \partial f(\mathbf{0})$ if and only if

$$f(\mathbf{y}) \geq f(\mathbf{0}) + \langle \mathbf{g}, \mathbf{y} - \mathbf{0} \rangle \text{ for all } \mathbf{y} \in \mathbb{E},$$

which is the same as

$$\|\mathbf{y}\| \geq \langle \mathbf{g}, \mathbf{y} \rangle \text{ for all } \mathbf{y} \in \mathbb{E}. \quad (3.3)$$

We will prove that the latter holds true if and only if $\|\mathbf{g}\|_* \leq 1$. Indeed, if $\|\mathbf{g}\|_* \leq 1$, then by the generalized Cauchy–Schwarz inequality (Lemma 1.4),

$$\langle \mathbf{g}, \mathbf{y} \rangle \leq \|\mathbf{g}\|_* \|\mathbf{y}\| \leq \|\mathbf{y}\| \text{ for any } \mathbf{y} \in \mathbb{E},$$

implying (3.3). In the reverse direction, assume that (3.3) holds. Taking the maximum of both sides of (3.3) over all \mathbf{y} satisfying $\|\mathbf{y}\| \leq 1$, we get

$$\|\mathbf{g}\|_* = \max_{\mathbf{y}: \|\mathbf{y}\| \leq 1} \langle \mathbf{g}, \mathbf{y} \rangle \leq \max_{\mathbf{y}: \|\mathbf{y}\| \leq 1} \|\mathbf{y}\| = 1.$$

We have thus established the equivalence between (3.3) and the inequality $\|\mathbf{g}\|_* \leq 1$, which is the same as the result (3.2). ■

Example 3.4 (subdifferential of the l_1 -norm at 0). Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}) = \|\mathbf{x}\|_1$. Then, since this is a special case of Example 3.3 with $\|\cdot\| = \|\cdot\|_1$, and since the l_∞ -norm is the dual of the l_1 -norm, it follows that

$$\partial f(0) = B_{\|\cdot\|_\infty}[\mathbf{0}, 1] = [-1, 1]^n.$$

In particular, when $n = 1$, then $f(x) = |x|$, and we have

$$\partial f(0) = [-1, 1].$$

The linear underestimators that correspond to -0.8 , -0.3 , and $0.7 \in \partial f(0)$, meaning $-0.8x$, $-0.3x$, and $0.7x$, are described in Figure 3.1. ■

For the next example, we need the definition of the *normal cone*. Given a set $S \subseteq \mathbb{E}$ and a point $\mathbf{x} \in S$, the normal cone of S at \mathbf{x} is defined as

$$N_S(\mathbf{x}) = \{\mathbf{y} \in \mathbb{E}^* : \langle \mathbf{y}, \mathbf{z} - \mathbf{x} \rangle \leq 0 \text{ for any } \mathbf{z} \in S\}.$$

The normal cone, in addition to being a cone, is closed and convex as an intersection of half-spaces. When $\mathbf{x} \notin S$, we define $N_S(\mathbf{x}) = \emptyset$.

Example 3.5 (subdifferential of indicator functions). Suppose that $S \subseteq \mathbb{E}$ is nonempty and consider the indicator function δ_S . Then for any $\mathbf{x} \in S$, we have that $\mathbf{y} \in \partial \delta_S(\mathbf{x})$ if and only if

$$\delta_S(\mathbf{z}) \geq \delta_S(\mathbf{x}) + \langle \mathbf{y}, \mathbf{z} - \mathbf{x} \rangle \text{ for all } \mathbf{z} \in S,$$

which is the same as

$$\langle \mathbf{y}, \mathbf{z} - \mathbf{x} \rangle \leq 0 \text{ for all } \mathbf{z} \in S.$$

Therefore, we have that

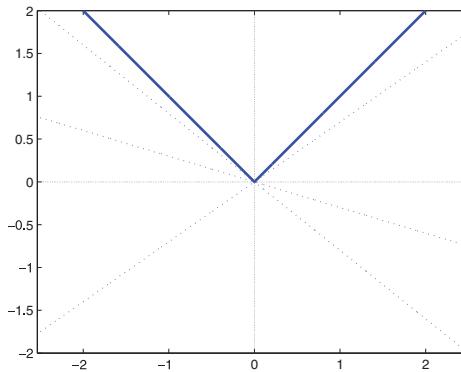


Figure 3.1. The linear underestimators of $|x|$ corresponding to $-0.8, -0.3, 0.7 \in \partial f(0)$; see Example 3.4.

$$\partial\delta_S(\mathbf{x}) = N_S(\mathbf{x}) \text{ for all } \mathbf{x} \in S. \quad (3.4)$$

For $\mathbf{x} \notin S$, $\partial\delta_S(\mathbf{x}) = N_S(\mathbf{x}) = \emptyset$ by convention, so we obtain that (3.4) holds also for $\mathbf{x} \notin S$. ■

Example 3.6 (subdifferential of the indicator function of the unit ball). As a special case of Example 3.5, let

$$S = B[\mathbf{0}, 1] = \{\mathbf{x} \in \mathbb{E} : \|\mathbf{x}\| \leq 1\}.$$

Then $\partial\delta_S(\mathbf{x}) = N_S(\mathbf{x})$, where $N_S(\mathbf{x})$ is given by

$$N_S(\mathbf{x}) = \{\mathbf{y} \in \mathbb{E}^* : \langle \mathbf{y}, \mathbf{z} - \mathbf{x} \rangle \leq 0 \text{ for all } \mathbf{z} \in S\}.$$

We will find a more explicit representation for $N_S(\mathbf{x})$. If $\mathbf{x} \notin S$, then $N_S(\mathbf{x}) = \emptyset$. Suppose that $\|\mathbf{x}\| \leq 1$. A vector $\mathbf{y} \in \mathbb{E}^*$ satisfies $\mathbf{y} \in N_S(\mathbf{x})$ if and only if

$$\langle \mathbf{y}, \mathbf{z} - \mathbf{x} \rangle \leq 0 \text{ for any } \mathbf{z} \text{ satisfying } \|\mathbf{z}\| \leq 1,$$

which is the same as the inequality,

$$\max_{\mathbf{z}: \|\mathbf{z}\| \leq 1} \langle \mathbf{y}, \mathbf{z} \rangle \leq \langle \mathbf{y}, \mathbf{x} \rangle.$$

Using the definition of the dual norm, we obtain that the latter can be rewritten as

$$\|\mathbf{y}\|_* \leq \langle \mathbf{y}, \mathbf{x} \rangle.$$

Therefore,

$$\partial\delta_{B[0,1]}(\mathbf{x}) = N_{B[0,1]}(\mathbf{x}) = \begin{cases} \{\mathbf{y} \in \mathbb{E}^* : \|\mathbf{y}\|_* \leq \langle \mathbf{y}, \mathbf{x} \rangle\}, & \|\mathbf{x}\| \leq 1, \\ \emptyset, & \|\mathbf{x}\| > 1. \end{cases}$$

■

Example 3.7 (subgradient of the dual function). Consider the minimization problem

$$\min\{f(\mathbf{x}) : \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \mathbf{x} \in X\}, \quad (3.5)$$

where $\emptyset \neq X \subseteq \mathbb{E}$, $f : \mathbb{E} \rightarrow \mathbb{R}$ and $\mathbf{g} : \mathbb{E} \rightarrow \mathbb{R}^m$ is a vector-valued function. The Lagrangian dual objective function is given by

$$q(\boldsymbol{\lambda}) = \min_{\mathbf{x} \in X} \left\{ L(\mathbf{x}; \boldsymbol{\lambda}) \equiv f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x}) \right\}.$$

The dual problem consists of maximizing q on its effective domain, which is given by

$$\text{dom}(-q) = \{\boldsymbol{\lambda} \in \mathbb{R}_+^m : q(\boldsymbol{\lambda}) > -\infty\}.$$

No matter whether the primal problem (3.5) is convex or not, the dual problem

$$\max_{\boldsymbol{\lambda} \in \mathbb{R}^m} \{q(\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \text{dom}(-q)\}$$

is always convex, meaning that q is a concave function and $\text{dom}(-q)$ is a convex set. Let $\boldsymbol{\lambda}_0 \in \text{dom}(-q)$ and assume that the minimum in the minimization problem defining $q(\boldsymbol{\lambda}_0)$,

$$q(\boldsymbol{\lambda}_0) = \min_{\mathbf{x} \in X} \left\{ f(\mathbf{x}) + \boldsymbol{\lambda}_0^T \mathbf{g}(\mathbf{x}) \right\},$$

is attained at $\mathbf{x}_0 \in X$, that is,

$$L(\mathbf{x}_0; \boldsymbol{\lambda}_0) = f(\mathbf{x}_0) + \boldsymbol{\lambda}_0^T \mathbf{g}(\mathbf{x}_0) = q(\boldsymbol{\lambda}_0).$$

We seek to find a subgradient of the convex function $-q$ at $\boldsymbol{\lambda}_0$. For that, note that for any $\boldsymbol{\lambda} \in \text{dom}(-q)$,

$$\begin{aligned} q(\boldsymbol{\lambda}) &= \min_{\mathbf{x} \in X} \left\{ f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x}) \right\} \\ &\leq f(\mathbf{x}_0) + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x}_0) \\ &= f(\mathbf{x}_0) + \boldsymbol{\lambda}_0^T \mathbf{g}(\mathbf{x}_0) + (\boldsymbol{\lambda} - \boldsymbol{\lambda}_0)^T \mathbf{g}(\mathbf{x}_0) \\ &= q(\boldsymbol{\lambda}_0) + \mathbf{g}(\mathbf{x}_0)^T (\boldsymbol{\lambda} - \boldsymbol{\lambda}_0). \end{aligned}$$

Thus,

$$-q(\boldsymbol{\lambda}) \geq -q(\boldsymbol{\lambda}_0) + (-\mathbf{g}(\mathbf{x}_0))^T (\boldsymbol{\lambda} - \boldsymbol{\lambda}_0) \text{ for any } \boldsymbol{\lambda} \in \text{dom}(-q),$$

concluding that

$$-\mathbf{g}(\mathbf{x}_0) \in \partial(-q)(\boldsymbol{\lambda}_0).$$

■

Example 3.8 (subgradient of the maximum eigenvalue function). Consider the function $f : \mathbb{S}^n \rightarrow \mathbb{R}$ given by $f(\mathbf{X}) = \lambda_{\max}(\mathbf{X})$ (recall that \mathbb{S}^n is the set of all $n \times n$ symmetric matrices). Let $\mathbf{X} \in \mathbb{S}^n$ and let \mathbf{v} be a normalized eigenvector of \mathbf{X} ($\|\mathbf{v}\|_2 = 1$) associated with the maximum eigenvalue of \mathbf{X} . We will establish the relation

$$\mathbf{v}\mathbf{v}^T \in \partial f(\mathbf{X}). \quad (3.6)$$

To show this, note that for any $\mathbf{Y} \in \mathbb{S}^n$,

$$\begin{aligned} \lambda_{\max}(\mathbf{Y}) &= \max_{\mathbf{u}} \{\mathbf{u}^T \mathbf{Y} \mathbf{u} : \|\mathbf{u}\|_2 = 1\} \\ &\geq \mathbf{v}^T \mathbf{Y} \mathbf{v} \\ &= \mathbf{v}^T \mathbf{X} \mathbf{v} + \mathbf{v}^T (\mathbf{Y} - \mathbf{X}) \mathbf{v} \\ &= \lambda_{\max}(\mathbf{X}) \|\mathbf{v}\|_2^2 + \text{Tr}(\mathbf{v}^T (\mathbf{Y} - \mathbf{X}) \mathbf{v}) \\ &= \lambda_{\max}(\mathbf{X}) + \text{Tr}(\mathbf{v} \mathbf{v}^T (\mathbf{Y} - \mathbf{X})) \\ &= \lambda_{\max}(\mathbf{X}) + \langle \mathbf{v} \mathbf{v}^T, \mathbf{Y} - \mathbf{X} \rangle, \end{aligned}$$

establishing (3.6). ■

There is an intrinsic difference between the results in Examples 3.7 and 3.8 and the results in Examples 3.3, 3.4, 3.5, and 3.6. Only one subgradient is computed in Examples 3.7 and 3.8; such results are referred to as *weak results*. On the other hand, in Examples 3.3, 3.4, 3.5, and 3.6 the entire subdifferential set is characterized—such results are called *strong results*.

3.2 Properties of the Subdifferential Set

Note that the subdifferential sets computed in the previous section are all closed and convex. This is not a coincidence. Subdifferential sets are *always* closed and convex.

Theorem 3.9 (closedness and convexity of the subdifferential set). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper function. Then the set $\partial f(\mathbf{x})$ is closed and convex for any $\mathbf{x} \in \mathbb{E}$.*

Proof. For any $\mathbf{x} \in \mathbb{E}$, the subdifferential set can be represented as

$$\partial f(\mathbf{x}) = \bigcap_{\mathbf{y} \in \mathbb{E}} H_{\mathbf{y}},$$

where $H_{\mathbf{y}} = \{\mathbf{g} \in \mathbb{E}^* : f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle\}$. Since the sets $H_{\mathbf{y}}$ are half-spaces and, in particular, closed and convex, it follows that $\partial f(\mathbf{x})$ is closed and convex. □

The subdifferential set $\partial f(\mathbf{x})$ may be empty. When it is nonempty at a given $\mathbf{x} \in \mathbb{E}$, the function f is called *subdifferentiable* at \mathbf{x} .

Definition 3.10 (subdifferentiability). *A proper function $f : \mathbb{E} \rightarrow (-\infty, \infty]$ is called **subdifferentiable** at $\mathbf{x} \in \text{dom}(f)$ if $\partial f(\mathbf{x}) \neq \emptyset$.*

The collection of points of subdifferentiability is denoted by $\text{dom}(\partial f)$:

$$\text{dom}(\partial f) = \{\mathbf{x} \in \mathbb{E} : \partial f(\mathbf{x}) \neq \emptyset\}.$$

We will now show that if a function is subdifferentiable at any point in its domain, which is assumed to be convex, then it is necessarily convex.

Lemma 3.11 (nonemptiness of subdifferential sets \Rightarrow convexity). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper function and assume that $\text{dom}(f)$ is convex. Suppose that for any $\mathbf{x} \in \text{dom}(f)$, the set $\partial f(\mathbf{x})$ is nonempty. Then f is convex.*

Proof. Let $\mathbf{x}, \mathbf{y} \in \text{dom}(f)$ and $\alpha \in [0, 1]$. Define $\mathbf{z}_\alpha = (1 - \alpha)\mathbf{x} + \alpha\mathbf{y}$. By the convexity of $\text{dom}(f)$, $\mathbf{z}_\alpha \in \text{dom}(f)$, and hence there exists $\mathbf{g} \in \partial f(\mathbf{z}_\alpha)$, which in particular implies the following two inequalities:

$$\begin{aligned} f(\mathbf{y}) &\geq f(\mathbf{z}_\alpha) + \langle \mathbf{g}, \mathbf{y} - \mathbf{z}_\alpha \rangle = f(\mathbf{z}_\alpha) + (1 - \alpha)\langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle, \\ f(\mathbf{x}) &\geq f(\mathbf{z}_\alpha) + \langle \mathbf{g}, \mathbf{x} - \mathbf{z}_\alpha \rangle = f(\mathbf{z}_\alpha) - \alpha\langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle. \end{aligned}$$

Multiplying the first inequality by α , the second by $1 - \alpha$, and summing them yields

$$f((1 - \alpha)\mathbf{x} + \alpha\mathbf{y}) = f(\mathbf{z}_\alpha) \leq (1 - \alpha)f(\mathbf{x}) + \alpha f(\mathbf{y}).$$

Since the latter holds for any $\mathbf{x}, \mathbf{y} \in \text{dom}(f)$ with $\text{dom}(f)$ being convex, it follows that the function f is convex. \square

We have thus shown that if a function is subdifferentiable at any point in its (assumed to be) convex domain, then it is convex. However, this does not mean that the reverse direction is correct. The next example describes a convex function, which is not subdifferentiable at one of the points in its domain.

Example 3.12. Consider the convex function $f : \mathbb{R} \rightarrow (-\infty, \infty]$ defined by

$$f(x) = \begin{cases} -\sqrt{x}, & x \geq 0, \\ \infty, & \text{else.} \end{cases}$$

The function is plotted in Figure 3.2. It is not subdifferentiable at $x = 0$. To show this, suppose by contradiction that there exists $g \in \mathbb{R}$ such that $g \in \partial f(0)$. Then

$$f(y) \geq f(0) + g(y - 0) \text{ for any } y \geq 0,$$

which is the same as

$$-\sqrt{y} \geq gy \text{ for any } y \geq 0. \quad (3.7)$$

The above is impossible since substituting $y = 1$, we obtain that $g \leq -1$ (and in particular $g < 0$), while substituting $y = \frac{1}{2g^2}$ in (3.7) yields the inequality

$$-\sqrt{1/(2g^2)} \geq 1/(2g),$$

which is equivalent to the impossible inequality (utilizing the fact that $g < 0$)

$$\frac{1}{2g^2} \leq \frac{1}{4g^2}. \quad \blacksquare$$

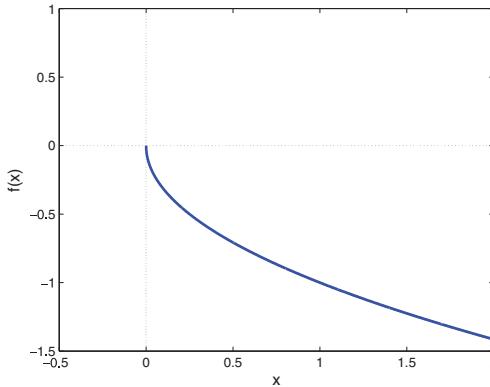


Figure 3.2. The function $f(x) = -\sqrt{x}$ with $\text{dom}(f) = \mathbb{R}_+$. The function is not subdifferentiable at $x = 0$.

Although, as demonstrated in Example 3.12, convex functions are not necessarily subdifferentiable at any point in their domain, they must be subdifferentiable at any point in the interior of their domain. This is stated in Theorem 3.14 below, which also shows the boundedness of the subdifferential set in this setting. The proof of the theorem strongly relies on the supporting hyperplane theorem stated explicitly below.

Theorem 3.13 (supporting hyperplane theorem [29, Proposition 2.4.1]). Let $\emptyset \neq C \subseteq \mathbb{E}$ be a convex set, and let $\mathbf{y} \notin \text{int}(C)$. Then there exists $\mathbf{0} \neq \mathbf{p} \in \mathbb{E}^*$ such that

$$\langle \mathbf{p}, \mathbf{x} \rangle \leq \langle \mathbf{p}, \mathbf{y} \rangle \text{ for any } \mathbf{x} \in C.$$

Theorem 3.14 (nonemptiness and boundedness of the subdifferential set at interior points of the domain). Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper convex function, and assume that $\tilde{\mathbf{x}} \in \text{int}(\text{dom}(f))$. Then $\partial f(\tilde{\mathbf{x}})$ is nonempty and bounded.

Proof. Recall that the inner product in the product space $\mathbb{E} \times \mathbb{R}$ is defined as (see Section 1.9)

$$\langle (\mathbf{y}_1, \beta_1), (\mathbf{y}_2, \beta_2) \rangle \equiv \langle \mathbf{y}_1, \mathbf{y}_2 \rangle + \beta_1 \beta_2, \quad (\mathbf{y}_1, \beta_1), (\mathbf{y}_2, \beta_2) \in \mathbb{E} \times \mathbb{R}.$$

Since $(\tilde{\mathbf{x}}, f(\tilde{\mathbf{x}}))$ is on the boundary of $\text{epi}(f) \subseteq \mathbb{E} \times \mathbb{R}$, it follows by the supporting hyperplane theorem (Theorem 3.13) that there exists a separating hyperplane between $(\tilde{\mathbf{x}}, f(\tilde{\mathbf{x}}))$ and $\text{epi}(f)$, meaning that there exists a nonzero vector $(\mathbf{p}, -\alpha) \in \mathbb{E}^* \times \mathbb{R}$ for which

$$\langle \mathbf{p}, \tilde{\mathbf{x}} \rangle - \alpha f(\tilde{\mathbf{x}}) \geq \langle \mathbf{p}, \mathbf{x} \rangle - \alpha t \text{ for any } (\mathbf{x}, t) \in \text{epi}(f). \quad (3.8)$$

Note that $\alpha \geq 0$ since $(\tilde{\mathbf{x}}, f(\tilde{\mathbf{x}}) + 1) \in \text{epi}(f)$, and hence plugging $\mathbf{x} = \tilde{\mathbf{x}}$ and $t = f(\tilde{\mathbf{x}}) + 1$ into (3.8) yields

$$\langle \mathbf{p}, \tilde{\mathbf{x}} \rangle - \alpha f(\tilde{\mathbf{x}}) \geq \langle \mathbf{p}, \tilde{\mathbf{x}} \rangle - \alpha(f(\tilde{\mathbf{x}}) + 1),$$

implying that $\alpha \geq 0$. Since $\tilde{\mathbf{x}} \in \text{int}(\text{dom}(f))$, it follows by the local Lipschitz continuity property of convex functions (Theorem 2.21) that there exist $\varepsilon > 0$ and $L > 0$ such that $B_{\|\cdot\|}[\tilde{\mathbf{x}}, \varepsilon] \subseteq \text{dom}(f)$ and

$$|f(\mathbf{x}) - f(\tilde{\mathbf{x}})| \leq L\|\mathbf{x} - \tilde{\mathbf{x}}\| \text{ for any } \mathbf{x} \in B_{\|\cdot\|}[\tilde{\mathbf{x}}, \varepsilon]. \quad (3.9)$$

Since $B_{\|\cdot\|}[\tilde{\mathbf{x}}, \varepsilon] \subseteq \text{dom}(f)$, it follows that $(\mathbf{x}, f(\mathbf{x})) \in \text{epi}(f)$ for any $\mathbf{x} \in B_{\|\cdot\|}[\tilde{\mathbf{x}}, \varepsilon]$. Therefore, plugging $t = f(\mathbf{x})$ into (3.8), yields that

$$\langle \mathbf{p}, \mathbf{x} - \tilde{\mathbf{x}} \rangle \leq \alpha(f(\mathbf{x}) - f(\tilde{\mathbf{x}})) \text{ for any } \mathbf{x} \in B_{\|\cdot\|}[\tilde{\mathbf{x}}, \varepsilon]. \quad (3.10)$$

Combining (3.9) and (3.10), we obtain that for any $\mathbf{x} \in B_{\|\cdot\|}[\tilde{\mathbf{x}}, \varepsilon]$,

$$\langle \mathbf{p}, \mathbf{x} - \tilde{\mathbf{x}} \rangle \leq \alpha(f(\mathbf{x}) - f(\tilde{\mathbf{x}})) \leq \alpha L\|\mathbf{x} - \tilde{\mathbf{x}}\|. \quad (3.11)$$

Take $\mathbf{p}^\dagger \in \mathbb{E}$ satisfying $\langle \mathbf{p}, \mathbf{p}^\dagger \rangle = \|\mathbf{p}\|_*$ and $\|\mathbf{p}^\dagger\| = 1$. Since $\tilde{\mathbf{x}} + \varepsilon \mathbf{p}^\dagger \in B_{\|\cdot\|}[\tilde{\mathbf{x}}, \varepsilon]$, we can plug $\mathbf{x} = \tilde{\mathbf{x}} + \varepsilon \mathbf{p}^\dagger$ into (3.11) and obtain that

$$\varepsilon\|\mathbf{p}\|_* = \varepsilon\langle \mathbf{p}, \mathbf{p}^\dagger \rangle \leq \alpha L\varepsilon\|\mathbf{p}^\dagger\| = \alpha L\varepsilon.$$

Therefore, $\alpha > 0$, since otherwise we would have $\alpha = 0$ and $\mathbf{p} = \mathbf{0}$, which is impossible by the fact that the vector (\mathbf{p}, α) is not the zeros vector. Taking $t = f(\mathbf{x})$ in (3.8) and dividing by α yields

$$f(\mathbf{x}) \geq f(\tilde{\mathbf{x}}) + \langle \mathbf{g}, \mathbf{x} - \tilde{\mathbf{x}} \rangle \text{ for all } \mathbf{x} \in \text{dom}(f), \quad (3.12)$$

where $\mathbf{g} = \mathbf{p}/\alpha$. Thus, $\mathbf{g} \in \partial f(\tilde{\mathbf{x}})$, establishing the nonemptiness of $\partial f(\tilde{\mathbf{x}})$. To show the boundedness of $\partial f(\tilde{\mathbf{x}})$, let $\mathbf{g} \in \partial f(\tilde{\mathbf{x}})$, meaning that (3.12) holds. Take $\mathbf{g}^\dagger \in \mathbb{E}$ for which $\|\mathbf{g}\|_* = \langle \mathbf{g}, \mathbf{g}^\dagger \rangle$ and $\|\mathbf{g}^\dagger\| = 1$. Then plugging $\mathbf{x} = \tilde{\mathbf{x}} + \varepsilon \mathbf{g}^\dagger$ in (3.12) yields

$$\varepsilon\|\mathbf{g}\|_* = \varepsilon\langle \mathbf{g}, \mathbf{g}^\dagger \rangle = \langle \mathbf{g}, \mathbf{x} - \tilde{\mathbf{x}} \rangle \leq f(\mathbf{x}) - f(\tilde{\mathbf{x}}) \stackrel{(3.9)}{\leq} L\|\mathbf{x} - \tilde{\mathbf{x}}\| = L\varepsilon,$$

showing that $\partial f(\tilde{\mathbf{x}}) \subseteq B_{\|\cdot\|_*}[0, L]$, and hence establishing the boundedness of $\partial f(\tilde{\mathbf{x}})$. \square

The result of Theorem 3.14 can be stated as the following inclusion relation:

$$\text{int}(\text{dom}(f)) \subseteq \text{dom}(\partial f).$$

A direct consequence of Theorem 3.14 is that real-valued convex functions (namely, convex functions f with $\text{dom}(f) = \mathbb{E}$) are subdifferentiable at any point.

Corollary 3.15 (subdifferentiability of real-valued convex functions). *Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be a convex function. Then f is subdifferentiable over \mathbb{E} .*

We can extend the boundedness result of Theorem 3.14 and show that subgradients of points in a given compact set contained in the interior of the domain are always bounded.

Theorem 3.16 (boundedness of subgradients over compact sets). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper convex function, and assume that $X \subseteq \text{int}(\text{dom}(f))$ is nonempty and compact. Then $Y = \bigcup_{\mathbf{x} \in X} \partial f(\mathbf{x})$ is nonempty and bounded.*

Proof. The set Y is nonempty, since by Theorem 3.14 $\partial f(\mathbf{x}) \neq \emptyset$ for any $\mathbf{x} \in X$. To prove the boundedness, assume by contradiction that there exists a sequence $\{\mathbf{x}_k\}_{k \geq 1} \subseteq X$ and $\mathbf{g}_k \in \partial f(\mathbf{x}_k)$ such that $\|\mathbf{g}_k\|_* \rightarrow \infty$ as $k \rightarrow \infty$. For any k , let \mathbf{g}_k^\dagger be a vector satisfying $\langle \mathbf{g}_k, \mathbf{g}_k^\dagger \rangle = \|\mathbf{g}_k\|_*$ and $\|\mathbf{g}_k^\dagger\| = 1$. Since X is compact, $(\text{int}(\text{dom}(f)))^c$ (the complement of $\text{int}(\text{dom}(f))$) is closed, and $X \cap (\text{int}(\text{dom}(f)))^c = \emptyset$, it follows that the distance between the two sets is nonempty, meaning in particular that there exists an $\varepsilon > 0$ for which¹⁰

$$\|\mathbf{x} - \mathbf{y}\| \geq \varepsilon \text{ for any } \mathbf{x} \in X, \mathbf{y} \notin \text{int}(\text{dom}(f)). \quad (3.13)$$

The relation $\mathbf{g}_k \in \partial f(\mathbf{x}_k)$ implies in particular that

$$f\left(\mathbf{x}_k + \frac{\varepsilon}{2}\mathbf{g}_k^\dagger\right) - f(\mathbf{x}_k) \geq \frac{\varepsilon}{2}\langle \mathbf{g}_k, \mathbf{g}_k^\dagger \rangle = \frac{\varepsilon}{2}\|\mathbf{g}_k\|_*, \quad (3.14)$$

where we used the fact that by (3.13), $\mathbf{x}_k + \frac{\varepsilon}{2}\mathbf{g}_k^\dagger \in \text{int}(\text{dom}(f))$. We will show that the left-hand side of (3.14) is bounded. Suppose by contradiction that it is not bounded. Then there exist subsequences $\{\mathbf{x}_k\}_{k \in T}$, $\{\mathbf{g}_k^\dagger\}_{k \in T}$ (T being the set of indices of the subsequences) for which

$$f\left(\mathbf{x}_k + \frac{\varepsilon}{2}\mathbf{g}_k^\dagger\right) - f(\mathbf{x}_k) \rightarrow \infty \text{ as } k \xrightarrow{T} \infty. \quad (3.15)$$

Since both $\{\mathbf{x}_k\}_{k \in T}$ and $\{\mathbf{g}_k^\dagger\}_{k \in T}$ are bounded, it follows that there exist convergent subsequences $\{\mathbf{x}_k\}_{k \in S}$, $\{\mathbf{g}_k^\dagger\}_{k \in S}$ ($S \subseteq T$) whose limits will be denoted by $\bar{\mathbf{x}}$ and $\bar{\mathbf{g}}$. Consequently, $\mathbf{x}_k + \frac{\varepsilon}{2}\mathbf{g}_k^\dagger \rightarrow \bar{\mathbf{x}} + \frac{\varepsilon}{2}\bar{\mathbf{g}}$ as $k \xrightarrow{S} \infty$. Since $\mathbf{x}_k, \mathbf{x}_k + \frac{\varepsilon}{2}\mathbf{g}_k^\dagger, \bar{\mathbf{x}} + \frac{\varepsilon}{2}\bar{\mathbf{g}}$ are all¹¹ in $\text{int}(\text{dom}(f))$, it follows by the continuity of f over $\text{int}(\text{dom}(f))$ (Theorem 2.21) that

$$f\left(\mathbf{x}_k + \frac{\varepsilon}{2}\mathbf{g}_k^\dagger\right) - f(\mathbf{x}_k) \rightarrow f\left(\bar{\mathbf{x}} + \frac{\varepsilon}{2}\bar{\mathbf{g}}^\dagger\right) - f(\bar{\mathbf{x}}) \text{ as } k \xrightarrow{S} \infty,$$

which is a contradiction of (3.15). We can thus conclude that the left-hand side of (3.14) is bounded and hence that the right-hand side of (3.14) is also bounded, in contradiction to our assumption that $\|\mathbf{g}_k\|_*$ goes to ∞ as $k \rightarrow \infty$. \square

Subdifferentiability can be guaranteed for points that are not necessarily in the interior of the domain but are in the interior of the domain w.r.t. its affine hull. This is the notion of *relative interior* that we now recall:

$$\text{ri}(S) = \{\mathbf{x} \in \text{aff}(S) : B[\mathbf{x}, \varepsilon] \cap \text{aff}(S) \subseteq S \text{ for some } \varepsilon > 0\}.$$

One key property of the relative interior is that it is nonempty for convex sets.

Theorem 3.17 (nonemptiness of the relative interior [108, Theorem 6.2]). *Let $C \subseteq \mathbb{E}$ be a nonempty convex set. Then $\text{ri}(C)$ is nonempty.*

¹⁰The proof of (3.13) is simple. Suppose by contradiction that there exist sequences $\{\mathbf{x}_k\}_{k \geq 1} \subseteq X$ and $\{\mathbf{y}_k\}_{k \geq 1} \subseteq (\text{int}(\text{dom}(f)))^c$ satisfying $\|\mathbf{x}_k - \mathbf{y}_k\| \rightarrow 0$ as $k \rightarrow \infty$. Since $\{\mathbf{x}_k\}_{k \geq 1}$ is bounded, there exists $M > 0$ for which $\|\mathbf{x}_k\| \leq M$ for all k . Therefore, $\|\mathbf{y}_k\| \leq \|\mathbf{x}_k - \mathbf{y}_k\| + \|\mathbf{x}_k\| \leq \|\mathbf{x}_k - \mathbf{y}_k\| + M$, and we can conclude by the boundedness of $\{\|\mathbf{x}_k - \mathbf{y}_k\|\}_{k \geq 1}$ that $\{\mathbf{y}_k\}_{k \geq 1}$ is bounded. By the Bolzano–Weierstrass theorem, there exist convergent subsequences $\mathbf{x}_{k_j} \rightarrow \bar{\mathbf{x}}, \mathbf{y}_{k_j} \rightarrow \bar{\mathbf{y}}$, and by the closedness of X and $(\text{int}(\text{dom}(f)))^c$, we have that $\bar{\mathbf{x}} \in X, \bar{\mathbf{y}} \in (\text{int}(\text{dom}(f)))^c$. The limit $\|\mathbf{x}_{k_j} - \mathbf{y}_{k_j}\| \rightarrow 0$ as $j \rightarrow \infty$ now brings us to the impossible equality $\bar{\mathbf{x}} = \bar{\mathbf{y}}$.

¹¹The fact that $\bar{\mathbf{x}} + \frac{\varepsilon}{2}\bar{\mathbf{g}} \in \text{int}(\text{dom}(f))$ follows by (3.13) and the relations $\bar{\mathbf{x}} \in X$ and $\|\bar{\mathbf{g}}\| = 1$.

A well-known result is that a proper convex function is always subdifferentiable at relative interior points of its domain. We state this result without a proof.

Theorem 3.18 (nonemptiness of the subdifferential set at relative interior points [108, Theorem 23.4]). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper convex function, and let $\tilde{\mathbf{x}} \in \text{ri}(\text{dom}(f))$. Then $\partial f(\tilde{\mathbf{x}})$ is nonempty.*

The result stated in Theorem 3.18 can be written as the inclusion

$$\text{ri}(\text{dom}(f)) \subseteq \text{dom}(\partial f).$$

Since the relative interior of $\text{dom}(f)$ is always nonempty (Theorem 3.17), we can conclude that there always exists a point in the domain for which the subdifferential set is nonempty.

Corollary 3.19. *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper convex function. Then there exists $\mathbf{x} \in \text{dom}(f)$ for which $\partial f(\mathbf{x})$ is nonempty.*

One instance in which the subdifferential set $\partial f(\mathbf{x})$ is guaranteed to be *unbounded* is when the dimension of the domain of the function is strictly smaller than the dimension of the underlying space \mathbb{E} .

Theorem 3.20 (unboundedness of the subdifferential set when $\dim(\text{dom}(f)) < \dim(\mathbb{E})$). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper convex function. Suppose that $\dim(\text{dom}(f)) < \dim(\mathbb{E})$ and let $\mathbf{x} \in \text{dom}(f)$. If $\partial f(\mathbf{x}) \neq \emptyset$, then $\partial f(\mathbf{x})$ is unbounded.*

Proof. Let $\boldsymbol{\eta}$ be an arbitrary vector in $\partial f(\mathbf{x})$. The set¹² $\mathbb{V} \equiv \text{aff}(\text{dom}(f)) - \{\mathbf{x}\}$ is a vector space. The dimension condition translates to $\dim(\mathbb{V}) < \dim(\mathbb{E})$, which in particular implies that there exists a nonzero vector $\mathbf{v} \in \mathbb{E}$ such that $\langle \mathbf{v}, \mathbf{w} \rangle = 0$ for any $\mathbf{w} \in \mathbb{V}$. Take any $\beta \in \mathbb{R}$. For any $\mathbf{y} \in \text{dom}(f)$,

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \boldsymbol{\eta}, \mathbf{y} - \mathbf{x} \rangle = f(\mathbf{x}) + \langle \boldsymbol{\eta} + \beta\mathbf{v}, \mathbf{y} - \mathbf{x} \rangle,$$

where the equality is due to the fact that $\mathbf{y} - \mathbf{x} \in \mathbb{V}$. We thus obtained that $\boldsymbol{\eta} + \beta\mathbf{v} \in \partial f(\mathbf{x})$ for any β , implying the unboundedness of $\partial f(\mathbf{x})$. \square

3.3 Directional Derivatives

3.3.1 Definition and Basic Properties

Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper function and let $\mathbf{x} \in \text{int}(\text{dom}(f))$. The *directional derivative* of f at \mathbf{x} in a given direction $\mathbf{d} \in \mathbb{E}$, if it exists, is defined by

$$f'(\mathbf{x}; \mathbf{d}) \equiv \lim_{\alpha \rightarrow 0^+} \frac{f(\mathbf{x} + \alpha\mathbf{d}) - f(\mathbf{x})}{\alpha}.$$

A well-known result states that convex functions have directional derivatives in all directions at points in the interior of their domains.

¹²Here the notation “ $-$ ” stands for the Minkowski difference.

Theorem 3.21 ([108, Theorem 23.1]¹³). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper convex function and let $\mathbf{x} \in \text{int}(\text{dom}(f))$. Then for any $\mathbf{d} \in \mathbb{E}$, the directional derivative $f'(\mathbf{x}; \mathbf{d})$ exists.*

It is important to establish some basic properties of the function $\mathbf{d} \mapsto f'(\mathbf{x}; \mathbf{d})$. The next theorem shows that it is convex and homogeneous of degree 1.

Lemma 3.22 (convexity and homogeneity of $\mathbf{d} \mapsto f'(\mathbf{x}; \mathbf{d})$). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper convex function and let $\mathbf{x} \in \text{int}(\text{dom}(f))$. Then*

- (a) *the function $\mathbf{d} \mapsto f'(\mathbf{x}; \mathbf{d})$ is convex;*
- (b) *for any $\lambda \geq 0$ and $\mathbf{d} \in \mathbb{E}$, it holds that $f'(\mathbf{x}; \lambda\mathbf{d}) = \lambda f'(\mathbf{x}; \mathbf{d})$.*

Proof. (a) To show that the function $g(\mathbf{d}) \equiv f'(\mathbf{x}; \mathbf{d})$ is convex, take $\mathbf{d}_1, \mathbf{d}_2 \in \mathbb{E}$ and $\lambda \in [0, 1]$. Then

$$\begin{aligned} & f'(\mathbf{x}; \lambda\mathbf{d}_1 + (1 - \lambda)\mathbf{d}_2) \\ &= \lim_{\alpha \rightarrow 0^+} \frac{f(\mathbf{x} + \alpha[\lambda\mathbf{d}_1 + (1 - \lambda)\mathbf{d}_2]) - f(\mathbf{x})}{\alpha} \\ &= \lim_{\alpha \rightarrow 0^+} \frac{f(\lambda(\mathbf{x} + \alpha\mathbf{d}_1) + (1 - \lambda)(\mathbf{x} + \alpha\mathbf{d}_2)) - f(\mathbf{x})}{\alpha} \\ &\leq \lim_{\alpha \rightarrow 0^+} \frac{\lambda f(\mathbf{x} + \alpha\mathbf{d}_1) + (1 - \lambda)f(\mathbf{x} + \alpha\mathbf{d}_2) - f(\mathbf{x})}{\alpha} \\ &= \lambda \lim_{\alpha \rightarrow 0^+} \frac{f(\mathbf{x} + \alpha\mathbf{d}_1) - f(\mathbf{x})}{\alpha} + (1 - \lambda) \lim_{\alpha \rightarrow 0^+} \frac{f(\mathbf{x} + \alpha\mathbf{d}_2) - f(\mathbf{x})}{\alpha} \\ &= \lambda f'(\mathbf{x}; \mathbf{d}_1) + (1 - \lambda) f'(\mathbf{x}; \mathbf{d}_2), \end{aligned}$$

where the inequality follows from Jensen's inequality for convex functions.

(b) If $\lambda = 0$, the claim is trivial. Take $\lambda > 0$. Then

$$f'(\mathbf{x}; \lambda\mathbf{d}) = \lim_{\alpha \rightarrow 0^+} \frac{f(\mathbf{x} + \alpha\lambda\mathbf{d}) - f(\mathbf{x})}{\alpha} = \lambda \lim_{\alpha \rightarrow 0^+} \frac{f(\mathbf{x} + \alpha\lambda\mathbf{d}) - f(\mathbf{x})}{\alpha\lambda} = \lambda f'(\mathbf{x}; \mathbf{d}).$$

□

The next result highlights a connection between function values and directional derivatives under a convexity assumption.

Lemma 3.23. *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper convex function, and let $\mathbf{x} \in \text{int}(\text{dom}(f))$. Then*

$$f(\mathbf{y}) \geq f(\mathbf{x}) + f'(\mathbf{x}; \mathbf{y} - \mathbf{x}) \text{ for all } \mathbf{y} \in \text{dom}(f).$$

¹³See also [10, Theorem 7.37].

Proof. By the definition of the directional derivative,

$$\begin{aligned} f'(\mathbf{x}; \mathbf{y} - \mathbf{x}) &= \lim_{\alpha \rightarrow 0^+} \frac{f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{\alpha} \\ &= \lim_{\alpha \rightarrow 0^+} \frac{f((1 - \alpha)\mathbf{x} + \alpha\mathbf{y}) - f(\mathbf{x})}{\alpha} \\ &\leq \lim_{\alpha \rightarrow 0^+} \frac{-\alpha f(\mathbf{x}) + \alpha f(\mathbf{y})}{\alpha} \\ &= f(\mathbf{y}) - f(\mathbf{x}), \end{aligned}$$

where the inequality follows by Jensen's inequality. \square

A useful “calculus” rule for directional derivatives shows how to compute the directional derivative of maximum of a finite collection of functions without any convexity assumptions.

Theorem 3.24 (directional derivative of maximum of functions). Suppose that $f(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})\}$, where $f_1, f_2, \dots, f_m : \mathbb{E} \rightarrow (-\infty, \infty]$ are proper functions. Let $\mathbf{x} \in \bigcap_{i=1}^m \text{int}(\text{dom}(f_i))$ and $\mathbf{d} \in \mathbb{E}$. Assume that $f'_i(\mathbf{x}; \mathbf{d})$ exist for any $i \in \{1, 2, \dots, m\}$. Then

$$f'(\mathbf{x}; \mathbf{d}) = \max_{i \in I(\mathbf{x})} f'_i(\mathbf{x}; \mathbf{d}),$$

where $I(\mathbf{x}) = \{i : f_i(\mathbf{x}) = f(\mathbf{x})\}$.

Proof. For any $i \in \{1, 2, \dots, m\}$,

$$\lim_{t \rightarrow 0^+} f_i(\mathbf{x} + t\mathbf{d}) = \lim_{t \rightarrow 0^+} \left[t \frac{f_i(\mathbf{x} + t\mathbf{d}) - f_i(\mathbf{x})}{t} + f_i(\mathbf{x}) \right] = 0 \cdot f'_i(\mathbf{x}; \mathbf{d}) + f_i(\mathbf{x}) = f_i(\mathbf{x}). \quad (3.16)$$

By the definition of $I(\mathbf{x})$, $f_i(\mathbf{x}) > f_j(\mathbf{x})$ for any $i \in I(\mathbf{x}), j \notin I(\mathbf{x})$. Utilizing (3.16), it follows that there exists an $\varepsilon > 0$ such that $f_i(\mathbf{x} + t\mathbf{d}) > f_j(\mathbf{x} + t\mathbf{d})$ for any $i \in I(\mathbf{x}), j \notin I(\mathbf{x})$ and $t \in (0, \varepsilon]$. Therefore, for any $t \in (0, \varepsilon]$,

$$f(\mathbf{x} + t\mathbf{d}) = \max_{i=1,2,\dots,m} f_i(\mathbf{x} + t\mathbf{d}) = \max_{i \in I(\mathbf{x})} f_i(\mathbf{x} + t\mathbf{d}).$$

Consequently, for any $t \in (0, \varepsilon]$,

$$\frac{f(\mathbf{x} + t\mathbf{d}) - f(\mathbf{x})}{t} = \frac{\max_{i \in I(\mathbf{x})} f_i(\mathbf{x} + t\mathbf{d}) - f(\mathbf{x})}{t} = \max_{i \in I(\mathbf{x})} \frac{f_i(\mathbf{x} + t\mathbf{d}) - f_i(\mathbf{x})}{t},$$

where the last equality follows from the fact that $f(\mathbf{x}) = f_i(\mathbf{x})$ for any $i \in I(\mathbf{x})$. Finally, taking $t \rightarrow 0^+$, we obtain that

$$\begin{aligned} f'(\mathbf{x}; \mathbf{d}) &= \lim_{t \rightarrow 0^+} \frac{f(\mathbf{x} + t\mathbf{d}) - f(\mathbf{x})}{t} \\ &= \lim_{t \rightarrow 0^+} \max_{i \in I(\mathbf{x})} \frac{f_i(\mathbf{x} + t\mathbf{d}) - f_i(\mathbf{x})}{t} \\ &= \max_{i \in I(\mathbf{x})} \lim_{t \rightarrow 0^+} \frac{f_i(\mathbf{x} + t\mathbf{d}) - f_i(\mathbf{x})}{t} \\ &= \max_{i \in I(\mathbf{x})} f'_i(\mathbf{x}; \mathbf{d}). \quad \square \end{aligned}$$

Note that an assumption of Theorem 3.24 is that the directional derivatives $f'_i(\mathbf{x}; \mathbf{d})$ exist. This assumption is automatically satisfied when the functions f_1, f_2, \dots, f_m are convex. We can thus write the next corollary that replaces the condition on the existence of the directional derivatives by a convexity assumption.

Corollary 3.25 (directional derivative of maximum of functions—convex case). *Suppose that $f(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})\}$, where $f_1, f_2, \dots, f_m : \mathbb{E} \rightarrow (-\infty, \infty]$ are proper convex functions. Let $\mathbf{x} \in \bigcap_{i=1}^m \text{int}(\text{dom}(f_i))$ and $\mathbf{d} \in \mathbb{E}$. Then*

$$f'(\mathbf{x}; \mathbf{d}) = \max_{i \in I(\mathbf{x})} f'_i(\mathbf{x}; \mathbf{d}),$$

where $I(\mathbf{x}) = \{i : f_i(\mathbf{x}) = f(\mathbf{x})\}$.

3.3.2 The Max Formula

We will now prove an extremely important and useful result, known as the *max formula*, that connects subgradients and directional derivatives.

Theorem 3.26 (max formula). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper convex function. Then for any $\mathbf{x} \in \text{int}(\text{dom}(f))$ and $\mathbf{d} \in \mathbb{E}$,*

$$f'(\mathbf{x}; \mathbf{d}) = \max \{\langle \mathbf{g}, \mathbf{d} \rangle : \mathbf{g} \in \partial f(\mathbf{x})\}. \quad (3.17)$$

Proof. Let $\mathbf{x} \in \text{int}(\text{dom}(f))$ and $\mathbf{d} \in \mathbb{E}$. By the subgradient inequality, we have that for any $\mathbf{g} \in \partial f(\mathbf{x})$,

$$f'(\mathbf{x}; \mathbf{d}) = \lim_{\alpha \rightarrow 0^+} \frac{1}{\alpha} (f(\mathbf{x} + \alpha \mathbf{d}) - f(\mathbf{x})) \geq \lim_{\alpha \rightarrow 0^+} \langle \mathbf{g}, \mathbf{d} \rangle = \langle \mathbf{g}, \mathbf{d} \rangle \quad (3.18)$$

and, consequently,

$$f'(\mathbf{x}; \mathbf{d}) \geq \max \{\langle \mathbf{g}, \mathbf{d} \rangle : \mathbf{g} \in \partial f(\mathbf{x})\}. \quad (3.19)$$

All that is left is to show the reverse direction of the above inequality. For that, define the function $h(\mathbf{w}) \equiv f'(\mathbf{x}; \mathbf{w})$. Then by Lemma 3.22(a), h is a real-valued convex function and is thus subdifferentiable over \mathbb{E} (Corollary 3.15). Let $\tilde{\mathbf{g}} \in \partial h(\mathbf{d})$. Then for any $\mathbf{v} \in \mathbb{E}$ and $\alpha \geq 0$, using the homogeneity of h (Lemma 3.22(b)),

$$\alpha f'(\mathbf{x}; \mathbf{v}) = f'(\mathbf{x}; \alpha \mathbf{v}) = h(\alpha \mathbf{v}) \geq h(\mathbf{d}) + \langle \tilde{\mathbf{g}}, \alpha \mathbf{v} - \mathbf{d} \rangle = f'(\mathbf{x}; \mathbf{d}) + \langle \tilde{\mathbf{g}}, \alpha \mathbf{v} - \mathbf{d} \rangle.$$

Therefore,

$$\alpha(f'(\mathbf{x}; \mathbf{v}) - \langle \tilde{\mathbf{g}}, \mathbf{v} \rangle) \geq f'(\mathbf{x}; \mathbf{d}) - \langle \tilde{\mathbf{g}}, \mathbf{d} \rangle. \quad (3.20)$$

Since the above inequality holds for any $\alpha \geq 0$, it follows that the coefficient of α in the left-hand side expression is nonnegative (otherwise, inequality (3.20) would be violated for large enough α), meaning that

$$f'(\mathbf{x}; \mathbf{v}) \geq \langle \tilde{\mathbf{g}}, \mathbf{v} \rangle.$$

Thus, by Lemma 3.23, for any $\mathbf{y} \in \text{dom}(f)$,

$$f(\mathbf{y}) \geq f(\mathbf{x}) + f'(\mathbf{x}; \mathbf{y} - \mathbf{x}) \geq f(\mathbf{x}) + \langle \tilde{\mathbf{g}}, \mathbf{y} - \mathbf{x} \rangle,$$

showing that $\tilde{\mathbf{g}} \in \partial f(\mathbf{x})$. Taking $\alpha = 0$ in (3.20), we have that $f'(\mathbf{x}; \mathbf{d}) \leq \langle \tilde{\mathbf{g}}, \mathbf{d} \rangle$, so that

$$f'(\mathbf{x}; \mathbf{d}) \leq \langle \tilde{\mathbf{g}}, \mathbf{d} \rangle \leq \max\{\langle \mathbf{g}, \mathbf{d} \rangle : \mathbf{g} \in \partial f(\mathbf{x})\},$$

establishing the desired result. \square

Remark 3.27. The max formula (3.17) can also be rewritten using the support function notation as follows:

$$f'(\mathbf{x}; \mathbf{d}) = \sigma_{\partial f(\mathbf{x})}(\mathbf{d}).$$

3.3.3 Differentiability

Definition 3.28 (differentiability). Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ and $\mathbf{x} \in \text{int}(\text{dom } f)$. The function f is said to be **differentiable** at \mathbf{x} if there exists $\mathbf{g} \in \mathbb{E}^*$ such that

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - \langle \mathbf{g}, \mathbf{h} \rangle}{\|\mathbf{h}\|} = 0. \quad (3.21)$$

The unique¹⁴ vector \mathbf{g} satisfying (3.21) is called the **gradient** of f at \mathbf{x} and is denoted by $\nabla f(\mathbf{x})$.

The above is actually a definition of Fréchet differentiability, which is the one used in this book.

If f is differentiable at $\mathbf{x} \in \text{int}(\text{dom } f)$, then the directional derivative has a simple formula.

Theorem 3.29 (directional derivatives at points of differentiability). Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be proper, and suppose that f is differentiable at $\mathbf{x} \in \text{int}(\text{dom } f)$. Then for any $\mathbf{d} \in \mathbb{E}$

$$f'(\mathbf{x}; \mathbf{d}) = \langle \nabla f(\mathbf{x}), \mathbf{d} \rangle. \quad (3.22)$$

Proof. The formula is obviously correct for $\mathbf{d} = \mathbf{0}$. Suppose that $\mathbf{d} \neq \mathbf{0}$. The differentiability of f implies that

$$\begin{aligned} 0 &= \lim_{\alpha \rightarrow 0^+} \frac{f(\mathbf{x} + \alpha \mathbf{d}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \alpha \mathbf{d} \rangle}{\|\alpha \mathbf{d}\|} \\ &= \lim_{\alpha \rightarrow 0^+} \left[\frac{f(\mathbf{x} + \alpha \mathbf{d}) - f(\mathbf{x})}{\alpha \|\mathbf{d}\|} - \frac{\langle \nabla f(\mathbf{x}), \mathbf{d} \rangle}{\|\mathbf{d}\|} \right]. \end{aligned}$$

Therefore,

$$\begin{aligned} f'(\mathbf{x}; \mathbf{d}) &= \lim_{\alpha \rightarrow 0^+} \frac{f(\mathbf{x} + \alpha \mathbf{d}) - f(\mathbf{x})}{\alpha} \\ &= \lim_{\alpha \rightarrow 0^+} \left\{ \|\mathbf{d}\| \left[\frac{f(\mathbf{x} + \alpha \mathbf{d}) - f(\mathbf{x})}{\alpha \|\mathbf{d}\|} - \frac{\langle \nabla f(\mathbf{x}), \mathbf{d} \rangle}{\|\mathbf{d}\|} \right] + \langle \nabla f(\mathbf{x}), \mathbf{d} \rangle \right\} \\ &= \langle \nabla f(\mathbf{x}), \mathbf{d} \rangle. \quad \square \end{aligned}$$

¹⁴The uniqueness can be shown by the following argument. Suppose that (3.21) is satisfied by both $\mathbf{g} = \mathbf{g}_1$ and $\mathbf{g} = \mathbf{g}_2$. Then by subtracting the two limits, we obtain that $\lim_{\mathbf{h} \rightarrow \mathbf{0}} \langle \mathbf{g}_1 - \mathbf{g}_2, \mathbf{h} \rangle / \|\mathbf{h}\| = 0$, which immediately shows that $\mathbf{g}_1 = \mathbf{g}_2$.

Example 3.30 (directional derivative of maximum of differentiable functions). Consider the function $f(\mathbf{x}) = \max_{i=1,2,\dots,m} f_i(\mathbf{x})$, where $f_i : \mathbb{E} \rightarrow (-\infty, \infty]$ are proper functions. Assume that f_1, f_2, \dots, f_m are differentiable at a given point $\mathbf{x} \in \cap_{i=1}^m \text{int}(\text{dom}(f_i))$. Then by Theorem 3.29, for any $\mathbf{d} \in \mathbb{E}$, $f'_i(\mathbf{x}; \mathbf{d}) = \langle \nabla f_i(\mathbf{x}), \mathbf{d} \rangle$. Therefore, invoking Theorem 3.24,

$$f'(\mathbf{x}; \mathbf{d}) = \max_{i \in I(\mathbf{x})} f'_i(\mathbf{x}; \mathbf{d}) = \max_{i \in I(\mathbf{x})} \langle \nabla f_i(\mathbf{x}), \mathbf{d} \rangle,$$

where $I(\mathbf{x}) = \{i : f_i(\mathbf{x}) = f(\mathbf{x})\}$. ■

Example 3.31 (gradient of $\frac{1}{2}d_C^2(\cdot)$).¹⁵ Suppose that \mathbb{E} is a Euclidean space, and let $C \subseteq \mathbb{E}$ be nonempty closed and convex set. Consider the function $\varphi_C : \mathbb{E} \rightarrow \mathbb{R}$ given by $\varphi_C(\mathbf{x}) \equiv \frac{1}{2}d_C^2(\mathbf{x}) = \frac{1}{2}\|\mathbf{x} - P_C(\mathbf{x})\|^2$, where P_C is the so-called *orthogonal projection* mapping defined by

$$P_C(\mathbf{x}) \equiv \operatorname{argmin}_{\mathbf{y} \in C} \|\mathbf{y} - \mathbf{x}\|.$$

It is well known that P_C is well defined (exists and unique) when the underlying set C is nonempty, closed, and convex.¹⁶ We will show that for any $\mathbf{x} \in \mathbb{E}$,

$$\nabla \varphi_C(\mathbf{x}) = \mathbf{x} - P_C(\mathbf{x}). \quad (3.23)$$

For that, fix $\mathbf{x} \in \mathbb{E}$ and define the function $g_{\mathbf{x}}$ by

$$g_{\mathbf{x}}(\mathbf{d}) \equiv \varphi_C(\mathbf{x} + \mathbf{d}) - \varphi_C(\mathbf{x}) - \langle \mathbf{d}, \mathbf{z}_{\mathbf{x}} \rangle,$$

where $\mathbf{z}_{\mathbf{x}} = \mathbf{x} - P_C(\mathbf{x})$. By the definition of the gradient, to show (3.23), it is enough to establish that

$$\frac{g_{\mathbf{x}}(\mathbf{d})}{\|\mathbf{d}\|} \rightarrow \mathbf{0} \text{ as } \mathbf{d} \rightarrow \mathbf{0}. \quad (3.24)$$

To prove (3.24), note that by the definition of the orthogonal projection, we have for any $\mathbf{d} \in \mathbb{E}$

$$\|\mathbf{x} + \mathbf{d} - P_C(\mathbf{x} + \mathbf{d})\|^2 \leq \|\mathbf{x} + \mathbf{d} - P_C(\mathbf{x})\|^2,$$

which implies that for any $\mathbf{d} \in \mathbb{E}$,

$$\begin{aligned} g_{\mathbf{x}}(\mathbf{d}) &= \frac{1}{2}\|\mathbf{x} + \mathbf{d} - P_C(\mathbf{x} + \mathbf{d})\|^2 - \frac{1}{2}\|\mathbf{x} - P_C(\mathbf{x})\|^2 - \langle \mathbf{d}, \mathbf{z}_{\mathbf{x}} \rangle \\ &\leq \frac{1}{2}\|\mathbf{x} + \mathbf{d} - P_C(\mathbf{x})\|^2 - \frac{1}{2}\|\mathbf{x} - P_C(\mathbf{x})\|^2 - \langle \mathbf{d}, \mathbf{z}_{\mathbf{x}} \rangle \\ &= \frac{1}{2}\|\mathbf{x} - P_C(\mathbf{x})\|^2 + \langle \mathbf{d}, \mathbf{x} - P_C(\mathbf{x}) \rangle + \frac{1}{2}\|\mathbf{d}\|^2 - \frac{1}{2}\|\mathbf{x} - P_C(\mathbf{x})\|^2 - \langle \mathbf{d}, \mathbf{z}_{\mathbf{x}} \rangle \\ &= \frac{1}{2}\|\mathbf{d}\|^2. \end{aligned} \quad (3.25)$$

In particular, we also have

$$g_{\mathbf{x}}(-\mathbf{d}) \leq \frac{1}{2}\|\mathbf{d}\|^2. \quad (3.26)$$

¹⁵The proof in Example 3.31 follows Beck and Teboulle [20, proof of Theorem 4.1].

¹⁶See, for example, [10, Theorem 8.8]. In addition, see Section 6.4.

Since φ_C is convex, so is $g_{\mathbf{x}}$. Therefore, by Jensen's inequality, and noting that $g_{\mathbf{x}}(\mathbf{0}) = 0$,

$$0 = g_{\mathbf{x}}(\mathbf{0}) = g_{\mathbf{x}}\left(\frac{\mathbf{d} + (-\mathbf{d})}{2}\right) \leq \frac{1}{2}(g_{\mathbf{x}}(\mathbf{d}) + g_{\mathbf{x}}(-\mathbf{d})). \quad (3.27)$$

Combining (3.26) and (3.27), we get

$$g_{\mathbf{x}}(\mathbf{d}) \geq -g_{\mathbf{x}}(-\mathbf{d}) \geq -\frac{1}{2}\|\mathbf{d}\|^2. \quad (3.28)$$

Finally, by (3.25) and (3.28), it follows that $|g_{\mathbf{x}}(\mathbf{d})| \leq \frac{1}{2}\|\mathbf{d}\|^2$, from which the limit (3.24) follows and hence also the desired result (3.23). ■

Remark 3.32 (what is the gradient?). We will now illustrate the fact that the gradient depends on the choice of the inner product in the underlying space. Let $\mathbb{E} = \mathbb{R}^n$ be endowed with the dot product. By Theorem 3.29 we know that when f is differentiable at \mathbf{x} , then

$$(\nabla f(\mathbf{x}))_i = \langle \nabla f(\mathbf{x}), \mathbf{e}_i \rangle = f'(\mathbf{x}; \mathbf{e}_i);$$

that is, in this case, the i th component of $\nabla f(\mathbf{x})$ is equal to $\frac{\partial f}{\partial x_i}(\mathbf{x}) = f'(\mathbf{x}; \mathbf{e}_i)$ —the i th partial derivative of f at \mathbf{x} —so that $\nabla f(\mathbf{x}) = D_f(\mathbf{x})$, where

$$D_f(\mathbf{x}) \equiv \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) \\ \frac{\partial f}{\partial x_2}(\mathbf{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}) \end{pmatrix}. \quad (3.29)$$

Note that the definition of the directional derivative does not depend on the choice of the inner product in the underlying space, so we can arbitrarily choose the inner product in the formula (3.22) as the dot product and obtain (recalling that in this case $\nabla f(\mathbf{x}) = D_f(\mathbf{x})$)

$$f'(\mathbf{x}; \mathbf{d}) = D_f(\mathbf{x})^T \mathbf{d} = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\mathbf{x}) d_i. \quad (3.30)$$

Formula (3.30) holds for any choice of inner product in the space. However, $\nabla f(\mathbf{x})$ is not necessarily equal to $D_f(\mathbf{x})$ when the endowed inner product is not the dot product. For example, suppose that the inner product is given by

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{H} \mathbf{y}, \quad (3.31)$$

where \mathbf{H} is a given $n \times n$ positive definite matrix. In this case,

$$\begin{aligned} (\nabla f(\mathbf{x}))_i &= \nabla f(\mathbf{x})^T \mathbf{e}_i = \nabla f(\mathbf{x})^T \mathbf{H} (\mathbf{H}^{-1} \mathbf{e}_i) \\ &= \langle \nabla f(\mathbf{x}), \mathbf{H}^{-1} \mathbf{e}_i \rangle \quad [\text{by (3.31)}] \\ &= f'(\mathbf{x}; \mathbf{H}^{-1} \mathbf{e}_i) \quad [\text{by (3.22)}] \\ &= D_f(\mathbf{x})^T \mathbf{H}^{-1} \mathbf{e}_i. \quad [\text{by (3.30)}] \end{aligned}$$

Hence, we obtain that with respect to the inner product (3.31), the gradient is actually a “scaled”/“weighted” gradient:

$$\nabla f(\mathbf{x}) = \mathbf{H}^{-1} D_f(\mathbf{x}).$$

Now consider the space $\mathbb{E} = \mathbb{R}^{m \times n}$ of all $m \times n$ real-valued matrices with the dot product as the endowed inner product:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \text{Tr}(\mathbf{x}^T \mathbf{y}) \text{ for any } \mathbf{x}, \mathbf{y} \in \mathbb{R}^{m \times n}.$$

Given a proper function $f : \mathbb{R}^{m \times n} \rightarrow (-\infty, \infty]$ and $\mathbf{x} \in \text{int}(\text{dom}(f))$, the gradient, if it exists, is given by $\nabla f(\mathbf{x}) = D_f(\mathbf{x})$, where $D_f(\mathbf{x})$ is the $m \times n$ matrix

$$D_f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_{ij}}(\mathbf{x}) \right)_{i,j}.$$

If the inner product is replaced by

$$\langle \mathbf{x}, \mathbf{y} \rangle = \text{Tr}(\mathbf{x}^T \mathbf{H} \mathbf{y}),$$

where \mathbf{H} is a given $m \times m$ positive definite matrix, then a similar argument to the one given previously shows that

$$\nabla f(\mathbf{x}) = \mathbf{H}^{-1} D_f(\mathbf{x}).$$

We will now show that when a convex function is differentiable at a point in the interior of its domain, then the subdifferential set is the singleton (i.e., a set containing a single vector) consisting of the gradient at the point. The reverse is also correct in the sense that if the subdifferential set is a singleton $\{\mathbf{g}\}$, then the function is differentiable at the given point with \mathbf{g} being its gradient.

Theorem 3.33 (the subdifferential at points of differentiability). Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper convex function, and let $\mathbf{x} \in \text{int}(\text{dom}(f))$. If f is differentiable at \mathbf{x} , then $\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$. Conversely, if f has a unique subgradient at \mathbf{x} , then it is differentiable at \mathbf{x} and $\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$.

Proof. Let $\mathbf{x} \in \text{int}(\text{dom}(f))$ and assume that f is differentiable at \mathbf{x} . Then by Theorem 3.29 it follows that for any $\mathbf{d} \in \mathbb{E}$,

$$f'(\mathbf{x}; \mathbf{d}) = \langle \nabla f(\mathbf{x}), \mathbf{d} \rangle. \quad (3.32)$$

Let $\mathbf{g} \in \partial f(\mathbf{x})$. We will show that $\mathbf{g} = \nabla f(\mathbf{x})$. Combining (3.32) with the max formula (Theorem 3.26) we have

$$\langle \nabla f(\mathbf{x}), \mathbf{d} \rangle = f'(\mathbf{x}; \mathbf{d}) \geq \langle \mathbf{g}, \mathbf{d} \rangle,$$

so that

$$\langle \mathbf{g} - \nabla f(\mathbf{x}), \mathbf{d} \rangle \leq 0.$$

Taking the maximum over all \mathbf{d} satisfying $\|\mathbf{d}\| \leq 1$, we obtain that $\|\mathbf{g} - \nabla f(\mathbf{x})\|_* \leq 0$ and consequently that $\nabla f(\mathbf{x}) = \mathbf{g}$. We have thus shown that the only possible

subgradient in $\partial f(\mathbf{x})$ is $\nabla f(\mathbf{x})$. Combining this with the fact that the subdifferential set is nonempty (Theorem 3.14) yields the desired result $\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$.

For the reverse direction, suppose that f has a unique subgradient \mathbf{g} at $\mathbf{x} \in \text{int}(\text{dom}(f))$. Consider the convex function

$$h(\mathbf{u}) \equiv f(\mathbf{x} + \mathbf{u}) - f(\mathbf{x}) - \langle \mathbf{g}, \mathbf{u} \rangle.$$

We will show that

$$\lim_{\mathbf{u} \rightarrow \mathbf{0}} \frac{h(\mathbf{u})}{\|\mathbf{u}\|} = 0.$$

This will establish (by definition) that $\mathbf{g} = \nabla f(\mathbf{x})$. Obviously, $\mathbf{0}$ is the unique subgradient of h at $\mathbf{0}$ and $\mathbf{0} \in \text{int}(\text{dom}(h))$, and hence by the max formula (Theorem 3.26), for any $\mathbf{d} \in \mathbb{E}$,

$$h'(\mathbf{0}; \mathbf{d}) = \sigma_{\partial h(\mathbf{0})}(\mathbf{d}) = 0.$$

We can thus conclude that for any $\mathbf{d} \in \mathbb{E}$,

$$0 = h'(\mathbf{0}; \mathbf{d}) = \lim_{\alpha \rightarrow 0^+} \frac{h(\alpha \mathbf{d}) - h(\mathbf{0})}{\alpha} = \lim_{\alpha \rightarrow 0^+} \frac{h(\alpha \mathbf{d})}{\alpha}. \quad (3.33)$$

Let $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ be an orthonormal basis of \mathbb{E} . Since $\mathbf{0} \in \text{int}(\text{dom}(h))$, there exists $\varepsilon \in (0, 1)$ such that $\varepsilon \mathbf{v}_i, -\varepsilon \mathbf{v}_i \in \text{dom}(h)$ for any $i = 1, 2, \dots, k$. Therefore, since $\text{dom}(h)$ is convex, the set

$$D = \text{conv}(\{\pm \varepsilon \mathbf{v}_i\}_{i=1}^k)$$

satisfies $D \subseteq \text{dom}(h)$. Let $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$ be the Euclidean norm corresponding to the endowed inner product on \mathbb{E} . Note that $B_{\|\cdot\|}[\mathbf{0}, \gamma] \subseteq D$, where $\gamma = \frac{\varepsilon}{k}$. Indeed, let $\mathbf{w} \in B_{\|\cdot\|}[\mathbf{0}, \gamma]$. Then since $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ is an orthonormal basis of \mathbb{E} , we have

$$\mathbf{w} = \sum_{i=1}^k \langle \mathbf{w}, \mathbf{v}_i \rangle \mathbf{v}_i$$

as well as

$$\|\mathbf{w}\|^2 = \sum_{i=1}^k \langle \mathbf{w}, \mathbf{v}_i \rangle^2. \quad (3.34)$$

Since $\|\mathbf{w}\|^2 \leq \gamma^2$, it follows by (3.34) that $|\langle \mathbf{w}, \mathbf{v}_i \rangle| \leq \gamma$, and hence

$$\mathbf{w} = \sum_{i=1}^k \langle \mathbf{w}, \mathbf{v}_i \rangle \mathbf{v}_i = \sum_{i=1}^k \frac{|\langle \mathbf{w}, \mathbf{v}_i \rangle|}{\varepsilon} [\text{sgn}(\langle \mathbf{w}, \mathbf{v}_i \rangle) \varepsilon \mathbf{v}_i] + \left(1 - \sum_{i=1}^k \frac{|\langle \mathbf{w}, \mathbf{v}_i \rangle|}{\varepsilon}\right) \cdot \mathbf{0} \in D,$$

where the membership in D follows by the fact that $\mathbf{0}, \pm \varepsilon \mathbf{v}_i \in D$ and $\sum_{i=1}^k \frac{|\langle \mathbf{w}, \mathbf{v}_i \rangle|}{\varepsilon} \leq \frac{k\gamma}{\varepsilon} = 1$. We have therefore established the inclusion $B_{\|\cdot\|}[\mathbf{0}, \gamma] \subseteq D$. Denote the $2k$ vectors $\{\pm \varepsilon \mathbf{v}_i\}_{i=1}^k$ by $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{2k}$. Take $\mathbf{0} \neq \mathbf{u} \in B_{\|\cdot\|}[\mathbf{0}, \gamma^2]$. We have that $\gamma \frac{\mathbf{u}}{\|\mathbf{u}\|} \in B_{\|\cdot\|}[\mathbf{0}, \gamma] \subseteq D$, and hence there exists $\boldsymbol{\lambda} \in \Delta_{2k}$ such that

$$\gamma \frac{\mathbf{u}}{\|\mathbf{u}\|} = \sum_{i=1}^{2k} \lambda_i \mathbf{z}_i.$$

Therefore,

$$\begin{aligned}
 \frac{h(\mathbf{u})}{\|\mathbf{u}\|} &= \frac{h\left(\frac{\|\mathbf{u}\|}{\gamma}\gamma\frac{\mathbf{u}}{\|\mathbf{u}\|}\right)}{\|\mathbf{u}\|} = \frac{h\left(\sum_{i=1}^{2k} \lambda_i \frac{\|\mathbf{u}\|}{\gamma} \mathbf{z}_i\right)}{\|\mathbf{u}\|} \\
 &\leq \sum_{i=1}^{2k} \lambda_i \frac{h\left(\|\mathbf{u}\| \frac{\mathbf{z}_i}{\gamma}\right)}{\|\mathbf{u}\|} \\
 &\leq \max_{i=1,2,\dots,2k} \left\{ \frac{h\left(\|\mathbf{u}\| \frac{\mathbf{z}_i}{\gamma}\right)}{\|\mathbf{u}\|} \right\},
 \end{aligned} \tag{3.35}$$

where the first inequality follows by the convexity of h and by the fact that $\|\mathbf{u}\| \frac{\mathbf{z}_i}{\gamma} \in B_{\|\cdot\|}[\mathbf{0}, \gamma] \subseteq D \subseteq \text{dom}(h)$. By (3.33),

$$\lim_{\mathbf{u} \rightarrow \mathbf{0}} \frac{h\left(\|\mathbf{u}\| \frac{\mathbf{z}_i}{\gamma}\right)}{\|\mathbf{u}\|} = \lim_{\|\mathbf{u}\| \rightarrow 0} \frac{h\left(\|\mathbf{u}\| \frac{\mathbf{z}_i}{\gamma}\right)}{\|\mathbf{u}\|} = \lim_{\alpha \rightarrow 0^+} \frac{h\left(\alpha \frac{\mathbf{z}_i}{\gamma}\right)}{\alpha} = 0,$$

which, combined with (3.35), implies that $\frac{h(\mathbf{u})}{\|\mathbf{u}\|} \rightarrow 0$ as $\mathbf{u} \rightarrow \mathbf{0}$, proving the desired result. \square

Example 3.34 (subdifferential of the l_2 -norm). Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}) = \|\mathbf{x}\|_2$. Then the subdifferential set of f at $\mathbf{x} = \mathbf{0}$ was already computed in Example 3.3. When $\mathbf{x} \neq \mathbf{0}$, the function is differentiable with gradient $\frac{\mathbf{x}}{\|\mathbf{x}\|_2}$. Thus, using Theorem 3.33, we can summarize and write the subdifferential set as

$$\partial f(\mathbf{x}) = \begin{cases} \left\{ \frac{\mathbf{x}}{\|\mathbf{x}\|_2} \right\}, & \mathbf{x} \neq \mathbf{0}, \\ B_{\|\cdot\|_2}[\mathbf{0}, 1], & \mathbf{x} = \mathbf{0}. \end{cases}$$

In particular, when considering the case $n = 1$, we obtain that for the one-dimensional function $g(x) = |x|$, we have

$$\partial g(x) = \begin{cases} \{\text{sgn}(x)\}, & x \neq 0, \\ [-1, 1], & x = 0. \end{cases} \quad \blacksquare$$

3.4 Computing Subgradients

This section establishes several useful calculus rules for subgradients and subdifferentials. Some of the results are “weak results” (rules for computing some of the subgradients in the subdifferential set), and some are “strong” (full characterization of the subdifferential set).

3.4.1 Multiplication by a Positive Scalar

Theorem 3.35. Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper function and let $\alpha > 0$. Then for any $\mathbf{x} \in \text{dom}(f)$

$$\partial(\alpha f)(\mathbf{x}) = \alpha \partial f(\mathbf{x}).$$

Proof. We have that $\mathbf{g} \in \partial f(\mathbf{x})$ if and only if

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle \text{ for any } \mathbf{y} \in \text{dom}(f).$$

Multiplying the inequality by $\alpha > 0$, we can conclude that the above inequality holds if and only if

$$\alpha f(\mathbf{y}) \geq \alpha f(\mathbf{x}) + \langle \alpha \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle \text{ for any } \mathbf{y} \in \text{dom}(\alpha f), \quad (3.36)$$

where we used the obvious fact that $\text{dom}(\alpha f) = \text{dom}(f)$. The statement (3.36) is equivalent to the relation $\alpha \mathbf{g} \in \partial(\alpha f)(\mathbf{x})$. \square

3.4.2 Summation

The following result contains both weak and strong results on the subdifferential set of a sum of functions. The weak result is also “weak” in the sense that its proof only requires the definition of the subgradient. The strong result utilizes the max formula.

Theorem 3.36. Let $f_1, f_2 : \mathbb{E} \rightarrow (-\infty, \infty]$ be proper convex functions, and let $\mathbf{x} \in \text{dom}(f_1) \cap \text{dom}(f_2)$.

(a) The following inclusion holds:

$$\partial f_1(\mathbf{x}) + \partial f_2(\mathbf{x}) \subseteq \partial(f_1 + f_2)(\mathbf{x}).$$

(b) If $\mathbf{x} \in \text{int}(\text{dom}(f_1)) \cap \text{int}(\text{dom}(f_2))$, then

$$\partial(f_1 + f_2)(\mathbf{x}) = \partial f_1(\mathbf{x}) + \partial f_2(\mathbf{x}).$$

Proof. (a) Let $\mathbf{g} \in \partial f_1(\mathbf{x}) + \partial f_2(\mathbf{x})$. Then there exist $\mathbf{g}_1 \in \partial f_1(\mathbf{x})$ and $\mathbf{g}_2 \in \partial f_2(\mathbf{x})$ such that $\mathbf{g} = \mathbf{g}_1 + \mathbf{g}_2$. By the definition of \mathbf{g}_1 and \mathbf{g}_2 , it follows that for any $\mathbf{y} \in \text{dom}(f_1) \cap \text{dom}(f_2)$,

$$\begin{aligned} f_1(\mathbf{y}) &\geq f_1(\mathbf{x}) + \langle \mathbf{g}_1, \mathbf{y} - \mathbf{x} \rangle, \\ f_2(\mathbf{y}) &\geq f_2(\mathbf{x}) + \langle \mathbf{g}_2, \mathbf{y} - \mathbf{x} \rangle. \end{aligned}$$

Summing the two inequalities, we obtain that for any $\mathbf{y} \in \text{dom}(f_1) \cap \text{dom}(f_2)$,

$$f_1(\mathbf{y}) + f_2(\mathbf{y}) \geq f_1(\mathbf{x}) + f_2(\mathbf{x}) + \langle \mathbf{g}_1 + \mathbf{g}_2, \mathbf{y} - \mathbf{x} \rangle,$$

that is, $\mathbf{g} = \mathbf{g}_1 + \mathbf{g}_2 \in \partial(f_1 + f_2)(\mathbf{x})$.

(b) Let $\mathbf{d} \in \mathbb{E}$ and define $f \equiv f_1 + f_2$. Then since $\mathbf{x} \in \text{int}(\text{dom}(f)) = \text{int}(\text{dom}(f_1)) \cap \text{int}(\text{dom}(f_2))$, it follows by the max formula (Theorem 3.26) that

$$\sigma_{\partial f(\mathbf{x})}(\mathbf{d}) = \max \{\langle \mathbf{g}, \mathbf{d} \rangle : \mathbf{g} \in \partial f(\mathbf{x})\} = f'(\mathbf{x}; \mathbf{d}).$$

Using the additivity of the directional derivative and the max formula (again), we also obtain

$$\begin{aligned} \sigma_{\partial f(\mathbf{x})}(\mathbf{d}) &= f'(\mathbf{x}; \mathbf{d}) \\ &= f'_1(\mathbf{x}; \mathbf{d}) + f'_2(\mathbf{x}; \mathbf{d}) \\ &= \max \{\langle \mathbf{g}_1, \mathbf{d} \rangle : \mathbf{g}_1 \in \partial f_1(\mathbf{x})\} + \max \{\langle \mathbf{g}_2, \mathbf{d} \rangle : \mathbf{g}_2 \in \partial f_2(\mathbf{x})\} \\ &= \max \{\langle \mathbf{g}_1 + \mathbf{g}_2, \mathbf{d} \rangle : \mathbf{g}_1 \in \partial f_1(\mathbf{x}), \mathbf{g}_2 \in \partial f_2(\mathbf{x})\} \\ &= \sigma_{\partial f_1(\mathbf{x}) + \partial f_2(\mathbf{x})}(\mathbf{d}). \end{aligned}$$

By Theorems 3.9 and 3.14, $\partial f(\mathbf{x})$, $\partial f_1(\mathbf{x})$, and $\partial f_2(\mathbf{x})$ are nonempty compact convex sets, which also implies (simple exercise) that $\partial f_1(\mathbf{x}) + \partial f_2(\mathbf{x})$ is nonempty compact and convex. Finally, invoking Lemma 2.34, it follows that $\partial f(\mathbf{x}) = \partial f_1(\mathbf{x}) + \partial f_2(\mathbf{x})$. \square

Remark 3.37. Note that the proof of part (a) of Theorem 3.36 does not require a convexity assumption on f_1 and f_2 .

A simple induction argument can be used to generalize the last result to an arbitrary number of functions.

Corollary 3.38. Let $f_1, f_2, \dots, f_m : \mathbb{E} \rightarrow (-\infty, \infty]$ be proper convex functions, and let $\mathbf{x} \in \cap_{i=1}^m \text{dom}(f_i)$.

(a) **(weak sum rule of subdifferential calculus)** The following inclusion holds:

$$\sum_{i=1}^m \partial f_i(\mathbf{x}) \subseteq \partial \left(\sum_{i=1}^m f_i \right) (\mathbf{x}).$$

(b) If $\mathbf{x} \in \cap_{i=1}^m \text{int}(\text{dom}(f_i))$, then

$$\partial \left(\sum_{i=1}^m f_i \right) (\mathbf{x}) = \sum_{i=1}^m \partial f_i(\mathbf{x}). \quad (3.37)$$

Another direct consequence is that if f_1, f_2, \dots, f_m are real-valued, meaning that their domain is the entire space \mathbb{E} , then the sum formula (3.37) holds.

Corollary 3.39. Let $f_1, f_2, \dots, f_m : \mathbb{E} \rightarrow \mathbb{R}$ be real-valued convex functions. Then for any $\mathbf{x} \in \mathbb{E}$

$$\partial \left(\sum_{i=1}^m f_i \right) (\mathbf{x}) = \sum_{i=1}^m \partial f_i(\mathbf{x}).$$

A result with a less restrictive assumption than the one in Corollary 3.38(b) states that if the intersection $\cap_{i=1}^m \text{ri}(\text{dom}(f_i))$ is nonempty, then the sum formula is correct at *any* point. We state this result without a proof.

Theorem 3.40 (sum rule of subdifferential calculus [108, Theorem 23.8]). *Let $f_1, f_2, \dots, f_m : \mathbb{E} \rightarrow (-\infty, \infty]$ be proper convex functions, and assume that $\cap_{i=1}^m \text{ri}(\text{dom}(f_i)) \neq \emptyset$. Then for any $\mathbf{x} \in \mathbb{E}$*

$$\partial \left(\sum_{i=1}^m f_i \right) (\mathbf{x}) = \sum_{i=1}^m \partial f_i(\mathbf{x}).$$

Example 3.41 (subdifferential set of the l_1 -norm function—strong result). Consider the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ given by $f(\mathbf{x}) = \|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$. Then $f = \sum_{i=1}^n f_i$, where $f_i(\mathbf{x}) \equiv |x_i|$. We have (see also Example 3.34)

$$\partial f_i(\mathbf{x}) = \begin{cases} \{\text{sgn}(x_i)\mathbf{e}_i\}, & x_i \neq 0, \\ [-\mathbf{e}_i, \mathbf{e}_i], & x_i = 0. \end{cases}$$

Thus, by Corollary 3.39,

$$\partial f(\mathbf{x}) = \sum_{i=1}^n \partial f_i(\mathbf{x}) = \sum_{i \in I_{\neq}(\mathbf{x})} \text{sgn}(x_i)\mathbf{e}_i + \sum_{i \in I_0(\mathbf{x})} [-\mathbf{e}_i, \mathbf{e}_i],$$

where

$$I_{\neq}(\mathbf{x}) = \{i : x_i \neq 0\}, \quad I_0(\mathbf{x}) = \{i : x_i = 0\},$$

and hence

$$\partial f(\mathbf{x}) = \{\mathbf{z} \in \mathbb{R}^n : z_i = \text{sgn}(x_i), i \in I_{\neq}(\mathbf{x}), |z_j| \leq 1, j \in I_0(\mathbf{x})\}.$$

Example 3.42 (a subgradient of the l_1 -norm function—weak result). Using the formula for the subdifferential set described in Example 3.41, we can readily conclude that

$$\text{sgn}(\mathbf{x}) \in \partial f(\mathbf{x}).$$

3.4.3 Affine Transformation

The following theorem states strong and weak results on the subdifferential set of a composition of a convex function with an affine transformation.

Theorem 3.43. *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper convex function and $\mathcal{A} : \mathbb{V} \rightarrow \mathbb{E}$ be a linear transformation. Let $h(\mathbf{x}) = f(\mathcal{A}(\mathbf{x}) + \mathbf{b})$ with $\mathbf{b} \in \mathbb{E}$. Assume that h is proper, meaning that*

$$\text{dom}(h) = \{\mathbf{x} \in \mathbb{V} : \mathcal{A}(\mathbf{x}) + \mathbf{b} \in \text{dom}(f)\} \neq \emptyset.$$

- (a) (**weak affine transformation rule of subdifferential calculus**) For any $\mathbf{x} \in \text{dom}(h)$,

$$\mathcal{A}^T(\partial f(\mathcal{A}(\mathbf{x}) + \mathbf{b})) \subseteq \partial h(\mathbf{x}).$$

- (b) (**affine transformation rule of subdifferential calculus**) If $\mathbf{x} \in \text{int}(\text{dom}(h))$ and $\mathcal{A}(\mathbf{x}) + \mathbf{b} \in \text{int}(\text{dom}(f))$, then

$$\partial h(\mathbf{x}) = \mathcal{A}^T(\partial f(\mathcal{A}(\mathbf{x}) + \mathbf{b})).$$

Proof. (a) Let $\mathbf{x} \in \text{dom}(h)$ and assume that $\mathbf{g} \in \mathcal{A}^T(\partial f(\mathcal{A}(\mathbf{x}) + \mathbf{b}))$. Then there exists $\mathbf{d} \in \mathbb{E}^*$ for which $\mathbf{g} = \mathcal{A}^T(\mathbf{d})$, where

$$\mathbf{d} \in \partial f(\mathcal{A}(\mathbf{x}) + \mathbf{b}). \quad (3.38)$$

For any $\mathbf{y} \in \text{dom}(h)$, we have $\mathcal{A}(\mathbf{y}) + \mathbf{b} \in \text{dom}(f)$, and hence, by (3.38),

$$f(\mathcal{A}(\mathbf{y}) + \mathbf{b}) \geq f(\mathcal{A}(\mathbf{x}) + \mathbf{b}) + \langle \mathbf{d}, \mathcal{A}(\mathbf{y}) + \mathbf{b} - \mathcal{A}(\mathbf{x}) - \mathbf{b} \rangle,$$

and therefore

$$h(\mathbf{y}) \geq h(\mathbf{x}) + \langle \mathcal{A}^T(\mathbf{d}), \mathbf{y} - \mathbf{x} \rangle \text{ for all } \mathbf{y} \in \text{dom}(h).$$

Hence, $\mathbf{g} = \mathcal{A}^T(\mathbf{d}) \in \partial h(\mathbf{x})$, proving that $\mathcal{A}^T(\partial f(\mathcal{A}(\mathbf{x}) + \mathbf{b})) \subseteq \partial h(\mathbf{x})$.

(b) Since $\mathbf{x} \in \text{int}(\text{dom}(h))$, then for any $\mathbf{d} \in \mathbb{V}$, by the max formula (Theorem 3.26),

$$h'(\mathbf{x}; \mathbf{d}) = \sigma_{\partial h(\mathbf{x})}(\mathbf{d}). \quad (3.39)$$

In addition, by the definition of the directional derivative, we have

$$\begin{aligned} h'(\mathbf{x}; \mathbf{d}) &= \lim_{\alpha \rightarrow 0^+} \frac{h(\mathbf{x} + \alpha \mathbf{d}) - h(\mathbf{x})}{\alpha} \\ &= \lim_{\alpha \rightarrow 0^+} \frac{f(\mathcal{A}(\mathbf{x}) + \mathbf{b} + \alpha \mathcal{A}(\mathbf{d})) - f(\mathcal{A}(\mathbf{x}) + \mathbf{b})}{\alpha} \\ &= f'(\mathcal{A}(\mathbf{x}) + \mathbf{b}; \mathcal{A}(\mathbf{d})), \end{aligned}$$

which, combined with (3.39), yields

$$\sigma_{\partial h(\mathbf{x})}(\mathbf{d}) = f'(\mathcal{A}(\mathbf{x}) + \mathbf{b}; \mathcal{A}(\mathbf{d})).$$

Therefore, using the max formula again and the assumption that $\mathcal{A}(\mathbf{x}) + \mathbf{b} \in \text{int}(\text{dom}(f))$, we obtain that

$$\begin{aligned} \sigma_{\partial h(\mathbf{x})}(\mathbf{d}) &= f'(\mathcal{A}(\mathbf{x}) + \mathbf{b}; \mathcal{A}(\mathbf{d})) \\ &= \max_{\mathbf{g}} \{ \langle \mathbf{g}, \mathcal{A}(\mathbf{d}) \rangle : \mathbf{g} \in \partial f(\mathcal{A}(\mathbf{x}) + \mathbf{b}) \} \\ &= \max_{\mathbf{g}} \{ \langle \mathcal{A}^T(\mathbf{g}), \mathbf{d} \rangle : \mathbf{g} \in \partial f(\mathcal{A}(\mathbf{x}) + \mathbf{b}) \} \\ &= \max_{\tilde{\mathbf{g}}} \{ \langle \tilde{\mathbf{g}}, \mathbf{d} \rangle : \tilde{\mathbf{g}} \in \mathcal{A}^T(\partial f(\mathcal{A}(\mathbf{x}) + \mathbf{b})) \} \\ &= \sigma_{\mathcal{A}^T(\partial f(\mathcal{A}(\mathbf{x}) + \mathbf{b}))}(\mathbf{d}). \end{aligned}$$

Since $\mathbf{x} \in \text{int}(\text{dom}(h))$, it follows by Theorems 3.9 and 3.14 that $\partial h(\mathbf{x})$ is nonempty compact and convex. Similarly, since $\mathcal{A}(\mathbf{x}) + \mathbf{b} \in \text{int}(\text{dom}(f))$, the set $\partial f(\mathcal{A}(\mathbf{x}) + \mathbf{b})$ is nonempty, compact, and convex, which implies that $\mathcal{A}^T(\partial f(\mathcal{A}(\mathbf{x}) + \mathbf{b}))$ is also nonempty, compact, and convex. Finally, invoking Lemma 2.34, we obtain that $\partial h(\mathbf{x}) = \mathcal{A}^T(\partial f(\mathcal{A}(\mathbf{x}) + \mathbf{b}))$. \square

Example 3.44 (subdifferential of $\|\mathbf{Ax} + \mathbf{b}\|_1$). Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be the function given by $f(\mathbf{x}) = \|\mathbf{Ax} + \mathbf{b}\|_1$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$. Then $f(\mathbf{x}) = g(\mathbf{Ax} + \mathbf{b})$ with $g : \mathbb{R}^m \rightarrow \mathbb{R}$ given by $g(\mathbf{y}) = \|\mathbf{y}\|_1$. By the affine transformation rule of subdifferential calculus (Theorem 3.43(b)), we have that

$$\partial f(\mathbf{x}) = \mathbf{A}^T \partial g(\mathbf{Ax} + \mathbf{b}). \quad (3.40)$$

Denote the i th row of \mathbf{A} by \mathbf{a}_i^T and define

$$\begin{aligned} I_{\neq}(\mathbf{x}) &= \{i : \mathbf{a}_i^T \mathbf{x} + b_i \neq 0\}, \\ I_0(\mathbf{x}) &= \{i : \mathbf{a}_i^T \mathbf{x} + b_i = 0\}. \end{aligned}$$

In this terminology, by Example 3.41,

$$\partial g(\mathbf{Ax} + \mathbf{b}) = \sum_{i \in I_{\neq}(\mathbf{x})} \text{sgn}(\mathbf{a}_i^T \mathbf{x} + b_i) \mathbf{e}_i + \sum_{i \in I_0(\mathbf{x})} [-\mathbf{e}_i, \mathbf{e}_i].$$

Thus, by (3.40),

$$\begin{aligned} \partial f(\mathbf{x}) &= \mathbf{A}^T \partial g(\mathbf{Ax} + \mathbf{b}) \\ &= \sum_{i \in I_{\neq}(\mathbf{x})} \text{sgn}(\mathbf{a}_i^T \mathbf{x} + b_i) \mathbf{A}^T \mathbf{e}_i + \sum_{i \in I_0(\mathbf{x})} [-\mathbf{A}^T \mathbf{e}_i, \mathbf{A}^T \mathbf{e}_i]. \end{aligned}$$

Using the relation $\mathbf{A}^T \mathbf{e}_i = \mathbf{a}_i$, we finally conclude that

$$\partial f(\mathbf{x}) = \sum_{i \in I_{\neq}(\mathbf{x})} \text{sgn}(\mathbf{a}_i^T \mathbf{x} + b_i) \mathbf{a}_i + \sum_{i \in I_0(\mathbf{x})} [-\mathbf{a}_i, \mathbf{a}_i].$$

The above is a strong result characterizing the entire subdifferential set. A weak result indicating one possible subgradient is

$$\mathbf{A}^T \text{sgn}(\mathbf{Ax} + \mathbf{b}) \in \partial f(\mathbf{x}). \quad \blacksquare$$

Example 3.45 (subdifferential of $\|\mathbf{Ax} + \mathbf{b}\|_2$). Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be the function $f(\mathbf{x}) = \|\mathbf{Ax} + \mathbf{b}\|_2$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$. Then $f(\mathbf{x}) = g(\mathbf{Ax} + \mathbf{b})$ with $g : \mathbb{R}^m \rightarrow \mathbb{R}$ given by $g(\mathbf{y}) = \|\mathbf{y}\|_2$. By Example 3.34,

$$\partial g(\mathbf{Ax} + \mathbf{b}) = \begin{cases} \frac{\mathbf{Ax} + \mathbf{b}}{\|\mathbf{Ax} + \mathbf{b}\|_2}, & \mathbf{Ax} + \mathbf{b} \neq \mathbf{0}, \\ B_{\|\cdot\|_2}[\mathbf{0}, 1], & \mathbf{Ax} + \mathbf{b} = \mathbf{0}. \end{cases}$$

Thus, by the affine transformation rule of subdifferential calculus (Theorem 3.43(b)),

$$\partial f(\mathbf{x}) = \mathbf{A}^T \partial g(\mathbf{Ax} + \mathbf{b}) = \begin{cases} \frac{\mathbf{A}^T(\mathbf{Ax} + \mathbf{b})}{\|\mathbf{Ax} + \mathbf{b}\|_2}, & \mathbf{Ax} + \mathbf{b} \neq \mathbf{0}, \\ \mathbf{A}^T B_{\|\cdot\|_2}[\mathbf{0}, 1], & \mathbf{Ax} + \mathbf{b} = \mathbf{0}. \end{cases}$$

Note that at a vector \mathbf{x} satisfying $\mathbf{Ax} + \mathbf{b} = \mathbf{0}$, the subdifferential set can be explicitly written as

$$\partial f(\mathbf{x}) = \mathbf{A}^T B_{\|\cdot\|_2}[\mathbf{0}, 1] = \{\mathbf{A}^T \mathbf{y} : \|\mathbf{y}\|_2 \leq 1\}.$$

If a weak result is required, then since $\mathbf{0} \in B_{\|\cdot\|_2}[\mathbf{0}, 1]$, we can write $\mathbf{0} = \mathbf{A}^T \mathbf{0} \in \partial f(\mathbf{x})$ for any \mathbf{x} satisfying $\mathbf{Ax} + \mathbf{b} = \mathbf{0}$. ■

3.4.4 Composition

The derivative of a composition of differentiable functions can be computed by using the well-known chain rule. We recall here the classical result on the derivative of the composition of two one-dimensional functions. The result is a small variation of the result from [112, Theorem 5.5].

Theorem 3.46. *Suppose that f is continuous on $[a, b]$ ($a < b$) and that $f'_+(a)$ exists. Let g be a function defined on an open interval I which contains the range of f , and assume that g is differentiable at $f(a)$. Then the function*

$$h(t) = g(f(t)) \quad (a \leq t \leq b)$$

is right differentiable at $t = a$ and

$$h'_+(a) = g'(f(a))f'_+(a).$$

Proof.

$$\begin{aligned} h'_+(a) &= \lim_{t \rightarrow a^+} \frac{g(f(t)) - g(f(a))}{t - a} \\ &= \lim_{t \rightarrow a^+} \frac{g(f(t)) - g(f(a))}{f(t) - f(a)} \cdot \frac{f(t) - f(a)}{t - a} = g'(f(a))f'_+(a). \quad \square \end{aligned}$$

We will now show how the one-dimensional chain rule can be used with the help of the max formula (Theorem 3.26) to show a multidimensional version of the chain rule.

Theorem 3.47 (chain rule of subdifferential calculus). *Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be a convex function and $g : \mathbb{R} \rightarrow \mathbb{R}$ be a nondecreasing convex function. Let $\mathbf{x} \in \mathbb{E}$, and suppose that g is differentiable at the point $f(\mathbf{x})$. Let $h = g \circ f$. Then*

$$\partial h(\mathbf{x}) = g'(f(\mathbf{x}))\partial f(\mathbf{x}).$$

Proof. For any $\mathbf{d} \in \mathbb{E}$, define the following one-dimensional functions:

$$f_{\mathbf{x}, \mathbf{d}}(t) = f(\mathbf{x} + t\mathbf{d}), \quad t \in \mathbb{R},$$

$$h_{\mathbf{x}, \mathbf{d}}(t) = h(\mathbf{x} + t\mathbf{d}), \quad t \in \mathbb{R}.$$

We have the following simple relation between $f_{\mathbf{x}, \mathbf{d}}$ and $h_{\mathbf{x}, \mathbf{d}}$:

$$h_{\mathbf{x}, \mathbf{d}}(t) = h(\mathbf{x} + t\mathbf{d}) = g(f(\mathbf{x} + t\mathbf{d})) = g(f_{\mathbf{x}, \mathbf{d}}(t)), \quad t \in \mathbb{R}. \quad (3.41)$$

The function f is convex by the premise of the theorem, and h is convex since it is a composition of a nondecreasing convex function with a convex function. Therefore, the directional derivatives of f and h exist in every direction (Theorem 3.21), and we have by the definition of the directional derivative that

$$(f_{\mathbf{x}, \mathbf{d}})'_+(0) = f'(\mathbf{x}; \mathbf{d}), \quad (3.42)$$

$$(h_{\mathbf{x}, \mathbf{d}})'_+(0) = h'(\mathbf{x}; \mathbf{d}). \quad (3.43)$$

Since $h_{\mathbf{x}, \mathbf{d}} = g \circ f_{\mathbf{x}, \mathbf{d}}$ (by (3.41)), $f_{\mathbf{x}, \mathbf{d}}$ is right differentiable at 0, and g is differentiable at $f_{\mathbf{x}, \mathbf{d}}(0) = f(\mathbf{x})$, it follows by the chain rule for one-dimensional functions (Theorem 3.46) that

$$(h_{\mathbf{x}, \mathbf{d}})'_+(0) = g'(f(\mathbf{x}))(f_{\mathbf{x}, \mathbf{d}})'_+(0).$$

Plugging (3.42) and (3.43) into the latter equality, we obtain

$$h'(\mathbf{x}; \mathbf{d}) = g'(f(\mathbf{x}))f'(\mathbf{x}; \mathbf{d}).$$

By the max formula (Theorem 3.26), since f and h are convex and $\mathbf{x} \in \text{int}(\text{dom}(f)) = \text{int}(\text{dom}(h)) = \mathbb{E}$,

$$h'(\mathbf{x}; \mathbf{d}) = \sigma_{\partial h(\mathbf{x})}(\mathbf{d}), \quad f'(\mathbf{x}; \mathbf{d}) = \sigma_{\partial f(\mathbf{x})}(\mathbf{d}),$$

and hence

$$\sigma_{\partial h(\mathbf{x})}(\mathbf{d}) = h'(\mathbf{x}; \mathbf{d}) = g'(f(\mathbf{x}))f'(\mathbf{x}; \mathbf{d}) = g'(f(\mathbf{x}))\sigma_{\partial f(\mathbf{x})}(\mathbf{d}) = \sigma_{g'(f(\mathbf{x}))\partial f(\mathbf{x})}(\mathbf{d}),$$

where the last equality is due to Lemma 2.24(c) and the fact that $g'(f(\mathbf{x})) \geq 0$. Finally, by Theorems 3.9 and 3.14 the sets $\partial h(\mathbf{x}), \partial f(\mathbf{x})$ are nonempty, closed, and convex, and thus by Lemma 2.34

$$\partial h(\mathbf{x}) = g'(f(\mathbf{x}))\partial f(\mathbf{x}). \quad \square$$

Example 3.48 (subdifferential of $\|\cdot\|_1^2$). Consider the function $h : \mathbb{R}^n \rightarrow \mathbb{R}$ given by $h(\mathbf{x}) = \|\mathbf{x}\|_1^2$, which can be written as the composition $h = g \circ f$, where $f(\mathbf{x}) = \|\mathbf{x}\|_1$ and $g(t) = [t]_+^2 = \max\{t, 0\}^2$. Both f and g are real-valued convex functions, and g is nondecreasing and differentiable over \mathbb{R} with derivative $g'(t) = 2[t]_+$. Therefore, by the chain rule of subdifferential calculus (Theorem 3.47), for any $\mathbf{x} \in \mathbb{R}^n$,

$$\partial h(\mathbf{x}) = g'(f(\mathbf{x}))\partial f(\mathbf{x}) = 2[\|\mathbf{x}\|_1]_+ \partial f(\mathbf{x}) = 2\|\mathbf{x}\|_1 \partial f(\mathbf{x}).$$

Using the general form of $\partial f(\mathbf{x})$ as derived in Example 3.41, we can write $\partial h(\mathbf{x})$ explicitly as follows:

$$\partial h(\mathbf{x}) = 2\|\mathbf{x}\|_1 \{ \mathbf{z} \in \mathbb{R}^n : z_i = \text{sgn}(x_i), i \in I_{\neq}(\mathbf{x}), |z_j| \leq 1, j \in I_0(\mathbf{x}) \},$$

where $I_{\neq}(\mathbf{x}) = \{i : x_i \neq 0\}$, $I_0(\mathbf{x}) = \{i : x_i = 0\}$.

Plugging $\mathbf{x} = \mathbf{0}$ into the above formula, we obtain that

$$\partial h(\mathbf{0}) = \{\mathbf{0}\}.$$

Since h has a unique subgradient at $\mathbf{x} = \mathbf{0}$, it follows by Theorem 3.33 that h is differentiable at $\mathbf{x} = \mathbf{0}$ and $\nabla h(\mathbf{0}) = \mathbf{0}$. Note that the function is obviously not differentiable over \mathbb{R}^n . For example, when $n = 2$, the nondifferentiability points are $\{(x_1, 0)^T : x_1 \neq 0\} \cup \{(0, x_2)^T : x_2 \neq 0\}$, as illustrated in Figure 3.3. ■

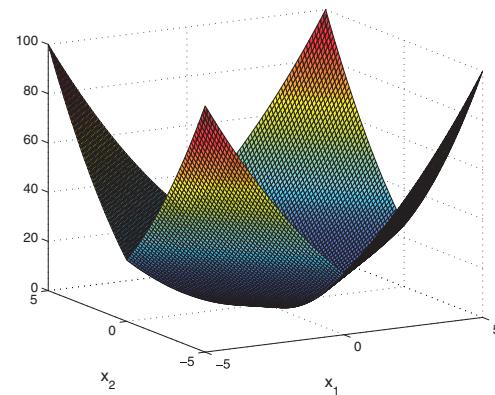


Figure 3.3. Surface plot of the function $f(x_1, x_2) = (|x_1| + |x_2|)^2$.

Example 3.49 (subdifferential of $d_C(\cdot)$). Suppose that \mathbb{E} is a Euclidean space, and let $C \subseteq \mathbb{E}$ be a nonempty closed and convex set (see Example 2.20). The distance function d_C is convex. We will show that

$$\partial d_C(\mathbf{x}) = \begin{cases} \left\{ \frac{\mathbf{x} - P_C(\mathbf{x})}{d_C(\mathbf{x})} \right\}, & \mathbf{x} \notin C, \\ N_C(\mathbf{x}) \cap B[\mathbf{0}, 1], & \mathbf{x} \in C. \end{cases}$$

By Example 3.31, we know that the function $\varphi_C(\mathbf{x}) = \frac{1}{2}d_C^2(\mathbf{x})$ is differentiable and

$$\partial \varphi_C(\mathbf{x}) = \{\mathbf{x} - P_C(\mathbf{x})\} \quad (3.44)$$

for any $\mathbf{x} \in \mathbb{E}$. Note that $\varphi_C = g \circ d_C$, where $g(t) = \frac{1}{2}[t]_+^2$ is a nonincreasing real-valued convex differentiable function. Then by the chain rule of subdifferential calculus (Theorem 3.47),

$$\partial \varphi_C(\mathbf{x}) = g'(d_C(\mathbf{x})) \partial d_C(\mathbf{x}) = [d_C(\mathbf{x})]_+ \partial d_C(\mathbf{x}) = d_C(\mathbf{x}) \partial d_C(\mathbf{x}). \quad (3.45)$$

If $\mathbf{x} \notin C$, then $d_C(\mathbf{x}) \neq 0$, and thus by (3.44) and (3.45),

$$\partial d_C(\mathbf{x}) = \left\{ \frac{\mathbf{x} - P_C(\mathbf{x})}{d_C(\mathbf{x})} \right\} \text{ for any } \mathbf{x} \notin C.$$

Since $\partial d_C(\mathbf{x})$ is a singleton for any $\mathbf{x} \notin C$, it follows in particular, by Theorem 3.33, that d_C is differentiable at points outside C .

Now assume that $\mathbf{x} \in C$. We will show that

$$\partial d_C(\mathbf{x}) = N_C(\mathbf{x}) \cap B[\mathbf{0}, 1].$$

Indeed, if $\mathbf{d} \in \partial d_C(\mathbf{x})$, then

$$d_C(\mathbf{y}) \geq \langle \mathbf{d}, \mathbf{y} - \mathbf{x} \rangle \text{ for any } \mathbf{y} \in \mathbb{E}. \quad (3.46)$$

This means in particular that for any $\mathbf{y} \in C$

$$\langle \mathbf{d}, \mathbf{y} - \mathbf{x} \rangle \leq 0,$$

that is, $\mathbf{d} \in N_C(\mathbf{x})$. In addition, taking $\mathbf{y} = \mathbf{x} + \mathbf{d}$ in (3.46), we get

$$\|\mathbf{d}\|^2 = \langle \mathbf{d}, \mathbf{x} + \mathbf{d} - \mathbf{x} \rangle \leq d_C(\mathbf{x} + \mathbf{d}) \leq \|\mathbf{x} + \mathbf{d} - \mathbf{x}\| = \|\mathbf{d}\|,$$

which readily implies that $\|\mathbf{d}\| \leq 1$. We conclude that $\partial d_C(\mathbf{x}) \subseteq N_C(\mathbf{x}) \cap B[\mathbf{0}, 1]$. To show the reverse direction, take $\mathbf{d} \in N_C(\mathbf{x}) \cap B[\mathbf{0}, 1]$. Then for any $\mathbf{y} \in \mathbb{E}$,

$$\langle \mathbf{d}, \mathbf{y} - \mathbf{x} \rangle = \langle \mathbf{d}, \mathbf{y} - P_C(\mathbf{y}) \rangle + \langle \mathbf{d}, P_C(\mathbf{y}) - \mathbf{x} \rangle. \quad (3.47)$$

Since $\mathbf{d} \in N_C(\mathbf{x})$ and $P_C(\mathbf{y}) \in C$, it follows by the definition of the normal cone that $\langle \mathbf{d}, P_C(\mathbf{y}) - \mathbf{x} \rangle \leq 0$, which, combined with (3.47), the Cauchy–Schwarz inequality, and the assertion that $\|\mathbf{d}\| \leq 1$, implies that for any $\mathbf{y} \in \mathbb{E}$

$$\langle \mathbf{d}, \mathbf{y} - \mathbf{x} \rangle \leq \langle \mathbf{d}, \mathbf{y} - P_C(\mathbf{y}) \rangle \leq \|\mathbf{d}\| \cdot \|\mathbf{y} - P_C(\mathbf{y})\| \leq \|\mathbf{y} - P_C(\mathbf{y})\| = d_C(\mathbf{y}),$$

and hence $\mathbf{d} \in \partial d_C(\mathbf{x})$. ■

3.4.5 Maximization

The following result shows how to compute the subdifferential set of a maximum of a finite collection of convex functions.

Theorem 3.50 (max rule of subdifferential calculus). *Let $f_1, f_2, \dots, f_m : \mathbb{E} \rightarrow (-\infty, \infty]$ be proper convex functions, and define*

$$f(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})\}.$$

Let $\mathbf{x} \in \bigcap_{i=1}^m \text{int}(\text{dom}(f_i))$. Then

$$\partial f(\mathbf{x}) = \text{conv} \left(\bigcup_{i \in I(\mathbf{x})} \partial f_i(\mathbf{x}) \right),$$

where $I(\mathbf{x}) = \{i \in \{1, 2, \dots, m\} : f_i(\mathbf{x}) = f(\mathbf{x})\}$.

Proof. First note that f , as a maximum of convex functions, is convex (see Theorem 2.16(c)) and that by Corollary 3.25 for any $\mathbf{d} \in \mathbb{E}$,

$$f'(\mathbf{x}; \mathbf{d}) = \max_{i \in I(\mathbf{x})} f'_i(\mathbf{x}; \mathbf{d}).$$

For the sake of simplicity of notation, we will assume that $I(\mathbf{x}) = \{1, 2, \dots, k\}$ for some $k \in \{1, 2, \dots, m\}$. Now, using the max formula (Theorem 3.26), we obtain

$$f'(\mathbf{x}; \mathbf{d}) = \max_{i=1, 2, \dots, k} f'_i(\mathbf{x}; \mathbf{d}) = \max_{i=1, 2, \dots, k} \max_{\mathbf{g}_i \in \partial f_i(\mathbf{x})} \langle \mathbf{g}_i, \mathbf{d} \rangle. \quad (3.48)$$

Using the fact that for any $a_1, a_2, \dots, a_k \in \mathbb{R}$ the identity

$$\max\{a_1, a_2, \dots, a_k\} = \max_{\boldsymbol{\lambda} \in \Delta_k} \sum_{i=1}^k \lambda_i a_i$$

holds, we can continue (3.48) and write

$$\begin{aligned} f'(\mathbf{x}; \mathbf{d}) &= \max_{\boldsymbol{\lambda} \in \Delta_k} \left\{ \sum_{i=1}^k \lambda_i \max\{ \langle \mathbf{g}_i, \mathbf{d} \rangle : \mathbf{g}_i \in \partial f_i(\mathbf{x}) \} \right\} \\ &= \max \left\{ \left\langle \sum_{i=1}^k \lambda_i \mathbf{g}_i, \mathbf{d} \right\rangle : \mathbf{g}_i \in \partial f_i(\mathbf{x}), \boldsymbol{\lambda} \in \Delta_k \right\} \\ &= \max \{ \langle \mathbf{g}, \mathbf{d} \rangle : \mathbf{g} \in \text{conv}(\cup_{i=1}^k \partial f_i(\mathbf{x})) \} \\ &= \sigma_A(\mathbf{d}), \end{aligned}$$

where $A = \text{conv}(\cup_{i=1}^k \partial f_i(\mathbf{x}))$. By the max formula (Theorem 3.26), since $\mathbf{x} \in \text{int}(\text{dom}(f))$,

$$f'(\mathbf{x}; \mathbf{d}) = \sigma_{\partial f(\mathbf{x})}(\mathbf{d}),$$

and hence

$$\sigma_A(\mathbf{d}) = \sigma_{\partial f(\mathbf{x})}(\mathbf{d}) \text{ for any } \mathbf{d} \in \mathbb{E}. \quad (3.49)$$

The set $\partial f(\mathbf{x})$ is closed and convex by Theorem 3.9, and since $\mathbf{x} \in \text{int}(\text{dom}(f))$, it is also nonempty and bounded by Theorem 3.14. Similarly, $\partial f_i(\mathbf{x})$, $i = 1, 2, \dots, k$, are nonempty and compact sets, and hence also is $\cup_{i=1}^k \partial f_i(\mathbf{x})$. We can conclude that the set $A = \text{conv}(\cup_{i=1}^k \partial f_i(\mathbf{x}))$, as a convex hull of a nonempty compact set, is also nonempty and compact.¹⁷ In addition, by the definition of the convex hull, A is convex.

To conclude, both A and $\partial f(\mathbf{x})$ are nonempty closed and convex, and thus (3.49) implies by Lemma 2.34 that

$$\partial f(\mathbf{x}) = A,$$

which is the desired result. \square

Example 3.51 (subdifferential of the max function). Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}) = \max\{x_1, x_2, \dots, x_n\}$. Obviously, $f(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_n(\mathbf{x})\}$, where $f_i(\mathbf{x}) = x_i$, and hence $\partial f_i(\mathbf{x}) = \{\mathbf{e}_i\}$ for any $i = 1, 2, \dots, n$. Denote

$$I(\mathbf{x}) = \{i : f(\mathbf{x}) = x_i\}.$$

¹⁷This follows by [10, Proposition 6.31].

Then by the max rule of subdifferential calculus (Theorem 3.50),

$$\partial f(\mathbf{x}) = \text{conv}(\cup_{i \in I(\mathbf{x})} \partial f_i(\mathbf{x})) = \text{conv}(\cup_{i \in I(\mathbf{x})} \{\mathbf{e}_i\}),$$

and hence

$$\boxed{\partial f(\mathbf{x}) = \left\{ \sum_{i \in I(\mathbf{x})} \lambda_i \mathbf{e}_i : \sum_{i \in I(\mathbf{x})} \lambda_i = 1, \lambda_j \geq 0, j \in I(\mathbf{x}) \right\}}.$$

In particular,

$$\partial f(\alpha \mathbf{e}) = \Delta_n \text{ for any } \alpha \in \mathbb{R}. \quad \blacksquare$$

Example 3.52 (subdifferential of the l_∞ -norm). Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}) = \|\mathbf{x}\|_\infty$. There are two options. If $\mathbf{x} = \mathbf{0}$, then by Example 3.3 $\partial f(\mathbf{0})$ is the dual-norm unit ball, and in this case,

$$\partial f(\mathbf{0}) = B_{\|\cdot\|_1}[\mathbf{0}, 1] = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_1 \leq 1\}.$$

Suppose that $\mathbf{x} \neq \mathbf{0}$. Note that $f(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_n(\mathbf{x})\}$ with $f_i(\mathbf{x}) = |x_i|$ and set

$$I(\mathbf{x}) = \{i : |x_i| = \|\mathbf{x}\|_\infty\}.$$

For any $i \in I(\mathbf{x})$ we have $x_i \neq 0$, and hence for any such i , $\partial f_i(\mathbf{x}) = \{\text{sgn}(x_i) \mathbf{e}_i\}$. Thus, by the max rule of subdifferential calculus (Theorem 3.50),

$$\begin{aligned} \partial f(\mathbf{x}) &= \text{conv}(\cup_{i \in I(\mathbf{x})} \partial f_i(\mathbf{x})) \\ &= \text{conv}(\cup_{i \in I(\mathbf{x})} \{\text{sgn}(x_i) \mathbf{e}_i\}) \\ &= \left\{ \sum_{i \in I(\mathbf{x})} \lambda_i \text{sgn}(x_i) \mathbf{e}_i : \sum_{i \in I(\mathbf{x})} \lambda_i = 1, \lambda_j \geq 0, j \in I(\mathbf{x}) \right\}. \end{aligned}$$

To conclude,

$$\boxed{\partial f(\mathbf{x}) = \begin{cases} B_{\|\cdot\|_1}[\mathbf{0}, 1], & \mathbf{x} = \mathbf{0}, \\ \left\{ \sum_{i \in I(\mathbf{x})} \lambda_i \text{sgn}(x_i) \mathbf{e}_i : \sum_{i \in I(\mathbf{x})} \lambda_i = 1, \lambda_j \geq 0, j \in I(\mathbf{x}) \right\}, & \mathbf{x} \neq \mathbf{0}. \end{cases}}$$

■

Example 3.53 (subdifferential of piecewise linear functions). Consider the piecewise linear function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ given by

$$f(\mathbf{x}) = \max_{i=1,2,\dots,m} \{\mathbf{a}_i^T \mathbf{x} + b_i\},$$

where $\mathbf{a}_i \in \mathbb{R}^n, b_i \in \mathbb{R}, i = 1, 2, \dots, m$. The function f can be written as $f(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})\}$, where $f_i(\mathbf{x}) = \mathbf{a}_i^T \mathbf{x} + b_i, i = 1, 2, \dots, m$. Obviously, $\partial f_i(\mathbf{x}) = \{\mathbf{a}_i\}$. Thus, by the max rule of subdifferential calculus (Theorem 3.50),

$$\partial f(\mathbf{x}) = \left\{ \sum_{i \in I(\mathbf{x})} \lambda_i \mathbf{a}_i : \sum_{i \in I(\mathbf{x})} \lambda_i = 1, \lambda_j \geq 0, j \in I(\mathbf{x}) \right\},$$

where $I(\mathbf{x}) = \{i : f(\mathbf{x}) = \mathbf{a}_i^T \mathbf{x} + b_i\}$.

■

Example 3.54 (subdifferential of $\|\mathbf{Ax} + \mathbf{b}\|_\infty$). Consider the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ given by $f(\mathbf{x}) = \|\mathbf{Ax} + \mathbf{b}\|_\infty$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Then $f(\mathbf{x}) = g(\mathbf{Ax} + \mathbf{b})$, where $g : \mathbb{R}^m \rightarrow \mathbb{R}$ is given by $g(\mathbf{y}) = \|\mathbf{y}\|_\infty$. By Example 3.52, we have, for any $\mathbf{y} \in \mathbb{R}^m$,

$$\partial g(\mathbf{y}) = \begin{cases} B_{\|\cdot\|_1}[\mathbf{0}, 1], & \mathbf{y} = \mathbf{0}, \\ \left\{ \sum_{i \in I(\mathbf{y})} \lambda_i \text{sgn}(y_i) \mathbf{e}_i : \sum_{i \in I(\mathbf{y})} \lambda_i = 1, \lambda_j \geq 0, j \in I(\mathbf{y}) \right\}, & \mathbf{y} \neq \mathbf{0}, \end{cases}$$

where

$$I(\mathbf{y}) = \{i \in \{1, 2, \dots, m\} : |y_i| = \|\mathbf{y}\|_\infty\}.$$

We can thus use the affine transformation rule of subdifferential calculus (Theorem 3.43(b)) to conclude that $\partial f(\mathbf{x}) = \mathbf{A}^T \partial g(\mathbf{Ax} + \mathbf{b})$ is given by

$$\partial f(\mathbf{x}) = \begin{cases} \mathbf{A}^T B_{\|\cdot\|_1}[\mathbf{0}, 1], & \mathbf{Ax} + \mathbf{b} = \mathbf{0}, \\ \left\{ \sum_{i \in I_{\mathbf{x}}} \lambda_i \text{sgn}(\mathbf{a}_i^T \mathbf{x} + b_i) \mathbf{a}_i : \sum_{i \in I_{\mathbf{x}}} \lambda_i = 1, \lambda_j \geq 0, j \in I_{\mathbf{x}} \right\}, & \mathbf{Ax} + \mathbf{b} \neq \mathbf{0}, \end{cases}$$

where $\mathbf{a}_1^T, \mathbf{a}_2^T, \dots, \mathbf{a}_m^T$ are the rows of \mathbf{A} and $I_{\mathbf{x}} = I(\mathbf{Ax} + \mathbf{b})$. ■

When the index set is arbitrary (for example, infinite), it is still possible to prove a weak subdifferential calculus rule.

Theorem 3.55 (weak maximum rule of subdifferential calculus). Let I be an arbitrary set, and suppose that any $i \in I$ is associated with a proper convex function $f_i : \mathbb{E} \rightarrow (-\infty, \infty]$. Let

$$f(\mathbf{x}) = \max_{i \in I} f_i(\mathbf{x}). \quad (3.50)$$

Then for any $\mathbf{x} \in \text{dom}(f)$

$$\text{conv}(\cup_{i \in I(\mathbf{x})} \partial f_i(\mathbf{x})) \subseteq \partial f(\mathbf{x}), \quad (3.51)$$

where $I(\mathbf{x}) = \{i \in I : f(\mathbf{x}) = f_i(\mathbf{x})\}$.

Proof. Let $\mathbf{x} \in \text{dom}(f)$. Then for any $\mathbf{z} \in \text{dom}(f)$, $i \in I(\mathbf{x})$ and $\mathbf{g} \in \partial f_i(\mathbf{x})$,

$$f(\mathbf{z}) \geq f_i(\mathbf{z}) \geq f_i(\mathbf{x}) + \langle \mathbf{g}, \mathbf{z} - \mathbf{x} \rangle = f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{z} - \mathbf{x} \rangle, \quad (3.52)$$

where the first inequality follows from (3.50), the second inequality is the subgradient inequality, and the equality is due to the assertion that $i \in I(\mathbf{x})$. Since (3.52) holds for any $\mathbf{z} \in \text{dom}(f)$, we can conclude that $\mathbf{g} \in \partial f(\mathbf{x})$. Thus, $\partial f_i(\mathbf{x}) \subseteq \partial f(\mathbf{x})$. Finally, by the convexity of $\partial f(\mathbf{x})$ (Theorem 3.9), the result (3.51) follows. \square

Example 3.56 (subgradient of $\lambda_{\max}(\mathbf{A}_0 + \sum_{i=1}^m x_i \mathbf{A}_i)$). Let $\mathbf{A}_0, \mathbf{A}_1, \dots, \mathbf{A}_m \in \mathbb{S}^n$. Let $\mathcal{A} : \mathbb{R}^m \rightarrow \mathbb{S}^n$ be the affine transformation given by

$$\mathcal{A}(\mathbf{x}) = \mathbf{A}_0 + \sum_{i=1}^m x_i \mathbf{A}_i \text{ for any } \mathbf{x} \in \mathbb{R}^m.$$

Consider the function $f : \mathbb{R}^m \rightarrow \mathbb{R}$ given by $f(\mathbf{x}) = \lambda_{\max}(\mathcal{A}(\mathbf{x}))$. Since for any $\mathbf{x} \in \mathbb{R}^m$,

$$f(\mathbf{x}) = \max_{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y}\|_2=1} \mathbf{y}^T \mathcal{A}(\mathbf{x}) \mathbf{y}, \quad (3.53)$$

and since the function

$$f_{\mathbf{y}}(\mathbf{x}) \equiv \mathbf{y}^T \mathcal{A}(\mathbf{x}) \mathbf{y} = \mathbf{y}^T \mathbf{A}_0 \mathbf{y} + \sum_{i=1}^m (\mathbf{y}^T \mathbf{A}_i \mathbf{y}) x_i$$

is affine in \mathbf{x} , and in particular convex in \mathbf{x} , it follows by Theorem 2.16(c) that f is convex. For a given $\mathbf{x} \in \mathbb{R}^n$, the maximum in (3.53) is attained at normalized eigenvectors which correspond to the maximum eigenvalue of $\mathcal{A}(\mathbf{x})$. Let $\tilde{\mathbf{y}}$ be such a normalized eigenvector. Then it follows by the weak maximum rule of subdifferential calculus (Theorem 3.55) that a subgradient of the affine function $f_{\tilde{\mathbf{y}}}$ at \mathbf{x} is a subgradient of f at \mathbf{x} , that is,

$$(\tilde{\mathbf{y}}^T \mathbf{A}_1 \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^T \mathbf{A}_2 \tilde{\mathbf{y}}, \dots, \tilde{\mathbf{y}}^T \mathbf{A}_m \tilde{\mathbf{y}})^T \in \partial f(\mathbf{x}), \quad (3.54)$$

where $\tilde{\mathbf{y}}$ is a normalized eigenvector of $\mathcal{A}(\mathbf{x})$ corresponding to the maximum eigenvalue.

It is interesting to note that the result (3.54) can also be deduced by the affine transformation rule of subdifferential calculus (Theorem 3.43(b)). Indeed, let $\tilde{\mathbf{y}}$ be as defined above. The function f can be written as $f(\mathbf{x}) = g(\mathcal{B}(\mathbf{x}) + \mathbf{A}_0)$, where $\mathcal{B}(\mathbf{x}) \equiv \sum_{i=1}^m x_i \mathbf{A}_i$ and $g(\mathbf{X}) \equiv \lambda_{\max}(\mathbf{X})$. Then by the affine transformation rule of subdifferential calculus,

$$\partial f(\mathbf{x}) = \mathcal{B}^T(\partial g(\mathcal{B}(\mathbf{x}) + \mathbf{A}_0)). \quad (3.55)$$

By Example 3.8, we know that $\tilde{\mathbf{y}} \tilde{\mathbf{y}}^T \in \partial g(\mathcal{B}(\mathbf{x}) + \mathbf{A}_0)$, and hence, by (3.55),

$$\mathcal{B}^T(\tilde{\mathbf{y}} \tilde{\mathbf{y}}^T) \in \partial f(\mathbf{x}).$$

The result now follows by noting that

$$\begin{aligned} \mathcal{B}^T(\tilde{\mathbf{y}} \tilde{\mathbf{y}}^T) &= (\text{Tr}(\mathbf{A}_1 \tilde{\mathbf{y}} \tilde{\mathbf{y}}^T), \text{Tr}(\mathbf{A}_2 \tilde{\mathbf{y}} \tilde{\mathbf{y}}^T), \dots, \text{Tr}(\mathbf{A}_m \tilde{\mathbf{y}} \tilde{\mathbf{y}}^T))^T \\ &= (\tilde{\mathbf{y}}^T \mathbf{A}_1 \tilde{\mathbf{y}}, \tilde{\mathbf{y}}^T \mathbf{A}_2 \tilde{\mathbf{y}}, \dots, \tilde{\mathbf{y}}^T \mathbf{A}_m \tilde{\mathbf{y}})^T. \quad \blacksquare \end{aligned}$$

3.5 The Value Function¹⁸

Consider the minimization problem

$$f_{\text{opt}} = \min_{\mathbf{x} \in X} \{f(\mathbf{x}) : g_i(\mathbf{x}) \leq 0, i = 1, 2, \dots, m, \mathbf{Ax} + \mathbf{b} = \mathbf{0}\}, \quad (3.56)$$

where $f, g_1, g_2, \dots, g_m : \mathbb{E} \rightarrow (-\infty, \infty]$ are extended real-valued functions, $X \subseteq \mathbb{E}$ a nonempty set, $\mathbf{A} \in \mathbb{R}^{p \times n}$, and $\mathbf{b} \in \mathbb{R}^p$. We will define the vector-valued function $\mathbf{g} : \mathbb{E} \rightarrow \mathbb{R}^m$ as

$$\mathbf{g}(\mathbf{x}) \equiv (g_1(\mathbf{x}), g_2(\mathbf{x}), \dots, g_m(\mathbf{x}))^T,$$

so that problem (3.56) can be rewritten more compactly as

$$\min_{\mathbf{x} \in X} \{f(\mathbf{x}) : \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \mathbf{Ax} + \mathbf{b} = \mathbf{0}\}.$$

The *value function* associated with problem (3.56) is the function $v : \mathbb{R}^m \times \mathbb{R}^p \rightarrow [-\infty, \infty]$ given by

$$v(\mathbf{u}, \mathbf{t}) = \min_{\mathbf{x} \in X} \{f(\mathbf{x}) : g(\mathbf{x}) \leq \mathbf{u}, \mathbf{Ax} + \mathbf{b} = \mathbf{t}\}. \quad (3.57)$$

The feasible set of the minimization problem in (3.57) will be denoted by

$$C(\mathbf{u}, \mathbf{t}) = \{\mathbf{x} \in X : g(\mathbf{x}) \leq \mathbf{u}, \mathbf{Ax} + \mathbf{b} = \mathbf{t}\},$$

so that the value function can also be rewritten as $v(\mathbf{u}, \mathbf{t}) = \min\{f(\mathbf{x}) : \mathbf{x} \in C(\mathbf{u}, \mathbf{t})\}$. By convention $v(\mathbf{u}, \mathbf{t}) = \infty$ if $C(\mathbf{u}, \mathbf{t})$ is empty. A simple property of the value function $v(\cdot, \cdot)$ is that it is monotone w.r.t. its first argument.

Lemma 3.57 (monotonicity of the value function). *Let $f, g_1, g_2, \dots, g_m : \mathbb{E} \rightarrow (-\infty, \infty]$ be extended real-valued functions, $X \subseteq \mathbb{E}$ a nonempty set, $\mathbf{A} \in \mathbb{R}^{p \times n}$, and $\mathbf{b} \in \mathbb{R}^p$. Let v be the value function given in (3.57). Then*

$$v(\mathbf{u}, \mathbf{t}) \geq v(\mathbf{w}, \mathbf{t}) \text{ for any } \mathbf{u}, \mathbf{w} \in \mathbb{R}^m, \mathbf{t} \in \mathbb{R}^p \text{ satisfying } \mathbf{u} \leq \mathbf{w}.$$

Proof. Follows by the obvious fact that $C(\mathbf{u}, \mathbf{t}) \subseteq C(\mathbf{w}, \mathbf{t})$ whenever $\mathbf{u} \leq \mathbf{w}$. \square

From now on we will also assume in addition that f, g_1, g_2, \dots, g_m , and X are convex. With these additional assumptions, we now show that the value function is convex as long as it is proper.

Lemma 3.58 (convexity of the value function). *Let $f, g_1, g_2, \dots, g_m : \mathbb{E} \rightarrow (-\infty, \infty]$ be convex functions, $X \subseteq \mathbb{E}$ a nonempty convex set, $\mathbf{A} \in \mathbb{R}^{p \times n}$, and $\mathbf{b} \in \mathbb{R}^p$. Suppose that the value function v given in (3.57) is proper. Then v is convex over $\mathbb{R}^m \times \mathbb{R}^p$.*

Proof. Let $(\mathbf{u}, \mathbf{t}), (\mathbf{w}, \mathbf{s}) \in \text{dom}(v)$ and $\lambda \in [0, 1]$. Since v is proper, to prove the convexity, we need to show that

$$v(\lambda\mathbf{u} + (1 - \lambda)\mathbf{w}, \lambda\mathbf{t} + (1 - \lambda)\mathbf{s}) \leq \lambda v(\mathbf{u}, \mathbf{t}) + (1 - \lambda)v(\mathbf{w}, \mathbf{s}).$$

¹⁸Section 3.5, excluding Theorem 3.60, follows Hiriart-Urruty and Lemaréchal [67, Section VII.3.3].

By the definition of the value function v , there exist sequences $\{\mathbf{x}_k\}_{k \geq 1}, \{\mathbf{y}_k\}_{k \geq 1}$ satisfying

$$\mathbf{x}_k \in C(\mathbf{u}, \mathbf{t}), \mathbf{y}_k \in C(\mathbf{w}, \mathbf{s}), f(\mathbf{x}_k) \rightarrow v(\mathbf{u}, \mathbf{t}), f(\mathbf{y}_k) \rightarrow v(\mathbf{w}, \mathbf{s}) \text{ as } k \rightarrow \infty.$$

Since $\mathbf{x}_k \in C(\mathbf{u}, \mathbf{t})$ and $\mathbf{y}_k \in C(\mathbf{w}, \mathbf{s})$, we have $\mathbf{g}(\mathbf{x}_k) \leq \mathbf{u}, \mathbf{g}(\mathbf{y}_k) \leq \mathbf{w}$. Therefore, by the convexity of the components of \mathbf{g} ,

$$\mathbf{g}(\lambda \mathbf{x}_k + (1 - \lambda) \mathbf{y}_k) \leq \lambda \mathbf{g}(\mathbf{x}_k) + (1 - \lambda) \mathbf{g}(\mathbf{y}_k) \leq \lambda \mathbf{u} + (1 - \lambda) \mathbf{w}. \quad (3.58)$$

Moreover,

$$\mathbf{A}(\lambda \mathbf{x}_k + (1 - \lambda) \mathbf{y}_k) + \mathbf{b} = \lambda(\mathbf{A}\mathbf{x}_k + \mathbf{b}) + (1 - \lambda)(\mathbf{A}\mathbf{y}_k + \mathbf{b}) = \lambda \mathbf{s} + (1 - \lambda) \mathbf{t}. \quad (3.59)$$

Combining (3.58) and (3.59), we conclude that

$$\lambda \mathbf{x}_k + (1 - \lambda) \mathbf{y}_k \in C(\lambda \mathbf{u} + (1 - \lambda) \mathbf{w}, \lambda \mathbf{s} + (1 - \lambda) \mathbf{t}). \quad (3.60)$$

By the convexity of f ,

$$f(\lambda \mathbf{x}_k + (1 - \lambda) \mathbf{y}_k) \leq \lambda f(\mathbf{x}_k) + (1 - \lambda) f(\mathbf{y}_k). \quad (3.61)$$

Since $\lambda f(\mathbf{x}_k) + (1 - \lambda) f(\mathbf{y}_k) \rightarrow \lambda v(\mathbf{u}, \mathbf{t}) + (1 - \lambda) v(\mathbf{w}, \mathbf{s})$ as $k \rightarrow \infty$, by (3.61) we have

$$\liminf_{k \rightarrow \infty} f(\lambda \mathbf{x}_k + (1 - \lambda) \mathbf{y}_k) \leq \lambda v(\mathbf{u}, \mathbf{t}) + (1 - \lambda) v(\mathbf{w}, \mathbf{s}). \quad (3.62)$$

Finally, since (3.60) holds, by the definition of v , for all k ,

$$v(\lambda \mathbf{u} + (1 - \lambda) \mathbf{w}, \lambda \mathbf{t} + (1 - \lambda) \mathbf{s}) \leq f(\lambda \mathbf{x}_k + (1 - \lambda) \mathbf{y}_k),$$

and hence

$$v(\lambda \mathbf{u} + (1 - \lambda) \mathbf{w}, \lambda \mathbf{t} + (1 - \lambda) \mathbf{s}) \leq \liminf_{k \rightarrow \infty} f(\lambda \mathbf{x}_k + (1 - \lambda) \mathbf{y}_k),$$

which, combined with (3.62), yields the inequality

$$v(\lambda \mathbf{u} + (1 - \lambda) \mathbf{w}, \lambda \mathbf{t} + (1 - \lambda) \mathbf{s}) \leq \lambda v(\mathbf{u}, \mathbf{t}) + (1 - \lambda) v(\mathbf{w}, \mathbf{s}),$$

establishing the convexity of v . \square

The dual objective function $q : \mathbb{R}_+^m \times \mathbb{R}^q \rightarrow [-\infty, \infty)$ of problem (3.56) is

$$q(\mathbf{y}, \mathbf{z}) = \min_{\mathbf{x} \in X} \{L(\mathbf{x}; \mathbf{y}, \mathbf{z}) = f(\mathbf{x}) + \mathbf{y}^T \mathbf{g}(\mathbf{x}) + \mathbf{z}^T (\mathbf{A}\mathbf{x} + \mathbf{b})\}, \mathbf{y} \in \mathbb{R}_+^m, \mathbf{z} \in \mathbb{R}^p.$$

The dual problem consists of maximizing q on its effective domain given by

$$\text{dom}(-q) = \{(\mathbf{y}, \mathbf{z}) \in \mathbb{R}_+^m \times \mathbb{R}^p : q(\mathbf{y}, \mathbf{z}) > -\infty\}.$$

The dual problem

$$q_{\text{opt}} = \max_{\mathbf{y} \in \mathbb{R}_+^m, \mathbf{z} \in \mathbb{R}^p} \{q(\mathbf{y}, \mathbf{z}) : (\mathbf{y}, \mathbf{z}) \in \text{dom}(-q)\} \quad (3.63)$$

is convex in the sense that it consists of maximizing the concave function q over the convex feasible set $\text{dom}(-q)$. We are now ready to show the main result of this section, which is a relation between the subdifferential set of the value function at the zeros vector and the set of optimal solutions of the dual problem. The result is established under the assumption that strong duality holds, meaning under the assumptions that the optimal values of the primal and dual problems are finite and equal ($f_{\text{opt}} = q_{\text{opt}}$) and the optimal set of the dual problem is nonempty. By the strong duality theorem stated as Theorem A.1 in the appendix, it follows that these assumptions are met if the optimal value of problem (3.56) is finite, and if there exists a feasible solution $\bar{\mathbf{x}}$ satisfying $\mathbf{g}(\bar{\mathbf{x}}) < \mathbf{0}$ and a vector $\hat{\mathbf{x}} \in \text{ri}(X)$ satisfying $\mathbf{A}\hat{\mathbf{x}} + \mathbf{b} = \mathbf{0}$.

Theorem 3.59 (characterization of the subdifferential of the value function at 0). *Let $f, g_1, g_2, \dots, g_m : \mathbb{E} \rightarrow (-\infty, \infty]$ be convex functions, $X \subseteq \mathbb{E}$ a nonempty convex set, $\mathbf{A} \in \mathbb{R}^{p \times n}$, and $\mathbf{b} \in \mathbb{R}^p$. Let v be the value function given by (3.57). Suppose that $f_{\text{opt}} = q_{\text{opt}} \in (-\infty, \infty)$ and that the optimal set of the dual problem is nonempty. Then*

- (a) v is proper and convex;
- (b) (\mathbf{y}, \mathbf{z}) is an optimal solution of problem (3.63) if and only if $-(\mathbf{y}, \mathbf{z}) \in \partial v(\mathbf{0}, \mathbf{0})$.

Proof. Let $(\mathbf{y}, \mathbf{z}) \in \text{dom}(-q)$ be an optimal solution of the dual problem. Then (recalling that $v(\mathbf{0}, \mathbf{0}) = f_{\text{opt}}$)

$$L(\mathbf{x}; \mathbf{y}, \mathbf{z}) \geq \min_{\mathbf{w} \in X} L(\mathbf{w}; \mathbf{y}, \mathbf{z}) = q(\mathbf{y}, \mathbf{z}) = q_{\text{opt}} = f_{\text{opt}} = v(\mathbf{0}, \mathbf{0}) \text{ for all } \mathbf{x} \in X.$$

Therefore, for any $\mathbf{x} \in C(\mathbf{u}, \mathbf{t})$,

$$\begin{aligned} v(\mathbf{0}, \mathbf{0}) - \mathbf{y}^T \mathbf{u} - \mathbf{z}^T \mathbf{t} &\leq L(\mathbf{x}; \mathbf{y}, \mathbf{z}) - \mathbf{y}^T \mathbf{u} - \mathbf{z}^T \mathbf{t} \\ &= f(\mathbf{x}) + \mathbf{y}^T \mathbf{g}(\mathbf{x}) + \mathbf{z}^T (\mathbf{Ax} + \mathbf{b}) - \mathbf{y}^T \mathbf{u} - \mathbf{z}^T \mathbf{t} \\ &= f(\mathbf{x}) + \mathbf{y}^T (\mathbf{g}(\mathbf{x}) - \mathbf{u}) + \mathbf{z}^T (\mathbf{Ax} + \mathbf{b} - \mathbf{t}) \\ &\leq f(\mathbf{x}), \end{aligned}$$

where the last inequality follows from the facts that $\mathbf{g}(\mathbf{x}) \leq \mathbf{u}$, $\mathbf{y} \geq \mathbf{0}$, and $\mathbf{Ax} + \mathbf{b} = \mathbf{t}$. We thus obtained the bound

$$f(\mathbf{x}) \geq v(\mathbf{0}, \mathbf{0}) - \mathbf{y}^T \mathbf{u} - \mathbf{z}^T \mathbf{t} \text{ for any } \mathbf{x} \in C(\mathbf{u}, \mathbf{t}).$$

Minimizing the left-hand side w.r.t. $\mathbf{x} \in C(\mathbf{u}, \mathbf{t})$ yields

$$v(\mathbf{u}, \mathbf{t}) \geq v(\mathbf{0}, \mathbf{0}) - \mathbf{y}^T \mathbf{u} - \mathbf{z}^T \mathbf{t}, \quad (3.64)$$

which is equivalent to saying that $-(\mathbf{y}, \mathbf{z}) \in \partial v(\mathbf{0}, \mathbf{0})$. We actually showed one direction of claim (b), as well as the properness of v since by (3.64), $v(\mathbf{u}, \mathbf{t}) > -\infty$ for any $(\mathbf{u}, \mathbf{t}) \in \mathbb{R}^m \times \mathbb{R}^p$, and by the premise of the theorem, $v(\mathbf{0}, \mathbf{0}) = f_{\text{opt}} < \infty$. Invoking Lemma 3.58, it follows that v is convex, establishing claim (a).

All that is left is to show the reverse direction of claim (b). Assume that $-(\mathbf{y}, \mathbf{z}) \in \partial v(\mathbf{0}, \mathbf{0})$, meaning that

$$v(\mathbf{u}, \mathbf{t}) \geq v(\mathbf{0}, \mathbf{0}) - \mathbf{y}^T \mathbf{u} - \mathbf{z}^T \mathbf{t} \text{ for any } (\mathbf{u}, \mathbf{t}) \in \mathbb{R}^m \times \mathbb{R}^p. \quad (3.65)$$

Let $\mathbf{x} \in X$. Then

$$f(\mathbf{x}) \geq v(\mathbf{g}(\mathbf{x}), \mathbf{Ax} + \mathbf{b}) \stackrel{(3.65)}{\geq} v(\mathbf{0}, \mathbf{0}) - \mathbf{y}^T \mathbf{g}(\mathbf{x}) - \mathbf{z}^T (\mathbf{Ax} + \mathbf{b}).$$

Therefore,

$$v(\mathbf{0}, \mathbf{0}) \leq f(\mathbf{x}) + \mathbf{y}^T \mathbf{g}(\mathbf{x}) + \mathbf{z}^T (\mathbf{Ax} + \mathbf{b}) = L(\mathbf{x}; \mathbf{y}, \mathbf{z}) \text{ for any } \mathbf{x} \in X.$$

Minimizing the right-hand side w.r.t. $\mathbf{x} \in X$ yields

$$v(\mathbf{0}, \mathbf{0}) \leq \min_{\mathbf{x} \in X} L(\mathbf{x}; \mathbf{y}, \mathbf{z}) = q(\mathbf{y}, \mathbf{z}). \quad (3.66)$$

Let $j \in \{1, 2, \dots, m\}$. Plugging $\mathbf{u} = \mathbf{e}_j, \mathbf{t} = \mathbf{0}$ into (3.65), we obtain

$$y_j \geq v(\mathbf{0}, \mathbf{0}) - v(\mathbf{e}_j, \mathbf{0}) \geq 0,$$

where the second inequality follows from the monotonicity property of the value function stated in Lemma 3.57. We thus obtained that $\mathbf{y} \geq \mathbf{0}$, and we can consequently write using (3.66)

$$q_{\text{opt}} = f_{\text{opt}} = v(\mathbf{0}, \mathbf{0}) \leq q(\mathbf{y}, \mathbf{z}) \leq q_{\text{opt}},$$

showing that $q(\mathbf{y}, \mathbf{z}) = q_{\text{opt}}$, meaning that (\mathbf{y}, \mathbf{z}) is an optimal solution of the dual problem. \square

Theorem 3.59 can be used to prove a result concerning an optimality measure of problem (3.56). Consider the following expression:

$$D(\mathbf{x}) \equiv f(\mathbf{x}) - f_{\text{opt}} + \rho_1 \|[\mathbf{g}(\mathbf{x})]_+\|_2 + \rho_2 \|\mathbf{Ax} + \mathbf{b}\|_2.$$

Now assume that

$$D(\tilde{\mathbf{x}}) \leq \delta \quad (3.67)$$

for some $\tilde{\mathbf{x}} \in X$ and a small $\delta > 0$. The question that now arises is whether (3.67) implies that the expressions $f(\tilde{\mathbf{x}}) - f_{\text{opt}}$ as well as $\|[\mathbf{g}(\tilde{\mathbf{x}})]_+\|_2$ and $\|\mathbf{Ax} + \mathbf{b}\|_2$ are also “small” in the sense that they are smaller than a constant times δ . In general, the answer is no. The vector $\tilde{\mathbf{x}}$ is not guaranteed to be feasible, and therefore, in principle, $f(\tilde{\mathbf{x}}) - f_{\text{opt}}$ might be very small (and negative), and $\|[\mathbf{g}(\tilde{\mathbf{x}})]_+\|_2$ and $\|\mathbf{Ax} + \mathbf{b}\|_2$ can be very large. However, we will show in the next theorem that if ρ_1 and ρ_2 are chosen to be large enough, then under the setting of Theorem 3.59, such a conclusion can be drawn.

Theorem 3.60.¹⁹ Let $f, g_1, g_2, \dots, g_m : \mathbb{E} \rightarrow (-\infty, \infty]$ be convex functions, $X \subseteq \mathbb{E}$ a nonempty convex set, $\mathbf{A} \in \mathbb{R}^{p \times n}$, and $\mathbf{b} \in \mathbb{R}^p$. Let f_{opt} and q_{opt} be the optimal values of the primal and dual problems (3.56) and (3.63), respectively. Suppose that $f_{\text{opt}} = q_{\text{opt}} \in (-\infty, \infty)$ and that the optimal set of the dual problem is nonempty. Let $(\mathbf{y}^*, \mathbf{z}^*)$ be an optimal solution of the dual problem. Assume that $\tilde{\mathbf{x}} \in X$ satisfies

$$f(\tilde{\mathbf{x}}) - f_{\text{opt}} + \rho_1 \|[\mathbf{g}(\tilde{\mathbf{x}})]_+\|_2 + \rho_2 \|\mathbf{Ax} + \mathbf{b}\|_2 \leq \delta, \quad (3.68)$$

¹⁹Theorem 3.60 is a slight extension of Lemma 6 from Lan [78].

where $\delta > 0$ and ρ_1, ρ_2 are constants satisfying $\rho_1 \geq 2\|\mathbf{y}^*\|_2, \rho_2 \geq 2\|\mathbf{z}^*\|_2$. Then

$$\begin{aligned} f(\tilde{\mathbf{x}}) - f_{\text{opt}} &\leq \delta, \\ \|[\mathbf{g}(\tilde{\mathbf{x}})]_+\|_2 &\leq \frac{2}{\rho_1} \delta, \\ \|\mathbf{A}\tilde{\mathbf{x}} + \mathbf{b}\|_2 &\leq \frac{2}{\rho_2} \delta. \end{aligned}$$

Proof. The inequality $f(\tilde{\mathbf{x}}) - f_{\text{opt}} \leq \delta$ trivially follows from (3.68) and the fact that the expressions $\rho_1\|[\mathbf{g}(\tilde{\mathbf{x}})]_+\|_2$ and $\rho_2\|\mathbf{A}\tilde{\mathbf{x}} + \mathbf{b}\|_2$ are nonnegative.

Define the function

$$v(\mathbf{u}, \mathbf{t}) = \min_{\mathbf{x} \in X} \{f(\mathbf{x}) : \mathbf{g}(\mathbf{x}) \leq \mathbf{u}, \mathbf{A}\mathbf{x} + \mathbf{b} = \mathbf{t}\}.$$

Since $(\mathbf{y}^*, \mathbf{z}^*)$ is an optimal solution of the dual problem, it follows by Theorem 3.59 that $(-\mathbf{y}^*, -\mathbf{z}^*) \in \partial v(\mathbf{0}, \mathbf{0})$. Therefore, for any $(\mathbf{u}, \mathbf{t}) \in \text{dom}(v)$,

$$v(\mathbf{u}, \mathbf{t}) - v(\mathbf{0}, \mathbf{0}) \geq \langle -\mathbf{y}^*, \mathbf{u} \rangle + \langle -\mathbf{z}^*, \mathbf{t} \rangle. \quad (3.69)$$

Plugging $\mathbf{u} = \tilde{\mathbf{u}} \equiv [\mathbf{g}(\tilde{\mathbf{x}})]_+$ and $\mathbf{t} = \tilde{\mathbf{t}} \equiv \mathbf{A}\tilde{\mathbf{x}} + \mathbf{b}$ into (3.69), while using the inequality $v(\tilde{\mathbf{u}}, \tilde{\mathbf{t}}) \leq f(\tilde{\mathbf{x}})$ and the equality $v(\mathbf{0}, \mathbf{0}) = f_{\text{opt}}$, we obtain

$$\begin{aligned} (\rho_1 - \|\mathbf{y}^*\|_2)\|\tilde{\mathbf{u}}\|_2 + (\rho_2 - \|\mathbf{z}^*\|_2)\|\tilde{\mathbf{t}}\|_2 &= -\|\mathbf{y}^*\|_2\|\tilde{\mathbf{u}}\|_2 - \|\mathbf{z}^*\|_2\|\tilde{\mathbf{t}}\|_2 + \rho_1\|\tilde{\mathbf{u}}\|_2 + \rho_2\|\tilde{\mathbf{t}}\|_2 \\ &\leq \langle -\mathbf{y}^*, \tilde{\mathbf{u}} \rangle + \langle -\mathbf{z}^*, \tilde{\mathbf{t}} \rangle + \rho_1\|\tilde{\mathbf{u}}\|_2 + \rho_2\|\tilde{\mathbf{t}}\|_2 \\ &\leq v(\tilde{\mathbf{u}}, \tilde{\mathbf{t}}) - v(\mathbf{0}, \mathbf{0}) + \rho_1\|\tilde{\mathbf{u}}\|_2 + \rho_2\|\tilde{\mathbf{t}}\|_2 \\ &\leq f(\tilde{\mathbf{x}}) - f_{\text{opt}} + \rho_1\|\tilde{\mathbf{u}}\|_2 + \rho_2\|\tilde{\mathbf{t}}\|_2 \\ &\leq \delta. \end{aligned}$$

Therefore, since both expressions $(\rho_1 - \|\mathbf{y}^*\|_2)\|\tilde{\mathbf{u}}\|_2$ and $(\rho_2 - \|\mathbf{z}^*\|_2)\|\tilde{\mathbf{t}}\|_2$ are non-negative, it follows that

$$\begin{aligned} (\rho_1 - \|\mathbf{y}^*\|_2)\|\tilde{\mathbf{u}}\|_2 &\leq \delta, \\ (\rho_2 - \|\mathbf{z}^*\|_2)\|\tilde{\mathbf{t}}\|_2 &\leq \delta, \end{aligned}$$

and hence, using the assumptions that $\rho_1 \geq 2\|\mathbf{y}^*\|_2$ and $\rho_2 \geq 2\|\mathbf{z}^*\|_2$,

$$\begin{aligned} \|[\mathbf{g}(\tilde{\mathbf{x}})]_+\|_2 &= \|\tilde{\mathbf{u}}\|_2 \leq \frac{\delta}{\rho_1 - \|\mathbf{y}^*\|_2} \leq \frac{2}{\rho_1} \delta, \\ \|\mathbf{A}\tilde{\mathbf{x}} + \mathbf{b}\|_2 &= \|\tilde{\mathbf{t}}\|_2 \leq \frac{\delta}{\rho_2 - \|\mathbf{z}^*\|_2} \leq \frac{2}{\rho_2} \delta. \quad \square \end{aligned}$$

3.6 Lipschitz Continuity and Boundedness of Subgradients

This section considers an important relation between Lipschitz continuity of a convex function and boundedness of its subgradients.

Theorem 3.61 (Lipschitz continuity and boundedness of the subdifferential sets). Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper and convex function. Suppose that $X \subseteq \text{int}(\text{dom}f)$. Consider the following two claims:

- (i) $|f(\mathbf{x}) - f(\mathbf{y})| \leq L\|\mathbf{x} - \mathbf{y}\|$ for any $\mathbf{x}, \mathbf{y} \in X$.
- (ii) $\|\mathbf{g}\|_* \leq L$ for any $\mathbf{g} \in \partial f(\mathbf{x}), \mathbf{x} \in X$.

Then

- (a) the implication (ii) \Rightarrow (i) holds;
- (b) if X is open, then (i) holds if and only if (ii) holds.

Proof. (a) Suppose that (ii) is satisfied and let $\mathbf{x}, \mathbf{y} \in X$. Let $\mathbf{g}_x \in \partial f(\mathbf{x})$ and $\mathbf{g}_y \in \partial f(\mathbf{y})$. The existence of these subgradients is guaranteed by Theorem 3.14. Then by the definitions of $\mathbf{g}_x, \mathbf{g}_y$ and the generalized Cauchy–Schwarz inequality (Lemma 1.4),

$$\begin{aligned} f(\mathbf{x}) - f(\mathbf{y}) &\leq \langle \mathbf{g}_x, \mathbf{x} - \mathbf{y} \rangle \leq \|\mathbf{g}_x\|_* \|\mathbf{x} - \mathbf{y}\| \leq L\|\mathbf{x} - \mathbf{y}\|, \\ f(\mathbf{y}) - f(\mathbf{x}) &\leq \langle \mathbf{g}_y, \mathbf{y} - \mathbf{x} \rangle \leq \|\mathbf{g}_y\|_* \|\mathbf{x} - \mathbf{y}\| \leq L\|\mathbf{x} - \mathbf{y}\|, \end{aligned}$$

showing the validity of (i).

(b) The implication (ii) \Rightarrow (i) was already shown. Now assume that (i) is satisfied. Take $\mathbf{x} \in X$ and $\mathbf{g} \in \partial f(\mathbf{x})$. We will show that $\|\mathbf{g}\|_* \leq L$. Define $\mathbf{g}^\dagger \in \mathbb{E}$ as a vector that satisfies $\|\mathbf{g}^\dagger\| = 1, \langle \mathbf{g}^\dagger, \mathbf{g} \rangle = \|\mathbf{g}\|_*$ (the existence of such a vector is warranted by the definition of the dual norm). Take $\varepsilon > 0$ small enough such that $\mathbf{x} + \varepsilon \mathbf{g}^\dagger \in X$. By the subgradient inequality we have

$$f(\mathbf{x} + \varepsilon \mathbf{g}^\dagger) \geq f(\mathbf{x}) + \langle \mathbf{g}, \varepsilon \mathbf{g}^\dagger \rangle.$$

Thus,

$$\varepsilon \|\mathbf{g}\|_* = \langle \mathbf{g}, \varepsilon \mathbf{g}^\dagger \rangle \leq f(\mathbf{x} + \varepsilon \mathbf{g}^\dagger) - f(\mathbf{x}) \leq L\|\mathbf{x} + \varepsilon \mathbf{g}^\dagger - \mathbf{x}\| = L\varepsilon,$$

showing that $\|\mathbf{g}\|_* \leq L$. \square

Recall that by Theorem 3.16, the subgradients of a given convex function f are bounded over compact sets contained in $\text{int}(\text{dom}(f))$. Combining this with Theorem 3.61, we can conclude that convex functions are always Lipschitz continuous over compact sets contained in the interior of their domain.

Corollary 3.62 (Lipschitz continuity of convex functions over compact domains). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper and convex function. Suppose that $X \subseteq \text{int}(\text{dom}(f))$ is compact. Then there exists $L > 0$ for which*

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq L\|\mathbf{x} - \mathbf{y}\| \text{ for any } \mathbf{x}, \mathbf{y} \in X.$$

3.7 Optimality Conditions

3.7.1 Fermat's Optimality Condition

Subdifferential sets are extremely useful in characterizing minima points. Perhaps the most basic optimality condition states that a point is a global minimum of a

proper extended real-valued convex function if and only if $\mathbf{0}$ belongs to the subdifferential set at the point. In a sense, this is a generalization of Fermat's optimality condition at points of differentiability (" $\nabla f(\mathbf{x}^*) = \mathbf{0}$ "). We will refer to this condition as *Fermat's optimality condition*.

Theorem 3.63 (Fermat's optimality condition). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper convex function. Then*

$$\mathbf{x}^* \in \operatorname{argmin}\{f(\mathbf{x}) : \mathbf{x} \in \mathbb{E}\} \quad (3.70)$$

if and only if $\mathbf{0} \in \partial f(\mathbf{x}^)$.*

Proof. Follows by the definition of the subgradient. Indeed, (3.70) is satisfied if and only if

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \langle \mathbf{0}, \mathbf{x} - \mathbf{x}^* \rangle \text{ for any } \mathbf{x} \in \operatorname{dom}(f),$$

which is the same as the inclusion $\mathbf{0} \in \partial f(\mathbf{x}^*)$. \square

Example 3.64 (minimizing piecewise linear functions). Consider the problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \left[f(\mathbf{x}) \equiv \max_{i=1,2,\dots,m} \{ \mathbf{a}_i^T \mathbf{x} + b_i \} \right], \quad (3.71)$$

where $\mathbf{a}_i \in \mathbb{R}^n, b_i \in \mathbb{R}, i = 1, 2, \dots, m$. Denote

$$I(\mathbf{x}) = \{i : f(\mathbf{x}) = \mathbf{a}_i^T \mathbf{x} + b_i\}.$$

Then, by Example 3.53,

$$\partial f(\mathbf{x}) = \left\{ \sum_{i \in I(\mathbf{x})} \lambda_i \mathbf{a}_i : \sum_{i \in I(\mathbf{x})} \lambda_i = 1, \lambda_j \geq 0, j \in I(\mathbf{x}) \right\}.$$

Therefore, since by Fermat's optimality condition \mathbf{x}^* is an optimal solution of (3.71) if and only if $\mathbf{0} \in \partial f(\mathbf{x}^*)$, it follows that \mathbf{x}^* is an optimal solution of problem (3.71) if and only if there exists $\boldsymbol{\lambda} \in \Delta_m$ such that

$$\mathbf{0} = \sum_{i=1}^m \lambda_i \mathbf{a}_i \text{ and } \lambda_j = 0 \text{ for any } j \notin I(\mathbf{x}^*). \quad (3.72)$$

We can rewrite this condition in a more compact way by denoting $\mathbf{A} \in \mathbb{R}^{m \times n}$ to be the matrix whose rows are $\mathbf{a}_1^T, \mathbf{a}_2^T, \dots, \mathbf{a}_m^T$. Then the optimality condition (3.72) can be written as

$$\exists \boldsymbol{\lambda} \in \Delta_m \text{ s.t. } \mathbf{A}^T \boldsymbol{\lambda} = \mathbf{0} \text{ and } \lambda_j (\mathbf{a}_j^T \mathbf{x}^* + b_j - f(\mathbf{x}^*)) = 0, j = 1, 2, \dots, m. \quad \blacksquare$$

Example 3.65 (medians). Suppose that we are given n different²⁰ and ordered numbers $a_1 < a_2 < \dots < a_n$. Denote $A = \{a_1, a_2, \dots, a_n\} \subseteq \mathbb{R}$. The *median* of A is a number β that satisfies

$$\#\{i : a_i \leq \beta\} \geq \frac{n}{2} \text{ and } \#\{i : a_i \geq \beta\} \geq \frac{n}{2}.$$

²⁰The assumption that these are different and ordered numbers is not essential and is made for the sake of simplicity of exposition.

That is, a median of A is a number that satisfies that at least half of the numbers in A are smaller or equal to it and that at least half are larger or equal. It is not difficult to see that if A has an odd number of elements, then the median is the middlemost number. For example, the median of $\{5, 8, 11, 60, 100\}$ is 11. If the number of elements in A is even, then there is no unique median. The set of medians comprises all numbers between the two middle values. For example, if $A = \{5, 8, 11, 20, 60, 100\}$, then the set of medians of A is the interval $[11, 20]$. In general,

$$\text{median}(A) = \begin{cases} a_{\frac{n+1}{2}}, & n \text{ odd}, \\ [a_{\frac{n}{2}}, a_{\frac{n}{2}+1}], & n \text{ even}. \end{cases}$$

From an optimization perspective, the set of possible medians is the optimal solution set of the problem

$$\min \left\{ f(x) \equiv \sum_{i=1}^n |x - a_i| \right\}. \quad (3.73)$$

To show this, denote $f_i(x) \equiv |x - a_i|$, so that $f(x) = f_1(x) + f_2(x) + \cdots + f_n(x)$, and note that for any $i \in \{1, 2, \dots, n\}$,

$$\partial f_i(x) = \begin{cases} 1, & x > a_i, \\ -1, & x < a_i, \\ [-1, 1], & x = a_i. \end{cases}$$

By the sum rule of subdifferential calculus (Theorem 3.40),

$$\begin{aligned} \partial f(x) &= \partial f_1(x) + \partial f_2(x) + \cdots + \partial f_n(x) \\ &= \begin{cases} \#\{i : a_i < x\} - \#\{i : a_i > x\}, & x \notin A, \\ \#\{i : a_i < x\} - \#\{i : a_i > x\} + [-1, 1], & x \in A. \end{cases} \end{aligned}$$

We can further elaborate and write

$$\partial f(x) = \begin{cases} 2i - n, & x \in (a_i, a_{i+1}), \\ 2i - 1 - n + [-1, 1], & x = a_i, \\ -n, & x < a_1, \\ n, & x > a_n. \end{cases} \quad (3.74)$$

Let $i \in \{1, 2, \dots, n\}$. By (3.74), $0 \in \partial f(a_i)$ if and only if $|2i - 1 - n| \leq 1$, which is equivalent to $\frac{n}{2} \leq i \leq \frac{n}{2} + 1$ and $0 \in \partial f(x)$ for some $x \in (a_i, a_{i+1})$ if and only if $i = \frac{n}{2}$. We can thus conclude that if n is odd, then the only optimal point is $a_{\frac{n+1}{2}}$, and when n is even, the optimal set is the interval $[a_{\frac{n}{2}}, a_{\frac{n}{2}+1}]$, establishing the fact that the optimal set of (3.73) is exactly the set of medians. ■

Example 3.66 (Fermat–Weber problem). Given m different points in \mathbb{R}^d , $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m\}$, and m positive weights $\omega_1, \omega_2, \dots, \omega_m$, the *Fermat–Weber problem* is given by

$$(FW) \quad \min_{\mathbf{x} \in \mathbb{R}^d} \left\{ f(\mathbf{x}) \equiv \sum_{i=1}^m \omega_i \|\mathbf{x} - \mathbf{a}_i\|_2 \right\}.$$

The Fermat–Weber problem is actually a weighted multidimensional version of the median problem (3.73) discussed in the previous example and is therefore also referred to in the literature as the *geometric median* problem. Let us write explicitly the optimality conditions for problem (FW). Denote $f_i(\mathbf{x}) = \omega_i g_i(\mathbf{x})$, where $g_i(\mathbf{x}) \equiv \|\mathbf{x} - \mathbf{a}_i\|_2$. Then for any $i \in \{1, 2, \dots, m\}$

$$\partial f_i(\mathbf{x}) = \begin{cases} \omega_i \frac{\mathbf{x} - \mathbf{a}_i}{\|\mathbf{x} - \mathbf{a}_i\|_2}, & \mathbf{x} \neq \mathbf{a}_i, \\ B_{\|\cdot\|_2}[\mathbf{0}, \omega_i], & \mathbf{x} = \mathbf{a}_i, \end{cases}$$

where here we used Theorems 3.35 (“multiplication by a positive scalar”), the affine transformation rule of subdifferential calculus (Theorem 3.43(b)), and Example 3.34, in which the subdifferential set of the l_2 -norm was computed. Obviously, $f = \sum_{i=1}^m f_i$, and hence, by the sum rule of subdifferential calculus (Theorem 3.40²¹), we obtain that

$$\partial f(\mathbf{x}) = \sum_{i=1}^m \partial f_i(\mathbf{x}) = \begin{cases} \sum_{i=1}^m \omega_i \frac{\mathbf{x} - \mathbf{a}_i}{\|\mathbf{x} - \mathbf{a}_i\|_2}, & \mathbf{x} \notin \mathcal{A}, \\ \sum_{i=1, i \neq j}^m \omega_i \frac{\mathbf{x} - \mathbf{a}_i}{\|\mathbf{x} - \mathbf{a}_i\|_2} + B[\mathbf{0}, \omega_j], & \mathbf{x} = \mathbf{a}_j (j = 1, 2, \dots, m). \end{cases}$$

Using Fermat’s optimality condition (Theorem 3.63), we can conclude that $\mathbf{x}^* \in \mathbb{R}^d$ is an optimal solution of problem (FW) if and only if either

$$\mathbf{x}^* \notin \mathcal{A} \text{ and } \sum_{i=1}^m \omega_i \frac{\mathbf{x}^* - \mathbf{a}_i}{\|\mathbf{x}^* - \mathbf{a}_i\|_2} = \mathbf{0}$$

or for some $j \in \{1, 2, \dots, m\}$

$$\mathbf{x}^* = \mathbf{a}_j \text{ and } \left\| \sum_{i=1, i \neq j}^m \omega_i \frac{\mathbf{x}^* - \mathbf{a}_i}{\|\mathbf{x}^* - \mathbf{a}_i\|_2} \right\|_2 \leq \omega_j. \quad \blacksquare$$

3.7.2 Convex Constrained Optimization

Consider the constrained optimization problem

$$\min\{f(\mathbf{x}) : \mathbf{x} \in C\}, \tag{3.75}$$

where f is an extended real-valued convex function and $C \subseteq \mathbb{E}$ is a convex set. Using Fermat’s optimality condition (Theorem 3.63) and the convexity assumptions, it is easy to write a necessary and sufficient optimality condition for problem (3.75) in terms of the subdifferential set of f and the normal cone of C .

²¹or by Corollary 3.39

Theorem 3.67 (necessary and sufficient optimality conditions for convex constrained optimization). Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper and convex function, and let $C \subseteq \mathbb{E}$ be a convex set for which $\text{ri}(\text{dom}(f)) \cap \text{ri}(C) \neq \emptyset$. Then $\mathbf{x}^* \in C$ is an optimal solution of (3.75) if and only if

$$\text{there exists } \mathbf{g} \in \partial f(\mathbf{x}^*) \text{ for which } -\mathbf{g} \in N_C(\mathbf{x}^*). \quad (3.76)$$

Proof. Problem (3.75) can be rewritten as

$$\min_{\mathbf{x} \in \mathbb{E}} f(\mathbf{x}) + \delta_C(\mathbf{x}).$$

Since $\text{ri}(\text{dom}(f)) \cap \text{ri}(C) \neq \emptyset$, it follows by the sum rule of subdifferential calculus (Theorem 3.40) that for any $\mathbf{x} \in C$,

$$\partial(f + \delta_C)(\mathbf{x}) = \partial f(\mathbf{x}) + \partial \delta_C(\mathbf{x}).$$

By Example 3.5, $\partial \delta_C(\mathbf{x}) = N_C(\mathbf{x})$, and consequently for any $\mathbf{x} \in C$,

$$\partial(f + \delta_C)(\mathbf{x}) = \partial f(\mathbf{x}) + N_C(\mathbf{x}).$$

Therefore, invoking Fermat's optimality condition (Theorem 3.63), $\mathbf{x}^* \in C$ is an optimal solution of (3.75) if and only if $\mathbf{0} \in \partial f(\mathbf{x}^*) + N_C(\mathbf{x}^*)$, that is, if and only if

$$(-\partial f(\mathbf{x}^*)) \cap N_C(\mathbf{x}^*) \neq \emptyset,$$

which is the same as condition (3.76). \square

Using the definition of the normal cone, we can write the optimality condition in a slightly more explicit manner.

Corollary 3.68 (necessary and sufficient optimality conditions for convex constrained optimization—second version). Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper and convex function, and let C be a convex set satisfying $\text{ri}(\text{dom}(f)) \cap \text{ri}(C) \neq \emptyset$. Then $\mathbf{x}^* \in C$ is an optimal solution of (3.75) if and only if

$$\text{there exists } \mathbf{g} \in \partial f(\mathbf{x}^*) \text{ for which } \langle \mathbf{g}, \mathbf{x} - \mathbf{x}^* \rangle \geq 0 \text{ for any } \mathbf{x} \in C. \quad (3.77)$$

Condition (3.77) is not particularly explicit. We will show in the next example how to write it in an explicit way for the case where $C = \Delta_n$.

Example 3.69 (optimality conditions over the unit simplex). Suppose that the assumptions in Corollary 3.68 hold and that $C = \Delta_n, \mathbb{E} = \mathbb{R}^n$. Given $\mathbf{x}^* \in \Delta_n$, we will show that the condition

$$(I) \quad \mathbf{g}^T(\mathbf{x} - \mathbf{x}^*) \geq 0 \text{ for all } \mathbf{x} \in \Delta_n$$

is satisfied if and only if the following condition is satisfied:

$$(II) \quad \text{there exist } \mu \in \mathbb{R} \text{ such that } g_i \begin{cases} = \mu, & x_i^* > 0, \\ \geq \mu, & x_i^* = 0. \end{cases}$$

Assume first that (II) is satisfied. Then for any $\mathbf{x} \in \Delta_n$,

$$\begin{aligned}\mathbf{g}^T(\mathbf{x} - \mathbf{x}^*) &= \sum_{i=1}^n g_i(x_i - x_i^*) \\ &= \sum_{i:x_i^*>0} g_i(x_i - x_i^*) + \sum_{i:x_i^*=0} g_i x_i \\ &\geq \sum_{i:x_i^*>0} \mu(x_i - x_i^*) + \mu \sum_{i:x_i^*=0} x_i \\ &= \mu \sum_{i=1}^n x_i - \mu \sum_{i:x_i^*>0} x_i^* = \mu - \mu = 0,\end{aligned}$$

proving that condition (I) is satisfied. To show the reverse direction, assume that (I) is satisfied. Let i and j be two different indices for which $x_i^* > 0$. Define the vector $\mathbf{x} \in \Delta_n$ as

$$x_k = \begin{cases} x_k^*, & k \notin \{i, j\}, \\ x_i^* - \frac{x_i^*}{2}, & k = i, \\ x_j^* + \frac{x_i^*}{2}, & k = j. \end{cases}$$

The inequality $\mathbf{g}^T(\mathbf{x} - \mathbf{x}^*) \geq 0$ then amounts to

$$-\frac{x_i^*}{2}g_i + \frac{x_i^*}{2}g_j \geq 0,$$

which by the fact that $x_i^* > 0$ implies that

$$g_i \leq g_j. \quad (3.78)$$

In particular, for any two indices $i \neq j$ for which $x_i^*, x_j^* > 0$, the two inequalities $g_i \leq g_j$ and $g_j \leq g_i$ hold, and hence $g_i = g_j$. Therefore, all the components of \mathbf{g} corresponding to positive components of \mathbf{x}^* have the same value, which we will denote by μ . Let i be any index for which $x_i^* > 0$. Then for any index j for which $x_j^* = 0$, the inequality (3.78) holds. Therefore, $g_j \geq \mu$, and condition (II) is thus established. ■

We summarize the discussion in Example 3.69 with the following corollary.

Corollary 3.70 (necessary and sufficient optimality conditions for convex problems over the unit simplex). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper and convex function. Suppose that $\text{ri}(\Delta_n) \cap \text{ri}(\text{dom}(f)) \neq \emptyset$. Then $\mathbf{x}^* \in \Delta_n$ is an optimal solution of*

$$\min\{f(\mathbf{x}) : \mathbf{x} \in \Delta_n\}$$

if and only if there exists $\mathbf{g} \in \partial f(\mathbf{x}^)$ and $\mu \in \mathbb{R}$ for which*

$$g_i \begin{cases} = \mu, & x_i^* > 0, \\ \geq \mu, & x_i^* = 0. \end{cases}$$

The following example illustrates one instance in which the optimal solution of a convex problem over the unit simplex can be found using Corollary 3.70.

Example 3.71. Consider the problem

$$\min_{\mathbf{x}} \left\{ \sum_{i=1}^n x_i \log x_i - \sum_{i=1}^n y_i x_i : \mathbf{x} \in \Delta_n \right\}, \quad (3.79)$$

where $\mathbf{y} \in \mathbb{R}^n$ is a given vector. Problem (3.79) can be written as

$$\min\{f(\mathbf{x}) : \mathbf{x} \in \Delta_n\},$$

where $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ is given by

$$f(\mathbf{x}) = \begin{cases} \sum_{i=1}^n x_i \log x_i - \sum_{i=1}^n y_i x_i, & \mathbf{x} \geq \mathbf{0}, \\ \infty & \text{else.} \end{cases}$$

Let us assume that there exists an optimal solution²² \mathbf{x}^* satisfying $\mathbf{x}^* > \mathbf{0}$. Then under this assumption, by Corollary 3.70 and the fact that f is differentiable at any positive vector, it follows that there exists $\mu \in \mathbb{R}$ such that for any i , $\frac{\partial f}{\partial x_i}(\mathbf{x}^*) = \mu$, which is the same as $\log x_i^* + 1 - y_i = \mu$. Therefore, for any i ,

$$x_i^* = e^{\mu-1+y_i} = \alpha e^{y_i}, \quad i = 1, 2, \dots, n$$

where $\alpha = e^{\mu-1}$. Since $\sum_{i=1}^n x_i^* = 1$, it follows that $\alpha = \frac{1}{\sum_{j=1}^n e^{y_j}}$. Therefore,

$$x_i^* = \frac{e^{y_i}}{\sum_{j=1}^n e^{y_j}}, \quad i = 1, 2, \dots, n.$$

This is indeed an optimal solution of problem (3.79) since it satisfies the conditions of Corollary 3.70, which are (also) sufficient conditions for optimality. ■

3.7.3 The Nonconvex Composite Model

It is also possible to write a necessary optimality condition for *nonconvex* problems in terms of subgradients. We will write such a condition for problems consisting of minimizing a composite function $f + g$, where f possesses some differentiability properties but is not assumed to be convex while g is convex but not assumed to have any special differentiability properties.

Theorem 3.72 (optimality conditions for the composite problem). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper function, and let $g : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper convex function such that $\text{dom}(g) \subseteq \text{int}(\text{dom}(f))$. Consider the problem*

$$(P) \quad \min_{\mathbf{x} \in \mathbb{E}} f(\mathbf{x}) + g(\mathbf{x}).$$

²²It is not difficult to show a priori that the problem has a unique solution.

- (a) (**necessary condition**) If $\mathbf{x}^* \in \text{dom}(g)$ is a local optimal solution of (P) and f is differentiable at \mathbf{x}^* , then

$$-\nabla f(\mathbf{x}^*) \in \partial g(\mathbf{x}^*). \quad (3.80)$$

- (b) (**necessary and sufficient condition for convex problems**) Suppose that f is convex. If f is differentiable at $\mathbf{x}^* \in \text{dom}(g)$, then \mathbf{x}^* is a global optimal solution of (P) if and only if (3.80) is satisfied.

Proof. (a) Let $\mathbf{y} \in \text{dom}(g)$. Then by the convexity of $\text{dom}(g)$, for any $\lambda \in (0, 1)$, the point $\mathbf{x}_\lambda = (1 - \lambda)\mathbf{x}^* + \lambda\mathbf{y}$ is in $\text{dom}(g)$, and by the local optimality of \mathbf{x}^* , it follows that, for small enough λ ,

$$f(\mathbf{x}_\lambda) + g(\mathbf{x}_\lambda) \geq f(\mathbf{x}^*) + g(\mathbf{x}^*).$$

That is,

$$f((1 - \lambda)\mathbf{x}^* + \lambda\mathbf{y}) + g((1 - \lambda)\mathbf{x}^* + \lambda\mathbf{y}) \geq f(\mathbf{x}^*) + g(\mathbf{x}^*).$$

Using the convexity of g , it follows that

$$f((1 - \lambda)\mathbf{x}^* + \lambda\mathbf{y}) + (1 - \lambda)g(\mathbf{x}^*) + \lambda g(\mathbf{y}) \geq f(\mathbf{x}^*) + g(\mathbf{x}^*),$$

which is the same as

$$\frac{f((1 - \lambda)\mathbf{x}^* + \lambda\mathbf{y}) - f(\mathbf{x}^*)}{\lambda} \geq g(\mathbf{x}^*) - g(\mathbf{y}).$$

Taking $\lambda \rightarrow 0^+$ in the last inequality yields

$$f'(\mathbf{x}^*; \mathbf{y} - \mathbf{x}^*) \geq g(\mathbf{x}^*) - g(\mathbf{y}),$$

where we used the fact that since f is differentiable at \mathbf{x}^* , its directional derivatives exist. In fact, by Theorem 3.29, we have $f'(\mathbf{x}^*; \mathbf{y} - \mathbf{x}^*) = \langle \nabla f(\mathbf{x}^*), \mathbf{y} - \mathbf{x}^* \rangle$, and hence for any $\mathbf{y} \in \text{dom}(g)$,

$$g(\mathbf{y}) \geq g(\mathbf{x}^*) + \langle -\nabla f(\mathbf{x}^*), \mathbf{y} - \mathbf{x}^* \rangle,$$

showing that indeed $-\nabla f(\mathbf{x}^*) \in \partial g(\mathbf{x}^*)$.

(b) Suppose in addition that f is convex. If \mathbf{x}^* is an optimal solution of (P), then we already proved in part (a) that (3.80) is satisfied. Suppose now that (3.80) is satisfied. Then for any $\mathbf{y} \in \text{dom}(g)$,

$$g(\mathbf{y}) \geq g(\mathbf{x}^*) + \langle -\nabla f(\mathbf{x}^*), \mathbf{y} - \mathbf{x}^* \rangle. \quad (3.81)$$

By the convexity of f , for any $\mathbf{y} \in \text{dom}(g)$,

$$f(\mathbf{y}) \geq f(\mathbf{x}^*) + \langle \nabla f(\mathbf{x}^*), \mathbf{y} - \mathbf{x}^* \rangle. \quad (3.82)$$

Adding (3.81) and (3.82), we obtain that

$$f(\mathbf{y}) + g(\mathbf{y}) \geq f(\mathbf{x}^*) + g(\mathbf{x}^*)$$

for any $\mathbf{y} \in \text{dom}(g)$, meaning that \mathbf{x}^* is an optimal solution of (P). \square

The condition (3.80) is an important optimality condition, and we will refer to it as the “stationarity” condition.

Definition 3.73 (stationarity). Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be proper and let $g : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper convex function such that $\text{dom}(g) \subseteq \text{int}(\text{dom}(f))$. Consider the problem

$$(P) \quad \min_{\mathbf{x} \in \mathbb{E}} f(\mathbf{x}) + g(\mathbf{x}).$$

A point \mathbf{x}^* in which f is differentiable is called a **stationary point** of (P) if

$$-\nabla f(\mathbf{x}^*) \in \partial g(\mathbf{x}^*).$$

Under the setting of Definition 3.73, \mathbf{x}^* is also called a *stationary point of the function $f + g$* .

We have shown in Theorem 3.72 that stationarity is a necessary local optimality condition for problem (P), and that if f is convex, then stationarity is a necessary and sufficient global optimality condition. The case $g = \delta_C$ deserves a separate discussion.

Example 3.74 (convex constrained nonconvex programming). When $g = \delta_C$ for a nonempty convex set $C \subseteq \mathbb{E}$, problem (P) becomes

$$\min\{f(\mathbf{x}) : \mathbf{x} \in C\},$$

which is a problem consisting of minimizing a (possibly) nonconvex function over a convex set. A point $\mathbf{x}^* \in C$ in which f is differentiable is a stationary point of (P) if and only if

$$-\nabla f(\mathbf{x}^*) \in \partial\delta_C(\mathbf{x}^*) = N_C(\mathbf{x}^*), \quad (3.83)$$

where the equality is due to Example 3.5. By the definition of the normal cone, condition (3.83) can be rewritten as

$$\langle -\nabla f(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle \leq 0 \text{ for any } \mathbf{x} \in C,$$

which is the same as

$$\langle \nabla f(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle \geq 0 \text{ for any } \mathbf{x} \in C. \quad \blacksquare$$

Example 3.75. Consider the problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) + \lambda \|\mathbf{x}\|_1, \quad (3.84)$$

where $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ is an extended real-valued function. A point $\mathbf{x}^* \in \text{int}(\text{dom}(f))$ in which f is differentiable is a stationary point of (3.84) if

$$-\nabla f(\mathbf{x}^*) \in \lambda \partial g(\mathbf{x}^*),$$

where $g(\cdot) = \|\cdot\|_1$. Using the expression for the subdifferential set of the l_1 -norm given in Example 3.41, we obtain that \mathbf{x}^* is a stationary point of problem (3.84) if

and only if

$$\frac{\partial f(\mathbf{x}^*)}{\partial x_i} \begin{cases} = -\lambda, & x_i^* > 0, \\ = \lambda, & x_i^* < 0, \\ \in [-\lambda, \lambda], & x_i^* = 0. \end{cases} \quad (3.85)$$

By Theorem 3.72, condition (3.85) is a necessary condition for \mathbf{x}^* to be a local minimum of problem (3.84). If f is also convex, then condition (3.85) is a necessary and sufficient condition for \mathbf{x}^* to be a global optimal solution of problem (3.84). ■

3.7.4 The KKT Conditions

In this section we will show that the KKT conditions for constrained convex problems can be directly deduced by Fermat's optimality condition. For that, we begin by establishing an equivalent reformulation of general inequality constrained problems.

Lemma 3.76. *Let $f, g_1, g_2, \dots, g_m : \mathbb{E} \rightarrow \mathbb{R}$ be real-valued functions. Consider the problem*

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \dots, m. \end{aligned} \quad (3.86)$$

Assume that the minimum value of problem (3.86) is finite and equal to \bar{f} . Define the function

$$F(\mathbf{x}) \equiv \max\{f(\mathbf{x}) - \bar{f}, g_1(\mathbf{x}), g_2(\mathbf{x}), \dots, g_m(\mathbf{x})\}. \quad (3.87)$$

Then the optimal set of problem (3.86) is the same as the set of minimizers of F .

Proof. Let X^* be the optimal set of problem (3.86). To establish the result, we will show that F satisfies the following two properties:

- (i) $F(\mathbf{x}) > 0$ for any $\mathbf{x} \notin X^*$.
- (ii) $F(\mathbf{x}) = 0$ for any $\mathbf{x} \in X^*$.

To prove property (i), let $\mathbf{x} \notin X^*$. There are two options. Either \mathbf{x} is not feasible, meaning that $g_i(\mathbf{x}) > 0$ for some i , and hence by its definition $F(\mathbf{x}) > 0$. If \mathbf{x} is feasible but not optimal, then $g_i(\mathbf{x}) \leq 0$ for all $i = 1, 2, \dots, m$ and $f(\mathbf{x}) > \bar{f}$, which also implies that $F(\mathbf{x}) > 0$. To prove (ii), suppose that $\mathbf{x} \in X^*$. Then $g_i(\mathbf{x}) \leq 0$ for all $i = 1, 2, \dots, m$ and $f(\mathbf{x}) = \bar{f}$, implying that $F(\mathbf{x}) = 0$. □

Using Lemma 3.76, we can conclude that problem (3.86) reduces to

$$\min_{\mathbf{x} \in \mathbb{E}} F(\mathbf{x}) \quad (3.88)$$

in the sense that the optimal sets of the two problems are the same. Using this equivalence, we can now establish under additional convexity assumptions the well-known Fritz-John optimality conditions for problem (3.86).

Theorem 3.77 (Fritz-John necessary optimality conditions). Consider the minimization problem

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \dots, m, \end{aligned} \tag{3.89}$$

where $f, g_1, g_2, \dots, g_m : \mathbb{E} \rightarrow \mathbb{R}$ are real-valued convex functions. Let \mathbf{x}^* be an optimal solution of (3.89). Then there exist $\lambda_0, \lambda_1, \dots, \lambda_m \geq 0$, not all zeros, for which

$$\mathbf{0} \in \lambda_0 \partial f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \partial g_i(\mathbf{x}^*) \tag{3.90}$$

$$\lambda_i g_i(\mathbf{x}^*) = 0, \quad i = 1, 2, \dots, m. \tag{3.91}$$

Proof. Let \mathbf{x}^* be an optimal solution of problem (3.89). Denote the optimal value of problem (3.89) by $\bar{f} = f(\mathbf{x}^*)$. Using Lemma 3.76, it follows that \mathbf{x}^* is an optimal solution of the problem

$$\min_{\mathbf{x} \in \mathbb{E}} \{F(\mathbf{x}) \equiv \max\{g_0(\mathbf{x}), g_1(\mathbf{x}), \dots, g_m(\mathbf{x})\}\},$$

where $g_0(\mathbf{x}) \equiv f(\mathbf{x}) - \bar{f}$. Obviously, $F(\mathbf{x}^*) = 0$. Since F is a maximum of convex functions, it is convex, and hence, using Fermat's optimality condition (Theorem 3.63),

$$\mathbf{0} \in \partial F(\mathbf{x}^*). \tag{3.92}$$

By the max rule of subdifferential calculus (Theorem 3.50),

$$\partial F(\mathbf{x}^*) = \text{conv}((\cup_{i \in I(\mathbf{x}^*)} \partial g_i(\mathbf{x}^*))), \tag{3.93}$$

where $I(\mathbf{x}^*) = \{i \in \{0, 1, \dots, m\} : g_i(\mathbf{x}^*) = 0\}$. Combining (3.92) and (3.93), we can deduce that there exists $\lambda_i \geq 0, i \in I(\mathbf{x}^*)$, such that $\sum_{i \in I(\mathbf{x}^*)} \lambda_i = 1$ for which

$$\mathbf{0} \in \sum_{i \in I(\mathbf{x}^*)} \lambda_i \partial g_i(\mathbf{x}^*). \tag{3.94}$$

Since $g_0(\mathbf{x}^*) = f(\mathbf{x}^*) - \bar{f} = 0$, it follows that $0 \in I(\mathbf{x}^*)$, and hence (3.94) can be rewritten as

$$\mathbf{0} \in \lambda_0 \partial f(\mathbf{x}^*) + \sum_{i \in I(\mathbf{x}^*) \setminus \{0\}} \lambda_i \partial g_i(\mathbf{x}^*).$$

Defining $\lambda_i = 0$ for any $i \in \{1, 2, \dots, m\} \setminus I(\mathbf{x}^*)$, we conclude that conditions (3.90) and (3.91) are satisfied. Finally, not all the λ_i 's are zeros since $\sum_{i \in I(\mathbf{x}^*)} \lambda_i = 1$. \square

We will now establish the KKT conditions, which are the same as the Fritz-John conditions, but with $\lambda_0 = 1$. The necessity of these conditions requires the following additional condition, which we refer to as *Slater's condition*:

$$\text{there exists } \bar{\mathbf{x}} \in \mathbb{E} \text{ for which } g_i(\bar{\mathbf{x}}) < 0, \quad i = 1, 2, \dots, m. \tag{3.95}$$

The sufficiency of the KKT conditions does not require any additional assumptions (besides convexity) and is actually easily derived without using the result on the Fritz-John conditions.

Theorem 3.78 (KKT conditions). *Consider the minimization problem*

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \dots, m, \end{aligned} \tag{3.96}$$

where $f, g_1, g_2, \dots, g_m : \mathbb{E} \rightarrow \mathbb{R}$ are real-valued convex functions.

- (a) Let \mathbf{x}^* be an optimal solution of (3.96), and assume that Slater's condition (3.95) is satisfied. Then there exist $\lambda_1, \dots, \lambda_m \geq 0$ for which

$$\mathbf{0} \in \partial f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \partial g_i(\mathbf{x}^*) \tag{3.97}$$

$$\lambda_i g_i(\mathbf{x}^*) = 0, \quad i = 1, 2, \dots, m. \tag{3.98}$$

- (b) If $\mathbf{x}^* \in \mathbb{E}$ satisfies conditions (3.97) and (3.98) for some $\lambda_1, \lambda_2, \dots, \lambda_m \geq 0$, then it is an optimal solution of problem (3.96).

Proof. (a) By the Fritz-John conditions (Theorem 3.77) there exist $\tilde{\lambda}_0, \tilde{\lambda}_1, \dots, \tilde{\lambda}_m \geq 0$, not all zeros, for which

$$\mathbf{0} \in \tilde{\lambda}_0 \partial f(\mathbf{x}^*) + \sum_{i=1}^m \tilde{\lambda}_i \partial g_i(\mathbf{x}^*), \tag{3.99}$$

$$\tilde{\lambda}_i g_i(\mathbf{x}^*) = 0, \quad i = 1, 2, \dots, m. \tag{3.100}$$

We will show that $\tilde{\lambda}_0 \neq 0$. Assume by contradiction that $\tilde{\lambda}_0 = 0$. Then, by (3.99),

$$\mathbf{0} \in \sum_{i=1}^m \tilde{\lambda}_i \partial g_i(\mathbf{x}^*);$$

that is, there exist $\xi_i \in \partial g_i(\mathbf{x}^*), i = 1, 2, \dots, m$, such that

$$\sum_{i=1}^m \tilde{\lambda}_i \xi_i = \mathbf{0}. \tag{3.101}$$

Let $\bar{\mathbf{x}}$ be a point satisfying Slater's condition (3.95). By the subgradient inequality employed on the pair of points $\bar{\mathbf{x}}, \mathbf{x}^*$ w.r.t. the functions $g_i, i = 1, 2, \dots, m$, we have

$$g_i(\mathbf{x}^*) + \langle \xi_i, \bar{\mathbf{x}} - \mathbf{x}^* \rangle \leq g_i(\bar{\mathbf{x}}), \quad i = 1, 2, \dots, m.$$

Multiplying the i th inequality by $\tilde{\lambda}_i \geq 0$ and summing over $i = 1, 2, \dots, m$ yields

$$\sum_{i=1}^m \tilde{\lambda}_i g_i(\mathbf{x}^*) + \left\langle \sum_{i=1}^m \tilde{\lambda}_i \xi_i, \bar{\mathbf{x}} - \mathbf{x}^* \right\rangle \leq \sum_{i=1}^m \tilde{\lambda}_i g_i(\bar{\mathbf{x}}), \quad i = 1, 2, \dots, m.$$

Using (3.100) and (3.101), we obtain the inequality $\sum_{i=1}^m \tilde{\lambda}_i g_i(\bar{\mathbf{x}}) \geq 0$, which is impossible since $\tilde{\lambda}_i \geq 0$ and $g_i(\bar{\mathbf{x}}) < 0$ for any i , and not all the $\tilde{\lambda}_i$'s are zeros. Therefore, $\tilde{\lambda}_0 > 0$, and we can thus divide both the relation (3.99) and the equalities (3.100) by $\tilde{\lambda}_0$ to obtain that (3.97) and (3.98) are satisfied with $\lambda_i = \frac{\tilde{\lambda}_i}{\tilde{\lambda}_0}, i = 1, 2, \dots, m$.

(b) Suppose then that \mathbf{x}^* satisfies (3.97) and (3.98) for some nonnegative numbers $\lambda_1, \lambda_2, \dots, \lambda_m$. Let $\hat{\mathbf{x}} \in \mathbb{E}$ be a feasible point of (3.96), meaning that $g_i(\hat{\mathbf{x}}) \leq 0, i = 1, 2, \dots, m$. We will show that $f(\hat{\mathbf{x}}) \geq f(\mathbf{x}^*)$. Define the function

$$h(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^m \lambda_i g_i(\mathbf{x}).$$

The function h is convex, and the condition (3.97) along with the sum rule of subdifferential calculus (Theorem 3.40) yields the relation

$$\mathbf{0} \in \partial h(\mathbf{x}^*),$$

which by Fermat's optimality condition (Theorem 3.63) implies that \mathbf{x}^* is a minimizer of h over \mathbb{E} . Combining this fact with (3.98) implies that

$$f(\mathbf{x}^*) = f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i g_i(\mathbf{x}^*) = h(\mathbf{x}^*) \leq h(\hat{\mathbf{x}}) = f(\hat{\mathbf{x}}) + \sum_{i=1}^m \lambda_i g_i(\hat{\mathbf{x}}) \leq f(\hat{\mathbf{x}}),$$

where the last inequality follows from the facts that $\lambda_i \geq 0$ and $g_i(\hat{\mathbf{x}}) \leq 0$ for $i = 1, 2, \dots, m$. We have thus proven that \mathbf{x}^* is an optimal solution of (3.96). \square

3.8 Summary of Weak and Strong Subgradient Calculus Results

This section contains a summary of most of the rules and results concerning the computation of subdifferential sets (strong results), as well as rules for computing specific subgradients in the subdifferential sets (weak results). Before that, we begin by summarizing the rules of subdifferential calculus.

- **Multiplication by a positive scalar**

$$\partial(\alpha f)(\mathbf{x}) = \alpha \partial f(\mathbf{x}).$$

Assumptions: $f : \mathbb{E} \rightarrow (-\infty, \infty]$ proper, $\alpha > 0$, $\mathbf{x} \in \text{dom}(f)$. [Theorem 3.35]

- **Differentiability**

f is differentiable at \mathbf{x} if and only if $\partial f(\mathbf{x})$ is a singleton, and in that case

$$\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}.$$

Assumptions: $f : \mathbb{E} \rightarrow (-\infty, \infty]$ proper convex, $\mathbf{x} \in \text{int}(\text{dom}(f))$. [Theorem 3.33]

- **Weak sum rule of subdifferential calculus**

$$\sum_{i=1}^m \partial f_i(\mathbf{x}) \subseteq \partial(\sum_{i=1}^m f_i)(\mathbf{x}).$$

Assumptions: $f_1, f_2, \dots, f_m : \mathbb{E} \rightarrow (-\infty, \infty]$ proper convex. [Corollary 3.38(a)]

- **Sum rule of subdifferential calculus**

$$\partial(\sum_{i=1}^m f_i)(\mathbf{x}) = \sum_{i=1}^m \partial f_i(\mathbf{x}).$$

Assumptions: $f_1, f_2, \dots, f_m : \mathbb{E} \rightarrow (-\infty, \infty]$ proper convex, $\cap_{i=1}^m \text{ri}(\text{dom}(f_i)) \neq \emptyset$. [Theorem 3.40]

- **Weak affine transformation rule of subdifferential calculus**

$$\mathcal{A}^T(\partial f(\mathcal{A}(\mathbf{x}) + \mathbf{b})) \subseteq \partial h(\mathbf{x}) \quad (h(\mathbf{x}) \equiv f(\mathcal{A}(\mathbf{x}) + \mathbf{b})).$$

Assumptions: $f, h : \mathbb{E} \rightarrow (-\infty, \infty]$ proper convex, $\mathbf{x} \in \text{dom}(h)$. [Theorem 3.43(a)]

- **Affine transformation rule of subdifferential calculus**

$$\partial h(\mathbf{x}) = \mathcal{A}^T(\partial f(\mathcal{A}(\mathbf{x}) + \mathbf{b})) \quad (h(\mathbf{x}) \equiv f(\mathcal{A}(\mathbf{x}) + \mathbf{b})).$$

Assumptions: $f, h : \mathbb{E} \rightarrow (-\infty, \infty]$ proper convex, $\mathbf{x} \in \text{int}(\text{dom}(h))$, $\mathcal{A}(\mathbf{x}) + \mathbf{b} \in \text{int}(\text{dom}(f))$. [Theorem 3.43(b)]

- **Chain rule of subdifferential calculus**

$$\partial h(\mathbf{x}) = g'(f(\mathbf{x}))\partial f(\mathbf{x}) \quad (h = g \circ f).$$

Assumptions: $f : \mathbb{E} \rightarrow \mathbb{R}$ convex, $g : \mathbb{R} \rightarrow \mathbb{R}$ nondecreasing, differentiable and convex. [Theorem 3.47]

- **Max rule of subdifferential calculus**

$$\partial(\max(f_1, f_2, \dots, f_m))(\mathbf{x}) = \text{conv}(\cup_{i \in I(\mathbf{x})} \partial f_i(\mathbf{x})),$$

where

$$I(\mathbf{x}) = \{i : f_i(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})\}\}.$$

Assumptions: f_1, f_2, \dots, f_m proper, convex, $\mathbf{x} \in \cap_{i=1}^m \text{int}(\text{dom}(f_i))$. [Theorem 3.50]

- **Weak max rule of subdifferential calculus**

$$\text{conv}(\cup_{i \in I(\mathbf{x})} \partial f_i(\mathbf{x})) \subseteq \partial(\max_{i \in I} f_i)(\mathbf{x}),$$

where

$$I(\mathbf{x}) = \{i \in I : f_i(\mathbf{x}) = \max_{i \in I} f_i(\mathbf{x})\}.$$

Assumptions: I = arbitrary index set. $f_i : \mathbb{E} \rightarrow (-\infty, \infty]$ ($i \in I$) proper, convex, $\mathbf{x} \in \cap_{i \in I} \text{dom}(f_i)$. [Theorem 3.55]

The table below contains the main examples from the chapter related to weak results of subgradients computations.

Function	Weak result	Setting	Reference
$-q =$ negative dual function	$-\mathbf{g}(\mathbf{x}_0) \in \partial(-q)(\boldsymbol{\lambda}_0)$	$q(\boldsymbol{\lambda}) = \min_{\mathbf{x} \in X} f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x})$, $f : \mathbb{E} \rightarrow \mathbb{R}$, $\mathbf{g} : \mathbb{E} \rightarrow \mathbb{R}^m$, $\mathbf{x}_0 =$ a minimizer of $f(\mathbf{x}) + \boldsymbol{\lambda}_0^T \mathbf{g}(\mathbf{x})$ over X	Example 3.7
$f(\mathbf{X}) = \lambda_{\max}(\mathbf{X})$	$\mathbf{v}\mathbf{v}^T \in \partial f(\mathbf{X})$	$f : \mathbb{S}^n \rightarrow \mathbb{R}$, \mathbf{v} = normalized maximum eigenvector of $X \in \mathbb{S}^n$	Example 3.8
$f(\mathbf{x}) = \ \mathbf{x}\ _1$	$\text{sgn}(\mathbf{x}) \in \partial f(\mathbf{x})$	$\mathbb{E} = \mathbb{R}^n$	Example 3.42
$f(\mathbf{x}) = \lambda_{\max}(\mathbf{A}_0 + \sum_{i=1}^m x_i \mathbf{A}_i)$	$(\tilde{\mathbf{y}}^T \mathbf{A}_i \tilde{\mathbf{y}})_{i=1}^m \in \partial f(\mathbf{x})$	$\tilde{\mathbf{y}}$ = normalized maximum eigenvector of $\mathbf{A}_0 + \sum_{i=1}^m x_i \mathbf{A}_i$	Example 3.56

The following table contains the main strong results of subdifferential sets computations derived in this chapter.

$f(\mathbf{x})$	$\partial f(\mathbf{x})$	Assumptions	Reference
$\ \mathbf{x}\ $	$B_{\ \cdot\ _*}[\mathbf{0}, 1]$	$\mathbf{x} = \mathbf{0}$	Example 3.3
$\ \mathbf{x}\ _1$	$\left\{ \sum_{i \in I_{\neq}(\mathbf{x})} \text{sgn}(x_i) \mathbf{e}_i + \sum_{i \in I_0(\mathbf{x})} [-\mathbf{e}_i, \mathbf{e}_i] \right\}$	$\mathbb{E} = \mathbb{R}^n, I_{\neq}(\mathbf{x}) = \{i : x_i \neq 0\}, I_0(\mathbf{x}) = \{i : x_i = 0\}$	Example 3.41
$\ \mathbf{x}\ _2$	$\begin{cases} \left\{ \frac{\mathbf{x}}{\ \mathbf{x}\ _2} \right\}, & \mathbf{x} \neq \mathbf{0}, \\ B_{\ \cdot\ _2}[\mathbf{0}, 1], & \mathbf{x} = \mathbf{0}. \end{cases}$	$\mathbb{E} = \mathbb{R}^n$	Example 3.34
$\ \mathbf{x}\ _\infty$	$\left\{ \sum_{i \in I(\mathbf{x})} \lambda_i \text{sgn}(x_i) \mathbf{e}_i : \begin{array}{l} \sum_{i \in I(\mathbf{x})} \lambda_i = 1 \\ \lambda_i \geq 0 \end{array} \right\}$	$\mathbb{E} = \mathbb{R}^n, I(\mathbf{x}) = \{i : \ \mathbf{x}\ _\infty = x_i \}, \mathbf{x} \neq \mathbf{0}$	Example 3.52
$\max(\mathbf{x})$	$\left\{ \sum_{i \in I(\mathbf{x})} \lambda_i \mathbf{e}_i : \sum_{i \in I(\mathbf{x})} \lambda_i = 1, \lambda_i \geq 0 \right\}$	$\mathbb{E} = \mathbb{R}^n, I(\mathbf{x}) = \{i : \max(\mathbf{x}) = x_i\}$	Example 3.51
$\max(\mathbf{x})$	Δ_n	$\mathbb{E} = \mathbb{R}^n, \mathbf{x} = \alpha \mathbf{e}$ for some $\alpha \in \mathbb{R}$	Example 3.51
$\delta_S(\mathbf{x})$	$N_S(\mathbf{x})$	$\emptyset \neq S \subseteq \mathbb{E}$	Example 3.5
$\delta_{B[\mathbf{0}, 1]}(\mathbf{x})$	$\begin{cases} \{\mathbf{y} \in \mathbb{E}^* : \ \mathbf{y}\ _* \leq \langle \mathbf{y}, \mathbf{x} \rangle\}, & \ \mathbf{x}\ \leq 1, \\ \emptyset, & \ \mathbf{x}\ > 1. \end{cases}$		Example 3.6
$\ \mathbf{Ax} + \mathbf{b}\ _1$	$\sum_{i \in I_{\neq}(\mathbf{x})} \text{sgn}(\mathbf{a}_i^T \mathbf{x} + b_i) \mathbf{a}_i + \sum_{i \in I_0(\mathbf{x})} [-\mathbf{a}_i, \mathbf{a}_i]$	$\mathbb{E} = \mathbb{R}^n, \mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{b} \in \mathbb{R}^m, I_{\neq}(\mathbf{x}) = \{i : \mathbf{a}_i^T \mathbf{x} + b_i \neq 0\}, I_0(\mathbf{x}) = \{i : \mathbf{a}_i^T \mathbf{x} + b_i = 0\}$	Example 3.44
$\ \mathbf{Ax} + \mathbf{b}\ _2$	$\begin{cases} \frac{\mathbf{A}^T(\mathbf{Ax} + \mathbf{b})}{\ \mathbf{Ax} + \mathbf{b}\ _2}, & \mathbf{Ax} + \mathbf{b} \neq \mathbf{0}, \\ \mathbf{A}^T B_{\ \cdot\ _2}[\mathbf{0}, 1], & \mathbf{Ax} + \mathbf{b} = \mathbf{0}. \end{cases}$	$\mathbb{E} = \mathbb{R}^n, \mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{b} \in \mathbb{R}^m$	Example 3.45
$\ \mathbf{Ax} + \mathbf{b}\ _\infty$	$\left\{ \sum_{i \in I_{\mathbf{x}}} \lambda_i \text{sgn}(\mathbf{a}_i^T \mathbf{x} + b_i) \mathbf{a}_i : \begin{array}{l} \sum_{i \in I_{\mathbf{x}}} \lambda_i = 1 \\ \lambda_i \geq 0 \end{array} \right\}$	$\mathbb{E} = \mathbb{R}^n, \mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{b} \in \mathbb{R}^m, I_{\mathbf{x}} = \{i : \ \mathbf{Ax} + \mathbf{b}\ _\infty = \mathbf{a}_i^T \mathbf{x} + b_i \}, \mathbf{Ax} + \mathbf{b} \neq \mathbf{0}$	Example 3.54
$\ \mathbf{Ax} + \mathbf{b}\ _\infty$	$\mathbf{A}^T B_{\ \cdot\ _1}[\mathbf{0}, 1]$	same as above but with $\mathbf{Ax} + \mathbf{b} = \mathbf{0}$	Example 3.54
$\max_i \{\mathbf{a}_i^T \mathbf{x} + b\}$	$\left\{ \sum_{i \in I(\mathbf{x})} \lambda_i \mathbf{a}_i : \sum_{i \in I(\mathbf{x})} \lambda_i = 1, \lambda_i \geq 0 \right\}$	$\mathbb{E} = \mathbb{R}^n, \mathbf{a}_i \in \mathbb{R}^n, b_i \in \mathbb{R}, I(\mathbf{x}) = \{i : f(\mathbf{x}) = \mathbf{a}_i^T \mathbf{x} + b_i\}$	Example 3.53
$\frac{1}{2} d_C^2(\mathbf{x})$	$\{\mathbf{x} - P_C(\mathbf{x})\}$	$C = \text{nonempty closed and convex}, \mathbb{E} = \text{Euclidean}$	Example 3.31
$d_C(\mathbf{x})$	$\begin{cases} \left\{ \frac{\mathbf{x} - P_C(\mathbf{x})}{d_C(\mathbf{x})} \right\}, & \mathbf{x} \notin C, \\ N_C(\mathbf{x}) \cap B[\mathbf{0}, 1] & \mathbf{x} \in C. \end{cases}$	$C = \text{nonempty closed and convex}, \mathbb{E} = \text{Euclidean}$	Example 3.49

Chapter 4

Conjugate Functions

4.1 Definition and Basic Properties

We begin with the definition of the conjugate function.

Definition 4.1 (conjugate functions). Let $f : \mathbb{E} \rightarrow [-\infty, \infty]$ be an extended real-valued function. The function $f^* : \mathbb{E}^* \rightarrow [-\infty, \infty]$, defined by

$$f^*(\mathbf{y}) = \max_{\mathbf{x} \in \mathbb{E}} \{\langle \mathbf{y}, \mathbf{x} \rangle - f(\mathbf{x})\}, \quad \mathbf{y} \in \mathbb{E}^*,$$

is called the conjugate function of f .

Example 4.2 (conjugate of indicator functions). Let $f = \delta_C$, where $C \subseteq \mathbb{E}$ is nonempty. Then for any $\mathbf{y} \in \mathbb{E}^*$

$$f^*(\mathbf{y}) = \max_{\mathbf{x} \in \mathbb{E}} \{\langle \mathbf{y}, \mathbf{x} \rangle - \delta_C(\mathbf{x})\} = \max_{\mathbf{x} \in C} \langle \mathbf{y}, \mathbf{x} \rangle = \sigma_C(\mathbf{y}).$$

That is, the conjugate of the indicator function is the support function of the same underlying set:

$$\delta_C^* = \sigma_C.$$

■

Two fundamental properties of conjugate functions are their convexity and closedness (regardless of whether the original function is closed or convex).

Theorem 4.3 (convexity and closedness of conjugate functions). Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be an extended real-valued function. Then the conjugate function f^* is closed and convex.

Proof. Note that f^* is the pointwise maximum of affine functions, which are convex and closed, and thus, invoking Theorems 2.16(c) and 2.7(c), it follows that f^* is closed and convex. □

Example 4.4 (conjugate of $\frac{1}{2}\|\cdot\|^2 + \delta_C$). Suppose that \mathbb{E} is Euclidean and that $C \subseteq \mathbb{E}$ is nonempty. Define $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2 + \delta_C(\mathbf{x})$. Then by Example 2.17 (specifically, (2.6)), it follows that

$$f^*(\mathbf{y}) = \frac{1}{2}\|\mathbf{y}\|^2 - \frac{1}{2}d_C^2(\mathbf{y}).$$

Note that while f is convex only if C is convex, the convexity of f^* is guaranteed regardless of whether C is convex or not. ■

The next result states that the conjugate function of a proper convex function is also proper.

Theorem 4.5 (properness of conjugate functions). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper convex function. Then f^* is proper.*

Proof. Since f is proper, it follows that there exists $\hat{\mathbf{x}} \in \mathbb{E}$ such that $f(\hat{\mathbf{x}}) < \infty$. By the definition of the conjugate function, for any $\mathbf{y} \in \mathbb{E}^*$,

$$f^*(\mathbf{y}) \geq \langle \mathbf{y}, \hat{\mathbf{x}} \rangle - f(\hat{\mathbf{x}}),$$

and hence $f^*(\mathbf{y}) > -\infty$. What remains in order to establish the properness of f^* is to show that there exists $\mathbf{g} \in \mathbb{E}^*$ such that $f^*(\mathbf{g}) < \infty$. By Corollary 3.19, there exists $\mathbf{x} \in \text{dom}(f)$ such that $\partial f(\mathbf{x}) \neq \emptyset$. Take $\mathbf{g} \in \partial f(\mathbf{x})$. Then by the definition of the subgradient, for any $\mathbf{z} \in \mathbb{E}$,

$$f(\mathbf{z}) \geq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{z} - \mathbf{x} \rangle.$$

Hence,

$$f^*(\mathbf{g}) = \max_{\mathbf{z} \in \mathbb{E}} \{ \langle \mathbf{g}, \mathbf{z} \rangle - f(\mathbf{z}) \} \leq \langle \mathbf{g}, \mathbf{x} \rangle - f(\mathbf{x}) < \infty,$$

concluding that f^* is a proper function. □

The following result, called *Fenchel's inequality*, is a trivial implication of the definition of conjugacy.

Theorem 4.6 (Fenchel's inequality). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be an extended real-valued proper function. Then for any $\mathbf{x} \in \mathbb{E}$ and $\mathbf{y} \in \mathbb{E}^*$,*

$$f(\mathbf{x}) + f^*(\mathbf{y}) \geq \langle \mathbf{y}, \mathbf{x} \rangle.$$

Proof. By the definition of the conjugate function we have that for any $\mathbf{x} \in \mathbb{E}$ and $\mathbf{y} \in \mathbb{E}^*$,

$$f^*(\mathbf{y}) \geq \langle \mathbf{y}, \mathbf{x} \rangle - f(\mathbf{x}). \quad (4.1)$$

Since f is proper, it follows that $f(\mathbf{x}), f^*(\mathbf{y}) > -\infty$. We can thus add $f(\mathbf{x})$ to both sides of (4.1) and obtain the desired result. □

4.2 The Biconjugate

The conjugacy operation can be invoked twice resulting in the biconjugate operation. Specifically, for a function $f : \mathbb{E} \rightarrow [-\infty, \infty]$ we define (recall that in this book \mathbb{E} and \mathbb{E}^* are considered to be identical)

$$f^{**}(\mathbf{x}) = \max_{\mathbf{y} \in \mathbb{E}^*} \{\langle \mathbf{x}, \mathbf{y} \rangle - f^*(\mathbf{y})\}, \quad \mathbf{x} \in \mathbb{E}.$$

The biconjugate function is always a lower bound on the original function, as the following result states.

Lemma 4.7 ($f^{} \leq f$).** *Let $f : \mathbb{E} \rightarrow [-\infty, \infty]$ be an extended real-valued function. Then $f(\mathbf{x}) \geq f^{**}(\mathbf{x})$ for any $\mathbf{x} \in \mathbb{E}$.*

Proof. By the definition of the conjugate function we have for any $\mathbf{x} \in \mathbb{E}$ and $\mathbf{y} \in \mathbb{E}^*$,

$$f^*(\mathbf{y}) \geq \langle \mathbf{y}, \mathbf{x} \rangle - f(\mathbf{x}).$$

Thus,

$$f(\mathbf{x}) \geq \langle \mathbf{y}, \mathbf{x} \rangle - f^*(\mathbf{y}),$$

implying that

$$f(\mathbf{x}) \geq \max_{\mathbf{y} \in \mathbb{E}^*} \{\langle \mathbf{y}, \mathbf{x} \rangle - f^*(\mathbf{y})\} = f^{**}(\mathbf{x}). \quad \square$$

If we assume that f is proper closed and convex, then the biconjugate is not just a lower bound on f —it is equal to f .

Theorem 4.8 ($f = f^{}$ for proper closed convex functions).** *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper closed and convex function. Then $f^{**} = f$.*

Proof. By Lemma 4.7, $f^{**} \leq f$. We thus need to show that $f^{**} \geq f$. Suppose by contradiction that for some $\mathbf{x} \in \mathbb{E}$ we have $f^{**}(\mathbf{x}) < f(\mathbf{x})$. This means that $(\mathbf{x}, f^{**}(\mathbf{x})) \notin \text{epi}(f) \subseteq \mathbb{E} \times \mathbb{R}$. We assume as usual that the product space $\mathbb{V} = \mathbb{E} \times \mathbb{R}$ is endowed with the inner product $\langle (\mathbf{u}, s), (\mathbf{v}, t) \rangle_{\mathbb{V}} = \langle \mathbf{u}, \mathbf{v} \rangle + st$, where $\langle \cdot, \cdot \rangle$ is the inner product associated with \mathbb{E} (see Section 1.9). Since f is proper closed and convex, the set $\text{epi}(f)$ is nonempty closed and convex, and hence, by the strict separation theorem (Theorem 2.33), there exist $\mathbf{a} \in \mathbb{E}^*, b, c_1, c_2 \in \mathbb{R}$ such that

$$\langle \mathbf{a}, \mathbf{z} \rangle + bs \leq c_1 < c_2 \leq \langle \mathbf{a}, \mathbf{x} \rangle + bf^{**}(\mathbf{x}) \text{ for all } (\mathbf{z}, s) \in \text{epi}(f).$$

We can thus conclude that

$$\langle \mathbf{a}, \mathbf{z} - \mathbf{x} \rangle + b(s - f^{**}(\mathbf{x})) \leq c_1 - c_2 \equiv c < 0 \text{ for all } (\mathbf{z}, s) \in \text{epi}(f). \quad (4.2)$$

The scalar b must be nonpositive, since otherwise, if it was positive, the inequality would have been violated by taking a fixed \mathbf{z} and large enough s . We will now consider two cases.

- If $b < 0$, then dividing (4.2) by $-b$ and taking $\mathbf{y} = -\frac{\mathbf{a}}{b}$, we get

$$\langle \mathbf{y}, \mathbf{z} - \mathbf{x} \rangle - s + f^{**}(\mathbf{x}) \leq \frac{c}{-b} < 0 \text{ for all } (\mathbf{z}, s) \in \text{epi}(f).$$

In particular, taking $s = f(\mathbf{z})$ (which is possible since $(\mathbf{z}, f(\mathbf{z})) \in \text{epi}(f)$), we obtain that

$$\langle \mathbf{y}, \mathbf{z} \rangle - f(\mathbf{z}) - \langle \mathbf{y}, \mathbf{x} \rangle + f^{**}(\mathbf{x}) \leq \frac{c}{-b} < 0 \text{ for all } \mathbf{z} \in \mathbb{E}.$$

Taking the maximum over \mathbf{z} yields the inequality

$$f^*(\mathbf{y}) - \langle \mathbf{y}, \mathbf{x} \rangle + f^{**}(\mathbf{x}) \leq \frac{c}{-b} < 0,$$

which is a contradiction of Fenchel's inequality (Theorem 4.6).

- If $b = 0$, then take some $\hat{\mathbf{y}} \in \text{dom}(f^*)$. Such a vector exists since f^* is proper by the properness and convexity of f (Theorem 4.5). Let $\varepsilon > 0$ and define $\hat{\mathbf{a}} = \mathbf{a} + \varepsilon \hat{\mathbf{y}}$ and $\hat{b} = -\varepsilon$. Then for any $\mathbf{z} \in \text{dom}(f)$,

$$\begin{aligned} \langle \hat{\mathbf{a}}, \mathbf{z} - \mathbf{x} \rangle + \hat{b}(f(\mathbf{z}) - f^{**}(\mathbf{x})) &= \langle \mathbf{a}, \mathbf{z} - \mathbf{x} \rangle + \varepsilon[\langle \hat{\mathbf{y}}, \mathbf{z} \rangle - f(\mathbf{z}) + f^{**}(\mathbf{x}) - \langle \hat{\mathbf{y}}, \mathbf{x} \rangle] \\ &\leq c + \varepsilon[\langle \hat{\mathbf{y}}, \mathbf{z} \rangle - f(\mathbf{z}) + f^{**}(\mathbf{x}) - \langle \hat{\mathbf{y}}, \mathbf{x} \rangle] \\ &\leq c + \varepsilon[f^*(\hat{\mathbf{y}}) - \langle \hat{\mathbf{y}}, \mathbf{x} \rangle + f^{**}(\mathbf{x})], \end{aligned}$$

where the first inequality is due to (4.2) and the second by the definition of $f^*(\hat{\mathbf{y}})$ as the maximum of $\langle \hat{\mathbf{y}}, \mathbf{z} \rangle - f(\mathbf{z})$ over all possible $\mathbf{z} \in \mathbb{E}$. We thus obtained the inequality

$$\langle \hat{\mathbf{a}}, \mathbf{z} - \mathbf{x} \rangle + \hat{b}(f(\mathbf{z}) - f^{**}(\mathbf{x})) \leq \hat{c}, \quad (4.3)$$

where $\hat{c} \equiv c + \varepsilon[f^*(\hat{\mathbf{y}}) - \langle \hat{\mathbf{y}}, \mathbf{x} \rangle + f^{**}(\mathbf{x})]$. Since $c < 0$, we can pick $\varepsilon > 0$ small enough to ensure that $\hat{c} < 0$. At this point we employ exactly the same argument used in the first case. Dividing (4.3) by $-\hat{b}$ and denoting $\tilde{\mathbf{y}} = -\frac{1}{\hat{b}}\hat{\mathbf{a}}$ yields the inequality

$$\langle \tilde{\mathbf{y}}, \mathbf{z} \rangle - f(\mathbf{z}) - \langle \tilde{\mathbf{y}}, \mathbf{x} \rangle + f^{**}(\mathbf{x}) \leq -\frac{\hat{c}}{\hat{b}} < 0 \text{ for any } \mathbf{z} \in \text{dom}(f).$$

Taking the maximum over \mathbf{z} results in

$$f^*(\tilde{\mathbf{y}}) - \langle \tilde{\mathbf{y}}, \mathbf{x} \rangle + f^{**}(\mathbf{x}) \leq -\frac{\hat{c}}{\hat{b}} < 0,$$

which, again, is a contradiction of Fenchel's inequality. \square

Example 4.9 (conjugate of support functions). We will now show how to exploit Theorem 4.8 in order to compute the conjugate of support functions. Suppose that $C \subseteq \mathbb{E}$ is a given nonempty set. Since $\text{cl}(\text{conv}(C))$ is closed and convex, it follows that $\delta_{\text{cl}(\text{conv}(C))}$ is closed and convex, and hence, by Example 4.2 and Theorem 4.8,

$$\sigma_{\text{cl}(\text{conv}(C))}^* = (\delta_{\text{cl}(\text{conv}(C))}^*)^* = \delta_{\text{cl}(\text{conv}(C))}^{**} = \delta_{\text{cl}(\text{conv}(C))}. \quad (4.4)$$

Finally, by Lemma 2.35,

$$\sigma_C = \sigma_{\text{cl}(\text{conv}(C))},$$

which, combined with (4.4), establishes the result

$$\sigma_C^* = \delta_{\text{cl}(\text{conv}(C))}.$$

Example 4.10 (conjugate of the max function). Consider the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ given by $f(\mathbf{x}) = \max\{x_1, x_2, \dots, x_n\}$. Note that the following elementary identity holds for any $\mathbf{x} \in \mathbb{R}^n$:

$$\max\{x_1, x_2, \dots, x_n\} = \max_{\mathbf{y} \in \Delta_n} \mathbf{y}^T \mathbf{x} = \sigma_{\Delta_n}(\mathbf{x}).$$

Therefore, using Example 4.9, we can conclude, exploiting the convexity and closedness of Δ_n , that

$$f^* = \delta_{\Delta_n}.$$

Example 4.11 (conjugate of $\frac{1}{2}\|\cdot\|^2 - d_C^2$). Let \mathbb{E} be Euclidean, and let $C \subseteq \mathbb{E}$ be a nonempty closed and convex set. Define $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2 - \frac{1}{2}d_C^2(\mathbf{x})$. By Example 4.4, $f = g^*$, where $g(\mathbf{y}) = \frac{1}{2}\|\mathbf{y}\|^2 + \delta_C(\mathbf{y})$. By the nonemptiness, closedness, and convexity of C , it follows that g is proper closed and convex, and hence, by Theorem 4.8,

$$f^*(\mathbf{y}) = g^{**}(\mathbf{y}) = g(\mathbf{y}) = \frac{1}{2}\|\mathbf{y}\|^2 + \delta_C(\mathbf{y}).$$

4.3 Conjugate Calculus Rules

In this section we present the basic calculus rules for computing conjugate functions. We begin with a very simple rule for separable functions.

Theorem 4.12 (conjugate of separable functions). Let $g : \mathbb{E}_1 \times \mathbb{E}_2 \times \dots \times \mathbb{E}_p \rightarrow (-\infty, \infty]$ be given by $g(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p) = \sum_{i=1}^p f_i(\mathbf{x}_i)$, where $f_i : \mathbb{E}_i \rightarrow (-\infty, \infty]$ is a proper function for any $i = 1, 2, \dots, p$. Then

$$g^*(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_p) = \sum_{i=1}^p f_i^*(\mathbf{y}_i) \text{ for any } \mathbf{y}_i \in \mathbb{E}_i^*, \quad i = 1, 2, \dots, p.$$

Proof. For any $(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_p) \in \mathbb{E}_1^* \times \mathbb{E}_2^* \times \dots \times \mathbb{E}_p^*$, it holds that

$$\begin{aligned} g^*(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_p) &= \max_{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p} \{ \langle (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_p), (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p) \rangle - g(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p) \} \\ &= \max_{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p} \left\{ \sum_{i=1}^p \langle \mathbf{y}_i, \mathbf{x}_i \rangle - \sum_{i=1}^p f_i(\mathbf{x}_i) \right\} \\ &= \sum_{i=1}^p \max_{\mathbf{x}_i} \{ \langle \mathbf{y}_i, \mathbf{x}_i \rangle - f_i(\mathbf{x}_i) \} \\ &= \sum_{i=1}^p f_i^*(\mathbf{y}_i). \quad \square \end{aligned}$$

The next result shows how the conjugate operation is affected by invertible affine change of variables as well as by addition of an affine function.

Theorem 4.13 (conjugate of $f(\mathcal{A}(\mathbf{x} - \mathbf{a})) + \langle \mathbf{b}, \mathbf{x} \rangle + c$). Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be an extended real-valued function, and let $\mathcal{A} : \mathbb{V} \rightarrow \mathbb{E}$ be an invertible linear transformation, $\mathbf{a} \in \mathbb{V}$, $\mathbf{b} \in \mathbb{V}^*$, and $c \in \mathbb{R}$. Then the conjugate of the function $g(\mathbf{x}) = f(\mathcal{A}(\mathbf{x} - \mathbf{a})) + \langle \mathbf{b}, \mathbf{x} \rangle + c$ is given by

$$g^*(\mathbf{y}) = f^*((\mathcal{A}^T)^{-1}(\mathbf{y} - \mathbf{b})) + \langle \mathbf{a}, \mathbf{y} \rangle - c - \langle \mathbf{a}, \mathbf{b} \rangle, \quad \mathbf{y} \in \mathbb{V}^*.$$

Proof. Making the change of variables $\mathbf{z} = \mathcal{A}(\mathbf{x} - \mathbf{a})$, which is equivalent to $\mathbf{x} = \mathcal{A}^{-1}(\mathbf{z}) + \mathbf{a}$, we can write for any $\mathbf{y} \in \mathbb{V}^*$,

$$\begin{aligned} g^*(\mathbf{y}) &= \max_{\mathbf{x}} \{ \langle \mathbf{y}, \mathbf{x} \rangle - g(\mathbf{x}) \} \\ &= \max_{\mathbf{x}} \{ \langle \mathbf{y}, \mathbf{x} \rangle - f(\mathcal{A}(\mathbf{x} - \mathbf{a})) - \langle \mathbf{b}, \mathbf{x} \rangle - c \} \\ &= \max_{\mathbf{z}} \{ \langle \mathbf{y}, \mathcal{A}^{-1}(\mathbf{z}) + \mathbf{a} \rangle - f(\mathbf{z}) - \langle \mathbf{b}, \mathcal{A}^{-1}(\mathbf{z}) + \mathbf{a} \rangle - c \} \\ &= \max_{\mathbf{z}} \{ \langle \mathbf{y} - \mathbf{b}, \mathcal{A}^{-1}(\mathbf{z}) \rangle - f(\mathbf{z}) + \langle \mathbf{a}, \mathbf{y} \rangle - \langle \mathbf{a}, \mathbf{b} \rangle - c \} \\ &= \max_{\mathbf{z}} \{ \langle (\mathcal{A}^{-1})^T(\mathbf{y} - \mathbf{b}), \mathbf{z} \rangle - f(\mathbf{z}) + \langle \mathbf{a}, \mathbf{y} \rangle - \langle \mathbf{a}, \mathbf{b} \rangle - c \} \\ &= f^*((\mathcal{A}^T)^{-1}(\mathbf{y} - \mathbf{b})) + \langle \mathbf{a}, \mathbf{y} \rangle - c - \langle \mathbf{a}, \mathbf{b} \rangle, \end{aligned}$$

where in the last equality we also used the fact that $(\mathcal{A}^{-1})^T = (\mathcal{A}^T)^{-1}$. \square

Theorem 4.14 (conjugate of $\alpha f(\cdot)$ and $\alpha f(\cdot/\alpha)$). Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be an extended real-valued function and let $\alpha \in \mathbb{R}_{++}$.

(a) The conjugate of the function $g(\mathbf{x}) = \alpha f(\mathbf{x})$ is given by

$$g^*(\mathbf{y}) = \alpha f^*\left(\frac{\mathbf{y}}{\alpha}\right), \quad \mathbf{y} \in \mathbb{E}^*.$$

(b) The conjugate of the function $h(\mathbf{x}) = \alpha f\left(\frac{\mathbf{x}}{\alpha}\right)$ is given by

$$h^*(\mathbf{y}) = \alpha f^*(\mathbf{y}), \quad \mathbf{y} \in \mathbb{E}^*.$$

Proof. For any $\mathbf{y} \in \mathbb{E}^*$,

$$\begin{aligned} g^*(\mathbf{y}) &= \max_{\mathbf{x}} \{ \langle \mathbf{y}, \mathbf{x} \rangle - g(\mathbf{x}) \} \\ &= \max_{\mathbf{x}} \{ \langle \mathbf{y}, \mathbf{x} \rangle - \alpha f(\mathbf{x}) \} \\ &= \alpha \max_{\mathbf{x}} \left\{ \left\langle \frac{\mathbf{y}}{\alpha}, \mathbf{x} \right\rangle - f(\mathbf{x}) \right\} \\ &= \alpha f^*\left(\frac{\mathbf{y}}{\alpha}\right), \end{aligned}$$

proving (a). The proof of (b) follows by the following chain of equalities:

$$\begin{aligned}
 h^*(\mathbf{y}) &= \max_{\mathbf{x}} \{ \langle \mathbf{y}, \mathbf{x} \rangle - h(\mathbf{x}) \} \\
 &= \max_{\mathbf{x}} \left\{ \langle \mathbf{y}, \mathbf{x} \rangle - \alpha f\left(\frac{\mathbf{x}}{\alpha}\right) \right\} \\
 &= \alpha \max_{\mathbf{x}} \left\{ \left\langle \mathbf{y}, \frac{\mathbf{x}}{\alpha} \right\rangle - f\left(\frac{\mathbf{x}}{\alpha}\right) \right\} \\
 &\stackrel{\mathbf{z} \leftarrow \frac{\mathbf{x}}{\alpha}}{=} \alpha \max_{\mathbf{z}} \{ \langle \mathbf{y}, \mathbf{z} \rangle - f(\mathbf{z}) \} \\
 &= \alpha f^*(\mathbf{y}). \quad \square
 \end{aligned}$$

The table below summarizes the four calculus rules discussed in this section.

$g(\mathbf{x})$	$g^*(\mathbf{y})$	Reference
$\sum_{i=1}^m f_i(\mathbf{x}_i)$	$\sum_{i=1}^m f_i^*(\mathbf{y}_i)$	Theorem 4.12
$\alpha f(\mathbf{x})$ ($\alpha > 0$)	$\alpha f^*(\mathbf{y}/\alpha)$	Theorem 4.14
$\alpha f(\mathbf{x}/\alpha)$ ($\alpha > 0$)	$\alpha f^*(\mathbf{y})$	Theorem 4.14
$f(\mathcal{A}(\mathbf{x} - \mathbf{a})) + \langle \mathbf{b}, \mathbf{x} \rangle + c$	$f^*((\mathcal{A}^T)^{-1}(\mathbf{y} - \mathbf{b})) + \langle \mathbf{a}, \mathbf{y} \rangle - c - \langle \mathbf{a}, \mathbf{b} \rangle$	Theorem 4.13

4.4 Examples

In this section we compute the conjugate functions of several fundamental convex functions. The first examples are one-dimensional, while the rest are multidimensional.

4.4.1 Exponent

Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be given by $f(x) = e^x$. Then for any $y \in \mathbb{R}$,

$$f^*(y) = \max_x \{ xy - e^x \}. \quad (4.5)$$

If $y < 0$, then the maximum value of the above problem is ∞ (easily seen by taking $x \rightarrow -\infty$). If $y = 0$, then obviously the maximal value (which is not attained) is 0. If $y > 0$, the unique maximizer of (4.5) is $x = \tilde{x} \equiv \log y$. Consequently, $f^*(y) = \tilde{x}y - e^{\tilde{x}} = y \log y - y$ for any $y > 0$. Using the convention $0 \log 0 \equiv 0$, we can finally deduce that

$$f^*(y) = \begin{cases} y \log y - y, & y \geq 0, \\ \infty & \text{else.} \end{cases}$$

4.4.2 Negative Log

Let $f : \mathbb{R} \rightarrow (-\infty, \infty]$ be given by

$$f(x) = \begin{cases} -\log(x), & x > 0, \\ \infty, & x \leq 0. \end{cases}$$

For any $y \in \mathbb{R}$,

$$f^*(y) = \max_{x>0} \{xy - f(x)\} = \max_{x>0} \{xy + \log(x)\}. \quad (4.6)$$

If $y \geq 0$, then the maximum value of the above problem is ∞ (since the objective function in (4.6) goes to ∞ as $x \rightarrow \infty$). If $y < 0$, the unique optimal solution of (4.6) is attained at $\tilde{x} = -\frac{1}{y}$, and hence for $y < 0$ we have $f^*(y) = \tilde{x}y + \log(\tilde{x}) = -1 - \log(-y)$. To conclude,

$$f^*(y) = \begin{cases} -1 - \log(-y), & y < 0, \\ \infty, & y \geq 0. \end{cases}$$

4.4.3 Hinge Loss

Consider the one-dimensional function $f : \mathbb{R} \rightarrow \mathbb{R}$ given by

$$f(x) = \max\{1 - x, 0\}.$$

Then for any $y \in \mathbb{R}$,

$$f^*(y) = \max_x [yx - \max\{1 - x, 0\}] = \max_x [\min\{(1+y)x - 1, yx\}]. \quad (4.7)$$

The objective function in the above maximization problem can be rewritten as

$$\min\{(1+y)x - 1, yx\} = \begin{cases} (1+y)x - 1, & x < 1, \\ yx, & x \geq 1. \end{cases}$$

Thus, the objective function is a continuous piecewise linear function comprising two pieces: a line with slope $1+y$ over $(-\infty, 1]$ and a line with slope y over $[1, \infty)$. Therefore, a maximizer exists if the slope of the left line is nonnegative ($1+y \geq 0$) and the slope of the right line is nonpositive ($y \leq 0$). Consequently, a maximizer exists for the problem in (4.7) if and only if $y \in [-1, 0]$, and in that case it is attained at $x = 1$, with y being the corresponding optimal value. To summarize,

$$f^*(y) = y + \delta_{[-1,0]}(y), \quad y \in \mathbb{R}.$$

4.4.4 $\frac{1}{p}|\cdot|^p$ ($p > 1$)

Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be given by $f(x) = \frac{1}{p}|x|^p$, where $p > 1$. For any $y \in \mathbb{R}$,

$$f^*(y) = \max_x \left\{ xy - \frac{1}{p}|x|^p \right\}. \quad (4.8)$$

Since the problem in (4.8) consists of maximizing a differentiable concave function over \mathbb{R} , its optimal solutions are the points \tilde{x} in which the derivative vanishes:

$$y - \text{sgn}(\tilde{x})|\tilde{x}|^{p-1} = 0.$$

Therefore, $\operatorname{sgn}(\tilde{x}) = \operatorname{sgn}(y)$ and $|\tilde{x}|^{p-1} = |y|$, implying that $\tilde{x} = \operatorname{sgn}(y)|y|^{\frac{1}{p-1}}$. Thus,

$$f^*(y) = \tilde{x}y - \frac{1}{p}|\tilde{x}|^p = |y|^{1+\frac{1}{p-1}} - \frac{1}{p}|y|^{\frac{p}{p-1}} = \left(1 - \frac{1}{p}\right)|y|^{\frac{p}{p-1}} = \frac{1}{q}|y|^q,$$

where q is the positive number satisfying $\frac{1}{p} + \frac{1}{q} = 1$. To summarize,

$$f^*(y) = \frac{1}{q}|y|^q, \quad y \in \mathbb{R}.$$

4.4.5 $-\frac{(\cdot)^p}{p}$ ($0 < p < 1$)

Let $f : \mathbb{R} \rightarrow (-\infty, \infty]$ be given by

$$f(x) = \begin{cases} -\frac{x^p}{p}, & x \geq 0, \\ \infty, & x < 0. \end{cases}$$

For any $y \in \mathbb{R}$,

$$f^*(y) = \max_x \{xy - f(x)\} = \max_{x \geq 0} \left\{ g(x) \equiv xy + \frac{x^p}{p} \right\}.$$

When $y \geq 0$, the value of the above problem is ∞ since $g(x) \rightarrow \infty$ as $x \rightarrow \infty$. If $y < 0$, then the derivative of $g(x)$ vanishes at $x = \tilde{x} \equiv (-y)^{\frac{1}{p-1}} > 0$, and since g is concave, it follows that \tilde{x} is a global maximizer of g . Therefore,

$$f^*(y) = \tilde{x}y + \frac{\tilde{x}^p}{p} = -(-y)^{\frac{p}{p-1}} + \frac{1}{p}(-y)^{\frac{p}{p-1}} = -\frac{(-y)^q}{q},$$

where q is the negative number for which $\frac{1}{p} + \frac{1}{q} = 1$. To summarize,

$$f^*(y) = \begin{cases} -\frac{(-y)^q}{q}, & y < 0, \\ \infty, & \text{else.} \end{cases}$$

4.4.6 Strictly Convex Quadratic Functions

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} + \mathbf{b}^T \mathbf{x} + c$, where $\mathbf{A} \in \mathbb{S}_{++}^n$, $\mathbf{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$. We use our convention that (unless otherwise stated) \mathbb{R}^n is endowed with the dot product, meaning that $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$. For any $\mathbf{y} \in \mathbb{R}^n$,

$$\begin{aligned} f^*(\mathbf{y}) &= \max_{\mathbf{x}} \{\mathbf{y}^T \mathbf{x} - f(\mathbf{x})\} \\ &= \max_{\mathbf{x}} \left\{ \mathbf{y}^T \mathbf{x} - \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{b}^T \mathbf{x} - c \right\} \\ &= \max_{\mathbf{x}} \left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} - (\mathbf{b} - \mathbf{y})^T \mathbf{x} - c \right\}. \end{aligned}$$

The maximum in the above problem is attained at $\mathbf{x} = \mathbf{A}^{-1}(\mathbf{y} - \mathbf{b})$, leading to the following expression for the conjugate function:

$$f^*(\mathbf{y}) = \frac{1}{2}(\mathbf{y} - \mathbf{b})^T \mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}) - c.$$

4.4.7 Convex Quadratic Functions

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} + \mathbf{b}^T \mathbf{x} + c$, where $\mathbf{A} \in \mathbb{S}_+^n$, $\mathbf{b} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. The only difference between this example and the previous one is the fact that here \mathbf{A} is not necessarily positive definite but is assumed to be only positive *semidefinite*. For any $\mathbf{y} \in \mathbb{R}^n$,

$$f^*(\mathbf{y}) = \max_{\mathbf{x}} \{\mathbf{y}^T \mathbf{x} - f(\mathbf{x})\} = \max_{\mathbf{x}} \left\{ g(\mathbf{x}) \equiv -\frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} + (\mathbf{y} - \mathbf{b})^T \mathbf{x} - c \right\}.$$

Since g is concave and differentiable over \mathbb{R}^n , it follows that the maximizers of the above problem are the points for which the gradient vanishes, namely, points \mathbf{x} satisfying

$$\mathbf{A}\mathbf{x} = \mathbf{y} - \mathbf{b}. \quad (4.9)$$

This system has a solution if and only if $\mathbf{y} \in \mathbf{b} + \text{Range}(\mathbf{A})$, and in that case we can choose one of the solutions to the system (4.9), for example, $\tilde{\mathbf{x}} = \mathbf{A}^\dagger(\mathbf{y} - \mathbf{b})$, where \mathbf{A}^\dagger is the Moore–Penrose pseudoinverse of \mathbf{A} . We can now compute $f^*(\mathbf{y})$ as follows:

$$\begin{aligned} f^*(\mathbf{y}) &= -\frac{1}{2}\tilde{\mathbf{x}}^T \mathbf{A}\tilde{\mathbf{x}} - (\mathbf{b} - \mathbf{y})^T \tilde{\mathbf{x}} - c \\ &= -\frac{1}{2}(\mathbf{y} - \mathbf{b})^T \mathbf{A}^\dagger \mathbf{A} \mathbf{A}^\dagger(\mathbf{y} - \mathbf{b}) - (\mathbf{b} - \mathbf{y})^T \mathbf{A}^\dagger(\mathbf{y} - \mathbf{b}) - c \\ &= \frac{1}{2}(\mathbf{y} - \mathbf{b})^T \mathbf{A}^\dagger(\mathbf{y} - \mathbf{b}) - c, \end{aligned}$$

where we used the fact that the Moore–Penrose pseudoinverse of a symmetric matrix is symmetric, as well as the known identity $\mathbf{A}^\dagger \mathbf{A} \mathbf{A}^\dagger = \mathbf{A}^\dagger$. We are left with the case where $\mathbf{y} - \mathbf{b} \notin \text{Range}(\mathbf{A})$. We will show that in this case $f^*(\mathbf{y}) = \infty$. Indeed, since $\text{Range}(\mathbf{A}) = \text{Null}(\mathbf{A})^\perp$, it follows that $\mathbf{y} - \mathbf{b} \notin \text{Null}(\mathbf{A})^\perp$, meaning that there exists a vector $\mathbf{v} \in \text{Null}(\mathbf{A})$ such that $(\mathbf{y} - \mathbf{b})^T \mathbf{v} > 0$. Note that for any $\alpha \in \mathbb{R}$,

$$g(\alpha \mathbf{v}) = \alpha(\mathbf{y} - \mathbf{b})^T \mathbf{v} - c,$$

and hence $g(\alpha \mathbf{v}) \rightarrow \infty$ as α tends to ∞ , establishing the fact that $f^*(\mathbf{y}) = \infty$ whenever $\mathbf{y} \notin \mathbf{b} + \text{Range}(\mathbf{A})$. To conclude,

$$f^*(\mathbf{y}) = \begin{cases} \frac{1}{2}(\mathbf{y} - \mathbf{b})^T \mathbf{A}^\dagger(\mathbf{y} - \mathbf{b}) - c, & \mathbf{y} \in \mathbf{b} + \text{Range}(\mathbf{A}), \\ \infty & \text{else.} \end{cases}$$

4.4.8 Negative Entropy

Let $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ be given by

$$f(\mathbf{x}) = \begin{cases} \sum_{i=1}^n x_i \log x_i, & \mathbf{x} \geq \mathbf{0}, \\ \infty & \text{else.} \end{cases}$$

Since the function is separable, it is enough to compute the conjugate of the scalar function g defined by $g(t) = t \log t$ for $t \geq 0$ and ∞ for $t < 0$. For any $s \in \mathbb{R}$,

$$g^*(s) = \max_t \{ts - g(t)\} = \max_{t \geq 0} \{ts - t \log t\}.$$

The maximum of the above problem is attained at $t = e^{s-1}$, and hence the conjugate is given by

$$g^*(s) = se^{s-1} - (s-1)e^{s-1} = e^{s-1}.$$

Since $f(\mathbf{x}) = \sum_{i=1}^n g(x_i)$, it follows by Theorem 4.12 that for any $\mathbf{y} \in \mathbb{R}^n$,

$$f^*(\mathbf{y}) = \sum_{i=1}^n g^*(y_i) = \sum_{i=1}^n e^{y_i-1}.$$

4.4.9 Negative Sum of Logs

Let $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ be given by

$$f(\mathbf{x}) = \begin{cases} -\sum_{i=1}^n \log x_i, & \mathbf{x} > \mathbf{0}, \\ \infty & \text{else.} \end{cases}$$

Note that $f(\mathbf{x}) = \sum_{i=1}^n g(x_i)$, where $g(t) = -\log t$ for $t > 0$ and ∞ for $t \leq 0$. Therefore, invoking Theorem 4.12,

$$f^*(\mathbf{x}) = \sum_{i=1}^n g^*(x_i).$$

By Section 4.4.2, $g^*(y) = -1 - \log(-y)$ for $y < 0$ and ∞ otherwise. Therefore,

$$f^*(\mathbf{y}) = \begin{cases} -n - \sum_{i=1}^n \log(-y_i), & \mathbf{y} < \mathbf{0}, \\ \infty & \text{else.} \end{cases}$$

4.4.10 Negative Entropy over the Unit Simplex

Let $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ be given by

$$f(\mathbf{x}) = \begin{cases} \sum_{i=1}^n x_i \log x_i, & \mathbf{x} \in \Delta_n, \\ \infty & \text{else.} \end{cases} \quad (4.10)$$

For any $\mathbf{y} \in \mathbb{R}^n$,

$$f^*(\mathbf{y}) = \max \left\{ \sum_{i=1}^n y_i x_i - \sum_{i=1}^n x_i \log x_i : \sum_{i=1}^n x_i = 1, x_1, x_2, \dots, x_n \geq 0 \right\}.$$

By Example 3.71, the optimal solution of the above maximization problem is

$$x_i^* = \frac{e^{y_i}}{\sum_{j=1}^n e^{y_j}}, \quad i = 1, 2, \dots, n,$$

with a corresponding optimal value of

$$f^*(\mathbf{y}) = \sum_{i=1}^n y_i x_i^* - \sum_{i=1}^n x_i^* \log x_i^* = \log \left(\sum_{j=1}^n e^{y_j} \right).$$

That is, the conjugate of the negative entropy is the log-sum-exp function.

4.4.11 log-sum-exp

Let $g : \mathbb{R}^n \rightarrow \mathbb{R}$ be given by

$$g(\mathbf{x}) = \log \left(\sum_{j=1}^n e^{x_j} \right).$$

By Section 4.4.10, $g = f^*$, where f is the negative entropy over the unit simplex given by (4.10). Since f is proper closed and convex, it follows by Theorem 4.8 that $f^{**} = f$, and hence

$$g^* = f^{**} = f,$$

meaning that

$$g^*(\mathbf{y}) = \begin{cases} \sum_{i=1}^n y_i \log y_i, & \mathbf{y} \in \Delta_n, \\ \infty & \text{else.} \end{cases}$$

4.4.12 Norms

Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}) = \|\mathbf{x}\|$. Then, by Example 2.31,

$$f = \sigma_{B_{\|\cdot\|_*}[\mathbf{0}, 1]},$$

where we used the fact that the bidual norm $\|\cdot\|_{**}$ is identical to the norm $\|\cdot\|$. Hence, by Example 4.9,

$$f^* = \delta_{\text{cl}(\text{conv}(B_{\|\cdot\|_*}[\mathbf{0}, 1]))},$$

but since $B_{\|\cdot\|_*}[\mathbf{0}, 1]$ is closed and convex, $\text{cl}(\text{conv}(B_{\|\cdot\|_*}[\mathbf{0}, 1])) = B_{\|\cdot\|_*}[\mathbf{0}, 1]$, and therefore for any $\mathbf{y} \in \mathbb{E}^*$,

$$f^*(\mathbf{y}) = \delta_{B_{\|\cdot\|_*}[\mathbf{0}, 1]}(\mathbf{y}) = \begin{cases} 0, & \|\mathbf{y}\|_* \leq 1, \\ \infty & \text{else.} \end{cases}$$

4.4.13 Ball-Pen

Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be given by

$$f(\mathbf{x}) = \begin{cases} -\sqrt{1 - \|\mathbf{x}\|^2}, & \|\mathbf{x}\| \leq 1, \\ \infty & \text{else.} \end{cases}$$

To compute the conjugate function, we begin by rewriting it in a double maximization form:

$$\begin{aligned} f^*(\mathbf{y}) &= \max_{\mathbf{x}} \left\{ \langle \mathbf{y}, \mathbf{x} \rangle + \sqrt{1 - \|\mathbf{x}\|^2} : \|\mathbf{x}\| \leq 1 \right\} \\ &= \max_{\alpha \in [0, 1]} \max_{\mathbf{x} : \|\mathbf{x}\| = \alpha} \left\{ \langle \mathbf{y}, \mathbf{x} \rangle + \sqrt{1 - \alpha^2} \right\}. \end{aligned}$$

By the definition of the dual norm, the optimal value of the inner maximization problem is $\alpha \|\mathbf{y}\|_* + \sqrt{1 - \alpha^2}$, and we can therefore write, for any $\mathbf{y} \in \mathbb{E}^*$,

$$f^*(\mathbf{y}) = \max_{\alpha \in [0, 1]} \left\{ g(\alpha) \equiv \alpha \|\mathbf{y}\|_* + \sqrt{1 - \alpha^2} \right\}. \quad (4.11)$$

It is easy to see that the maximizer of g over $[0, 1]$ is

$$\tilde{\alpha} = \frac{\|\mathbf{y}\|_*}{\sqrt{\|\mathbf{y}\|_*^2 + 1}}.$$

Plugging $\alpha = \tilde{\alpha}$ into (4.11), we finally obtain that for any $\mathbf{y} \in \mathbb{E}^*$,

$$f^*(\mathbf{y}) = \sqrt{\|\mathbf{y}\|_*^2 + 1}.$$

It is also possible to generalize the result to functions of the form

$$f_\alpha(\mathbf{x}) = \begin{cases} -\sqrt{\alpha^2 - \|\mathbf{x}\|^2}, & \|\mathbf{x}\| \leq \alpha, \\ \infty & \text{else,} \end{cases}$$

where $\alpha \in \mathbb{R}_{++}$. In this notation, $f = f_1$. To compute f_α^* , note that $f_\alpha(\mathbf{x}) = \alpha f\left(\frac{\mathbf{x}}{\alpha}\right)$, and hence by Theorem 4.14(b) it follows that for any $\mathbf{y} \in \mathbb{E}^*$,

$$f_\alpha^*(\mathbf{y}) = \alpha f^*(\mathbf{y}) = \alpha \sqrt{1 + \|\mathbf{y}\|_*^2}.$$

4.4.14 $\sqrt{\alpha^2 + \|\cdot\|^2}$

Consider the function $g_\alpha : \mathbb{E} \rightarrow \mathbb{R}$ given by $g_\alpha(\mathbf{x}) = \sqrt{\alpha^2 + \|\mathbf{x}\|^2}$, where $\alpha > 0$. Then $g_\alpha(\mathbf{x}) = \alpha g\left(\frac{\mathbf{x}}{\alpha}\right)$, where $g(\mathbf{x}) = \sqrt{1 + \|\mathbf{x}\|^2}$. By Section 4.4.13, it follows that $g = f^*$, where f is given by

$$f(\mathbf{y}) = \begin{cases} -\sqrt{1 - \|\mathbf{y}\|_*^2}, & \|\mathbf{y}\|_* \leq 1, \\ \infty & \text{else.} \end{cases}$$

Since f is proper closed and convex, it follows by Theorem 4.8 that

$$g^* = f^{**} = f.$$

Finally, invoking Theorem 4.14(b), we conclude that for any $\mathbf{y} \in \mathbb{E}^*$,

$$g_\alpha^*(\mathbf{y}) = \alpha g^*(\mathbf{y}) = \alpha f(\mathbf{y}) = \begin{cases} -\alpha\sqrt{1 - \|\mathbf{y}\|_*^2}, & \|\mathbf{y}\|_* \leq 1, \\ \infty & \text{else.} \end{cases}$$

4.4.15 Squared Norm

Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2$, where $\|\cdot\|$ is the norm associated with the space \mathbb{E} . For any $\mathbf{y} \in \mathbb{E}^*$, we can write $f^*(\mathbf{y})$ as the optimal value of the following double maximization problem:

$$f^*(\mathbf{y}) = \max_{\mathbf{x} \in \mathbb{E}} \left\{ \langle \mathbf{y}, \mathbf{x} \rangle - \frac{1}{2}\|\mathbf{x}\|^2 \right\} = \max_{\alpha \geq 0} \max_{\mathbf{x}: \|\mathbf{x}\|=\alpha} \left\{ \langle \mathbf{y}, \mathbf{x} \rangle - \frac{1}{2}\alpha^2 \right\}.$$

Using the definition of the dual norm, it follows that

$$\max_{\mathbf{x} \in \mathbb{E}: \|\mathbf{x}\|=\alpha} \langle \mathbf{y}, \mathbf{x} \rangle = \alpha \|\mathbf{y}\|_*,$$

Hence,

$$f^*(\mathbf{y}) = \max_{\alpha \geq 0} \left\{ \alpha \|\mathbf{y}\|_* - \frac{1}{2}\alpha^2 \right\} = \frac{1}{2}\|\mathbf{y}\|_*^2.$$

4.4.16 Summary of Conjugate Computations

The table below summarizes all the computations of conjugate functions described in this chapter.

$f(\mathbf{x})$	$\text{dom}(f)$	f^*	Assumptions	Reference
e^x	\mathbb{R}	$y \log y - y$ ($\text{dom}(f^*) = \mathbb{R}_+$)	—	Section 4.4.1
$-\log x$	\mathbb{R}_{++}	$-1 - \log(-y)$ ($\text{dom}(f^*) = \mathbb{R}_{--}$)	—	Section 4.4.2
$\max\{1 - x, 0\}$	\mathbb{R}	$y + \delta_{[-1,0]}(y)$	—	Section 4.4.3
$\frac{1}{p} x ^p$	\mathbb{R}	$\frac{1}{q} y ^q$	$p > 1, \frac{1}{p} + \frac{1}{q} = 1$	Section 4.4.4
$-\frac{x^p}{p}$	\mathbb{R}_+	$-\frac{(-y)^q}{q}$ ($\text{dom}(f^*) = \mathbb{R}_{--}$)	$0 < p < 1, \frac{1}{p} + \frac{1}{q} = 1$	Section 4.4.5
$\frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} + \mathbf{b}^T \mathbf{x} + c$	\mathbb{R}^n	$\frac{1}{2}(\mathbf{y} - \mathbf{b})^T \mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}) - c$	$\mathbf{A} \in \mathbb{S}_{++}^n, \mathbf{b} \in \mathbb{R}^n, c \in \mathbb{R}$	Section 4.4.6
$\frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} + \mathbf{b}^T \mathbf{x} + c$	\mathbb{R}^n	$\frac{1}{2}(\mathbf{y} - \mathbf{b})^T \mathbf{A}^\dagger(\mathbf{y} - \mathbf{b}) - c$ ($\text{dom}(f^*) = \mathbf{b} + \text{Range}(\mathbf{A})$)	$\mathbf{A} \in \mathbb{S}_+^n, \mathbf{b} \in \mathbb{R}^n, c \in \mathbb{R}$	Section 4.4.7
$\sum_{i=1}^n x_i \log x_i$	\mathbb{R}_+^n	$\sum_{i=1}^n e^{y_i-1}$	—	Section 4.4.8
$\sum_{i=1}^n x_i \log x_i$	Δ_n	$\log(\sum_{i=1}^n e^{y_i})$	—	Section 4.4.10
$-\sum_{i=1}^n \log x_i$	\mathbb{R}_{++}^n	$-n - \sum_{i=1}^n \log(-y_i)$ ($\text{dom}(f^*) = \mathbb{R}_{--}^n$)	—	Section 4.4.9
$\log(\sum_{i=1}^n e^{x_i})$	\mathbb{R}^n	$\sum_{i=1}^n y_i \log y_i$ ($\text{dom}(f^*) = \Delta_n$)	—	Section 4.4.11
$\max_i\{x_i\}$	\mathbb{R}^n	$\delta_{\Delta_n}(\mathbf{y})$	—	Example 4.10
$\delta_C(\mathbf{x})$	C	$\sigma_C(\mathbf{y})$	$\emptyset \neq C \subseteq \mathbb{E}$	Example 4.2
$\sigma_C(\mathbf{x})$	$\text{dom}(\sigma_C)$	$\delta_{\text{cl}(\text{conv}(C))}(\mathbf{y})$	$\emptyset \neq C \subseteq \mathbb{E}$	Example 4.9
$\ \mathbf{x}\ $	\mathbb{E}	$\delta_{B_{\ \cdot\ _*}[\mathbf{0}, 1]}(\mathbf{y})$	—	Section 4.4.12
$-\sqrt{\alpha^2 - \ \mathbf{x}\ ^2}$	$B[\mathbf{0}, \alpha]$	$\alpha\sqrt{\ \mathbf{y}\ _*^2 + 1}$	$\alpha > 0$	Section 4.4.13
$\sqrt{\alpha^2 + \ \mathbf{x}\ ^2}$	\mathbb{E}	$-\alpha\sqrt{1 - \ \mathbf{y}\ _*^2}$ ($\text{dom}f^* = B_{\ \cdot\ _*}[\mathbf{0}, 1]$)	$\alpha > 0$	Section 4.4.14
$\frac{1}{2}\ \mathbf{x}\ ^2$	\mathbb{E}	$\frac{1}{2}\ \mathbf{y}\ _*^2$	—	Section 4.4.15
$\frac{1}{2}\ \mathbf{x}\ ^2 + \delta_C(\mathbf{x})$	C	$\frac{1}{2}\ \mathbf{y}\ ^2 - \frac{1}{2}d_C^2(\mathbf{y})$	$\emptyset \neq C \subseteq \mathbb{E}$, \mathbb{E} Euclidean	Example 4.4
$\frac{1}{2}\ \mathbf{x}\ ^2 - \frac{1}{2}d_C^2(\mathbf{x})$	\mathbb{E}	$\frac{1}{2}\ \mathbf{y}\ ^2 + \delta_C(\mathbf{y})$	$\emptyset \neq C \subseteq \mathbb{E}$ closed convex. \mathbb{E} Euclidean	Example 4.11

4.4.17 Fenchel's Duality Theorem

Conjugate functions naturally appear in dual problems most prominently in the celebrated *Fenchel's duality theorem*, which we now recall. Consider the problem

$$(P) \min_{\mathbf{x} \in \mathbb{E}} f(\mathbf{x}) + g(\mathbf{x}).$$

We begin by rewriting the problem as

$$\min_{\mathbf{x}, \mathbf{z} \in \mathbb{E}} \{f(\mathbf{x}) + g(\mathbf{z}) : \mathbf{x} = \mathbf{z}\}$$

and then constructing the Lagrangian

$$L(\mathbf{x}, \mathbf{z}; \mathbf{y}) = f(\mathbf{x}) + g(\mathbf{z}) + \langle \mathbf{y}, \mathbf{z} - \mathbf{x} \rangle = -[\langle \mathbf{y}, \mathbf{x} \rangle - f(\mathbf{x})] - [\langle -\mathbf{y}, \mathbf{z} \rangle - g(\mathbf{z})].$$

The dual objective function is computed by minimizing the Lagrangian w.r.t. the primal variables \mathbf{x}, \mathbf{z} :

$$q(\mathbf{y}) = \min_{\mathbf{x}, \mathbf{z}} L(\mathbf{x}, \mathbf{z}; \mathbf{y}) = -f^*(\mathbf{y}) - g^*(-\mathbf{y}).$$

We thus obtain the following dual problem, which is also called *Fenchel's dual*:

$$(D) \max_{\mathbf{y} \in \mathbb{E}^*} \{-f^*(\mathbf{y}) - g^*(-\mathbf{y})\}.$$

Fenchel's duality theorem, which we recall below, provides conditions under which strong duality holds for the pair of problems (P) and (D).

Theorem 4.15 (Fenchel's duality theorem [108, Theorem 31.1]). *Let $f, g : \mathbb{E} \rightarrow (-\infty, \infty]$ be proper convex functions. If $\text{ri}(\text{dom}(f)) \cap \text{ri}(\text{dom}(g)) \neq \emptyset$, then*

$$\min_{\mathbf{x} \in \mathbb{E}} \{f(\mathbf{x}) + g(\mathbf{x})\} = \max_{\mathbf{y} \in \mathbb{E}^*} \{-f^*(\mathbf{y}) - g^*(-\mathbf{y})\},$$

and the maximum in the right-hand problem is attained whenever it is finite.

4.5 Infimal Convolution and Conjugacy

We will now show that in some sense the operations of addition and infimal convolution are dual to each other under the conjugacy operation. The first result holds under the very mild condition of properness of the functions.

Theorem 4.16 (conjugate of infimal convolution). *For two proper functions $h_1, h_2 : \mathbb{E} \rightarrow (-\infty, \infty]$ it holds that*

$$(h_1 \square h_2)^* = h_1^* + h_2^*.$$

Proof. For every $\mathbf{y} \in \mathbb{E}^*$ one has

$$\begin{aligned}
(h_1 \square h_2)^*(\mathbf{y}) &= \max_{\mathbf{x} \in \mathbb{E}} \{ \langle \mathbf{y}, \mathbf{x} \rangle - (h_1 \square h_2)(\mathbf{x}) \} \\
&= \max_{\mathbf{x} \in \mathbb{E}} \{ \langle \mathbf{y}, \mathbf{x} \rangle - \min_{\mathbf{u} \in \mathbb{E}} \{ h_1(\mathbf{u}) + h_2(\mathbf{x} - \mathbf{u}) \} \} \\
&= \max_{\mathbf{x} \in \mathbb{E}} \max_{\mathbf{u} \in \mathbb{E}} \{ \langle \mathbf{y}, \mathbf{x} \rangle - h_1(\mathbf{u}) - h_2(\mathbf{x} - \mathbf{u}) \} \\
&= \max_{\mathbf{x} \in \mathbb{E}} \max_{\mathbf{u} \in \mathbb{E}} \{ \langle \mathbf{y}, \mathbf{x} - \mathbf{u} \rangle + \langle \mathbf{y}, \mathbf{u} \rangle - h_1(\mathbf{u}) - h_2(\mathbf{x} - \mathbf{u}) \} \\
&= \max_{\mathbf{u} \in \mathbb{E}} \max_{\mathbf{x} \in \mathbb{E}} \{ \langle \mathbf{y}, \mathbf{x} - \mathbf{u} \rangle + \langle \mathbf{y}, \mathbf{u} \rangle - h_1(\mathbf{u}) - h_2(\mathbf{x} - \mathbf{u}) \} \\
&= \max_{\mathbf{u} \in \mathbb{E}} \{ h_2^*(\mathbf{y}) + \langle \mathbf{y}, \mathbf{u} \rangle - h_1(\mathbf{u}) \} \\
&= h_1^*(\mathbf{y}) + h_2^*(\mathbf{y}). \quad \square
\end{aligned}$$

The second “direction” is a much deeper result requiring additional assumptions like convexity of the functions under consideration.

Theorem 4.17 (conjugate of sum). *Let $h_1 : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper convex function and $h_2 : \mathbb{E} \rightarrow \mathbb{R}$ be a real-valued convex function. Then*

$$(h_1 + h_2)^* = h_1^* \square h_2^*.$$

Proof. For any $\mathbf{y} \in \mathbb{E}^*$,

$$\begin{aligned}
(h_1 + h_2)^*(\mathbf{y}) &= \max_{\mathbf{x} \in \mathbb{E}} \{ \langle \mathbf{y}, \mathbf{x} \rangle - h_1(\mathbf{x}) - h_2(\mathbf{x}) \} \\
&= - \min_{\mathbf{x} \in \mathbb{E}} \{ h_1(\mathbf{x}) + h_2(\mathbf{x}) - \langle \mathbf{y}, \mathbf{x} \rangle \} \\
&= - \min_{\mathbf{x} \in \mathbb{E}} \{ h_1(\mathbf{x}) + g(\mathbf{x}) \},
\end{aligned} \tag{4.12}$$

where $g(\mathbf{x}) \equiv h_2(\mathbf{x}) - \langle \mathbf{y}, \mathbf{x} \rangle$. Note that

$$\text{ri}(\text{dom}(h_1)) \cap \text{ri}(\text{dom}(g)) = \text{ri}(\text{dom}(h_1)) \cap \mathbb{E} = \text{ri}(\text{dom}(h_1)) \neq \emptyset,$$

and we can thus employ Fenchel’s duality theorem (Theorem 4.15) and obtain the following equality:

$$\min_{\mathbf{x} \in \mathbb{E}} \{ h_1(\mathbf{x}) + g(\mathbf{x}) \} = \max_{\mathbf{z} \in \mathbb{E}^*} \{ -h_1^*(\mathbf{z}) - g^*(-\mathbf{z}) \} = \max_{\mathbf{z} \in \mathbb{E}^*} \{ -h_1^*(\mathbf{z}) - h_2^*(\mathbf{y} - \mathbf{z}) \}. \tag{4.13}$$

Combining (4.12) and (4.13), we finally obtain that for any $\mathbf{y} \in \mathbb{E}^*$,

$$(h_1 + h_2)^*(\mathbf{y}) = \min_{\mathbf{z} \in \mathbb{E}^*} \{ h_1^*(\mathbf{z}) + h_2^*(\mathbf{y} - \mathbf{z}) \} = (h_1^* \square h_2^*)(\mathbf{y}),$$

establishing the desired result. \square

Corollary 4.18. *Let $h_1 : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper closed convex function and $h_2 : \mathbb{E} \rightarrow \mathbb{R}$ be a real-valued convex function. Then*

$$h_1 + h_2 = (h_1^* \square h_2^*)^*.$$

Proof. The function $h_1 + h_2$ is obviously proper and is closed by the closedness of h_1, h_2 (Theorem 2.7(b)). Therefore, by Theorem 4.8, $(h_1 + h_2)^{**} = h_1 + h_2$, which, combined with Theorem 4.17, yields

$$h_1 + h_2 = (h_1 + h_2)^{**} = [(h_1 + h_2)^*]^* = (h_1^* \square h_2^*)^*. \quad \square$$

The next result shows a representation of the infimal convolution in terms of the corresponding conjugate functions.

Theorem 4.19 (representation of the infimal convolution by conjugates). *Let $h_1 : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper convex function, and let $h_2 : \mathbb{E} \rightarrow \mathbb{R}$ be a real-valued convex function. Suppose that $h_1 \square h_2$ is a real-valued function. Then*

$$h_1 \square h_2 = (h_1^* + h_2^*)^*. \quad (4.14)$$

Proof. By Theorem 4.16,

$$(h_1 \square h_2)^* = h_1^* + h_2^*. \quad (4.15)$$

Since h_1 is proper and convex and h_2 is real-valued and convex, it follows by Theorem 2.19 that $h_1 \square h_2$ is convex. Since $h_1 \square h_2$ is real-valued, it is in particular proper and closed. Therefore, by Theorem 4.8, $(h_1 \square h_2)^{**} = h_1 \square h_2$. Hence, taking the conjugate of both sides of (4.15), the identity (4.14) follows. \square

4.6 Subdifferentials of Conjugate Functions

The main result concerning the subdifferential of a conjugate function is the so-called *conjugate subgradient theorem*.

Theorem 4.20 (conjugate subgradient theorem). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be proper and convex. The following two claims are equivalent for any $\mathbf{x} \in \mathbb{E}, \mathbf{y} \in \mathbb{E}^*$:*

$$(i) \quad \langle \mathbf{x}, \mathbf{y} \rangle = f(\mathbf{x}) + f^*(\mathbf{y}).$$

$$(ii) \quad \mathbf{y} \in \partial f(\mathbf{x}).$$

If in addition f is closed, then (i) and (ii) are equivalent to

$$(iii) \quad \mathbf{x} \in \partial f^*(\mathbf{y}).$$

Proof. The relation $\mathbf{y} \in \partial f(\mathbf{x})$ holds if and only if

$$f(\mathbf{z}) \geq f(\mathbf{x}) + \langle \mathbf{y}, \mathbf{z} - \mathbf{x} \rangle \text{ for all } \mathbf{z} \in \mathbb{E},$$

which is the same as

$$\langle \mathbf{y}, \mathbf{x} \rangle - f(\mathbf{x}) \geq \langle \mathbf{y}, \mathbf{z} \rangle - f(\mathbf{z}) \text{ for all } \mathbf{z} \in \mathbb{E}. \quad (4.16)$$

Taking the maximum over \mathbf{z} , we obtain that (4.16) is the same as

$$\langle \mathbf{y}, \mathbf{x} \rangle - f(\mathbf{x}) \geq f^*(\mathbf{y}),$$

which by Fenchel's inequality (Theorem 4.6) is equivalent to the equality $\langle \mathbf{x}, \mathbf{y} \rangle = f(\mathbf{x}) + f^*(\mathbf{y})$. We have thus established the equivalence between (i) and (ii). Assume now that in addition f is closed. Then by Theorem 4.8, $f^{**} = f$, which in particular implies that (i) is equivalent to

$$\langle \mathbf{x}, \mathbf{y} \rangle = g(\mathbf{y}) + g^*(\mathbf{x}),$$

where $g = f^*$. By the same equivalence that was already established between (i) and (ii) (but here employed on g), we conclude that (i) is equivalent to $\mathbf{x} \in \partial g(\mathbf{y}) = \partial f^*(\mathbf{y})$. \blacksquare

By the definition of the conjugate function, claim (i) in Theorem 4.20 can be rewritten as

$$\mathbf{x} \in \operatorname{argmax}_{\tilde{\mathbf{x}} \in \mathbb{E}} \{ \langle \mathbf{y}, \tilde{\mathbf{x}} \rangle - f(\tilde{\mathbf{x}}) \},$$

and, when f is closed, also as

$$\mathbf{y} \in \operatorname{argmax}_{\tilde{\mathbf{y}} \in \mathbb{E}^*} \{ \langle \mathbf{x}, \tilde{\mathbf{y}} \rangle - f^*(\tilde{\mathbf{y}}) \}.$$

Equipped with the above observation, we can conclude that the conjugate subgradient theorem, in the case where f is closed, can also be equivalently formulated as follows.

Corollary 4.21 (conjugate subgradient theorem—second formulation). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper closed convex function. Then for any $\mathbf{x} \in \mathbb{E}, \mathbf{y} \in \mathbb{E}^*$,*

$$\partial f(\mathbf{x}) = \operatorname{argmax}_{\tilde{\mathbf{y}} \in \mathbb{E}^*} \{ \langle \mathbf{x}, \tilde{\mathbf{y}} \rangle - f^*(\tilde{\mathbf{y}}) \}$$

and

$$\partial f^*(\mathbf{y}) = \operatorname{argmax}_{\tilde{\mathbf{x}} \in \mathbb{E}} \{ \langle \mathbf{y}, \tilde{\mathbf{x}} \rangle - f(\tilde{\mathbf{x}}) \}.$$

In particular, we can also conclude that for any proper closed convex function f ,

$$\partial f(0) = \operatorname{argmin}_{\mathbf{y} \in \mathbb{E}^*} f^*(\mathbf{y})$$

and

$$\partial f^*(0) = \operatorname{argmin}_{\mathbf{x} \in \mathbb{E}} f(\mathbf{x}).$$

Example 4.22. Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}) = \|\mathbf{x}\|$. Obviously, f is proper, closed, and convex. By Example 2.31, $f = \sigma_{B_{\|\cdot\|_*}[\mathbf{0}, 1]}$. Therefore, by Example 4.9, $f^* = \delta_{B_{\|\cdot\|_*}[\mathbf{0}, 1]}$. We can now use the conjugate subgradient theorem (Corollary 4.21) and compute the subdifferential set of f at $\mathbf{0}$ as follows:

$$\partial f(\mathbf{0}) = \operatorname{argmin}_{\mathbf{y} \in \mathbb{E}^*} f^*(\mathbf{y}) = \operatorname{argmin}_{\mathbf{y} \in \mathbb{E}^*} \delta_{B_{\|\cdot\|_*}[\mathbf{0}, 1]} = B_{\|\cdot\|_*}[\mathbf{0}, 1].$$

This result was already established in Example 3.3. \blacksquare

A relation between Lipschitz continuity of a function and the boundedness of its subgradients over a given set was established in Theorem 3.61. We end this chapter with a related result showing that Lipschitz continuity over the entire space is also equivalent to boundedness of the domain of the conjugate.

Theorem 4.23 (Lipschitz continuity and boundedness of the domain of the conjugate). *Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be convex. Then the following three claims are equivalent for a given constant $L > 0$:*

- (i) $|f(\mathbf{x}) - f(\mathbf{y})| \leq L\|\mathbf{x} - \mathbf{y}\|$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{E}$.
- (ii) $\|\mathbf{g}\|_* \leq L$ for any $\mathbf{g} \in \partial f(\mathbf{x}), \mathbf{x} \in \mathbb{E}$.
- (iii) $\text{dom}(f^*) \subseteq B_{\|\cdot\|_*}[\mathbf{0}, L]$.

Proof. The equivalence between (i) and (ii) follows from Theorem 3.61. We will show that (iii) implies (ii). Indeed, assume that (iii) holds, that is, $\text{dom}(f^*) \subseteq B_{\|\cdot\|_*}[\mathbf{0}, L]$. Since by the conjugate subgradient theorem (Corollary 4.21) for any $\mathbf{x} \in \mathbb{E}$,

$$\partial f(\mathbf{x}) = \operatorname{argmax}_{\mathbf{y} \in \mathbb{E}^*} \{\langle \mathbf{x}, \mathbf{y} \rangle - f^*(\mathbf{y})\},$$

it follows that $\partial f(\mathbf{x}) \subseteq \text{dom}(f^*)$, and hence in particular $\partial f(\mathbf{x}) \subseteq B_{\|\cdot\|_*}[\mathbf{0}, L]$ for any $\mathbf{x} \in \mathbb{E}$, establishing (ii). In the reverse direction, we will show that the implication (i) \Rightarrow (iii) holds. Suppose that (i) holds. Then in particular

$$f(\mathbf{x}) - f(\mathbf{0}) \leq |f(\mathbf{x}) - f(\mathbf{0})| \leq L\|\mathbf{x}\|,$$

and hence

$$-f(\mathbf{x}) \geq -f(\mathbf{0}) - L\|\mathbf{x}\|.$$

Therefore, for any $\mathbf{y} \in \mathbb{E}^*$,

$$f^*(\mathbf{y}) = \max_{\mathbf{x} \in \mathbb{E}} \{\langle \mathbf{x}, \mathbf{y} \rangle - f(\mathbf{x})\} \geq \max_{\mathbf{x} \in \mathbb{E}} \{\langle \mathbf{x}, \mathbf{y} \rangle - f(\mathbf{0}) - L\|\mathbf{x}\|\}. \quad (4.17)$$

To show (iii), we take $\tilde{\mathbf{y}} \in \mathbb{E}^*$ that satisfies $\|\tilde{\mathbf{y}}\|_* > L$ and show that $\tilde{\mathbf{y}} \notin \text{dom}(f^*)$. Take a vector $\mathbf{y}^\dagger \in \mathbb{E}$ satisfying $\|\mathbf{y}^\dagger\| = 1$ for which $\langle \tilde{\mathbf{y}}, \mathbf{y}^\dagger \rangle = \|\tilde{\mathbf{y}}\|_*$ (such a vector exists by the definition of the dual norm). Define $C = \{\alpha \mathbf{y}^\dagger : \alpha \geq 0\} \subseteq \mathbb{E}$. We can now continue (4.17) (with $\mathbf{y} = \tilde{\mathbf{y}}$) and write

$$\begin{aligned} f^*(\tilde{\mathbf{y}}) &\geq \max_{\mathbf{x} \in \mathbb{E}} \{\langle \mathbf{x}, \tilde{\mathbf{y}} \rangle - f(\mathbf{0}) - L\|\mathbf{x}\|\} \\ &\geq \max_{\mathbf{x} \in C} \{\langle \mathbf{x}, \tilde{\mathbf{y}} \rangle - f(\mathbf{0}) - L\|\mathbf{x}\|\} \\ &= \max_{\alpha \geq 0} \{\langle \alpha \mathbf{y}^\dagger, \tilde{\mathbf{y}} \rangle - f(\mathbf{0}) - L\alpha \|\mathbf{y}^\dagger\|\} \\ &= \max_{\alpha \geq 0} \{\alpha \|\tilde{\mathbf{y}}\|_* - f(\mathbf{0}) - L\alpha\} \\ &= \max_{\alpha \geq 0} \{\alpha(\|\tilde{\mathbf{y}}\|_* - L) - f(\mathbf{0})\} \\ &\stackrel{\|\tilde{\mathbf{y}}\|_* > L}{=} \infty, \end{aligned}$$

thus showing that $\tilde{\mathbf{y}} \notin \text{dom}(f^*)$, establishing claim (iii). \square

Chapter 5

Smoothness and Strong Convexity

5.1 *L*-Smooth Functions

We begin with the definition of *L-smoothness*.

Definition 5.1 (*L*-smoothness). Let $L \geq 0$. A function $f : \mathbb{E} \rightarrow (-\infty, \infty]$ is said to be *L-smooth* over a set $D \subseteq \mathbb{E}$ if it is differentiable over D and satisfies

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_* \leq L\|\mathbf{x} - \mathbf{y}\| \text{ for all } \mathbf{x}, \mathbf{y} \in D.$$

The constant L is called the **smoothness parameter**.

Obviously, by the definition of differentiability, if f is *L-smooth* over a set $D \subseteq \mathbb{E}$, this means in particular that $D \subseteq \text{int}(\text{dom } f)$. If a function is *L-smooth* over \mathbb{E} , then we will just refer to it as *L-smooth* (without mentioning the entire space). Another frequent terminology in the literature refers to an *L-smooth* function over D as “a function with Lipschitz gradient with constant L .” The class of *L-smooth* functions is denoted by $C_L^{1,1}(D)$. When $D = \mathbb{E}$, the class is often denoted by $C_L^{1,1}$ instead of $C_L^{1,1}(\mathbb{E})$. The class of functions which are *L-smooth* for some $L \geq 0$ is denoted by $\bar{C}^{1,1}$.

By the definition of *L-smoothness*, it is clear that if a function is L_1 -smooth, then it is also L_2 -smooth for any $L_2 \geq L_1$. It is therefore sometimes interesting to discuss the value of the *smallest* possible smoothness parameter of a given function.

Example 5.2 (smoothness of quadratic functions). Consider the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ given by $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} + \mathbf{b}^T \mathbf{x} + c$, where $\mathbf{A} \in \mathbb{S}^n$, $\mathbf{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$. We assume that \mathbb{R}^n is endowed with the l_p -norm ($1 \leq p \leq \infty$). Then, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_q = \|\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{y}\|_q \leq \|\mathbf{A}\|_{p,q} \|\mathbf{x} - \mathbf{y}\|_p,$$

where $\|\cdot\|_{p,q}$ is the induced norm given by (see also Section 1.8.2)

$$\|\mathbf{A}\|_{p,q} = \max\{\|\mathbf{A}\mathbf{x}\|_q : \|\mathbf{x}\|_p \leq 1\},$$

with $q \in [1, \infty]$ satisfying $\frac{1}{p} + \frac{1}{q} = 1$. We can thus conclude that f is $\|\mathbf{A}\|_{p,q}$ -smooth. We will show that $\|\mathbf{A}\|_{p,q}$ is the smallest smoothness parameter. For that, assume that f is L -smooth. Take a vector $\tilde{\mathbf{x}}$ satisfying $\|\tilde{\mathbf{x}}\|_p = 1$ and $\|\mathbf{A}\tilde{\mathbf{x}}\|_q = \|\mathbf{A}\|_{p,q}$. The existence of such a vector is guaranteed by the definition the induced matrix norm. Then

$$\|\mathbf{A}\|_{p,q} = \|\mathbf{A}\tilde{\mathbf{x}}\|_q = \|\nabla f(\tilde{\mathbf{x}}) - \nabla f(\mathbf{0})\|_q \leq L\|\tilde{\mathbf{x}} - \mathbf{0}\|_p = L.$$

We thus showed that if f is L -smooth, then $L \geq \|\mathbf{A}\|_{p,q}$, proving that $\|\mathbf{A}\|_{p,q}$ is indeed the smallest possible smoothness parameter. ■

Example 5.3 (0-smoothness of affine functions). Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}) = \langle \mathbf{b}, \mathbf{x} \rangle + c$, where $\mathbf{b} \in \mathbb{E}^*$ and $c \in \mathbb{R}$. For any $\mathbf{x}, \mathbf{y} \in \mathbb{E}$,

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_* = \|\mathbf{b} - \mathbf{b}\|_* = 0 \leq 0\|\mathbf{x} - \mathbf{y}\|,$$

showing that affine functions are 0-smooth. ■

The next example will utilize a well-known result on the orthogonal projection operator, which was introduced in Example 3.31. A more general result will be shown later on in Theorem 6.42.

Theorem 5.4 (see [10, Theorem 9.9]). *Let \mathbb{E} be a Euclidean space, and let $C \subseteq \mathbb{E}$ be a nonempty closed and convex set. Then*

(a) **(firm nonexpansiveness)** *For any $\mathbf{v}, \mathbf{w} \in \mathbb{E}$,*

$$\langle P_C(\mathbf{v}) - P_C(\mathbf{w}), \mathbf{v} - \mathbf{w} \rangle \geq \|P_C(\mathbf{v}) - P_C(\mathbf{w})\|^2. \quad (5.1)$$

(b) **(nonexpansiveness)** *For any $\mathbf{v}, \mathbf{w} \in \mathbb{E}$,*

$$\|P_C(\mathbf{v}) - P_C(\mathbf{w})\| \leq \|\mathbf{v} - \mathbf{w}\|. \quad (5.2)$$

Example 5.5 (1-smoothness of $\frac{1}{2}d_C^2$). Suppose that \mathbb{E} is a Euclidean space, and let $C \subseteq \mathbb{E}$ be a nonempty closed and convex set. Consider the function $\varphi_C(\mathbf{x}) = \frac{1}{2}d_C^2(\mathbf{x})$. By Example 3.31, φ_C is differentiable over \mathbb{E} and $\nabla \varphi_C(\mathbf{x}) = \mathbf{x} - P_C(\mathbf{x})$. We will show that φ_C is 1-smooth. Indeed, for any $\mathbf{x}, \mathbf{y} \in \mathbb{E}$,

$$\begin{aligned} \|\nabla \varphi_C(\mathbf{x}) - \nabla \varphi_C(\mathbf{y})\|^2 &= \|\mathbf{x} - \mathbf{y} - P_C(\mathbf{x}) + P_C(\mathbf{y})\|^2 \\ &= \|\mathbf{x} - \mathbf{y}\|^2 - 2\langle P_C(\mathbf{x}) - P_C(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle + \|P_C(\mathbf{x}) - P_C(\mathbf{y})\|^2 \\ &\stackrel{(*)}{\leq} \|\mathbf{x} - \mathbf{y}\|^2 - 2\|P_C(\mathbf{x}) - P_C(\mathbf{y})\|^2 + \|P_C(\mathbf{x}) - P_C(\mathbf{y})\|^2 \\ &= \|\mathbf{x} - \mathbf{y}\|^2 - \|P_C(\mathbf{x}) - P_C(\mathbf{y})\|^2 \\ &\leq \|\mathbf{x} - \mathbf{y}\|^2, \end{aligned}$$

where the inequality $(*)$ follows by the firm nonexpansivity of the orthogonal projection operator (Theorem 5.4(a)). ■

Example 5.6 (1-smoothness of $\frac{1}{2}\|\cdot\|^2 - \frac{1}{2}d_C^2$). Suppose that \mathbb{E} is a Euclidean space, and let $C \subseteq \mathbb{E}$ be a nonempty closed convex set. Consider the function

$\psi_C(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2 - \frac{1}{2}d_C^2(\mathbf{x})$. By Example 2.17, ψ_C is convex.²³ We will now show that it is 1-smooth. By Example 3.31, $\frac{1}{2}d_C^2(\mathbf{x})$ is differentiable over \mathbb{E} , and its gradient is given by $\mathbf{x} - P_C(\mathbf{x})$. Therefore,

$$\nabla\psi_C(\mathbf{x}) = \mathbf{x} - (\mathbf{x} - P_C(\mathbf{x})) = P_C(\mathbf{x}).$$

The 1-smoothness of ψ_C now follows by the nonexpansivity of the projection operator (Theorem 5.4(b))—for any $\mathbf{x}, \mathbf{y} \in \mathbb{E}$,

$$\|\nabla\psi_C(\mathbf{x}) - \nabla\psi_C(\mathbf{y})\| = \|P_C(\mathbf{x}) - P_C(\mathbf{y})\| \leq \|\mathbf{x} - \mathbf{y}\|. \quad \blacksquare$$

5.1.1 The Descent Lemma

An extremely useful result on L -smooth functions is the descent lemma, which states that they can be upper bounded by a certain quadratic function.

Lemma 5.7 (descent lemma). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be an L -smooth function ($L \geq 0$) over a given convex set D . Then for any $\mathbf{x}, \mathbf{y} \in D$,*

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|^2. \quad (5.3)$$

Proof. By the fundamental theorem of calculus,

$$f(\mathbf{y}) - f(\mathbf{x}) = \int_0^1 \langle \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})), \mathbf{y} - \mathbf{x} \rangle dt.$$

Therefore,

$$f(\mathbf{y}) - f(\mathbf{x}) = \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \int_0^1 \langle \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle dt.$$

Thus,

$$\begin{aligned} |f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle| &= \left| \int_0^1 \langle \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle dt \right| \\ &\leq \int_0^1 |\langle \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle| dt \\ &\stackrel{(*)}{\leq} \int_0^1 \|\nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - \nabla f(\mathbf{x})\|_* \cdot \|\mathbf{y} - \mathbf{x}\| dt \\ &\leq \int_0^1 tL\|\mathbf{y} - \mathbf{x}\|^2 dt \\ &= \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|^2, \end{aligned}$$

where in $(*)$ we used the generalized Cauchy–Schwarz inequality (Lemma 1.4). \square

²³The convexity of ψ_C actually does not require the convexity of C ; see Example 2.17.

5.1.2 Characterizations of L -Smooth Functions

When f is convex, the next result gives several different and equivalent characterizations of the L -smoothness property of f over the entire space. Note that property (5.3) from the descent lemma is one of the mentioned equivalent properties.

Theorem 5.8 (characterizations of L -smoothness). *Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be a convex function, differentiable over \mathbb{E} , and let $L > 0$. Then the following claims are equivalent:*

- (i) f is L -smooth.
- (ii) $f(\mathbf{y}) \leq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{L}{2} \|\mathbf{x} - \mathbf{y}\|^2$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{E}$.
- (iii) $f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{1}{2L} \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_*^2$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{E}$.
- (iv) $\langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geq \frac{1}{L} \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_*^2$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{E}$.
- (v) $f(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) \geq \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}) - \frac{L}{2} \lambda(1 - \lambda) \|\mathbf{x} - \mathbf{y}\|^2$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{E}$ and $\lambda \in [0, 1]$.

Proof. (i) \Rightarrow (ii). The fact that (i) implies (ii) is just the descent lemma (Lemma 5.7).

(ii) \Rightarrow (iii). Suppose that (ii) is satisfied. We can assume that $\nabla f(\mathbf{x}) \neq \nabla f(\mathbf{y})$ since otherwise the inequality (iii) is trivial by the convexity of f . For a fixed $\mathbf{x} \in \mathbb{E}$ consider the function

$$g_{\mathbf{x}}(\mathbf{y}) = f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle, \quad \mathbf{y} \in \mathbb{E}.$$

The function $g_{\mathbf{x}}$ also satisfies property (ii). Indeed, for any $\mathbf{y}, \mathbf{z} \in \mathbb{E}$,

$$\begin{aligned} g_{\mathbf{x}}(\mathbf{z}) &= f(\mathbf{z}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{z} - \mathbf{x} \rangle \\ &\leq f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{z} - \mathbf{y} \rangle + \frac{L}{2} \|\mathbf{z} - \mathbf{y}\|^2 - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{z} - \mathbf{x} \rangle \\ &= f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \langle \nabla f(\mathbf{y}) - \nabla f(\mathbf{x}), \mathbf{z} - \mathbf{y} \rangle + \frac{L}{2} \|\mathbf{z} - \mathbf{y}\|^2 \\ &= g_{\mathbf{x}}(\mathbf{y}) + \langle \nabla g_{\mathbf{x}}(\mathbf{y}), \mathbf{z} - \mathbf{y} \rangle + \frac{L}{2} \|\mathbf{z} - \mathbf{y}\|^2, \end{aligned} \tag{5.4}$$

where we used in the last equality the fact that $\nabla g_{\mathbf{x}}(\mathbf{y}) = \nabla f(\mathbf{y}) - \nabla f(\mathbf{x})$ for any $\mathbf{y} \in \mathbb{E}$. In particular, $\nabla g_{\mathbf{x}}(\mathbf{x}) = \mathbf{0}$, which by the convexity of $g_{\mathbf{x}}$ implies that \mathbf{x} is a global minimizer of g , meaning that

$$g_{\mathbf{x}}(\mathbf{x}) \leq g_{\mathbf{x}}(\mathbf{z}) \text{ for all } \mathbf{z} \in \mathbb{E}. \tag{5.5}$$

Let $\mathbf{y} \in \mathbb{E}$, and let $\mathbf{v} \in \mathbb{E}$ be a vector satisfying $\|\mathbf{v}\| = 1$ and $\langle \nabla g_{\mathbf{x}}(\mathbf{y}), \mathbf{v} \rangle = \|\nabla g_{\mathbf{x}}(\mathbf{y})\|_*$. Substituting

$$\mathbf{z} = \mathbf{y} - \frac{\|\nabla g_{\mathbf{x}}(\mathbf{y})\|_*}{L} \mathbf{v} \tag{5.6}$$

into (5.5) yields

$$0 = g_{\mathbf{x}}(\mathbf{x}) \leq g_{\mathbf{x}} \left(\mathbf{y} - \frac{\|\nabla g_{\mathbf{x}}(\mathbf{y})\|_*}{L} \mathbf{v} \right).$$

Combining the last inequality with (5.4) (using the specific choice of \mathbf{z} given in (5.6)), we obtain

$$\begin{aligned} 0 &= g_{\mathbf{x}}(\mathbf{x}) \\ &\leq g_{\mathbf{x}}(\mathbf{y}) - \frac{\|\nabla g_{\mathbf{x}}(\mathbf{y})\|_*}{L} \langle \nabla g_{\mathbf{x}}(\mathbf{y}), \mathbf{v} \rangle + \frac{1}{2L} \|\nabla g_{\mathbf{x}}(\mathbf{y})\|_*^2 \cdot \|\mathbf{v}\|^2 \\ &= g_{\mathbf{x}}(\mathbf{y}) - \frac{1}{2L} \|\nabla g_{\mathbf{x}}(\mathbf{y})\|_*^2 \\ &= f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle - \frac{1}{2L} \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_*^2, \end{aligned}$$

which is claim (iii).

(iii) \Rightarrow (iv). Writing the inequality (iii) for the two pairs $(\mathbf{x}, \mathbf{y}), (\mathbf{y}, \mathbf{x})$ yields

$$\begin{aligned} f(\mathbf{y}) &\geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{1}{2L} \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_*^2, \\ f(\mathbf{x}) &\geq f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle + \frac{1}{2L} \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_*^2. \end{aligned}$$

Adding the two inequalities and rearranging terms results in (iv).

(iv) \Rightarrow (i). The Lipschitz condition

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_* \leq L \|\mathbf{x} - \mathbf{y}\|$$

is trivial when $\nabla f(\mathbf{x}) = \nabla f(\mathbf{y})$. We will therefore assume that $\nabla f(\mathbf{x}) \neq \nabla f(\mathbf{y})$. By (iv) and the generalized Cauchy–Schwarz inequality (Lemma 1.4) we have for any $\mathbf{x}, \mathbf{y} \in \mathbb{E}$,

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_* \cdot \|\mathbf{x} - \mathbf{y}\| \geq \langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geq \frac{1}{L} \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_*^2.$$

Dividing by $\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_*$ and multiplying by L , (i) is obtained.

We have just shown the equivalence between (i), (ii), (iii), and (iv). To prove that (v) is also equivalent to each of these four claims, we will establish the equivalence (ii) \Leftrightarrow (v).

(ii) \Rightarrow (v). Let $\mathbf{x}, \mathbf{y} \in \mathbb{E}$ and $\lambda \in [0, 1]$. Denote $\mathbf{x}_\lambda = \lambda\mathbf{x} + (1 - \lambda)\mathbf{y}$. Then by (ii),

$$\begin{aligned} f(\mathbf{x}) &\leq f(\mathbf{x}_\lambda) + \langle \nabla f(\mathbf{x}_\lambda), \mathbf{x} - \mathbf{x}_\lambda \rangle + \frac{L}{2} \|\mathbf{x} - \mathbf{x}_\lambda\|^2, \\ f(\mathbf{y}) &\leq f(\mathbf{x}_\lambda) + \langle \nabla f(\mathbf{x}_\lambda), \mathbf{y} - \mathbf{x}_\lambda \rangle + \frac{L}{2} \|\mathbf{y} - \mathbf{x}_\lambda\|^2, \end{aligned}$$

which is the same as

$$\begin{aligned} f(\mathbf{x}) &\leq f(\mathbf{x}_\lambda) + (1 - \lambda) \langle \nabla f(\mathbf{x}_\lambda), \mathbf{x} - \mathbf{y} \rangle + \frac{L(1 - \lambda)^2}{2} \|\mathbf{x} - \mathbf{y}\|^2 \\ f(\mathbf{y}) &\leq f(\mathbf{x}_\lambda) + \lambda \langle \nabla f(\mathbf{x}_\lambda), \mathbf{y} - \mathbf{x} \rangle + \frac{L\lambda^2}{2} \|\mathbf{x} - \mathbf{y}\|^2. \end{aligned}$$

Multiplying the first inequality by λ and the second by $1 - \lambda$ and adding them yields the inequality (iv).

(v) \Rightarrow (ii). Rearranging terms in the inequality (v), we obtain that it is equivalent to

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \frac{f(\mathbf{x} + (1 - \lambda)(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{1 - \lambda} + \frac{L}{2}\lambda\|\mathbf{x} - \mathbf{y}\|^2.$$

Taking $\lambda \rightarrow 1^-$, the last inequality becomes

$$f(\mathbf{y}) \leq f(\mathbf{x}) + f'(\mathbf{x}; \mathbf{y} - \mathbf{x}) + \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|^2,$$

which, by the fact that $f'(\mathbf{x}; \mathbf{y} - \mathbf{x}) = \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle$ (see Theorem 3.29), implies (ii). \square

Remark 5.9 (necessity of convexity in Theorem 5.8). *The convexity assumption in Theorem 5.8 is essential. Consider, for example, the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ given by $f(\mathbf{x}) = -\frac{1}{2}\|\mathbf{x}\|_2^2$, which is 1-smooth w.r.t. the l_2 -norm but is not L -smooth for $L < 1$ (see Example 5.2). However, f is concave, and hence*

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle,$$

which implies that property (ii) of Theorem 5.8 is satisfied with $L = 0$, although the function is obviously not 0-smooth.

The next example will require the linear approximation theorem, which we now recall.

Theorem 5.10 (linear approximation theorem, [10, Theorem 1.24], [101, Fact 3.3.10]). *Let $f : U \rightarrow \mathbb{R}$ be a twice continuously differentiable function²⁴ over an open set $U \subseteq \mathbb{R}^n$, and let $\mathbf{x} \in U, r > 0$ satisfy $B(\mathbf{x}, r) \subseteq U$. Then for any $\mathbf{y} \in B(\mathbf{x}, r)$ there exists $\xi \in [\mathbf{x}, \mathbf{y}]$ such that*

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) + \frac{1}{2}(\mathbf{y} - \mathbf{x})^T \nabla^2 f(\xi)(\mathbf{y} - \mathbf{x}).$$

Example 5.11 (($p - 1$)-smoothness of the half-squared l_p -norm function).²⁵ Consider the convex function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ given by

$$f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_p^2 = \frac{1}{2} \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{2}{p}},$$

where $p \in [2, \infty)$. We assume that \mathbb{R}^n is endowed with the l_p -norm and show that f is $(p - 1)$ -smooth w.r.t. the l_p -norm. The result was already established for the case $p = 2$ in Example 5.2, and we will henceforth assume that $p > 2$. We begin by computing the partial derivatives:

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = \begin{cases} \text{sgn}(x_i) \frac{|x_i|^{p-1}}{\|\mathbf{x}\|_p^{p-2}}, & \mathbf{x} \neq \mathbf{0}, \\ 0, & \mathbf{x} = \mathbf{0}, \end{cases}$$

²⁴By “twice continuously differentiable over U ,” we mean that the function has second-order partial derivatives, which are continuous over U .

²⁵The analysis in Example 5.11 follows the derivation of Ben-Tal, Margalit, and Nemirovski [24, Appendix 1].

The partial derivatives are continuous over \mathbb{R}^n , and hence f is differentiable over \mathbb{R}^n (in the sense of Definition 3.28).²⁶ The second-order partial derivatives exist for any $\mathbf{x} \neq \mathbf{0}$ and are given by

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) = \begin{cases} (2-p)\operatorname{sgn}(x_i)\operatorname{sgn}(x_j) \frac{|x_i|^{p-1}|x_j|^{p-1}}{\|\mathbf{x}\|_p^{2p-2}}, & i \neq j, \\ (p-1) \frac{|x_i|^{p-2}}{\|\mathbf{x}\|_p^{p-2}} + (2-p) \frac{|x_i|^{2p-2}}{\|\mathbf{x}\|_p^{2p-2}}, & i = j. \end{cases}$$

It is easy to see that the second-order partial derivatives are continuous for any $\mathbf{x} \neq \mathbf{0}$. We will show that property (ii) of Theorem 5.8 is satisfied with $L = p - 1$. Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ be such that $\mathbf{0} \notin [\mathbf{x}, \mathbf{y}]$. Then by the linear approximation theorem (Theorem 5.10)—taking U to be some open set containing $[\mathbf{x}, \mathbf{y}]$ but not containing $\mathbf{0}$ —there exists $\boldsymbol{\xi} \in [\mathbf{x}, \mathbf{y}]$ for which

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) + \frac{1}{2}(\mathbf{y} - \mathbf{x})^T \nabla^2 f(\boldsymbol{\xi})(\mathbf{y} - \mathbf{x}). \quad (5.7)$$

We will show that $\mathbf{d}^T \nabla^2 f(\boldsymbol{\xi}) \mathbf{d} \leq (p-1)\|\mathbf{d}\|_p^2$ for any $\mathbf{d} \in \mathbb{R}^n$. Since $\nabla^2 f(t\boldsymbol{\xi}) = \nabla^2 f(\boldsymbol{\xi})$ for any $t \in \mathbb{R}$, we can assume without loss of generality that $\|\boldsymbol{\xi}\|_p = 1$. Now, for any $\mathbf{d} \in \mathbb{R}^n$,

$$\begin{aligned} \mathbf{d}^T \nabla^2 f(\boldsymbol{\xi}) \mathbf{d} &= (2-p)\|\boldsymbol{\xi}\|_p^{2-2p} \left(\sum_{i=1}^n |\xi_i|^{p-1} \operatorname{sgn}(\xi_i) d_i \right)^2 + (p-1)\|\boldsymbol{\xi}\|_p^{2-p} \sum_{i=1}^n |\xi_i|^{p-2} d_i^2 \\ &\leq (p-1)\|\boldsymbol{\xi}\|_p^{2-p} \sum_{i=1}^n |\xi_i|^{p-2} d_i^2, \end{aligned} \quad (5.8)$$

where the last inequality follows by the fact that $p > 2$. Using the generalized Cauchy–Schwarz inequality (Lemma 1.4) with $\|\cdot\| = \|\cdot\|_{\frac{p}{p-2}}$, we have

$$\begin{aligned} \sum_{i=1}^n |\xi_i|^{p-2} d_i^2 &\leq \left(\sum_{i=1}^n (|\xi_i|^{p-2})^{\frac{p}{p-2}} \right)^{\frac{p-2}{p}} \left(\sum_{i=1}^n (d_i^2)^{\frac{p}{2}} \right)^{\frac{2}{p}} \\ &= \left(\sum_{i=1}^n |\xi_i|^p \right)^{\frac{p-2}{p}} \left(\sum_{i=1}^n |d_i|^p \right)^{\frac{2}{p}} \\ &= \|\mathbf{d}\|_p^2. \end{aligned} \quad (5.9)$$

Combining (5.8) and (5.9), we obtain that for any $\mathbf{d} \in \mathbb{R}^n$,

$$\mathbf{d}^T \nabla^2 f(\boldsymbol{\xi}) \mathbf{d} \leq (p-1)\|\mathbf{d}\|_p^2,$$

and specifically, for $\mathbf{d} = \mathbf{x} - \mathbf{y}$,

$$(\mathbf{y} - \mathbf{x})^T \nabla^2 f(\boldsymbol{\xi})(\mathbf{y} - \mathbf{x}) \leq (p-1)\|\mathbf{x} - \mathbf{y}\|_p^2.$$

Plugging the above inequality into (5.7) implies the inequality

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) + \frac{p-1}{2}\|\mathbf{x} - \mathbf{y}\|_p^2. \quad (5.10)$$

²⁶See, for example, [112, Theorem 9.21] for a precise statement of this result.

The inequality (5.10) was proven for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ that satisfy $\mathbf{0} \notin [\mathbf{x}, \mathbf{y}]$. We can show that it holds for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ using a continuity argument. Indeed, assume that $\mathbf{0} \in [\mathbf{x}, \mathbf{y}]$. Then we can find a sequence $\{\mathbf{y}_k\}_{k \geq 0}$ converging to \mathbf{y} for which $\mathbf{0} \notin [\mathbf{x}, \mathbf{y}_k]$. Thus, by what was already proven, for any $k \geq 0$,

$$f(\mathbf{y}_k) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}_k - \mathbf{x}) + \frac{p-1}{2}\|\mathbf{x} - \mathbf{y}_k\|_p^2.$$

Taking $k \rightarrow \infty$ in the last inequality and using the continuity of f , we obtain that (5.10) holds. To conclude, we established that (5.10) holds for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, and thus by Theorem 5.8 (equivalence between properties (i) and (ii)) and the convexity of f , it follows that f is $(p-1)$ -smooth w.r.t. the l_p -norm. ■

5.1.3 Second-Order Characterization

We will now consider the space $\mathbb{E} = \mathbb{R}^n$ endowed with the l_p -norm ($p \geq 1$). For twice continuously differentiable functions, it is possible to fully characterize the property of L -smoothness via the norm of the Hessian matrix.

Theorem 5.12 (L -smoothness and boundedness of the Hessian). *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a twice continuously differentiable function over \mathbb{R}^n . Then for a given $L \geq 0$, the following two claims are equivalent:*

- (i) f is L -smooth w.r.t. the l_p -norm ($p \in [1, \infty]$).
- (ii) $\|\nabla^2 f(\mathbf{x})\|_{p,q} \leq L$ for any $\mathbf{x} \in \mathbb{R}^n$, where $q \in [1, \infty]$ satisfies $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. (ii) \Rightarrow (i). Suppose that $\|\nabla^2 f(\mathbf{x})\|_{p,q} \leq L$ for any $\mathbf{x} \in \mathbb{R}^n$. Then by the fundamental theorem of calculus, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$\begin{aligned} \nabla f(\mathbf{y}) &= \nabla f(\mathbf{x}) + \int_0^1 \nabla^2 f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x}) dt \\ &= \nabla f(\mathbf{x}) + \left(\int_0^1 \nabla^2 f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) dt \right) \cdot (\mathbf{y} - \mathbf{x}). \end{aligned}$$

Then

$$\begin{aligned} \|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|_q &= \left\| \left(\int_0^1 \nabla^2 f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) dt \right) \cdot (\mathbf{y} - \mathbf{x}) \right\|_q \\ &\leq \left\| \int_0^1 \nabla^2 f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) dt \right\|_{p,q} \|\mathbf{y} - \mathbf{x}\|_p \\ &\leq \left(\int_0^1 \|\nabla^2 f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))\|_{p,q} dt \right) \|\mathbf{y} - \mathbf{x}\|_p \\ &\leq L \|\mathbf{y} - \mathbf{x}\|_p, \end{aligned}$$

establishing (i).

(i) \Rightarrow (ii). Suppose now that f is L -smooth w.r.t. the l_p -norm. Then by the fundamental theorem of calculus, for any $\mathbf{d} \in \mathbb{R}^n$ and $\alpha > 0$,

$$\nabla f(\mathbf{x} + \alpha \mathbf{d}) - \nabla f(\mathbf{x}) = \int_0^\alpha \nabla^2 f(\mathbf{x} + t \mathbf{d}) \mathbf{d} dt.$$

Thus,

$$\left\| \left(\int_0^\alpha \nabla^2 f(\mathbf{x} + t\mathbf{d}) dt \right) \mathbf{d} \right\|_q = \|\nabla f(\mathbf{x} + \alpha\mathbf{d}) - \nabla f(\mathbf{x})\|_q \leq \alpha L \|\mathbf{d}\|_p.$$

Dividing by α and taking the limit $\alpha \rightarrow 0^+$, we obtain

$$\|\nabla^2 f(\mathbf{x})\mathbf{d}\|_q \leq L \|\mathbf{d}\|_p \text{ for any } \mathbf{d} \in \mathbb{R}^n,$$

implying that $\|\nabla^2 f(\mathbf{x})\|_{p,q} \leq L$. \square

A direct consequence is that for twice continuously differentiable convex functions, L -smoothness w.r.t. the l_2 -norm is equivalent to the property that the maximum eigenvalue of the Hessian matrix is smaller than or equal to L .

Corollary 5.13. *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a twice continuously differentiable convex function over \mathbb{R}^n . Then f is L -smooth w.r.t. the l_2 -norm if and only if $\lambda_{\max}(\nabla^2 f(\mathbf{x})) \leq L$ for any $\mathbf{x} \in \mathbb{R}^n$.*

Proof. Since f is convex, it follows that $\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}$ for any $\mathbf{x} \in \mathbb{R}^n$. Therefore, in this case,

$$\|\nabla^2 f(\mathbf{x})\|_{2,2} = \sqrt{\lambda_{\max}((\nabla^2 f(\mathbf{x}))^2)} = \lambda_{\max}(\nabla^2 f(\mathbf{x})),$$

which, combined with Theorem 5.12, establishes the desired result. \square

Example 5.14 (1-smoothness of $\sqrt{1 + \|\cdot\|_2^2}$ w.r.t. the l_2 -norm). Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be the convex function given by

$$f(\mathbf{x}) = \sqrt{1 + \|\mathbf{x}\|_2^2}.$$

We will show that f is 1-smooth w.r.t. the l_2 -norm. For any $\mathbf{x} \in \mathbb{R}^n$,

$$\nabla f(\mathbf{x}) = \frac{\mathbf{x}}{\sqrt{\|\mathbf{x}\|_2^2 + 1}}$$

and

$$\nabla^2 f(\mathbf{x}) = \frac{1}{\sqrt{\|\mathbf{x}\|_2^2 + 1}} \mathbf{I} - \frac{\mathbf{x}\mathbf{x}^T}{(\|\mathbf{x}\|_2^2 + 1)^{3/2}} \preceq \frac{1}{\sqrt{\|\mathbf{x}\|_2^2 + 1}} \mathbf{I} \preceq \mathbf{I}.$$

Therefore, $\lambda_{\max}(\nabla^2 f(\mathbf{x})) \leq 1$ for all $\mathbf{x} \in \mathbb{R}^n$, and hence by Corollary 5.13 it follows that f is 1-smooth w.r.t. the l_2 -norm. \blacksquare

Example 5.15 (1-smoothness of the log-sum-exp function w.r.t. the l_2, l_∞ norms). Consider the log-sum-exp function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ given by

$$f(\mathbf{x}) = \log(e^{x_1} + e^{x_2} + \cdots + e^{x_n}).$$

We will first show that it is 1-smooth w.r.t. the l_2 -norm. The partial derivatives of f are

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = \frac{e^{x_i}}{\sum_{k=1}^n e^{x_k}}, \quad i = 1, 2, \dots, n,$$

and the second-order partial derivatives are

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) = \begin{cases} -\frac{e^{x_i} e^{x_j}}{(\sum_{k=1}^n e^{x_k})^2}, & i \neq j, \\ -\frac{e^{x_i} e^{x_i}}{(\sum_{k=1}^n e^{x_k})^2} + \frac{e^{x_i}}{\sum_{k=1}^n e^{x_k}}, & i = j. \end{cases}$$

We can thus write the Hessian matrix as

$$\nabla^2 f(\mathbf{x}) = \text{diag}(\mathbf{w}) - \mathbf{w}\mathbf{w}^T,$$

where $w_i = \frac{e^{x_i}}{\sum_{k=1}^n e^{x_k}}$. To show that f is 1-smooth w.r.t. the l_2 -norm, note that for any $\mathbf{x} \in \mathbb{R}^n$,

$$\nabla^2 f(\mathbf{x}) = \text{diag}(\mathbf{w}) - \mathbf{w}\mathbf{w}^T \preceq \text{diag}(\mathbf{w}) \preceq \mathbf{I},$$

and hence $\lambda_{\max}(\nabla^2 f(\mathbf{x})) \leq 1$ for any $\mathbf{x} \in \mathbb{R}^n$. Noting that the log-sum-exp function is convex, we can invoke Corollary 5.13 and conclude that f is 1-smooth w.r.t. the l_2 -norm.

We will show that f is 1-smooth also w.r.t. the l_∞ -norm. For that, we begin by proving that for any $\mathbf{d} \in \mathbb{R}^n$,

$$\mathbf{d}^T \nabla^2 f(\mathbf{x}) \mathbf{d} \leq \|\mathbf{d}\|_\infty^2. \quad (5.11)$$

Indeed,

$$\begin{aligned} \mathbf{d}^T \nabla^2 f(\mathbf{x}) \mathbf{d} &= \mathbf{d}^T (\text{diag}(\mathbf{w}) - \mathbf{w}\mathbf{w}^T) \mathbf{d} = \mathbf{d}^T \text{diag}(\mathbf{w}) \mathbf{d} - (\mathbf{w}^T \mathbf{d})^2 \\ &\leq \mathbf{d}^T \text{diag}(\mathbf{w}) \mathbf{d} \\ &= \sum_{i=1}^n w_i d_i^2 \\ &\leq \|\mathbf{d}\|_\infty^2 \sum_{i=1}^n w_i \\ &= \|\mathbf{d}\|_\infty^2. \end{aligned}$$

Now, since f is twice continuously differentiable over \mathbb{R}^n , it follows by the linear approximation theorem (Theorem 5.10) that for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ there exists $\boldsymbol{\xi} \in [\mathbf{x}, \mathbf{y}]$ for which

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + \frac{1}{2} (\mathbf{y} - \mathbf{x})^T \nabla^2 f(\boldsymbol{\xi}) (\mathbf{y} - \mathbf{x}). \quad (5.12)$$

Combining (5.12) (taking $\mathbf{d} = \mathbf{y} - \mathbf{x}$) and (5.11), we obtain the inequality

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_\infty^2,$$

which by Theorem 5.8 (equivalence between properties (i) and (ii)) implies the 1-smoothness of f w.r.t. the l_∞ -norm. ■

5.1.4 Summary of Smoothness Parameter Computations

The table below summarizes the smoothness parameters of the functions discussed in this section. The last function will only be discussed later on in Example 6.62.

$f(\mathbf{x})$	$\text{dom}(f)$	Parameter	Norm	Reference
$\frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} + \mathbf{b}^T \mathbf{x} + c$ ($\mathbf{A} \in \mathbb{S}^n$, $\mathbf{b} \in \mathbb{R}^n$, $c \in \mathbb{R}$)	\mathbb{R}^n	$\ \mathbf{A}\ _{p,q}$	l_p	Example 5.2
$\langle \mathbf{b}, \mathbf{x} \rangle + c$ ($\mathbf{b} \in \mathbb{E}^*$, $c \in \mathbb{R}$)	\mathbb{E}	0	any norm	Example 5.3
$\frac{1}{2}\ \mathbf{x}\ _p^2$, $p \in [2, \infty)$	\mathbb{R}^n	$p - 1$	l_p	Example 5.11
$\sqrt{1 + \ \mathbf{x}\ _2^2}$	\mathbb{R}^n	1	l_2	Example 5.14
$\log(\sum_{i=1}^n e^{x_i})$	\mathbb{R}^n	1	l_2, l_∞	Example 5.15
$\frac{1}{2}d_C^2(\mathbf{x})$ ($\emptyset \neq C \subseteq \mathbb{E}$ closed convex)	\mathbb{E}	1	Euclidean	Example 5.5
$\frac{1}{2}\ \mathbf{x}\ ^2 - \frac{1}{2}d_C^2(\mathbf{x})$ ($\emptyset \neq C \subseteq \mathbb{E}$ closed convex)	\mathbb{E}	1	Euclidean	Example 5.6
$H_\mu(\mathbf{x})$ ($\mu > 0$)	\mathbb{E}	$\frac{1}{\mu}$	Euclidean	Example 6.62

5.2 Strong Convexity

Definition 5.16 (strong convexity). A function $f : \mathbb{E} \rightarrow (-\infty, \infty]$ is called **σ -strongly convex** for a given $\sigma > 0$ if $\text{dom}(f)$ is convex and the following inequality holds for any $\mathbf{x}, \mathbf{y} \in \text{dom}(f)$ and $\lambda \in [0, 1]$:

$$f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}) - \frac{\sigma}{2}\lambda(1 - \lambda)\|\mathbf{x} - \mathbf{y}\|^2. \quad (5.13)$$

We will sometimes use the terminology “strongly convex with parameter σ ” instead of “ σ -strongly convex.” It is important to note that the strong convexity parameter σ depends on the underlying norm, and we will therefore sometimes refer to it as the “strong convexity parameter w.r.t. $\|\cdot\|$.” Obviously, strongly convex functions are necessarily also convex since their domain is assumed to be convex and inequality (5.13) implies that for any $\mathbf{x}, \mathbf{y} \in \text{dom}(f)$ and $\lambda \in [0, 1]$, Jensen’s inequality is satisfied:

$$f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}).$$

When the underlying set \mathbb{E} is Euclidean, meaning that $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ for any $\mathbf{x} \in \mathbb{E}$, we can write a different and simple property that is equivalent to strong convexity.

Theorem 5.17. Let \mathbb{E} be a Euclidean space. Then $f : \mathbb{E} \rightarrow (-\infty, \infty]$ is a σ -strongly convex function ($\sigma > 0$) if and only if the function $f(\cdot) - \frac{\sigma}{2}\|\cdot\|^2$ is convex.

Proof. The function $g(\mathbf{x}) \equiv f(\mathbf{x}) - \frac{\sigma}{2}\|\mathbf{x}\|^2$ is convex if and only if its domain $\text{dom}(g) = \text{dom}(f)$ is convex and for any $\mathbf{x}, \mathbf{y} \in \text{dom}(f)$ and $\lambda \in [0, 1]$,

$$g(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda g(\mathbf{x}) + (1 - \lambda)g(\mathbf{y}).$$

The latter inequality is the same as

$$f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}) + \frac{\sigma}{2} [\|\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}\|^2 - \lambda\|\mathbf{x}\|^2 - (1 - \lambda)\|\mathbf{y}\|^2]. \quad (5.14)$$

Now, using the identity (which holds since the norm is assumed to be Euclidean)

$$\|\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}\|^2 - \lambda\|\mathbf{x}\|^2 - (1 - \lambda)\|\mathbf{y}\|^2 = -\lambda(1 - \lambda)\|\mathbf{x} - \mathbf{y}\|^2,$$

combined with (5.14), we can conclude that the convexity of g is equivalent to the convexity of $\text{dom}(f)$ and the validity of the inequality

$$f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}) - \frac{\sigma}{2}\lambda(1 - \lambda)\|\mathbf{x} - \mathbf{y}\|^2$$

for any $\mathbf{x}, \mathbf{y} \in \text{dom}(f)$ and $\lambda \in [0, 1]$, namely, to the σ -strong convexity of f . \square

Remark 5.18. *The assumption that the underlying space is Euclidean is essential in Theorem 5.17. As an example, consider the negative entropy function over the unit simplex*

$$f(\mathbf{x}) \equiv \begin{cases} \sum_{i=1}^n x_i \log x_i, & \mathbf{x} \in \Delta_n, \\ \infty & \text{else.} \end{cases}$$

We will later show (in Example 5.27) that f is a 1-strongly convex function with respect to the l_1 -norm. Regardless of this fact, note that the function

$$g(\mathbf{x}) = f(\mathbf{x}) - \alpha\|\mathbf{x}\|_1^2$$

is convex for any $\alpha > 0$ since over the domain of f , we have that $\|\mathbf{x}\|_1 = 1$. Obviously, it is impossible that a function will be α -strongly convex for any $\alpha > 0$. Therefore, the characterization of strong convexity in Theorem 5.17 is not correct for any norm.

Note that if a function f is σ_1 -strongly convex ($\sigma_1 > 0$), then it is necessarily also σ_2 -strongly convex for any $\sigma_2 \in (0, \sigma_1)$. An interesting problem is to find the largest possible strong convexity parameter of a given function.

Example 5.19 (strong convexity of quadratic functions). Suppose that $\mathbb{E} = \mathbb{R}^n$ is endowed with the l_2 -norm, and consider the quadratic function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ given by

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} + \mathbf{b}^T \mathbf{x} + c,$$

where $\mathbf{A} \in \mathbb{S}^n$, $\mathbf{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$. Then by Theorem 5.17, f is strongly convex with parameter $\sigma > 0$ if and only if the function $\frac{1}{2}\mathbf{x}^T (\mathbf{A} - \sigma\mathbf{I})\mathbf{x} + \mathbf{b}^T \mathbf{x} + c$ is convex, which is equivalent to the matrix inequality $\mathbf{A} - \sigma\mathbf{I} \succeq \mathbf{0}$, namely, to the inequality $\lambda_{\min}(\mathbf{A}) \geq \sigma$. Thus, f is strongly convex if and only if \mathbf{A} is positive definite, and in that case, $\lambda_{\min}(\mathbf{A})$ is its largest possible strong convexity parameter. \blacksquare

A simple result is that the sum of a strongly convex function and a convex function is always a strongly convex function.

Lemma 5.20. *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a σ -strongly convex function ($\sigma > 0$), and let $g : \mathbb{E} \rightarrow (-\infty, \infty]$ be convex. Then $f + g$ is σ -strongly convex.*

Proof. Follows directly from the definitions of strong convexity and convexity. Since f and g are convex, both $\text{dom}(f)$ and $\text{dom}(g)$ are convex sets, and hence also $\text{dom}(f + g) = \text{dom}(f) \cap \text{dom}(g)$ is a convex set. Let $\mathbf{x}, \mathbf{y} \in \text{dom}(f) \cap \text{dom}(g)$ and $\lambda \in [0, 1]$. Then by the σ -strong convexity of f ,

$$f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}) - \frac{\sigma}{2}\lambda(1 - \lambda)\|\mathbf{x} - \mathbf{y}\|^2.$$

Since g is convex,

$$g(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda g(\mathbf{x}) + (1 - \lambda)g(\mathbf{y}).$$

Adding the two inequalities, we obtain

$$(f + g)(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda(f + g)(\mathbf{x}) + (1 - \lambda)(f + g)(\mathbf{y}) - \frac{\sigma}{2}\lambda(1 - \lambda)\|\mathbf{x} - \mathbf{y}\|^2,$$

showing that $f + g$ is σ -strongly convex. \square

Example 5.21 (strong convexity of $\frac{1}{2}\|\cdot\|^2 + \delta_C$). Suppose that \mathbb{E} is a Euclidean space, and let $C \subseteq \mathbb{E}$ be a nonempty convex set. The function $\frac{1}{2}\|\mathbf{x}\|^2$ is 1-strongly convex (Example 5.19), and by the convexity of C , δ_C is convex. Therefore, by Lemma 5.20, the function $\frac{1}{2}\|\mathbf{x}\|^2 + \delta_C(\mathbf{x})$ is 1-strongly convex. \blacksquare

Theorem 5.24 below describes two properties that are equivalent to σ -strong convexity. The two properties are of a first-order nature in the sense that they are written in terms of the function and its subgradients. The proof uses the following version of the mean-value theorem for one-dimensional functions.

Lemma 5.22 (see [67, p. 26]). *Let $f : \mathbb{R} \rightarrow (-\infty, \infty]$ be a closed convex function, and let $[a, b] \subseteq \text{dom}(f)$ ($a < b$). Then*

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

where $h : (a, b) \rightarrow \mathbb{R}$ satisfies $h(t) \in \partial f(t)$ for any $t \in (a, b)$.

Another technical lemma that is being used in the proof is the so-called line segment principle.

Lemma 5.23 (line segment principle [108, Theorem 6.1]). *Let C be a convex set. Suppose that $\mathbf{x} \in \text{ri}(C)$, $\mathbf{y} \in \text{cl}(C)$, and let $\lambda \in (0, 1]$. Then $\lambda\mathbf{x} + (1 - \lambda)\mathbf{y} \in \text{ri}(C)$.*

Theorem 5.24 (first-order characterizations of strong convexity). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper closed and convex function. Then for a given $\sigma > 0$, the following three claims are equivalent:*

(i) f is σ -strongly convex.

(ii)

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle + \frac{\sigma}{2} \|\mathbf{y} - \mathbf{x}\|^2$$

for any $\mathbf{x} \in \text{dom}(\partial f), \mathbf{y} \in \text{dom}(f)$ and $\mathbf{g} \in \partial f(\mathbf{x})$.

(iii)

$$\langle \mathbf{g}_x - \mathbf{g}_y, \mathbf{x} - \mathbf{y} \rangle \geq \sigma \|\mathbf{x} - \mathbf{y}\|^2 \quad (5.15)$$

for any $\mathbf{x}, \mathbf{y} \in \text{dom}(\partial f)$, and $\mathbf{g}_x \in \partial f(\mathbf{x}), \mathbf{g}_y \in \partial f(\mathbf{y})$.

Proof. (ii) \Rightarrow (i). Assume that (ii) is satisfied. To show (i), take $\mathbf{x}, \mathbf{y} \in \text{dom}(f)$ and $\lambda \in (0, 1)$. Take some $\mathbf{z} \in \text{ri}(\text{dom}(f))$. Then for any $\alpha \in (0, 1]$, by the line segment principle (Lemma 5.23), the vector $\tilde{\mathbf{x}} = (1 - \alpha)\mathbf{x} + \alpha\mathbf{z}$ is in $\text{ri}(\text{dom}(f))$. At this point we fix α . Using the notation $\mathbf{x}_\lambda = \lambda\tilde{\mathbf{x}} + (1 - \lambda)\mathbf{y}$, we obtain that $\mathbf{x}_\lambda \in \text{ri}(\text{dom}(f))$ for any $\lambda \in (0, 1)$, and hence, by Theorem 3.18, $\partial f(\mathbf{x}_\lambda) \neq \emptyset$, meaning that $\mathbf{x}_\lambda \in \text{dom}(\partial f)$. Take $\mathbf{g} \in \partial f(\mathbf{x}_\lambda)$. Then by (ii),

$$f(\tilde{\mathbf{x}}) \geq f(\mathbf{x}_\lambda) + \langle \mathbf{g}, \tilde{\mathbf{x}} - \mathbf{x}_\lambda \rangle + \frac{\sigma}{2} \|\tilde{\mathbf{x}} - \mathbf{x}_\lambda\|^2,$$

which is the same as

$$f(\tilde{\mathbf{x}}) \geq f(\mathbf{x}_\lambda) + (1 - \lambda) \langle \mathbf{g}, \tilde{\mathbf{x}} - \mathbf{y} \rangle + \frac{\sigma(1 - \lambda)^2}{2} \|\mathbf{y} - \tilde{\mathbf{x}}\|^2. \quad (5.16)$$

Similarly,

$$f(\mathbf{y}) \geq f(\mathbf{x}_\lambda) + \lambda \langle \mathbf{g}, \mathbf{y} - \tilde{\mathbf{x}} \rangle + \frac{\sigma\lambda^2}{2} \|\mathbf{y} - \tilde{\mathbf{x}}\|^2. \quad (5.17)$$

Multiplying (5.16) by λ and (5.17) by $1 - \lambda$ and adding the two resulting inequalities, we obtain that

$$f(\lambda\tilde{\mathbf{x}} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\tilde{\mathbf{x}}) + (1 - \lambda) f(\mathbf{y}) - \frac{\sigma\lambda(1 - \lambda)}{2} \|\tilde{\mathbf{x}} - \mathbf{y}\|^2.$$

Plugging the expression for $\tilde{\mathbf{x}}$ in the above inequality, we obtain that

$$g_1(\alpha) \leq \lambda g_2(\alpha) + (1 - \lambda) f(\mathbf{y}) - \frac{\sigma\lambda(1 - \lambda)}{2} \|(1 - \alpha)\mathbf{x} + \alpha\mathbf{z} - \mathbf{y}\|^2, \quad (5.18)$$

where $g_1(\alpha) \equiv f(\lambda(1 - \alpha)\mathbf{x} + (1 - \lambda)\mathbf{y} + \lambda\alpha\mathbf{z})$ and $g_2(\alpha) \equiv f((1 - \alpha)\mathbf{x} + \alpha\mathbf{z})$. The functions g_1 and g_2 are one-dimensional proper closed and convex functions, and consequently, by Theorem 2.22, they are also continuous over their domain. Thus, taking $\alpha \rightarrow 0^+$ in (5.18), it follows that

$$g_1(0) \leq \lambda g_2(0) + (1 - \lambda) f(\mathbf{y}) - \frac{\sigma\lambda(1 - \lambda)}{2} \|\mathbf{x} - \mathbf{y}\|^2.$$

Finally, since $g_1(0) = f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y})$ and $g_2(0) = f(\mathbf{x})$, we obtain the inequality

$$f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}) - \frac{\sigma\lambda(1 - \lambda)}{2} \|\mathbf{x} - \mathbf{y}\|^2,$$

establishing the σ -strong convexity of f .

(i) \Rightarrow (iii). Assume that (i) is satisfied. Let $\mathbf{x}, \mathbf{y} \in \text{dom}(\partial f)$ and $\mathbf{g}_x \in \partial f(\mathbf{x}), \mathbf{g}_y \in \partial f(\mathbf{y})$. We will show that inequality (5.15) is satisfied. Let $\lambda \in [0, 1]$ and denote $\mathbf{x}_\lambda = \lambda\mathbf{x} + (1 - \lambda)\mathbf{y}$. By condition (i),

$$f(\mathbf{x}_\lambda) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}) - \frac{\sigma}{2}\lambda(1 - \lambda)\|\mathbf{x} - \mathbf{y}\|^2,$$

which is the same as

$$\frac{f(\mathbf{x}_\lambda) - f(\mathbf{x})}{1 - \lambda} \leq f(\mathbf{y}) - f(\mathbf{x}) - \frac{\sigma}{2}\lambda\|\mathbf{x} - \mathbf{y}\|^2, \quad (5.19)$$

Since $\mathbf{g}_x \in \partial f(\mathbf{x})$,

$$\frac{f(\mathbf{x}_\lambda) - f(\mathbf{x})}{1 - \lambda} \geq \frac{\langle \mathbf{g}_x, \mathbf{x}_\lambda - \mathbf{x} \rangle}{1 - \lambda} = \langle \mathbf{g}_x, \mathbf{y} - \mathbf{x} \rangle,$$

which, combined with (5.19), yields the inequality

$$\langle \mathbf{g}_x, \mathbf{y} - \mathbf{x} \rangle \leq f(\mathbf{y}) - f(\mathbf{x}) - \frac{\sigma\lambda}{2}\|\mathbf{x} - \mathbf{y}\|^2. \quad (5.20)$$

Inequality (5.20) holds for any $\lambda \in [0, 1)$. Taking the limit $\lambda \rightarrow 1^-$, we conclude that

$$\langle \mathbf{g}_x, \mathbf{y} - \mathbf{x} \rangle \leq f(\mathbf{y}) - f(\mathbf{x}) - \frac{\sigma}{2}\|\mathbf{x} - \mathbf{y}\|^2. \quad (5.21)$$

Changing the roles of \mathbf{x} and \mathbf{y} yields the inequality

$$\langle \mathbf{g}_y, \mathbf{x} - \mathbf{y} \rangle \leq f(\mathbf{x}) - f(\mathbf{y}) - \frac{\sigma}{2}\|\mathbf{x} - \mathbf{y}\|^2. \quad (5.22)$$

Adding inequalities (5.21) and (5.22), we can finally conclude that

$$\langle \mathbf{g}_x - \mathbf{g}_y, \mathbf{x} - \mathbf{y} \rangle \geq \sigma\|\mathbf{x} - \mathbf{y}\|^2,$$

which is the desired inequality.

(iii) \Rightarrow (ii) Suppose that (iii) is satisfied, and let $\mathbf{x} \in \text{dom}(\partial f), \mathbf{y} \in \text{dom}(f)$, and $\mathbf{g} \in \partial f(\mathbf{x})$. Let \mathbf{z} be any vector in $\text{ri}(\text{dom}(f))$, and define $\tilde{\mathbf{y}} = (1 - \alpha)\mathbf{y} + \alpha\mathbf{z}$ for some $\alpha \in (0, 1)$, which at this point we fix. By the line segment principle (Lemma 5.23), $\tilde{\mathbf{y}} \in \text{ri}(\text{dom}(f))$. Consider now the one-dimensional function

$$\varphi(\lambda) = f(\mathbf{x}_\lambda), \quad \lambda \in [0, 1],$$

where $\mathbf{x}_\lambda = (1 - \lambda)\mathbf{x} + \lambda\tilde{\mathbf{y}}$. For any $\lambda \in (0, 1)$, let $\mathbf{g}_\lambda \in \partial f(\mathbf{x}_\lambda)$ (whose existence is guaranteed since $\mathbf{x}_\lambda \in \text{ri}(\text{dom}(f))$ by the line segment principle). Then $\langle \mathbf{g}_\lambda, \tilde{\mathbf{y}} - \mathbf{x} \rangle \in \partial \varphi(\lambda)$, and hence by the mean-value theorem (Lemma 5.22),

$$f(\tilde{\mathbf{y}}) - f(\mathbf{x}) = \varphi(1) - \varphi(0) = \int_0^1 \langle \mathbf{g}_\lambda, \tilde{\mathbf{y}} - \mathbf{x} \rangle d\lambda. \quad (5.23)$$

Since $\mathbf{g} \in \partial f(\mathbf{x})$ and $\mathbf{g}_\lambda \in \partial f(\mathbf{x}_\lambda)$, by property (iii),

$$\langle \mathbf{g}_\lambda - \mathbf{g}, \mathbf{x}_\lambda - \mathbf{x} \rangle \geq \sigma\|\mathbf{x}_\lambda - \mathbf{x}\|^2,$$

which is equivalent to

$$\langle \mathbf{g}_\lambda, \tilde{\mathbf{y}} - \mathbf{x} \rangle \geq \langle \mathbf{g}, \tilde{\mathbf{y}} - \mathbf{x} \rangle + \sigma \lambda \|\tilde{\mathbf{y}} - \mathbf{x}\|^2.$$

Plugging the last inequality into (5.23), we obtain that

$$\begin{aligned} f(\tilde{\mathbf{y}}) - f(\mathbf{x}) &\geq \int_0^1 [\langle \mathbf{g}, \tilde{\mathbf{y}} - \mathbf{x} \rangle + \sigma \lambda \|\tilde{\mathbf{y}} - \mathbf{x}\|^2] d\lambda \\ &= \langle \mathbf{g}, \tilde{\mathbf{y}} - \mathbf{x} \rangle + \frac{\sigma}{2} \|\tilde{\mathbf{y}} - \mathbf{x}\|^2. \end{aligned}$$

Recalling the definition of $\tilde{\mathbf{y}}$, we obtain that for any $\alpha \in (0, 1)$,

$$f((1-\alpha)\mathbf{y} + \alpha\mathbf{z}) \geq f(\mathbf{x}) + \langle \mathbf{g}, (1-\alpha)\mathbf{y} + \alpha\mathbf{z} - \mathbf{x} \rangle + \frac{\sigma}{2} \|(1-\alpha)\mathbf{y} + \alpha\mathbf{z} - \mathbf{x}\|^2.$$

Taking $\alpha \rightarrow 0^+$ and using the continuity of the one-dimensional function $\alpha \mapsto f((1-\alpha)\mathbf{y} + \alpha\mathbf{z})$ over $[0, 1]$ (follows by invoking Theorem 2.22 and recalling that the one-dimensional function is closed and convex), we obtain the desired result

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle + \frac{\sigma}{2} \|\mathbf{y} - \mathbf{x}\|^2. \quad \square$$

The next theorem states that a proper closed and strongly convex function has a unique minimizer and that it satisfies a certain growth property around the minimizer.

Theorem 5.25 (existence and uniqueness of a minimizer of closed strongly convex functions). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper closed and σ -strongly convex function ($\sigma > 0$). Then*

- (a) *f has a unique minimizer;*
- (b) *$f(\mathbf{x}) - f(\mathbf{x}^*) \geq \frac{\sigma}{2} \|\mathbf{x} - \mathbf{x}^*\|^2$ for all $\mathbf{x} \in \text{dom}(f)$, where \mathbf{x}^* is the unique minimizer of f .*

Proof. (a) Since $\text{dom}(f)$ is nonempty and convex, it follows that there exists $\mathbf{x}_0 \in \text{ri}(\text{dom}(f))$ (Theorem 3.17), and consequently, by Theorem 3.18, $\partial f(\mathbf{x}_0) \neq \emptyset$. Let $\mathbf{g} \in \partial f(\mathbf{x}_0)$. Then by the equivalence between σ -strong convexity and property (ii) of Theorem 5.24, it follows that

$$f(\mathbf{x}) \geq f(\mathbf{x}_0) + \langle \mathbf{g}, \mathbf{x} - \mathbf{x}_0 \rangle + \frac{\sigma}{2} \|\mathbf{x} - \mathbf{x}_0\|^2 \text{ for all } \mathbf{x} \in \mathbb{E}.$$

Since all norms in finite dimensional spaces are equivalent, there exists a constant $C > 0$ such that

$$\|\mathbf{y}\| \geq \sqrt{C} \|\mathbf{y}\|_a,$$

where $\|\cdot\|_a \equiv \sqrt{\langle \cdot, \cdot \rangle}$ denotes the Euclidean norm associated with the inner product of the space \mathbb{E} (which might be different than the endowed norm $\|\cdot\|$). Therefore,

$$f(\mathbf{x}) \geq f(\mathbf{x}_0) + \langle \mathbf{g}, \mathbf{x} - \mathbf{x}_0 \rangle + \frac{C\sigma}{2} \|\mathbf{x} - \mathbf{x}_0\|_a^2 \text{ for any } \mathbf{x} \in \mathbb{E},$$

which is the same as

$$f(\mathbf{x}) \geq f(\mathbf{x}_0) - \frac{1}{2C\sigma} \|\mathbf{g}\|_a^2 + \frac{C\sigma}{2} \left\| \mathbf{x} - \left(\mathbf{x}_0 - \frac{1}{C\sigma} \mathbf{g} \right) \right\|_a^2 \text{ for any } \mathbf{x} \in \mathbb{E}.$$

In particular, it follows that

$$\text{Lev}(f, f(\mathbf{x}_0)) \subseteq B_{\|\cdot\|_a} \left[\mathbf{x}_0 - \frac{1}{C\sigma} \mathbf{g}, \frac{1}{C\sigma} \|\mathbf{g}\|_a \right].$$

Since f is closed, the above level set is closed (Theorem 2.6), and since it is contained in a ball, it is also bounded. Therefore, $\text{Lev}(f, f(\mathbf{x}_0))$ is compact. We can thus deduce that the optimal set of the problem of minimizing f over $\text{dom}(f)$ is the same as the optimal set of the problem of minimizing f over the nonempty compact set $\text{Lev}(f, f(\mathbf{x}_0))$. Invoking Weierstrass theorem for closed functions (Theorem 2.12), it follows that a minimizer exists. To show the uniqueness, assume that $\tilde{\mathbf{x}}$ and $\hat{\mathbf{x}}$ are minimizers of f . Then $f(\tilde{\mathbf{x}}) = f(\hat{\mathbf{x}}) = f_{\text{opt}}$, where f_{opt} is the minimal value of f . Then by the definition of σ -strong convexity of f ,

$$f_{\text{opt}} \leq f\left(\frac{1}{2}\tilde{\mathbf{x}} + \frac{1}{2}\hat{\mathbf{x}}\right) \leq \frac{1}{2}f(\tilde{\mathbf{x}}) + \frac{1}{2}f(\hat{\mathbf{x}}) - \frac{\sigma}{8}\|\tilde{\mathbf{x}} - \hat{\mathbf{x}}\|^2 = f_{\text{opt}} - \frac{\sigma}{8}\|\tilde{\mathbf{x}} - \hat{\mathbf{x}}\|^2,$$

implying that $\tilde{\mathbf{x}} = \hat{\mathbf{x}}$ and hence establishing the uniqueness of the minimizer of f .

(b) Let \mathbf{x}^* be the unique minimizer of f . Then by Fermat's optimality condition (Theorem 3.63), $\mathbf{0} \in \partial f(\mathbf{x}^*)$ and hence by using the equivalence between σ -strong convexity and property (ii) of Theorem 5.24, it follows that

$$f(\mathbf{x}) - f(\mathbf{x}^*) \geq \langle \mathbf{0}, \mathbf{x} - \mathbf{x}^* \rangle + \frac{\sigma}{2}\|\mathbf{x} - \mathbf{x}^*\|^2 = \frac{\sigma}{2}\|\mathbf{x} - \mathbf{x}^*\|^2 \quad (5.24)$$

for any $\mathbf{x} \in \mathbb{E}$, establishing claim (b). \square

5.3 Smoothness and Strong Convexity Correspondence

5.3.1 The Conjugate Correspondence Theorem

An extremely useful connection between smoothness and strong convexity is given in the conjugate correspondence theorem that, loosely speaking, states that f is σ -strongly convex if and only if f^* is $\frac{1}{\sigma}$ -smooth.

Theorem 5.26 (conjugate correspondence theorem). *Let $\sigma > 0$. Then*

- (a) *If $f : \mathbb{E} \rightarrow \mathbb{R}$ is a $\frac{1}{\sigma}$ -smooth convex function, then f^* is σ -strongly convex w.r.t. the dual norm $\|\cdot\|_*$.*
- (b) *If $f : \mathbb{E} \rightarrow (-\infty, \infty]$ is a proper closed σ -strongly convex function, then $f^* : \mathbb{E}^* \rightarrow \mathbb{R}$ is $\frac{1}{\sigma}$ -smooth.*

Proof. (a) Suppose that $f : \mathbb{E} \rightarrow \mathbb{R}$ is a $\frac{1}{\sigma}$ -smooth convex function. To prove that f^* is σ -strongly convex, take $\mathbf{y}_1, \mathbf{y}_2 \in \text{dom}(\partial f^*)$ and $\mathbf{v}_1 \in \partial f^*(\mathbf{y}_1), \mathbf{v}_2 \in \partial f^*(\mathbf{y}_2)$. Then by the conjugate subgradient theorem (Theorem 4.20), using also the properness closedness and convexity of f , it follows that $\mathbf{y}_1 \in \partial f(\mathbf{v}_1)$ and $\mathbf{y}_2 \in \partial f(\mathbf{v}_2)$, which, by the differentiability of f , implies that $\mathbf{y}_1 = \nabla f(\mathbf{v}_1)$ and $\mathbf{y}_2 = \nabla f(\mathbf{v}_2)$ (see Theorem 3.33). By the equivalence between properties (i) and (iv) in Theorem 5.8, we can write

$$\langle \mathbf{y}_1 - \mathbf{y}_2, \mathbf{v}_1 - \mathbf{v}_2 \rangle \geq \sigma \|\mathbf{y}_1 - \mathbf{y}_2\|_*^2.$$

Since the last inequality holds for any $\mathbf{y}_1, \mathbf{y}_2 \in \text{dom}(\partial f^*)$ and $\mathbf{v}_1 \in \partial f^*(\mathbf{y}_1), \mathbf{v}_2 \in \partial f^*(\mathbf{y}_2)$, it follows by the equivalence between σ -strong convexity and property (iii) of Theorem 5.24 that f^* is a σ -strongly convex function.

(b) Suppose that f is a proper closed σ -strongly convex function. By the conjugate subgradient theorem (Corollary 4.21),

$$\partial f^*(\mathbf{y}) = \operatorname{argmax}_{\mathbf{x} \in \mathbb{E}} \{\langle \mathbf{x}, \mathbf{y} \rangle - f(\mathbf{x})\} \text{ for any } \mathbf{y} \in \mathbb{E}^*.$$

Thus, by the strong convexity and closedness of f , along with Theorem 5.25(a), it follows that $\partial f^*(\mathbf{y})$ is a singleton for any $\mathbf{y} \in \mathbb{E}^*$. Therefore, by Theorem 3.33, f^* is differentiable over the entire dual space \mathbb{E}^* . To show the $\frac{1}{\sigma}$ -smoothness of f^* , take $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{E}^*$ and denote $\mathbf{v}_1 = \nabla f^*(\mathbf{y}_1), \mathbf{v}_2 = \nabla f^*(\mathbf{y}_2)$. These relations, by the conjugate subgradient theorem (Theorem 4.20), are equivalent to $\mathbf{y}_1 \in \partial f(\mathbf{v}_1), \mathbf{y}_2 \in \partial f(\mathbf{v}_2)$. Therefore, by Theorem 5.24 (equivalence between properties (i) and (iii)),

$$\langle \mathbf{y}_1 - \mathbf{y}_2, \mathbf{v}_1 - \mathbf{v}_2 \rangle \geq \sigma \|\mathbf{v}_1 - \mathbf{v}_2\|^2,$$

that is,

$$\langle \mathbf{y}_1 - \mathbf{y}_2, \nabla f^*(\mathbf{y}_1) - \nabla f^*(\mathbf{y}_2) \rangle \geq \sigma \|\nabla f^*(\mathbf{y}_1) - \nabla f^*(\mathbf{y}_2)\|^2,$$

which, combined with the generalized Cauchy–Schwarz inequality (Lemma 1.4), implies the inequality

$$\|\nabla f^*(\mathbf{y}_1) - \nabla f^*(\mathbf{y}_2)\| \leq \frac{1}{\sigma} \|\mathbf{y}_1 - \mathbf{y}_2\|_*,$$

proving the $\frac{1}{\sigma}$ -smoothness of f^* . \square

5.3.2 Examples of Strongly Convex Functions

We can use the conjugate correspondence theorem (Theorem 5.26) to conclude several results on the strong convexity of several important functions.

Example 5.27 (negative entropy over the unit simplex). Consider the function $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ given by

$$f(\mathbf{x}) = \begin{cases} \sum_{i=1}^n x_i \log x_i, & \mathbf{x} \in \Delta_n, \\ \infty & \text{else.} \end{cases}$$

Then, by Section 4.4.10, the conjugate of this function is the log-sum-exp function $f^*(\mathbf{y}) = \log(\sum_{i=1}^n e^{y_i})$, which, by Example 5.15, is a 1-smooth function w.r.t. both the l_∞ - and l_2 -norms. Consequently, by the conjugate correspondence theorem, f is 1-strongly convex w.r.t. both the l_1 - and l_2 -norms. ■

Example 5.28 (squared p -norm for $p \in (1, 2]$). Consider the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ given by $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_p^2$ ($p \in (1, 2]$). Then, by Section 4.4.15, $f^*(\mathbf{y}) = \frac{1}{2}\|\mathbf{y}\|_q^2$, where $q \geq 2$ is determined by the relation $\frac{1}{p} + \frac{1}{q} = 1$. By Example 5.11, f^* is a $(q-1)$ -smooth function w.r.t. the l_q -norm, which, by the conjugate correspondence theorem, implies that the function f is $\frac{1}{q-1}$ -strongly convex w.r.t. the l_p -norm. Since $\frac{1}{q-1} = p - 1$, we conclude that f is $(p-1)$ -strongly convex w.r.t. the l_p -norm. ■

Example 5.29 (l_2 ball-pen function). Consider the ball-pen function $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ given by

$$f(\mathbf{x}) = \begin{cases} -\sqrt{1 - \|\mathbf{x}\|_2^2}, & \|\mathbf{x}\|_2 \leq 1, \\ \infty & \text{else.} \end{cases}$$

By Section 4.4.13, the conjugate of f is

$$f^*(\mathbf{y}) = \sqrt{\|\mathbf{y}\|_2^2 + 1},$$

which, by Example 5.14, is known to be 1-smooth w.r.t. the l_2 -norm, and hence, by the conjugate correspondence theorem, f is 1-strongly convex w.r.t. the l_2 -norm. ■

The table below contains all the strongly convex functions described in this chapter.

$f(\mathbf{x})$	$\text{dom}(f)$	Strong convexity parameter	Norm	Reference
$\frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} + 2\mathbf{b}^T \mathbf{x} + c$ ($\mathbf{A} \in \mathbb{S}_{++}^n$, $\mathbf{b} \in \mathbb{R}^n$, $c \in \mathbb{R}$)	\mathbb{R}^n	$\lambda_{\min}(\mathbf{A})$	l_2	Example 5.19
$\frac{1}{2}\ \mathbf{x}\ ^2 + \delta_C(\mathbf{x})$ ($\emptyset \neq C \subseteq \mathbb{E}$ convex)	C	1	Euclidean	Example 5.21
$-\sqrt{1 - \ \mathbf{x}\ _2^2}$	$B_{\ \cdot\ _2}[\mathbf{0}, 1]$	1	l_2	Example 5.29
$\frac{1}{2}\ \mathbf{x}\ _p^2$ ($p \in (1, 2]$)	\mathbb{R}^n	$p - 1$	l_p	Example 5.28
$\sum_{i=1}^n x_i \log x_i$	Δ_n	1	l_2 or l_1	Example 5.27

5.3.3 Smoothness and Differentiability of the Infimal Convolution

We will now show that under appropriate conditions, the infimal convolution of a convex function and an L -smooth convex function is also L -smooth; in addition, we will derive an expression for the gradient. The proof of the result is based on the conjugate correspondence theorem.

Theorem 5.30 (smoothness of the infimal convolution). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper closed and convex function, and let $\omega : \mathbb{E} \rightarrow \mathbb{R}$ be an L -smooth convex function. Assume that $f \square \omega$ is real-valued. Then the following hold:*

(a) *$f \square \omega$ is L -smooth.*

(b) *Let $\mathbf{x} \in \mathbb{E}$, and assume that $\mathbf{u}(\mathbf{x})$ is a minimizer of*

$$\min_{\mathbf{u}} \{f(\mathbf{u}) + \omega(\mathbf{x} - \mathbf{u})\}. \quad (5.25)$$

Then $\nabla(f \square \omega)(\mathbf{x}) = \nabla\omega(\mathbf{x} - \mathbf{u}(\mathbf{x}))$.

Proof. (a) By Theorem 4.19,

$$f \square \omega = (f^* + \omega^*)^*.$$

Since f and ω are proper closed and convex, then so are f^*, ω^* (Theorems 4.3, 4.5). In addition, by the conjugate correspondence theorem (Theorem 5.26), ω^* is $\frac{1}{L}$ -strongly convex. Therefore, by Lemma 5.20, $f^* + \omega^*$ is $\frac{1}{L}$ -strongly convex, and it is also closed as a sum of closed functions; we will prove that it is also proper. Indeed, by Theorem 4.16,

$$(f \square \omega)^* = f^* + \omega^*.$$

Since $f \square \omega$ is convex (by Theorem 2.19) and proper, it follows that $f^* + \omega^*$ is proper as a conjugate of a proper and convex function (Theorem 4.5). Thus, since $f^* + \omega^*$ is proper closed and $\frac{1}{L}$ -strongly convex function, by the conjugate correspondence theorem, it follows that $f \square \omega = (f^* + \omega^*)^*$ is L -smooth.

(b) Let $\mathbf{x} \in \mathbb{E}$ be such that $\mathbf{u}(\mathbf{x})$ is a minimizer of (5.25), namely,

$$(f \square \omega)(\mathbf{x}) = f(\mathbf{u}(\mathbf{x})) + \omega(\mathbf{x} - \mathbf{u}(\mathbf{x})). \quad (5.26)$$

For convenience, define $\mathbf{z} \equiv \nabla\omega(\mathbf{x} - \mathbf{u}(\mathbf{x}))$. Our objective is to show that $\nabla(f \square \omega)(\mathbf{x}) = \mathbf{z}$. This means that we have to show that for any $\boldsymbol{\xi} \in \mathbb{E}$, $\lim_{\|\boldsymbol{\xi}\| \rightarrow 0} |\phi(\boldsymbol{\xi})|/\|\boldsymbol{\xi}\| = 0$, where $\phi(\boldsymbol{\xi}) \equiv (f \square \omega)(\mathbf{x} + \boldsymbol{\xi}) - (f \square \omega)(\mathbf{x}) - \langle \boldsymbol{\xi}, \mathbf{z} \rangle$. By the definition of the infimal convolution,

$$(f \square \omega)(\mathbf{x} + \boldsymbol{\xi}) \leq f(\mathbf{u}(\mathbf{x})) + \omega(\mathbf{x} + \boldsymbol{\xi} - \mathbf{u}(\mathbf{x})), \quad (5.27)$$

which, combined with (5.26), yields

$$\begin{aligned}
 \phi(\xi) &= (f \square \omega)(\mathbf{x} + \xi) - (f \square \omega)(\mathbf{x}) - \langle \xi, \mathbf{z} \rangle \\
 &\leq \omega(\mathbf{x} + \xi - \mathbf{u}(\mathbf{x})) - \omega(\mathbf{x} - \mathbf{u}(\mathbf{x})) - \langle \xi, \mathbf{z} \rangle && [(5.26), (5.27)] \\
 &\leq \langle \xi, \nabla \omega(\mathbf{x} + \xi - \mathbf{u}(\mathbf{x})) \rangle - \langle \xi, \mathbf{z} \rangle, && [\text{gradient inequality for } \omega] \\
 &= \langle \xi, \nabla \omega(\mathbf{x} + \xi - \mathbf{u}(\mathbf{x})) - \nabla \omega(\mathbf{x} - \mathbf{u}(\mathbf{x})) \rangle && [\text{substitution of } \mathbf{z}] \\
 &\leq \|\xi\| \cdot \|\nabla \omega(\mathbf{x} + \xi - \mathbf{u}(\mathbf{x})) - \nabla \omega(\mathbf{x} - \mathbf{u}(\mathbf{x}))\|_* && [\text{generalized Cauchy-Schwarz}] \\
 &\leq L\|\xi\|^2. && [L\text{-smoothness of } \omega]
 \end{aligned}$$

To complete the proof, it is enough to show that we also have $\phi(\xi) \geq -L\|\xi\|^2$. Since $f \square \omega$ is convex, so is ϕ , which, along the fact that $\phi(\mathbf{0}) = 0$, implies that $\phi(\xi) \geq -\phi(-\xi)$, and hence the desired result follows. \square

Example 5.31 (revisiting the 1-smoothness of $\frac{1}{2}d_C^2$). Suppose that \mathbb{E} is a Euclidean space and let $C \subseteq \mathbb{E}$ be a nonempty closed and convex set. Consider the function $\varphi_C(\mathbf{x}) = \frac{1}{2}d_C^2(\mathbf{x})$. We have already shown in Example 5.5 that it is 1-smooth. We will provide here a second proof for this result, which is based on Theorem 5.30. Note that $\varphi_C = \delta_C \square h$, where $h(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2$. Since h is a real-valued 1-smooth convex function, and since δ_C is proper closed and convex, it follows by Theorem 5.30 that φ_C is 1-smooth. \blacksquare

Chapter 6

The Proximal Operator

Underlying Space: In this chapter \mathbb{E} is a Euclidean space, meaning a finite dimensional space endowed with an inner product $\langle \cdot, \cdot \rangle$ and the Euclidean norm $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$.

This chapter is devoted to the study of the proximal mapping, which will be fundamental in many of the algorithms that will be explored later in the book. The operator and its properties were first studied by Moreau, and hence it is also referred to as “Moreau’s proximal mapping.”

6.1 Definition, Existence, and Uniqueness

Definition 6.1 (proximal mapping). *Given a function $f : \mathbb{E} \rightarrow (-\infty, \infty]$, the proximal mapping of f is the operator given by*

$$\text{prox}_f(\mathbf{x}) = \operatorname{argmin}_{\mathbf{u} \in \mathbb{E}} \left\{ f(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \right\} \text{ for any } \mathbf{x} \in \mathbb{E}.$$

We will often use the term “prox” instead of “proximal.” The mapping prox_f takes a vector $\mathbf{x} \in \mathbb{E}$ and maps it into a subset of \mathbb{E} , which might be empty, a singleton, or a set with multiple vectors as the following example illustrates.

Example 6.2. Consider the following three functions from \mathbb{R} to \mathbb{R} :

$$g_1(x) \equiv 0,$$

$$g_2(x) = \begin{cases} 0, & x \neq 0, \\ -\lambda, & x = 0, \end{cases}$$

$$g_3(x) = \begin{cases} 0, & x \neq 0, \\ \lambda, & x = 0, \end{cases}$$

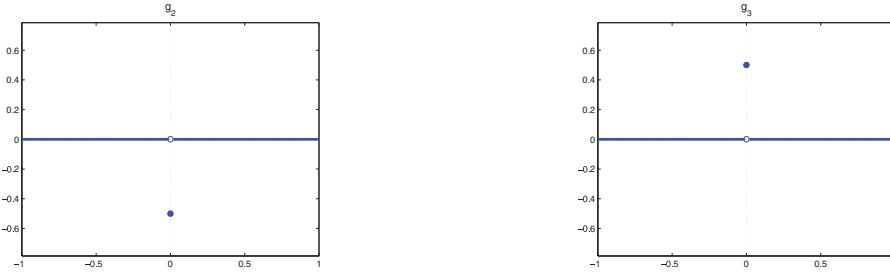


Figure 6.1. The left and right images are the plots of the functions g_2 and g_3 , respectively, with $\lambda = 0.5$ from Example 6.2.

where $\lambda > 0$ is a given constant. The plots of the noncontinuous functions g_2 and g_3 are given in Figure 6.1. The prox of g_1 can be computed as follows:

$$\text{prox}_{g_1}(x) = \operatorname{argmin}_{u \in \mathbb{R}} \left\{ g_1(u) + \frac{1}{2}(u - x)^2 \right\} = \operatorname{argmin}_{u \in \mathbb{R}} \left\{ \frac{1}{2}(u - x)^2 \right\} = \{x\}.$$

To compute the prox of g_2 , note that $\text{prox}_{g_2}(x) = \operatorname{argmin}_{u \in \mathbb{R}} \tilde{g}_2(u, x)$, where

$$\tilde{g}_2(u, x) \equiv g_2(u) + \frac{1}{2}(u - x)^2 = \begin{cases} -\lambda + \frac{x^2}{2}, & u = 0, \\ \frac{1}{2}(u - x)^2, & u \neq 0. \end{cases}$$

For $x \neq 0$, the minimum of $\frac{1}{2}(u - x)^2$ over $\mathbb{R} \setminus \{0\}$ is attained at $u = x (\neq 0)$ with a minimal value of 0. Therefore, in this case, if $0 > -\lambda + \frac{x^2}{2}$, then the unique minimizer of $\tilde{g}_2(\cdot, x)$ is $u = 0$, and if $0 < -\lambda + \frac{x^2}{2}$, then $u = x$ is the unique minimizer of $\tilde{g}_2(\cdot, x)$; finally, if $0 = -\lambda + \frac{x^2}{2}$, then 0 and x are the two minimizers $\tilde{g}_2(\cdot, x)$. When $x = 0$, the minimizer of $\tilde{g}_2(\cdot, 0)$ is $u = 0$. To conclude,

$$\text{prox}_{g_2}(x) = \begin{cases} \{0\}, & |x| < \sqrt{2\lambda}, \\ \{x\}, & |x| > \sqrt{2\lambda}, \\ \{0, x\}, & |x| = \sqrt{2\lambda}. \end{cases}$$

Similar arguments show that

$$\text{prox}_{g_3}(x) = \begin{cases} \{x\}, & x \neq 0, \\ \emptyset, & x = 0. \end{cases} \blacksquare$$

The next theorem, called the *first prox theorem*, states that if f is proper closed and convex, then $\text{prox}_f(\mathbf{x})$ is always a singleton, meaning that the prox exists and is unique. This is the reason why in the last example only g_1 , which was proper closed and convex, had a unique prox at any point.

Theorem 6.3 (first prox theorem). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper closed and convex function. Then $\text{prox}_f(\mathbf{x})$ is a singleton for any $\mathbf{x} \in \mathbb{E}$.*

Proof. For any $\mathbf{x} \in \mathbb{E}$,

$$\text{prox}_f(\mathbf{x}) = \operatorname{argmin}_{\mathbf{u} \in \mathbb{E}} \tilde{f}(\mathbf{u}, \mathbf{x}), \quad (6.1)$$

where $\tilde{f}(\mathbf{u}, \mathbf{x}) \equiv f(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2$. The function $\tilde{f}(\cdot, \mathbf{x})$ is a closed and strongly convex function as a sum of the closed and strongly convex function $\frac{1}{2}\|\cdot - \mathbf{x}\|^2$ and the closed and convex function f (see Lemma 5.20 and Theorem 2.7(b)). The properness of $\tilde{f}(\cdot, \mathbf{x})$ immediately follows from the properness of f . Therefore, by Theorem 5.25(a), there exists a unique minimizer to the problem in (6.1). \square

When f is proper closed and convex, the last result shows that $\text{prox}_f(\mathbf{x})$ is a singleton for any $\mathbf{x} \in \mathbb{E}$. In these cases, which will constitute the vast majority of cases that will be discussed in this chapter, we will treat prox_f as a single-valued mapping from \mathbb{E} to \mathbb{E} , meaning that we will write $\text{prox}_f(\mathbf{x}) = \mathbf{y}$ and not $\text{prox}_f(\mathbf{x}) = \{\mathbf{y}\}$.

If we relax the assumptions in the first prox theorem and only require closedness of the function, then it is possible to show under some coerciveness assumptions that $\text{prox}_f(\mathbf{x})$ is never an empty set.

Theorem 6.4 (nonemptiness of the prox under closedness and coerciveness). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper closed function, and assume that the following condition is satisfied:*

$$\text{the function } \mathbf{u} \mapsto f(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2 \text{ is coercive for any } \mathbf{x} \in \mathbb{E}. \quad (6.2)$$

Then $\text{prox}_f(\mathbf{x})$ is nonempty for any $\mathbf{x} \in \mathbb{E}$.

Proof. For any $\mathbf{x} \in \mathbb{E}$, the proper function $h(\mathbf{u}) \equiv f(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2$ is closed as a sum of two closed functions. Since by the premise of the theorem it is also coercive, it follows by Theorem 2.14 (with $S = \mathbb{E}$) that $\text{prox}_f(\mathbf{x})$, which consists of the minimizers of h , is nonempty. \square

Example 6.2 actually gave an illustration of Theorem 6.4 since although both g_2 and g_3 satisfy the coercivity assumption (6.2), only g_2 was closed, and thus the fact that $\text{prox}_{g_3}(x)$ was empty for a certain value of x , as opposed to $\text{prox}_{g_2}(x)$, which was never empty, is not surprising.

6.2 First Set of Examples of Proximal Mappings

Equipped just with the definition of the proximal mapping, we will now compute the proximal mapping of several proper closed and convex functions.

6.2.1 Constant

If $f \equiv c$ for some $c \in \mathbb{R}$, then

$$\text{prox}_f(\mathbf{x}) = \operatorname{argmin}_{\mathbf{u} \in \mathbb{E}} \left\{ c + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \right\} = \mathbf{x}.$$

Therefore,

$$\boxed{\text{prox}_f(\mathbf{x}) = \mathbf{x}}$$

is the identity mapping.

6.2.2 Affine

Let $f(\mathbf{x}) = \langle \mathbf{a}, \mathbf{x} \rangle + b$, where $\mathbf{a} \in \mathbb{E}$ and $b \in \mathbb{R}$. Then

$$\begin{aligned} \text{prox}_f(\mathbf{x}) &= \operatorname{argmin}_{\mathbf{u} \in \mathbb{E}} \left\{ \langle \mathbf{a}, \mathbf{u} \rangle + b + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \right\} \\ &= \operatorname{argmin}_{\mathbf{u} \in \mathbb{E}} \left\{ \langle \mathbf{a}, \mathbf{x} \rangle + b - \frac{1}{2} \|\mathbf{a}\|^2 + \frac{1}{2} \|\mathbf{u} - (\mathbf{x} - \mathbf{a})\|^2 \right\} \\ &= \mathbf{x} - \mathbf{a}. \end{aligned}$$

Therefore,

$$\boxed{\text{prox}_f(\mathbf{x}) = \mathbf{x} - \mathbf{a}}$$

is a translation mapping.

6.2.3 Convex Quadratic

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x} + c$, where $\mathbf{A} \in \mathbb{S}_+^n$, $\mathbf{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$. The vector $\text{prox}_f(\mathbf{x})$ is the minimizer of the problem

$$\min_{\mathbf{u} \in \mathbb{E}} \left\{ \frac{1}{2} \mathbf{u}^T \mathbf{A} \mathbf{u} + \mathbf{b}^T \mathbf{u} + c + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \right\}.$$

The optimal solution of the last problem is attained when the gradient of the objective function vanishes:

$$\mathbf{A} \mathbf{u} + \mathbf{b} + \mathbf{u} - \mathbf{x} = \mathbf{0},$$

that is, when

$$(\mathbf{A} + \mathbf{I}) \mathbf{u} = \mathbf{x} - \mathbf{b},$$

and hence

$$\boxed{\text{prox}_f(\mathbf{x}) = (\mathbf{A} + \mathbf{I})^{-1}(\mathbf{x} - \mathbf{b})}.$$

6.2.4 One-Dimensional Examples

The following lemma contains several prox computations of one-dimensional functions.

Lemma 6.5. *The following are pairs of proper closed and convex functions and their prox mappings:*

$$\begin{aligned}
 g_1(x) &= \begin{cases} \mu x, & x \geq 0, \\ \infty, & x < 0, \end{cases} & \text{prox}_{g_1}(x) &= [x - \mu]_+, \\
 g_2(x) &= \lambda|x|, & \text{prox}_{g_2}(x) &= [|x| - \lambda]_+ \text{sgn}(x), \\
 g_3(x) &= \begin{cases} \lambda x^3, & x \geq 0, \\ \infty, & x < 0, \end{cases} & \text{prox}_{g_3}(x) &= \frac{-1 + \sqrt{1 + 12\lambda[x]_+}}{6\lambda}, \\
 g_4(x) &= \begin{cases} -\lambda \log x, & x > 0, \\ \infty, & x \leq 0, \end{cases} & \text{prox}_{g_4}(x) &= \frac{x + \sqrt{x^2 + 4\lambda}}{2}, \\
 g_5(x) &= \delta_{[0,\eta] \cap \mathbb{R}}(x), & \text{prox}_{g_5}(x) &= \min\{\max\{x, 0\}, \eta\},
 \end{aligned}$$

where $\lambda \in \mathbb{R}_+$, $\eta \in [0, \infty]$, and $\mu \in \mathbb{R}$.

Proof. The proofs repeatedly use the following trivial arguments: (i) if $f'(u) = 0$ for a convex function f , then u must be one of its minimizers; (ii) if a minimizer of a convex function exists and is *not* attained at any point of differentiability, then it must be attained at a point of nondifferentiability.

[prox of g_1] By definition, $\text{prox}_{g_1}(x)$ is the minimizer of the function

$$f(u) = \begin{cases} \infty, & u < 0, \\ f_1(u), & u \geq 0, \end{cases}$$

where $f_1(u) = \mu u + \frac{1}{2}(u - x)^2$. First note that $f'_1(u) = 0$ if and only if $u = x - \mu$. If $x > \mu$, then $f'(x - \mu) = f'_1(x - \mu) = 0$, implying that in this case $\text{prox}_{g_1}(x) = x - \mu$. Otherwise, if $x \leq \mu$, the minimizer of f is not attained at a point of differentiability, meaning that it has to be attained at 0, which is the only point of nondifferentiability in the domain of f , so that $\text{prox}_{g_1}(x) = 0$.

[prox of g_2] $\text{prox}_{g_2}(x)$ is the minimizer of the function

$$h(u) = \begin{cases} h_1(u) \equiv \lambda u + \frac{1}{2}(u - x)^2, & u > 0, \\ h_2(u) \equiv -\lambda u + \frac{1}{2}(u - x)^2, & u \leq 0. \end{cases}$$

If the minimizer is attained at $u > 0$, then $0 = h'_1(u) = \lambda + u - x$, meaning that $u = x - \lambda$. Therefore, if $x > \lambda$, then $\text{prox}_{g_2}(x) = x - \lambda$. The same argument shows that if $x < -\lambda$, then $\text{prox}_{g_2}(x) = x + \lambda$. If $|x| \leq \lambda$, then $\text{prox}_{g_2}(x)$ must be the only point of nondifferentiability of h , namely, 0.

[prox of g_3] $\text{prox}_{g_3}(x)$ is the minimizer of the function

$$s(u) = \begin{cases} \lambda u^3 + \frac{1}{2}(u - x)^2, & u \geq 0, \\ \infty, & u < 0. \end{cases}$$

If the minimizer is positive, then $\tilde{u} = \text{prox}_{g_3}(x)$ satisfies $s'(\tilde{u}) = 0$, that is,

$$3\lambda\tilde{u}^2 + \tilde{u} - x = 0.$$

The above equation has a positive root if and only if $x > 0$, and in this case the (unique) positive root is $\text{prox}_{g_3}(x) = \tilde{u} = \frac{-1+\sqrt{1+12\lambda x}}{6\lambda}$. If $x \leq 0$, the minimizer of s is attained at the only point of nondifferentiability of s in its domain, that is, at 0.

[prox of g_4] $\tilde{u} = \text{prox}_{g_4}(x)$ is a minimizer over \mathbb{R}_{++} of

$$t(u) = -\lambda \log u + \frac{1}{2}(u - x)^2,$$

which is determined by the condition that the derivative vanishes:

$$-\frac{\lambda}{\tilde{u}} + (\tilde{u} - x) = 0,$$

that is,

$$\tilde{u}^2 - \tilde{u}x - \lambda = 0.$$

Therefore (taking the positive root),

$$\text{prox}_{g_4}(x) = \tilde{u} = \frac{x + \sqrt{x^2 + 4\lambda}}{2}.$$

[prox of g_5] We will first assume that $\eta < \infty$. Note that $\tilde{u} = \text{prox}_{g_5}(x)$ is the minimizer of

$$w(u) = \frac{1}{2}(u - x)^2$$

over $[0, \eta]$. The minimizer of w over \mathbb{R} is $u = x$. Therefore, if $0 \leq x \leq \eta$, then $\tilde{u} = x$. If $x < 0$, then w is increasing over $[0, \eta]$, and hence $\tilde{u} = 0$. Finally, if $x > \eta$, then w is decreasing over $[0, \eta]$, and thus $\tilde{u} = \eta$. To conclude,

$$\text{prox}_{g_5}(x) = \tilde{u} = \begin{cases} x, & 0 \leq x \leq \eta, \\ 0, & x < 0, \\ \eta, & x > \eta, \end{cases} = \min\{\max\{x, 0\}, \eta\}.$$

For $\eta = \infty$, $g_5(x) = \delta_{[0,\infty)}(x)$, and in this case, g_5 is identical to g_1 with $\mu = 0$, implying that $\text{prox}_{g_5}(x) = [x]_+$, which can also be written as

$$\text{prox}_{g_5}(x) = \min\{\max\{x, 0\}, \infty\}. \quad \square$$

6.3 Prox Calculus Rules

In this section we gather several important results on the calculus of proximal mappings. Note that some of the results do not require any convexity/closedness assumptions.

Theorem 6.6 (prox of separable functions). *Suppose that $f : \mathbb{E}_1 \times \mathbb{E}_2 \times \cdots \times \mathbb{E}_m \rightarrow (-\infty, \infty]$ is given by*

$$f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m) = \sum_{i=1}^m f_i(\mathbf{x}_i) \text{ for any } \mathbf{x}_i \in \mathbb{E}_i, \quad i = 1, 2, \dots, m.$$

Then for any $\mathbf{x}_1 \in \mathbb{E}_1, \mathbf{x}_2 \in \mathbb{E}_2, \dots, \mathbf{x}_m \in \mathbb{E}_m$,

$$\text{prox}_f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m) = \text{prox}_{f_1}(\mathbf{x}_1) \times \text{prox}_{f_2}(\mathbf{x}_2) \times \cdots \times \text{prox}_{f_m}(\mathbf{x}_m). \quad (6.3)$$

Proof. Formula (6.3) is a result of the following chain of equalities:

$$\begin{aligned} \text{prox}_f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m) &= \operatorname{argmin}_{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m} \sum_{i=1}^m \left[\frac{1}{2} \|\mathbf{y}_i - \mathbf{x}_i\|^2 + f_i(\mathbf{y}_i) \right] \\ &= \prod_{i=1}^m \operatorname{argmin}_{\mathbf{y}_i} \left[\frac{1}{2} \|\mathbf{y}_i - \mathbf{x}_i\|^2 + f_i(\mathbf{y}_i) \right] \\ &= \prod_{i=1}^m \text{prox}_{f_i}(\mathbf{x}_i). \quad \square \end{aligned}$$

Remark 6.7. *If $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is proper closed convex and separable,*

$$f(\mathbf{x}) = \sum_{i=1}^n f_i(x_i),$$

with f_i being proper closed and convex univariate functions, then the result of Theorem 6.6 can be rewritten as

$$\text{prox}_f(\mathbf{x}) = (\text{prox}_{f_i}(x_i))_{i=1}^n.$$

Example 6.8 (l_1 -norm). Suppose that $g : \mathbb{R}^n \rightarrow \mathbb{R}$ is given by $g(\mathbf{x}) = \lambda \|\mathbf{x}\|_1$, where $\lambda > 0$. Then

$$g(\mathbf{x}) = \sum_{i=1}^n \varphi(x_i), \quad (6.4)$$

where $\varphi(t) = \lambda|t|$. By Lemma 6.5 (computation of prox_{g_2}), $\text{prox}_\varphi(s) = \mathcal{T}_\lambda(s)$, where \mathcal{T}_λ is defined as

$$\mathcal{T}_\lambda(y) = [|y| - \lambda]_+ \operatorname{sgn}(y) = \begin{cases} y - \lambda, & y \geq \lambda, \\ 0, & |y| < \lambda, \\ y + \lambda, & y \leq -\lambda. \end{cases}$$

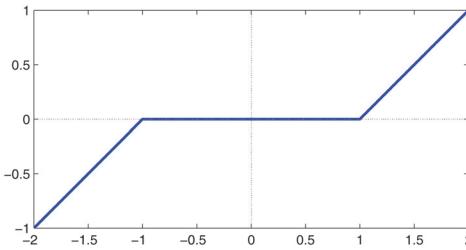


Figure 6.2. The soft thresholding function \mathcal{T}_1 .

The function \mathcal{T}_λ is called the *soft thresholding* function, and its description is given in Figure 6.2.

By Theorem 6.6,

$$\text{prox}_g(\mathbf{x}) = (\mathcal{T}_\lambda(x_j))_{j=1}^n.$$

We will expand the definition of the soft thresholding function for vectors by applying it componentwise, that is, for any $\mathbf{x} \in \mathbb{R}^n$,

$$\mathcal{T}_\lambda(\mathbf{x}) \equiv (\mathcal{T}_\lambda(x_j))_{j=1}^n = [|x| - \lambda \mathbf{e}]_+ \odot \text{sgn}(\mathbf{x}).$$

In this notation,

$$\text{prox}_g(\mathbf{x}) = \mathcal{T}_\lambda(\mathbf{x}).$$

■

Example 6.9 (negative sum of logs). Let $g : \mathbb{R}^n \rightarrow (-\infty, \infty]$ be given by

$$g(\mathbf{x}) = \begin{cases} -\lambda \sum_{j=1}^n \log x_j, & \mathbf{x} > \mathbf{0}, \\ \infty & \text{else,} \end{cases}$$

where $\lambda > 0$. Then $g(\mathbf{x}) = \sum_{i=1}^n \varphi(x_i)$, where

$$\varphi(t) = \begin{cases} -\lambda \log t, & t > 0, \\ \infty, & t < 0. \end{cases}$$

By Lemma 6.5 (computation of prox_{g_4}),

$$\text{prox}_\varphi(s) = \frac{s + \sqrt{s^2 + 4\lambda}}{2}.$$

Thus, by Theorem 6.6,

$$\text{prox}_g(\mathbf{x}) = (\text{prox}_\varphi(x_j))_{j=1}^n = \left(\frac{x_j + \sqrt{x_j^2 + 4\lambda}}{2} \right)_{j=1}^n.$$

■

Example 6.10 (l_0 -norm). Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}) = \lambda \|\mathbf{x}\|_0$, where $\lambda > 0$ and $\|\mathbf{x}\|_0 = \#\{i : x_i \neq 0\}$ is the l_0 -norm discussed in Example 2.11. For any $\mathbf{x} \in \mathbb{R}^n$,

$$f(\mathbf{x}) = \sum_{i=1}^n I(x_i),$$

where

$$I(t) = \begin{cases} \lambda, & t \neq 0, \\ 0, & t = 0. \end{cases}$$

Note that $I(\cdot) = J(\cdot) + \lambda$, where

$$J(t) = \begin{cases} 0, & t \neq 0, \\ -\lambda, & t = 0, \end{cases}$$

and that by Example 6.2,

$$\text{prox}_J(s) = \begin{cases} \{0\}, & |s| < \sqrt{2\lambda}, \\ \{s\}, & |s| > \sqrt{2\lambda}, \\ \{0, s\}, & |s| = \sqrt{2\lambda}. \end{cases} \quad (6.5)$$

We can write the above as $\text{prox}_J(s) = \mathcal{H}_{\sqrt{2\lambda}}(s)$, where \mathcal{H}_α is the so-called *hard thresholding* operator defined by

$$\mathcal{H}_\alpha(s) = \begin{cases} \{0\}, & |s| < \alpha, \\ \{s\}, & |s| > \alpha, \\ \{0, s\}, & |s| = \alpha. \end{cases}$$

The operators prox_J and prox_I are the same since for any $s \in \mathbb{R}$,

$$\begin{aligned} \text{prox}_I(s) &= \operatorname{argmin}_t \left\{ I(t) + \frac{1}{2}(t - s)^2 \right\} \\ &= \operatorname{argmin}_t \left\{ J(t) + \lambda + \frac{1}{2}(t - s)^2 \right\} \\ &= \operatorname{argmin}_t \left\{ J(t) + \frac{1}{2}(t - s)^2 \right\} \\ &= \text{prox}_J(s). \end{aligned}$$

Thus, invoking Theorem 6.6, it follows that²⁷

$$\text{prox}_g(\mathbf{x}) = \mathcal{H}_{\sqrt{2\lambda}}(x_1) \times \mathcal{H}_{\sqrt{2\lambda}}(x_2) \times \cdots \times \mathcal{H}_{\sqrt{2\lambda}}(x_n).$$

■

Theorem 6.11 (scaling and translation). *Let $g : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper function. Let $\lambda \neq 0$ and $\mathbf{a} \in \mathbb{E}$. Define $f(\mathbf{x}) = g(\lambda\mathbf{x} + \mathbf{a})$. Then*

$$\text{prox}_f(\mathbf{x}) = \frac{1}{\lambda} [\text{prox}_{\lambda^2 g}(\lambda\mathbf{x} + \mathbf{a}) - \mathbf{a}]. \quad (6.6)$$

Proof. By definition of the prox,

$$\begin{aligned} \text{prox}_f(\mathbf{x}) &= \operatorname{argmin}_{\mathbf{u}} \left\{ f(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \right\} \\ &= \operatorname{argmin}_{\mathbf{u}} \left\{ g(\lambda\mathbf{u} + \mathbf{a}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \right\}. \end{aligned} \quad (6.7)$$

Making the change of variables

$$\mathbf{z} = \lambda\mathbf{u} + \mathbf{a}, \quad (6.8)$$

the objective function in the minimization problem (6.7) becomes

$$g(\mathbf{z}) + \frac{1}{2} \left\| \frac{1}{\lambda} (\mathbf{z} - \mathbf{a}) - \mathbf{x} \right\|^2 = \frac{1}{\lambda^2} \left[\lambda^2 g(\mathbf{z}) + \frac{1}{2} \|\mathbf{z} - (\lambda\mathbf{x} + \mathbf{a})\|^2 \right]. \quad (6.9)$$

The minimizer of (6.9) is $\mathbf{z} = \text{prox}_{\lambda^2 g}(\lambda\mathbf{x} + \mathbf{a})$, and hence by (6.8), it follows that (6.6) holds. □

Theorem 6.12 (prox of $\lambda g(\cdot/\lambda)$). *Let $g : \mathbb{E} \rightarrow (-\infty, \infty]$ be proper, and let $\lambda \neq 0$. Define $f(\mathbf{x}) = \lambda g(\mathbf{x}/\lambda)$. Then*

$$\text{prox}_f(\mathbf{x}) = \lambda \text{prox}_{g/\lambda}(\mathbf{x}/\lambda).$$

Proof. Note that

$$\text{prox}_f(\mathbf{x}) = \operatorname{argmin}_{\mathbf{u}} \left\{ f(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \right\} = \operatorname{argmin}_{\mathbf{u}} \left\{ \lambda g\left(\frac{\mathbf{u}}{\lambda}\right) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \right\}.$$

²⁷Actually, $\text{prox}_g(\mathbf{x})$ should be a subset of \mathbb{R}^n , meaning the space of n -length *column* vectors, but here we practice some abuse of notation and represent $\text{prox}_g(\mathbf{x})$ as a set of n -length *row* vectors.

Making the change of variables $\mathbf{z} = \frac{\mathbf{u}}{\lambda}$, we can continue to write

$$\begin{aligned}\text{prox}_f(\mathbf{x}) &= \lambda \operatorname{argmin}_{\mathbf{z}} \left\{ \lambda g(\mathbf{z}) + \frac{1}{2} \|\lambda \mathbf{z} - \mathbf{x}\|^2 \right\} \\ &= \lambda \operatorname{argmin}_{\mathbf{z}} \left\{ \lambda^2 \left[\frac{g(\mathbf{z})}{\lambda} + \frac{1}{2} \left\| \mathbf{z} - \frac{\mathbf{x}}{\lambda} \right\|^2 \right] \right\} \\ &= \lambda \operatorname{argmin}_{\mathbf{z}} \left\{ \frac{g(\mathbf{z})}{\lambda} + \frac{1}{2} \left\| \mathbf{z} - \frac{\mathbf{x}}{\lambda} \right\|^2 \right\} \\ &= \lambda \text{prox}_{g/\lambda}(\mathbf{x}/\lambda). \quad \square\end{aligned}$$

Theorem 6.13 (quadratic perturbation). Let $g : \mathbb{E} \rightarrow (-\infty, \infty]$ be proper, and let $f(\mathbf{x}) = g(\mathbf{x}) + \frac{c}{2} \|\mathbf{x}\|^2 + \langle \mathbf{a}, \mathbf{x} \rangle + \gamma$, where $c > 0$, $\mathbf{a} \in \mathbb{E}$, and $\gamma \in \mathbb{R}$. Then

$$\text{prox}_f(\mathbf{x}) = \text{prox}_{\frac{1}{c+1}g} \left(\frac{\mathbf{x} - \mathbf{a}}{c+1} \right).$$

Proof. Follows by the following simple computation:

$$\begin{aligned}\text{prox}_f(\mathbf{x}) &= \operatorname{argmin}_{\mathbf{u}} \left\{ f(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \right\} \\ &= \operatorname{argmin}_{\mathbf{u}} \left\{ g(\mathbf{u}) + \frac{c}{2} \|\mathbf{u}\|^2 + \langle \mathbf{a}, \mathbf{u} \rangle + \gamma + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \right\} \\ &= \operatorname{argmin}_{\mathbf{u}} \left\{ g(\mathbf{u}) + \frac{c+1}{2} \left\| \mathbf{u} - \left(\frac{\mathbf{x} - \mathbf{a}}{c+1} \right) \right\|^2 \right\} \\ &= \text{prox}_{\frac{1}{c+1}g} \left(\frac{\mathbf{x} - \mathbf{a}}{c+1} \right). \quad \square\end{aligned}$$

Example 6.14. Consider the function $f : \mathbb{R} \rightarrow (-\infty, \infty]$ given for any $x \in \mathbb{R}$ by

$$f(x) = \begin{cases} \mu x, & 0 \leq x \leq \alpha, \\ \infty & \text{else,} \end{cases}$$

where $\mu \in \mathbb{R}$ and $\alpha \in [0, \infty]$. To compute the prox of f , note first that f can be represented as

$$f(x) = \delta_{[0,\alpha] \cap \mathbb{R}}(x) + \mu x.$$

By Lemma 6.5 (computation of prox_{g_5}), $\text{prox}_{\delta_{[0,\alpha] \cap \mathbb{R}}}(x) = \min\{\max\{x, 0\}, \alpha\}$. Therefore, using Theorem 6.13 with $c = 0$, $\mathbf{a} = \mu$, $\gamma = 0$, we obtain that for any $x \in \mathbb{R}$,

$$\text{prox}_f(x) = \text{prox}_g(x - \mu) = \min\{\max\{x - \mu, 0\}, \alpha\}.$$



Unfortunately, there is no useful calculus rule for computing the prox mapping of a composition of a function with a general affine mapping. However, if the associated linear transformation satisfies a certain orthogonality condition, such a rule exists.

Theorem 6.15 (composition with an affine mapping). *Let $g : \mathbb{R}^m \rightarrow (-\infty, \infty]$ be a proper closed convex function, and let $f(\mathbf{x}) = g(\mathcal{A}(\mathbf{x}) + \mathbf{b})$, where $\mathbf{b} \in \mathbb{R}^m$ and $\mathcal{A} : \mathbb{V} \rightarrow \mathbb{R}^m$ is a linear transformation satisfying²⁸ $\mathcal{A} \circ \mathcal{A}^T = \alpha \mathcal{I}$ for some constant $\alpha > 0$. Then for any $\mathbf{x} \in \mathbb{V}$,*

$$\text{prox}_f(\mathbf{x}) = \mathbf{x} + \frac{1}{\alpha} \mathcal{A}^T (\text{prox}_{\alpha g}(\mathcal{A}(\mathbf{x}) + \mathbf{b}) - \mathcal{A}(\mathbf{x}) - \mathbf{b}).$$

Proof. By definition, $\text{prox}_f(\mathbf{x})$ is the optimal solution of

$$\min_{\mathbf{u} \in \mathbb{V}} \left\{ f(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \right\},$$

which can be rewritten as

$$\min_{\mathbf{u} \in \mathbb{V}} \left\{ g(\mathcal{A}(\mathbf{u}) + \mathbf{b}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \right\}.$$

The above problem can be formulated as the following constrained problem:

$$\begin{aligned} \min_{\mathbf{u} \in \mathbb{V}, \mathbf{z} \in \mathbb{R}^m} \quad & g(\mathbf{z}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \\ \text{s.t.} \quad & \mathbf{z} = \mathcal{A}(\mathbf{u}) + \mathbf{b}. \end{aligned} \tag{6.10}$$

Denote the optimal solution of (6.10) by $(\tilde{\mathbf{z}}, \tilde{\mathbf{u}})$ (the existence and uniqueness of $\tilde{\mathbf{z}}$ and $\tilde{\mathbf{u}}$ follow by the underlying assumption that g is proper closed and convex). Note that $\tilde{\mathbf{u}} = \text{prox}_f(\mathbf{x})$. Fixing $\mathbf{z} = \tilde{\mathbf{z}}$, we obtain that $\tilde{\mathbf{u}}$ is the optimal solution of

$$\begin{aligned} \min_{\mathbf{u} \in \mathbb{V}} \quad & \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \\ \text{s.t.} \quad & \mathcal{A}(\mathbf{u}) = \tilde{\mathbf{z}} - \mathbf{b}. \end{aligned} \tag{6.11}$$

Since strong duality holds for problem (6.11) (see Theorem A.1), by Theorem A.2, it follows that there exists $\mathbf{y} \in \mathbb{R}^m$ for which

$$\tilde{\mathbf{u}} \in \operatorname{argmin}_{\mathbf{u} \in \mathbb{V}} \left\{ \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 + \langle \mathbf{y}, \mathcal{A}(\mathbf{u}) - \tilde{\mathbf{z}} + \mathbf{b} \rangle \right\} \tag{6.12}$$

$$\mathcal{A}(\tilde{\mathbf{u}}) = \tilde{\mathbf{z}} - \mathbf{b}. \tag{6.13}$$

By (6.12),

$$\tilde{\mathbf{u}} = \mathbf{x} - \mathcal{A}^T(\mathbf{y}). \tag{6.14}$$

²⁸The identity transformation \mathcal{I} was defined in Section 1.10.

Substituting this expression of $\tilde{\mathbf{u}}$ into (6.13), we obtain

$$\mathcal{A}(\mathbf{x} - \mathcal{A}^T(\mathbf{y})) = \tilde{\mathbf{z}} - \mathbf{b},$$

and hence, using the assumption that $\mathcal{A} \circ \mathcal{A}^T = \alpha\mathcal{I}$,

$$\alpha\mathbf{y} = \mathcal{A}(\mathbf{x}) + \mathbf{b} - \tilde{\mathbf{z}},$$

which, combined with (6.14), yields an explicit expression for $\tilde{\mathbf{u}} = \text{prox}_f(\mathbf{x})$ in terms of $\tilde{\mathbf{z}}$:

$$\text{prox}_f(\mathbf{x}) = \tilde{\mathbf{u}} = \mathbf{x} + \frac{1}{\alpha}\mathcal{A}^T(\tilde{\mathbf{z}} - \mathcal{A}(\mathbf{x}) - \mathbf{b}). \quad (6.15)$$

Substituting $\mathbf{u} = \tilde{\mathbf{u}}$ in the minimization problem (6.10), we obtain that $\tilde{\mathbf{z}}$ is given by

$$\begin{aligned} \tilde{\mathbf{z}} &= \operatorname{argmin}_{\mathbf{z} \in \mathbb{R}^m} \left\{ g(\mathbf{z}) + \frac{1}{2} \left\| \mathbf{x} + \frac{1}{\alpha}\mathcal{A}^T(\mathbf{z} - \mathcal{A}(\mathbf{x}) - \mathbf{b}) - \mathbf{x} \right\|^2 \right\} \\ &= \operatorname{argmin}_{\mathbf{z} \in \mathbb{R}^m} \left\{ g(\mathbf{z}) + \frac{1}{2\alpha^2} \|\mathcal{A}^T(\mathbf{z} - \mathcal{A}(\mathbf{x}) - \mathbf{b})\|^2 \right\} \\ &\stackrel{(*)}{=} \operatorname{argmin}_{\mathbf{z} \in \mathbb{R}^m} \left\{ \alpha g(\mathbf{z}) + \frac{1}{2} \|\mathbf{z} - \mathcal{A}(\mathbf{x}) - \mathbf{b}\|^2 \right\} \\ &= \text{prox}_{\alpha g}(\mathcal{A}(\mathbf{x}) + \mathbf{b}), \end{aligned}$$

where the equality $(*)$ uses the assumption that $\mathcal{A} \circ \mathcal{A}^T = \alpha\mathcal{I}$. Plugging the expression for $\tilde{\mathbf{z}}$ into (6.15) produces the desired result. \square

Example 6.16. Let $g : \mathbb{E} \rightarrow (-\infty, \infty]$ be proper closed and convex where $\mathbb{E} = \mathbb{R}^d$, and let $f : \mathbb{E}^m \rightarrow (-\infty, \infty]$ be defined as

$$f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m) = g(\mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_m).$$

The above can be written as $f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m) = g(\mathcal{A}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m))$, where $\mathcal{A} : \mathbb{E}^m \rightarrow \mathbb{E}$ is the linear transformation

$$\mathcal{A}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m) = \mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_m.$$

Obviously, the adjoint operator $\mathcal{A}^T : \mathbb{E} \rightarrow \mathbb{E}^m$ is given by

$$\mathcal{A}^T(\mathbf{x}) = (\mathbf{x}, \mathbf{x}, \dots, \mathbf{x}),$$

and for any $\mathbf{x} \in \mathbb{E}$,

$$\mathcal{A}(\mathcal{A}^T(\mathbf{x})) = m\mathbf{x}.$$

Thus, the conditions of Theorem 6.15 are satisfied with $\alpha = m$ and $\mathbf{b} = 0$, and consequently, for any $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m) \in \mathbb{E}^m$,

$$\text{prox}_f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m)_j = \mathbf{x}_j + \frac{1}{m} \left(\text{prox}_{mg} \left(\sum_{i=1}^m \mathbf{x}_i \right) - \sum_{i=1}^m \mathbf{x}_i \right), \quad j = 1, 2, \dots, m.$$

Example 6.17. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}) = |\mathbf{a}^T \mathbf{x}|$, where $\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$. We can write f as $f(\mathbf{x}) = g(\mathbf{a}^T \mathbf{x})$, where $g(t) = |t|$. By Lemma 6.5 (prox_{g_2} computation), $\text{prox}_{\lambda g} = \mathcal{T}_\lambda$, with $\mathcal{T}_\lambda(x) = [|x| - \lambda]_+ \text{sgn}(x)$ being the soft thresholding operator defined in Example 6.8. Invoking Theorem 6.15 with $\alpha = \|\mathbf{a}\|^2$, $\mathbf{b} = \mathbf{0}$, and \mathcal{A} defined as the transformation $\mathbf{x} \mapsto \mathbf{a}^T \mathbf{x}$, we obtain that

$$\text{prox}_f(\mathbf{x}) = \mathbf{x} + \frac{1}{\|\mathbf{a}\|^2} (\mathcal{T}_{\|\mathbf{a}\|^2}(\mathbf{a}^T \mathbf{x}) - \mathbf{a}^T \mathbf{x}) \mathbf{a}.$$

Theorem 6.18 (norm composition). Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}) = g(\|\mathbf{x}\|)$, where $g : \mathbb{R} \rightarrow (-\infty, \infty]$ is a proper closed and convex function satisfying $\text{dom}(g) \subseteq [0, \infty)$. Then

$$\text{prox}_f(\mathbf{x}) = \begin{cases} \text{prox}_g(\|\mathbf{x}\|) \frac{\mathbf{x}}{\|\mathbf{x}\|}, & \mathbf{x} \neq \mathbf{0}, \\ \{\mathbf{u} \in \mathbb{E} : \|\mathbf{u}\| = \text{prox}_g(0)\}, & \mathbf{x} = \mathbf{0}. \end{cases} \quad (6.16)$$

Proof. By definition, $\text{prox}_f(\mathbf{0})$ is the set of minimizers of the problem

$$\min_{\mathbf{u} \in \mathbb{E}} \left\{ f(\mathbf{u}) + \frac{1}{2} \|\mathbf{u}\|^2 \right\} = \min_{\mathbf{u} \in \mathbb{E}} \left\{ g(\|\mathbf{u}\|) + \frac{1}{2} \|\mathbf{u}\|^2 \right\}.$$

Making the change of variables $w = \|\mathbf{u}\|$, the problem reduces to (recalling that $\text{dom}(g) \subseteq [0, \infty)$)

$$\min_{w \in \mathbb{R}} \left\{ g(w) + \frac{1}{2} w^2 \right\}.$$

The optimal set of the above problem is $\text{prox}_g(0)$, and hence $\text{prox}_f(\mathbf{0})$ is the set of vectors \mathbf{u} satisfying $\|\mathbf{u}\| = \text{prox}_g(0)$. We will now compute $\text{prox}_f(\mathbf{x})$ for $\mathbf{x} \neq \mathbf{0}$. The optimization problem associated with the prox computation can be rewritten as the following double minimization problem:

$$\begin{aligned} \min_{\mathbf{u} \in \mathbb{E}} \left\{ g(\|\mathbf{u}\|) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \right\} &= \min_{\mathbf{u} \in \mathbb{E}} \left\{ g(\|\mathbf{u}\|) + \frac{1}{2} \|\mathbf{u}\|^2 - \langle \mathbf{u}, \mathbf{x} \rangle + \frac{1}{2} \|\mathbf{x}\|^2 \right\} \\ &= \min_{\alpha \in \mathbb{R}_+} \min_{\mathbf{u} \in \mathbb{E} : \|\mathbf{u}\| = \alpha} \left\{ g(\alpha) + \frac{1}{2} \alpha^2 - \langle \mathbf{u}, \mathbf{x} \rangle + \frac{1}{2} \|\mathbf{x}\|^2 \right\}. \end{aligned}$$

Using the Cauchy–Schwarz inequality, it is easy to see that the minimizer of the inner minimization problem is

$$\mathbf{u} = \alpha \frac{\mathbf{x}}{\|\mathbf{x}\|}, \quad (6.17)$$

and the corresponding optimal value is

$$g(\alpha) + \frac{1}{2}\alpha^2 - \alpha\|\mathbf{x}\| + \frac{1}{2}\|\mathbf{x}\|^2 = g(\alpha) + \frac{1}{2}(\alpha - \|\mathbf{x}\|)^2.$$

Therefore, $\text{prox}_f(\mathbf{x})$ is given by \mathbf{u} in (6.17) with α given by

$$\begin{aligned} \alpha &= \operatorname{argmin}_{\alpha \in \mathbb{R}_+} \left\{ g(\alpha) + \frac{1}{2}(\alpha - \|\mathbf{x}\|)^2 \right\} \\ &= \operatorname{argmin}_{\alpha \in \mathbb{R}} \left\{ g(\alpha) + \frac{1}{2}(\alpha - \|\mathbf{x}\|)^2 \right\} \\ &= \text{prox}_g(\|\mathbf{x}\|), \end{aligned}$$

where the second equality is due to the assumption that $\text{dom}(g) \subseteq [0, \infty)$. Thus, $\text{prox}_f(\mathbf{x}) = \text{prox}_g(\|\mathbf{x}\|) \frac{\mathbf{x}}{\|\mathbf{x}\|}$. \square

Example 6.19 (prox of Euclidean norm). Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}) = \lambda\|\mathbf{x}\|$, where $\lambda > 0$ and $\|\cdot\|$ is the underlying Euclidean norm (recall that in this section we assume that the underlying space is Euclidean). Then $f(\mathbf{x}) = g(\|\mathbf{x}\|)$, where

$$g(t) = \begin{cases} \lambda t, & t \geq 0, \\ \infty, & t < 0. \end{cases}$$

Then by Theorem 6.18, for any $\mathbf{x} \in \mathbb{E}$,

$$\text{prox}_f(\mathbf{x}) = \begin{cases} \text{prox}_g(\|\mathbf{x}\|) \frac{\mathbf{x}}{\|\mathbf{x}\|}, & \mathbf{x} \neq \mathbf{0}, \\ \{\mathbf{u} \in \mathbb{E} : \|\mathbf{u}\| = \text{prox}_g(0)\}, & \mathbf{x} = \mathbf{0}. \end{cases}$$

By Lemma 6.5 (computation of prox_{g_1}), $\text{prox}_g(t) = [t - \lambda]_+$. Thus, $\text{prox}_g(0) = 0$ and $\text{prox}_g(\|\mathbf{x}\|) = [\|\mathbf{x}\| - \lambda]_+$, and therefore

$$\text{prox}_f(\mathbf{x}) = \begin{cases} [\|\mathbf{x}\| - \lambda]_+ \frac{\mathbf{x}}{\|\mathbf{x}\|}, & \mathbf{x} \neq \mathbf{0}, \\ \mathbf{0}, & \mathbf{x} = \mathbf{0}. \end{cases}$$

Finally, we can write the above formula in the following compact form:

$$\text{prox}_{\lambda\|\cdot\|}(\mathbf{x}) = \left(1 - \frac{\lambda}{\max\{\|\mathbf{x}\|, \lambda\}}\right) \mathbf{x}.$$

Example 6.20 (prox of cubic Euclidean norm). Let $f(\mathbf{x}) = \lambda \|\mathbf{x}\|^3$, where $\lambda > 0$. Then $f(\mathbf{x}) = \lambda g(\|\mathbf{x}\|)$, where

$$g(t) = \begin{cases} t^3, & t \geq 0, \\ \infty, & t < 0. \end{cases}$$

Then by Theorem 6.18, for any $\mathbf{x} \in \mathbb{R}$,

$$\text{prox}_f(\mathbf{x}) = \begin{cases} \text{prox}_g(\|\mathbf{x}\|) \frac{\mathbf{x}}{\|\mathbf{x}\|}, & \mathbf{x} \neq \mathbf{0}, \\ \{\mathbf{u} \in \mathbb{E} : \|\mathbf{u}\| = \text{prox}_g(0)\}, & \mathbf{x} = \mathbf{0}. \end{cases}$$

By Lemma 6.5 (computation of prox_{g_3}), $\text{prox}_g(t) = \frac{-1 + \sqrt{1 + 12\lambda[t]_+}}{6\lambda}$. Therefore, $\text{prox}_g(0) = 0$ and

$$\text{prox}_f(\mathbf{x}) = \begin{cases} \frac{-1 + \sqrt{1 + 12\lambda\|\mathbf{x}\|}}{6\lambda} \frac{\mathbf{x}}{\|\mathbf{x}\|}, & \mathbf{x} \neq \mathbf{0}, \\ \mathbf{0}, & \mathbf{x} = \mathbf{0}, \end{cases}$$

and thus

$$\text{prox}_{\lambda\|\cdot\|^3}(\mathbf{x}) = \frac{2}{1 + \sqrt{1 + 12\lambda\|\mathbf{x}\|}} \mathbf{x}.$$

■

Example 6.21 (prox of negative Euclidean norm). Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}) = -\lambda \|\mathbf{x}\|$, where $\lambda > 0$. Since f is not convex, we do not expect the prox to be a single-valued mapping. However, since f is closed, and since the function $\mathbf{u} \mapsto f(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2$ is coercive for any $\mathbf{x} \in \mathbb{E}$, it follows by Theorem 6.4 that the set $\text{prox}_f(\mathbf{x})$ is always nonempty. To compute the prox, note that $f(\mathbf{x}) = g(\|\mathbf{x}\|)$, where

$$g(t) = \begin{cases} -\lambda t, & t \geq 0, \\ \infty, & t < 0. \end{cases}$$

By Theorem 6.18, for any $\mathbf{x} \in \mathbb{R}$,

$$\text{prox}_f(\mathbf{x}) = \begin{cases} \text{prox}_g(\|\mathbf{x}\|) \frac{\mathbf{x}}{\|\mathbf{x}\|}, & \mathbf{x} \neq \mathbf{0}, \\ \{\mathbf{u} \in \mathbb{E} : \|\mathbf{u}\| = \text{prox}_g(0)\}, & \mathbf{x} = \mathbf{0}. \end{cases}$$

By Lemma 6.5 (computation of prox_{g_1}), $\text{prox}_g(t) = [t + \lambda]_+$. Therefore, $\text{prox}_g(0) = \lambda$ and

$$\text{prox}_{-\lambda\|\cdot\|}(\mathbf{x}) = \begin{cases} \left(1 + \frac{\lambda}{\|\mathbf{x}\|}\right) \mathbf{x}, & \mathbf{x} \neq \mathbf{0}, \\ \{\mathbf{u} : \|\mathbf{u}\| = \lambda\}, & \mathbf{x} = \mathbf{0}. \end{cases}$$

■

Example 6.22 (prox of absolute value over symmetric intervals). Consider the function $f : \mathbb{R} \rightarrow (-\infty, \infty]$ given by

$$f(x) = \begin{cases} \lambda|x|, & |x| \leq \alpha, \\ \infty & \text{else,} \end{cases}$$

where $\lambda \in [0, \infty)$ and $\alpha \in [0, \infty]$. Then $f(x) = g(|x|)$, where

$$g(x) = \begin{cases} \lambda x, & 0 \leq x \leq \alpha, \\ \infty & \text{else.} \end{cases}$$

Thus, by Theorem 6.18, for any x ,

$$\text{prox}_f(\mathbf{x}) = \begin{cases} \text{prox}_g(|x|) \frac{x}{|x|}, & x \neq 0, \\ \{u \in \mathbb{R} : |u| = \text{prox}_g(0)\}, & x = 0. \end{cases} \quad (6.18)$$

By Example 6.14, $\text{prox}_g(x) = \min\{\max\{x - \lambda, 0\}, \alpha\}$, which, combined with (6.18) and the fact that $\frac{x}{|x|} = \text{sgn}(x)$ for any $x \neq 0$, yields the formula

$$\text{prox}_{\lambda|\cdot| + \delta_{[-\alpha, \alpha]}}(x) = \min\{\max\{|x| - \lambda, 0\}, \alpha\} \text{sgn}(x).$$

■

Using the previous example, we can compute the prox of weighted l_1 -norms over boxes.

Example 6.23 (prox of weighted l_1 over a box). Consider the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ given by

$$f(\mathbf{x}) = \begin{cases} \sum_{i=1}^n \omega_i |x_i|, & -\boldsymbol{\alpha} \leq \mathbf{x} \leq \boldsymbol{\alpha}, \\ \infty, & \text{else,} \end{cases}$$

for any $\mathbf{x} \in \mathbb{R}^n$, where $\boldsymbol{\omega} \in \mathbb{R}_+^n$ and $\boldsymbol{\alpha} \in [0, \infty]^n$. Then $f = \sum_{i=1}^n f_i$, where

$$f_i(x) = \begin{cases} w_i |x|, & -\alpha_i \leq x \leq \alpha_i, \\ \infty, & \text{else.} \end{cases}$$

Using Example 6.22 and invoking Theorem 6.6, we finally obtain that

$$\text{prox}_f(\mathbf{x}) = (\min\{\max\{|x_i| - \omega_i, 0\}, \alpha_i\} \text{sgn}(x_i))_{i=1}^n.$$

■

The table below summarizes the main prox calculus rules discussed in this section.

$f(\mathbf{x})$	$\text{prox}_f(\mathbf{x})$	Assumptions	Reference
$\sum_{i=1}^m f_i(\mathbf{x}_i)$	$\text{prox}_{f_1}(\mathbf{x}_1) \times \cdots \times \text{prox}_{f_m}(\mathbf{x}_m)$		Theorem 6.6
$g(\lambda \mathbf{x} + \mathbf{a})$	$\frac{1}{\lambda} [\text{prox}_{\lambda^2 g}(\lambda \mathbf{x} + \mathbf{a}) - \mathbf{a}]$	$\lambda \neq 0, \mathbf{a} \in \mathbb{E}, g$ proper	Theorem 6.11
$\lambda g(\mathbf{x}/\lambda)$	$\lambda \text{prox}_{g/\lambda}(\mathbf{x}/\lambda)$	$\lambda \neq 0, g$ proper	Theorem 6.12
$g(\mathbf{x}) + \frac{c}{2} \ \mathbf{x}\ ^2 + \langle \mathbf{a}, \mathbf{x} \rangle + \gamma$	$\text{prox}_{\frac{1}{c+1} g}(\frac{\mathbf{x}-\mathbf{a}}{c+1})$	$\mathbf{a} \in \mathbb{E}, c > 0,$ $\gamma \in \mathbb{R}, g$ proper	Theorem 6.13
$g(\mathcal{A}(\mathbf{x}) + \mathbf{b})$	$\mathbf{x} + \frac{1}{\alpha} \mathcal{A}^T (\text{prox}_{\alpha g}(\mathcal{A}(\mathbf{x}) + \mathbf{b}) - \mathcal{A}(\mathbf{x}) - \mathbf{b})$	$\mathbf{b} \in \mathbb{R}^m,$ $\mathcal{A} : \mathbb{V} \rightarrow \mathbb{R}^m,$ g proper closed convex, $\mathcal{A} \circ \mathcal{A}^T = \alpha I,$ $\alpha > 0$	Theorem 6.15
$g(\ \mathbf{x}\)$	$\text{prox}_g(\ \mathbf{x}\) \frac{\mathbf{x}}{\ \mathbf{x}\ }, \quad \mathbf{x} \neq \mathbf{0}$ $\{\mathbf{u} : \ \mathbf{u}\ = \text{prox}_g(0)\}, \quad \mathbf{x} = \mathbf{0}$	g proper closed convex, $\text{dom}(g) \subseteq [0, \infty)$	Theorem 6.18

6.4 Prox of Indicators—Orthogonal Projections

6.4.1 The First Projection Theorem

Let $g : \mathbb{E} \rightarrow (-\infty, \infty]$ be given by $g(\mathbf{x}) = \delta_C(\mathbf{x})$, where C is a nonempty set. Then

$$\text{prox}_g(\mathbf{x}) = \operatorname{argmin}_{\mathbf{u} \in \mathbb{E}} \left\{ \delta_C(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \right\} = \operatorname{argmin}_{\mathbf{u} \in C} \|\mathbf{u} - \mathbf{x}\|^2 = P_C(\mathbf{x}).$$

Thus, the proximal mapping of the indicator function of a given set is the orthogonal projection²⁹ operator onto the same set.

Theorem 6.24. *Let $C \subseteq \mathbb{E}$ be nonempty. Then $\text{prox}_{\delta_C}(\mathbf{x}) = P_C(\mathbf{x})$ for any $\mathbf{x} \in \mathbb{E}$.*

If C is closed and convex, in addition to being nonempty, the indicator function δ_C is proper closed and convex, and hence by the first prox theorem (Theorem 6.3), the orthogonal projection mapping (which coincides with the proximal mapping) exists and is unique. This is the first projection theorem.

²⁹The orthogonal projection operator was introduced in Example 3.31.

Theorem 6.25 (first projection theorem). *Let $C \subseteq \mathbb{E}$ be a nonempty closed convex set. Then $P_C(\mathbf{x})$ is a singleton for any $\mathbf{x} \in \mathbb{E}$.*

6.4.2 First Examples in \mathbb{R}^n

We begin by recalling³⁰ several known expressions for the orthogonal projection onto some basic subsets of \mathbb{R}^n . Since the assumption made throughout the book is that (unless otherwise stated) \mathbb{R}^n is endowed with the dot product, and since the standing assumption in this chapter is that the underlying space is Euclidean, it follows that the endowed norm is the l_2 -norm.

Lemma 6.26 (projection onto subsets of \mathbb{R}^n). *Following are pairs of nonempty closed and convex sets and their corresponding orthogonal projections:*

nonnegative orthant	$C_1 = \mathbb{R}_+^n$,	$[\mathbf{x}]_+$,
box	$C_2 = \text{Box}[\boldsymbol{\ell}, \mathbf{u}]$,	$(\min\{\max\{x_i, \ell_i\}, u_i\})_{i=1}^n$,
affine set	$C_3 = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} = \mathbf{b}\}$,	$\mathbf{x} - \mathbf{A}^T(\mathbf{AA}^T)^{-1}(\mathbf{Ax} - \mathbf{b})$,
l_2 ball	$C_4 = B_{\ \cdot\ _2}[\mathbf{c}, r]$,	$\mathbf{c} + \frac{r}{\max\{\ \mathbf{x}-\mathbf{c}\ _2, r\}}(\mathbf{x} - \mathbf{c})$,
half-space	$C_5 = \{\mathbf{x} : \mathbf{a}^T \mathbf{x} \leq \alpha\}$,	$\mathbf{x} - \frac{[\mathbf{a}^T \mathbf{x} - \alpha]_+}{\ \mathbf{a}\ ^2} \mathbf{a}$,

where $\boldsymbol{\ell} \in [-\infty, \infty)^n$, $\mathbf{u} \in (-\infty, \infty]^n$ are such that $\boldsymbol{\ell} \leq \mathbf{u}$, $\mathbf{A} \in \mathbb{R}^{m \times n}$ has full row rank, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{c} \in \mathbb{R}^n$, $r > 0$, $\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, and $\alpha \in \mathbb{R}$.

Note that we extended the definition of box sets given in Section 1.7.1 to include unbounded intervals, meaning that $\text{Box}[\boldsymbol{\ell}, \mathbf{u}]$ is also defined when the components of $\boldsymbol{\ell}$ might also take the value $-\infty$, and the components of \mathbf{u} might take the value ∞ . However, boxes are always subsets of \mathbb{R}^n , and the formula

$$\text{Box}[\boldsymbol{\ell}, \mathbf{u}] = \{\mathbf{x} \in \mathbb{R}^n : \boldsymbol{\ell} \leq \mathbf{x} \leq \mathbf{u}\}$$

still holds. For example, $\text{Box}[\mathbf{0}, \infty \mathbf{e}] = \mathbb{R}_+^n$.

6.4.3 Projection onto the Intersection of a Hyperplane and a Box

The next result develops an expression for the orthogonal projection onto another subset of \mathbb{R}^n —the intersection of an hyperplane and a box.

Theorem 6.27 (projection onto the intersection of a hyperplane and a box). *Let $C \subseteq \mathbb{R}^n$ be given by*

$$C = H_{\mathbf{a}, b} \cap \text{Box}[\boldsymbol{\ell}, \mathbf{u}] = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T \mathbf{x} = b, \boldsymbol{\ell} \leq \mathbf{x} \leq \mathbf{u}\},$$

where $\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, $b \in \mathbb{R}$, $\boldsymbol{\ell} \in [-\infty, \infty)^n$, $\mathbf{u} \in (-\infty, \infty]^n$. Assume that $C \neq \emptyset$. Then

$$P_C(\mathbf{x}) = P_{\text{Box}[\boldsymbol{\ell}, \mathbf{u}]}(\mathbf{x} - \mu^* \mathbf{a}),$$

³⁰The derivations of the orthogonal projection expressions in Lemma 6.26 can be found, for example, in [10].

where $\text{Box}[\ell, \mathbf{u}] = \{\mathbf{y} \in \mathbb{R}^n : \ell_i \leq y_i \leq u_i, i = 1, 2, \dots, n\}$ and μ^* is a solution of the equation

$$\mathbf{a}^T P_{\text{Box}[\ell, \mathbf{u}]}(\mathbf{x} - \mu \mathbf{a}) = b. \quad (6.19)$$

Proof. The orthogonal projection of \mathbf{x} onto C is the unique optimal solution of

$$\min_{\mathbf{y}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 : \mathbf{a}^T \mathbf{y} = b, \ell \leq \mathbf{y} \leq \mathbf{u} \right\}. \quad (6.20)$$

A Lagrangian of the problem is

$$L(\mathbf{y}; \mu) = \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \mu(\mathbf{a}^T \mathbf{y} - b) = \frac{1}{2} \|\mathbf{y} - (\mathbf{x} - \mu \mathbf{a})\|_2^2 - \frac{\mu^2}{2} \|\mathbf{a}\|_2^2 + \mu(\mathbf{a}^T \mathbf{x} - b). \quad (6.21)$$

Since strong duality holds for problem (6.20) (see Theorem A.1), it follows by Theorem A.2 that \mathbf{y}^* is an optimal solution of problem (6.20) if and only if there exists $\mu^* \in \mathbb{R}$ (which will actually be an optimal solution of the dual problem) for which

$$\mathbf{y}^* \in \operatorname{argmin}_{\ell \leq \mathbf{y} \leq \mathbf{u}} L(\mathbf{y}; \mu^*), \quad (6.22)$$

$$\mathbf{a}^T \mathbf{y}^* = b. \quad (6.23)$$

Using the expression of the Lagrangian given in (6.21), the relation (6.22) can be equivalently written as

$$\mathbf{y}^* = P_{\text{Box}[\ell, \mathbf{u}]}(\mathbf{x} - \mu^* \mathbf{a}).$$

The feasibility condition (6.23) can then be rewritten as

$$\mathbf{a}^T P_{\text{Box}[\ell, \mathbf{u}]}(\mathbf{x} - \mu^* \mathbf{a}) = b. \quad \square$$

Remark 6.28. The projection onto the box $\text{Box}[\ell, \mathbf{u}]$ is extremely simple and is done component-wise as described in Lemma 6.26. Note also that (6.19) actually consists in finding a root of the nonincreasing function $\varphi(\mu) = \mathbf{a}^T P_{\text{Box}}(\mathbf{x} - \mu \mathbf{a}) - b$, which is a task that can be performed efficiently even by simple procedures such as bisection. The fact that φ is nonincreasing follows from the observation that $\varphi(\mu) = \sum_{i=1}^n a_i \min\{\max\{x_i - \mu a_i, \ell_i\}, u_i\} - b$ and the fact that $\mu \mapsto a_i \min\{\max\{x_i - \mu a_i, \ell_i\}, u_i\}$ is a nonincreasing function for any i .

A direct consequence of Theorem 6.27 is an expression for the orthogonal projection onto the unit simplex.

Corollary 6.29 (orthogonal projection onto the unit simplex). For any $\mathbf{x} \in \mathbb{R}^n$,

$$P_{\Delta_n}(\mathbf{x}) = [\mathbf{x} - \mu^* \mathbf{e}]_+,$$

where μ^* is a root of the equation

$$\mathbf{e}^T [\mathbf{x} - \mu^* \mathbf{e}]_+ - 1 = 0.$$

Proof. Invoking Theorem 6.27 with $\mathbf{a} = \mathbf{e}$, $b = 1$, $\ell_i = 0$, $u_i = \infty$, $i = 1, 2, \dots, n$, and noting that in this case $P_{\text{Box}[\ell, \mathbf{u}]}(\mathbf{x}) = [\mathbf{x}]_+$, the result follows. \square

In order to expand the variety of sets on which we will be able to find simple expressions for the orthogonal projection mapping, in the next two subsections, we will discuss how to project onto level sets and epigraphs.

6.4.4 Projection onto Level Sets

Theorem 6.30 (orthogonal projection onto level sets). Let $C = \text{Lev}(f, \alpha) = \{\mathbf{x} \in \mathbb{E} : f(\mathbf{x}) \leq \alpha\}$, where $f : \mathbb{E} \rightarrow (-\infty, \infty]$ is proper closed and convex, and $\alpha \in \mathbb{R}$. Assume that there exists $\hat{\mathbf{x}} \in \mathbb{E}$ for which $f(\hat{\mathbf{x}}) < \alpha$. Then

$$P_C(\mathbf{x}) = \begin{cases} P_{\text{dom}(f)}(\mathbf{x}), & f(P_{\text{dom}(f)}(\mathbf{x})) \leq \alpha, \\ \text{prox}_{\lambda^* f}(\mathbf{x}) & \text{else,} \end{cases} \quad (6.24)$$

where λ^* is any positive root of the equation

$$\varphi(\lambda) \equiv f(\text{prox}_{\lambda f}(\mathbf{x})) - \alpha = 0.$$

In addition, the function φ is nonincreasing.

Proof. The orthogonal projection of \mathbf{x} onto C is an optimal solution of the problem

$$\min_{\mathbf{y} \in \mathbb{E}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|^2 : f(\mathbf{y}) \leq \alpha, \mathbf{y} \in X \right\},$$

where $X = \text{dom}(f)$. A Lagrangian of the problem is ($\lambda \geq 0$)

$$L(\mathbf{y}; \lambda) = \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|^2 + \lambda f(\mathbf{y}) - \alpha \lambda. \quad (6.25)$$

Since the problem is convex and satisfies Slater's condition, strong duality holds (see Theorem A.1), and therefore it follows by the optimality conditions in Theorem A.2 that \mathbf{y}^* is an optimal solution of problem (6.25) if and only if there exists $\lambda^* \in \mathbb{R}_+$ for which

$$\mathbf{y}^* \in \operatorname{argmin}_{\mathbf{y} \in X} L(\mathbf{y}; \lambda^*), \quad (6.26)$$

$$f(\mathbf{y}^*) \leq \alpha, \quad (6.27)$$

$$\lambda^*(f(\mathbf{y}^*) - \alpha) = 0. \quad (6.28)$$

There are two cases. If $P_X(\mathbf{x})$ exists and $f(P_X(\mathbf{x})) \leq \alpha$, then $\mathbf{y}^* = P_X(\mathbf{x})$, and $\lambda^* = 0$ is a solution to the system (6.26), (6.27), (6.28). Otherwise, if $P_X(\mathbf{x})$ does not exist or $f(P_X(\mathbf{x})) > \alpha$, then $\lambda^* > 0$, and in this case the system (6.26), (6.27), (6.28) reduces to $\mathbf{y}^* = \text{prox}_{\lambda^* f}(\mathbf{x})$ and $f(\text{prox}_{\lambda^* f}(\mathbf{x})) = \alpha$, which yields the formula (6.24).

To prove that φ is nonincreasing, recall that

$$\text{prox}_{\lambda f}(\mathbf{x}) = \operatorname{argmin}_{\mathbf{y} \in X} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|^2 + \lambda(f(\mathbf{y}) - \alpha) \right\}.$$

Take $0 \leq \lambda_1 < \lambda_2$. Then denoting $\mathbf{v}_1 = \text{prox}_{\lambda_1 f}(\mathbf{x})$ and $\mathbf{v}_2 = \text{prox}_{\lambda_2 f}(\mathbf{x})$, we have

$$\begin{aligned} & \frac{1}{2} \|\mathbf{v}_2 - \mathbf{x}\|^2 + \lambda_2(f(\mathbf{v}_2) - \alpha) \\ &= \frac{1}{2} \|\mathbf{v}_2 - \mathbf{x}\|^2 + \lambda_1(f(\mathbf{v}_2) - \alpha) + (\lambda_2 - \lambda_1)(f(\mathbf{v}_2) - \alpha) \\ &\geq \frac{1}{2} \|\mathbf{v}_1 - \mathbf{x}\|^2 + \lambda_1(f(\mathbf{v}_1) - \alpha) + (\lambda_2 - \lambda_1)(f(\mathbf{v}_2) - \alpha) \\ &= \frac{1}{2} \|\mathbf{v}_1 - \mathbf{x}\|^2 + \lambda_2(f(\mathbf{v}_1) - \alpha) + (\lambda_2 - \lambda_1)(f(\mathbf{v}_2) - f(\mathbf{v}_1)) \\ &\geq \frac{1}{2} \|\mathbf{v}_2 - \mathbf{x}\|^2 + \lambda_2(f(\mathbf{v}_2) - \alpha) + (\lambda_2 - \lambda_1)(f(\mathbf{v}_2) - f(\mathbf{v}_1)). \end{aligned}$$

Therefore, $(\lambda_2 - \lambda_1)(f(\mathbf{v}_2) - f(\mathbf{v}_1)) \leq 0$. Since $\lambda_1 < \lambda_2$, we can conclude that $f(\mathbf{v}_2) \leq f(\mathbf{v}_1)$. Finally,

$$\varphi(\lambda_2) = f(\mathbf{v}_2) - \alpha \leq f(\mathbf{v}_1) - \alpha = \varphi(\lambda_1),$$

establishing the monotonicity of φ . \square

Remark 6.31. Note that in Theorem 6.30 f is assumed to be closed, but this does not necessarily imply that $\text{dom}(f)$ is closed. In cases where $\text{dom}(f)$ is not closed, it might happen that $P_{\text{dom}(f)}(\mathbf{x})$ does not exist and formula (6.24) amounts to $P_C(\mathbf{x}) = \text{prox}_{\lambda^* f}(\mathbf{x})$.

Example 6.32 (projection onto the intersection of a half-space and a box). Consider the set

$$C = H_{\mathbf{a}, b}^- \cap \text{Box}[\boldsymbol{\ell}, \mathbf{u}] = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T \mathbf{x} \leq b, \boldsymbol{\ell} \leq \mathbf{x} \leq \mathbf{u}\},$$

where $\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, $b \in \mathbb{R}$, $\boldsymbol{\ell} \in [-\infty, \infty)^n$ and $\mathbf{u} \in (-\infty, \infty]^n$. Assume that $C \neq \emptyset$. Then $C = \text{Lev}(f, b)$, where $f(\mathbf{x}) = \mathbf{a}^T \mathbf{x} + \delta_{\text{Box}[\boldsymbol{\ell}, \mathbf{u}]}(\mathbf{x})$. For any $\lambda > 0$,

$$\text{prox}_{\lambda f}(\mathbf{x}) = \text{prox}_{\lambda \mathbf{a}^T (\cdot) + \delta_{\text{Box}[\boldsymbol{\ell}, \mathbf{u}]}(\cdot)}(\mathbf{x}) \stackrel{(*)}{=} \text{prox}_{\delta_{\text{Box}[\boldsymbol{\ell}, \mathbf{u}]}}(\mathbf{x} - \lambda \mathbf{a}) = P_{\text{Box}[\boldsymbol{\ell}, \mathbf{u}]}(\mathbf{x} - \lambda \mathbf{a}),$$

where in the equality $(*)$ we used Theorem 6.13. Invoking Theorem 6.30, we obtain the following formula for the projection on C :

$$P_C(\mathbf{x}) = \begin{cases} P_{\text{Box}[\boldsymbol{\ell}, \mathbf{u}]}(\mathbf{x}), & \mathbf{a}^T P_{\text{Box}[\boldsymbol{\ell}, \mathbf{u}]}(\mathbf{x}) \leq b, \\ P_{\text{Box}[\boldsymbol{\ell}, \mathbf{u}]}(\mathbf{x} - \lambda^* \mathbf{a}), & \mathbf{a}^T P_{\text{Box}[\boldsymbol{\ell}, \mathbf{u}]}(\mathbf{x}) > b, \end{cases}$$

where λ^* is any positive root of the nonincreasing function

$$\varphi(\lambda) = \mathbf{a}^T P_{\text{Box}[\boldsymbol{\ell}, \mathbf{u}]}(\mathbf{x} - \lambda \mathbf{a}) - b.$$

Example 6.33 (projection onto the l_1 ball). Let $C = B_{\|\cdot\|_1}[\mathbf{0}, \alpha] = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_1 \leq \alpha\}$, where $\alpha > 0$. Then $C = \text{Lev}(f, \alpha)$ with $f(\mathbf{x}) = \|\mathbf{x}\|_1$. The prox of $\lambda f = \lambda \|\cdot\|_1$ for any $\lambda > 0$ was computed in Example 6.8, where it was shown that

$$\text{prox}_{\lambda f}(\mathbf{x}) = \mathcal{T}_\lambda(\mathbf{x}) \text{ for all } \mathbf{x} \in \mathbb{R}^n$$

with \mathcal{T}_λ being the soft thresholding operator given by $\mathcal{T}_\lambda(\mathbf{x}) = [\mathbf{x} - \lambda \mathbf{e}]_+ \odot \text{sgn}(\mathbf{x})$. Invoking Theorem 6.30, we obtain that

$$P_{B_{\|\cdot\|_1}[\mathbf{0}, \alpha]}(\mathbf{x}) = \begin{cases} \mathbf{x}, & \|\mathbf{x}\|_1 \leq \alpha, \\ \mathcal{T}_{\lambda^*}(\mathbf{x}), & \|\mathbf{x}\|_1 > \alpha, \end{cases}$$

where λ^* is any positive root of the nonincreasing function

$$\varphi(\lambda) = \|\mathcal{T}_\lambda(\mathbf{x})\|_1 - \alpha.$$

■

The next example uses a generalization of the soft thresholding mapping, which will be called the *two-sided soft thresholding* operator, and is defined for any $\mathbf{a}, \mathbf{b} \in (-\infty, \infty]^n$ as

$$\mathcal{S}_{\mathbf{a}, \mathbf{b}}(\mathbf{x}) = (\min\{\max\{|x_i| - a_i, 0\}, b_i\} \text{sgn}(x_i))_{i=1}^n, \quad \mathbf{x} \in \mathbb{R}^n.$$

Obviously,

$$\mathcal{S}_{\lambda \mathbf{e}, \infty \mathbf{e}} = \mathcal{T}_\lambda.$$

Here $\infty \mathbf{e}$ is the n -dimensional column vector whose elements are all ∞ . A plot of the function $t \mapsto \mathcal{S}_{1,2}(t)$ is given in Figure 6.3.

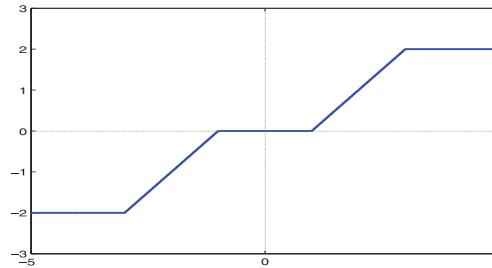


Figure 6.3. The two-sided soft thresholding function $t \mapsto \mathcal{S}_{1,2}(t) = \min\{\max\{|t| - 1, 0\}, 2\} \text{sgn}(t)$.

Example 6.34 (projection onto the intersection of weighted l_1 ball and a box). Let $C \subseteq \mathbb{R}^n$ be given by

$$C = \left\{ \mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n \omega_i |x_i| \leq \beta, -\boldsymbol{\alpha} \leq \mathbf{x} \leq \boldsymbol{\alpha} \right\},$$

where $\omega \in \mathbb{R}_+^n$, $\alpha \in [0, \infty]^n$, and $\beta \in \mathbb{R}_{++}$. Then obviously $C = \text{Lev}(f, \beta)$, where

$$f(\mathbf{x}) = \omega^T |\mathbf{x}| + \delta_{\text{Box}[-\alpha, \alpha]}(\mathbf{x}) = \begin{cases} \sum_{i=1}^n \omega_i |x_i|, & -\alpha \leq \mathbf{x} \leq \alpha, \\ \infty & \text{else} \end{cases}$$

for any $\mathbf{x} \in \mathbb{R}^n$. By Example 6.23, for any $\lambda > 0$ and $\mathbf{x} \in \mathbb{R}^n$,

$$\text{prox}_{\lambda f}(\mathbf{x}) = (\min\{\max\{|x_i| - \lambda \omega_i, 0\}, \alpha_i\} \text{sgn}(x_i))_{i=1}^n = \mathcal{S}_{\lambda \omega, \alpha}(\mathbf{x}).$$

Therefore, invoking Theorem 6.30, we obtain that

$$P_C(\mathbf{x}) = \begin{cases} P_{\text{Box}[-\alpha, \alpha]}(\mathbf{x}), & \omega^T |P_{\text{Box}[-\alpha, \alpha]}(\mathbf{x})| \leq \beta, \\ \mathcal{S}_{\lambda^* \omega, \alpha}(\mathbf{x}), & \omega^T |P_{\text{Box}[-\alpha, \alpha]}(\mathbf{x})| > \beta, \end{cases}$$

where λ^* is any positive root of the nonincreasing function

$$\varphi(\lambda) = \omega^T |\mathcal{S}_{\lambda \omega, \alpha}(\mathbf{x})| - \beta.$$

■

As a final illustration of Theorem 6.30, we give an example in which the domain of f is not closed.

Example 6.35. Let

$$C = \{\mathbf{x} \in \mathbb{R}_{++}^n : \prod_{i=1}^n x_i \geq \alpha\},$$

where $\alpha > 0$. The key property that will allow us to compute the orthogonal projection onto C is the fact that it can be rewritten as

$$C = \left\{ \mathbf{x} \in \mathbb{R}_{++}^n : -\sum_{i=1}^n \log x_i \leq -\log \alpha \right\}.$$

Thus, $C = \text{Lev}(f, -\log \alpha)$, where $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ is the negative sum of logs function:

$$f(\mathbf{x}) = \begin{cases} -\sum_{i=1}^n \log x_i, & \mathbf{x} \in \mathbb{R}_{++}^n, \\ \infty & \text{else.} \end{cases}$$

In Example 6.9 it was shown that for any $\mathbf{x} \in \mathbb{R}^n$,

$$\text{prox}_{\lambda f}(\mathbf{x}) = \left(\frac{x_j + \sqrt{x_j^2 + 4\lambda}}{2} \right)_{j=1}^n.$$

We can now invoke Theorem 6.30 to obtain a formula (up to a single parameter that can be found by a one-dimensional search) for the projection onto C , but there is one issue that needs to be treated delicately. If $\mathbf{x} \notin \mathbb{R}_{++}^n$, meaning that it has at least one nonpositive element, then $P_{\mathbb{R}_{++}^n}(\mathbf{x})$ does not exist. In this case only the second part of (6.24) is relevant, meaning that $P_C(\mathbf{x}) = \text{prox}_{\lambda^* f}(\mathbf{x})$. To conclude,

$$P_C(\mathbf{x}) = \begin{cases} \mathbf{x}, & \mathbf{x} \in C, \\ \left(\frac{x_j + \sqrt{x_j^2 + 4\lambda^*}}{2} \right)_{j=1}^n, & \mathbf{x} \notin C, \end{cases}$$

where λ^* is any positive root of the nonincreasing function

$$\varphi(\lambda) = - \sum_{j=1}^n \log \left(\frac{x_j + \sqrt{x_j^2 + 4\lambda}}{2} \right) + \log \alpha.$$

■

6.4.5 Projection onto Epigraphs

We can use Theorem 6.30 to prove a theorem on the projection onto epigraphs of convex functions.

Theorem 6.36 (orthogonal projection onto epigraphs). *Let*

$$C = \text{epi}(g) = \{(\mathbf{x}, t) \in \mathbb{E} \times \mathbb{R} : g(\mathbf{x}) \leq t\},$$

where $g : \mathbb{E} \rightarrow \mathbb{R}$ is convex. Then

$$P_C((\mathbf{x}, s)) = \begin{cases} (\mathbf{x}, s), & g(\mathbf{x}) \leq s, \\ (\text{prox}_{\lambda^* g}(\mathbf{x}), s + \lambda^*), & g(\mathbf{x}) > s, \end{cases}$$

where λ^* is any positive root of the function

$$\psi(\lambda) = g(\text{prox}_{\lambda g}(\mathbf{x})) - \lambda - s.$$

In addition, ψ is nonincreasing.

Proof. Define $f : \mathbb{E} \times \mathbb{R} \rightarrow \mathbb{R}$ as $f(\mathbf{x}, t) \equiv g(\mathbf{x}) - t$. By definition of the prox,

$$\begin{aligned} \text{prox}_{\lambda f}(\mathbf{x}, s) &= \underset{\mathbf{y}, t}{\text{argmin}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|^2 + \frac{1}{2}(t - s)^2 + \lambda f(\mathbf{y}, t) \right\} \\ &= \underset{\mathbf{y}, t}{\text{argmin}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|^2 + \frac{1}{2}(t - s)^2 + \lambda g(\mathbf{y}) - \lambda t \right\}. \end{aligned}$$

The above problem is separable in \mathbf{y} and t , and thus

$$\begin{aligned} \text{prox}_{\lambda f}(\mathbf{x}, s) &= \left(\underset{\mathbf{y}}{\text{argmin}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|^2 + \lambda g(\mathbf{y}) \right\}, \underset{t}{\text{argmin}} \left\{ \frac{1}{2}(t - s)^2 - \lambda t \right\} \right) \\ &= (\text{prox}_{\lambda g}(\mathbf{x}), \text{prox}_{\lambda h}(s)), \end{aligned}$$

where $h(t) \equiv -t$. Since λh is linear, then by Section 6.2.2, $\text{prox}_{\lambda h}(z) = z + \lambda$ for any $z \in \mathbb{R}$. Thus,

$$\text{prox}_{\lambda f}(\mathbf{x}, s) = (\text{prox}_{\lambda g}(\mathbf{x}), s + \lambda).$$

Since $\text{epi}(g) = \text{Lev}(f, 0)$, we can invoke Theorem 6.30 (noting that $\text{dom}(f) = \mathbb{E}$) and obtain that

$$P_C((\mathbf{x}, s)) = \begin{cases} (\mathbf{x}, s), & g(\mathbf{x}) \leq s, \\ (\text{prox}_{\lambda^* g}(\mathbf{x}), s + \lambda), & g(\mathbf{x}) > s, \end{cases}$$

where λ^* is any positive root of the function

$$\psi(\lambda) = g(\text{prox}_{\lambda g}(\mathbf{x})) - \lambda - s,$$

which by Theorem 6.30 is nonincreasing. \square

Example 6.37 (projection onto the Lorentz cone). Consider the *Lorentz cone*, which is given by $L^n = \{(\mathbf{x}, t) \in \mathbb{R}^n \times \mathbb{R} : \|\mathbf{x}\|_2 \leq t\}$. We will show that for any $(\mathbf{x}, s) \in \mathbb{R}^n \times \mathbb{R}$,

$$P_{L^n}(\mathbf{x}, s) = \begin{cases} \left(\frac{\|\mathbf{x}\|_2+s}{2\|\mathbf{x}\|_2}\mathbf{x}, \frac{\|\mathbf{x}\|_2+s}{2}\right), & \|\mathbf{x}\|_2 \geq |s|, \\ (\mathbf{0}, 0), & s < \|\mathbf{x}\|_2 < -s, \\ (\mathbf{x}, s), & \|\mathbf{x}\|_2 \leq s. \end{cases}$$

To show the above,³¹ we invoke Theorem 6.36 to obtain the formula

$$P_{L^n}((\mathbf{x}, s)) = \begin{cases} (\mathbf{x}, s), & \|\mathbf{x}\|_2 \leq s, \\ (\text{prox}_{\lambda^* \|\cdot\|_2}(\mathbf{x}), s + \lambda^*), & \|\mathbf{x}\|_2 > s, \end{cases}$$

where λ^* is any positive root of the nonincreasing function

$$\psi(\lambda) = \|\text{prox}_{\lambda \|\cdot\|_2}(\mathbf{x})\|_2 - \lambda - s. \quad (6.29)$$

Let $(\mathbf{x}, s) \in \mathbb{R}^n \times \mathbb{R}$ be such that $\|\mathbf{x}\|_2 > s$. Recall that by Example 6.19,

$$\text{prox}_{\lambda \|\cdot\|_2}(\mathbf{x}) = \left[1 - \frac{\lambda}{\max\{\|\mathbf{x}\|_2, \lambda\}}\right]_+ \mathbf{x}.$$

Plugging the above into the expression of ψ in (6.29) yields

$$\psi(\lambda) = \begin{cases} \|\mathbf{x}\|_2 - 2\lambda - s, & \lambda \leq \|\mathbf{x}\|_2, \\ -\lambda - s, & \lambda \geq \|\mathbf{x}\|_2. \end{cases}$$

The unique positive root λ^* of the piecewise linear function ψ is

$$\lambda^* = \begin{cases} \frac{\|\mathbf{x}\|_2-s}{2}, & \|\mathbf{x}\|_2 \geq -s, \\ -s, & \|\mathbf{x}\|_2 < -s. \end{cases}$$

³¹Actually, the formula for $P_C(\mathbf{x})$ when $\|\mathbf{x}\|_2 = s$ appears twice in the formula, but in both cases it amounts to (\mathbf{x}, s) .

Thus, in the case $\|\mathbf{x}\|_2 > s$ (noting that $\|\mathbf{x}\|_2 \geq -s$ corresponds to the case where $\|\mathbf{x}\|_2 \geq \lambda^*$ and $\|\mathbf{x}\|_2 < -s$ corresponds to $\|\mathbf{x}\|_2 \leq \lambda^*$),

$$\begin{aligned} (\text{prox}_{\lambda^* \|\cdot\|_2}(\mathbf{x}), s + \lambda^*) &= \left(\left[1 - \frac{\lambda^*}{\max\{\|\mathbf{x}\|_2, \lambda^*\}} \right]_+ \mathbf{x}, s + \lambda^* \right), \\ &= \begin{cases} \left(\left[1 - \frac{\|\mathbf{x}\|_2 - s}{2\|\mathbf{x}\|_2} \right]_+ \mathbf{x}, \frac{\|\mathbf{x}\|_2 + s}{2} \right), & \|\mathbf{x}\|_2 \geq -s, \\ (\mathbf{0}, 0), & \|\mathbf{x}\|_2 < -s. \end{cases} \\ &= \begin{cases} \left(\frac{\|\mathbf{x}\|_2 + s}{2\|\mathbf{x}\|_2} \mathbf{x}, \frac{\|\mathbf{x}\|_2 + s}{2} \right), & \|\mathbf{x}\|_2 \geq -s, \\ (\mathbf{0}, 0), & \|\mathbf{x}\|_2 < -s. \end{cases} \end{aligned}$$

Recalling that $\|\mathbf{x}\|_2 > s$, we have thus established that $P_{L^n}(\mathbf{x}, s) = (\mathbf{0}, 0)$ when $s < \|\mathbf{x}\|_2 < -s$ and that whenever (\mathbf{x}, s) satisfies $\|\mathbf{x}\|_2 > s$ and $\|\mathbf{x}\|_2 \geq -s$, the formula

$$P_{L^n}(\mathbf{x}, s) = \left(\frac{\|\mathbf{x}\|_2 + s}{2\|\mathbf{x}\|_2} \mathbf{x}, \frac{\|\mathbf{x}\|_2 + s}{2} \right) \quad (6.30)$$

holds. The result now follows by noting that

$$\{(\mathbf{x}, s) : \|\mathbf{x}\|_2 \geq |s|\} = \{(\mathbf{x}, s) : \|\mathbf{x}\|_2 > s, \|\mathbf{x}\|_2 \geq -s\} \cup \{(\mathbf{x}, s) : \|\mathbf{x}\|_2 = s\},$$

and that formula (6.30) is trivial for the case $\|\mathbf{x}\|_2 = s$ (amounts to $P_{L^n}(\mathbf{x}, s) = (\mathbf{x}, s)$). ■

Example 6.38 (projection onto the epigraph of the l_1 -norm). Let

$$C = \{(\mathbf{y}, t) \in \mathbb{R}^n \times \mathbb{R} : \|\mathbf{y}\|_1 \leq t\}.$$

Invoking Theorem 6.36 and recalling that for any $\lambda > 0$, $\text{prox}_{\lambda \|\cdot\|_1} = \mathcal{T}_\lambda$, where \mathcal{T}_λ is the soft thresholding operator (see Example 6.8), it follows that

$$P_C((\mathbf{x}, s)) = \begin{cases} (\mathbf{x}, s), & \|\mathbf{x}\|_1 \leq s, \\ (\mathcal{T}_{\lambda^*}(\mathbf{x}), s + \lambda^*), & \|\mathbf{x}\|_1 > s, \end{cases}$$

where λ^* is any positive root of the nonincreasing function

$$\varphi(\lambda) = \|\mathcal{T}_\lambda(\mathbf{x})\|_1 - \lambda - s.$$

■

6.4.6 Summary of Orthogonal Projection Computations

Table 6.1 describes all the examples of orthogonal projection computations onto subsets of \mathbb{R}^n and $\mathbb{R}^n \times \mathbb{R}$ that were discussed so far.

Table 6.1. The following notation is used in the table. $[\mathbf{x}]_+$ is the non-negative part of \mathbf{x} , $\mathcal{T}_\lambda(\mathbf{y}) = ([|y_i| - \lambda]_+ \text{sgn}(y_i))_{i=1}^n$, and $\mathcal{S}_{\mathbf{a}, \mathbf{b}}(\mathbf{x}) = (\min\{\max\{|x_i| - a_i, 0\}, b_i\} \text{sgn}(x_i))_{i=1}^n$.

set (C)	$P_C(\mathbf{x})$	Assumptions	Reference
\mathbb{R}_+^n	$[\mathbf{x}]_+$	—	Lemma 6.26
$\text{Box}[\ell, \mathbf{u}]$	$P_{\text{Box}}(\mathbf{x})_i = \min\{\max\{x_i, \ell_i\}, u_i\}$	$\ell_i \leq u_i$	Lemma 6.26
$B_{\ \cdot\ _2}[\mathbf{c}, r]$	$\mathbf{c} + \frac{r}{\max\{\ \mathbf{x}-\mathbf{c}\ _2, r\}}(\mathbf{x} - \mathbf{c})$	$\mathbf{c} \in \mathbb{R}^n, r > 0$	Lemma 6.26
$\{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}\}$	$\mathbf{x} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{x} - \mathbf{b})$	$\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, \mathbf{A} full row rank	Lemma 6.26
$\{\mathbf{x} : \mathbf{a}^T \mathbf{x} \leq b\}$	$\mathbf{x} - \frac{[\mathbf{a}^T \mathbf{x} - b]_+}{\ \mathbf{a}\ ^2} \mathbf{a}$	$\mathbf{0} \neq \mathbf{a} \in \mathbb{R}^n, b \in \mathbb{R}$	Lemma 6.26
Δ_n	$[\mathbf{x} - \mu^* \mathbf{e}]_+$ where $\mu^* \in \mathbb{R}$ satisfies $\mathbf{e}^T [\mathbf{x} - \mu^* \mathbf{e}]_+ = 1$	—	Corollary 6.29
$H_{\mathbf{a}, b} \cap \text{Box}[\ell, \mathbf{u}]$	$P_{\text{Box}[\ell, \mathbf{u}]}(\mathbf{x} - \mu^* \mathbf{a})$ where $\mu^* \in \mathbb{R}$ satisfies $\mathbf{a}^T P_{\text{Box}[\ell, \mathbf{u}]}(\mathbf{x} - \mu^* \mathbf{a}) = b$	$\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, $b \in \mathbb{R}$	Theorem 6.27
$H_{\mathbf{a}, b}^- \cap \text{Box}[\ell, \mathbf{u}]$	$\begin{cases} P_{\text{Box}[\ell, \mathbf{u}]}(\mathbf{x}), & \mathbf{a}^T \mathbf{v}_x \leq b, \\ P_{\text{Box}[\ell, \mathbf{u}]}(\mathbf{x} - \lambda^* \mathbf{a}), & \mathbf{a}^T \mathbf{v}_x > b, \end{cases}$ $\mathbf{v}_x = P_{\text{Box}[\ell, \mathbf{u}]}(\mathbf{x}),$ $\mathbf{a}^T P_{\text{Box}[\ell, \mathbf{u}]}(\mathbf{x} - \lambda^* \mathbf{a}) = b, \lambda^* > 0$	$\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, $b \in \mathbb{R}$	Example 6.32
$B_{\ \cdot\ _1}[\mathbf{0}, \alpha]$	$\begin{cases} \mathbf{x}, & \ \mathbf{x}\ _1 \leq \alpha, \\ \mathcal{T}_{\lambda^*}(\mathbf{x}), & \ \mathbf{x}\ _1 > \alpha, \\ \ \mathcal{T}_{\lambda^*}(\mathbf{x})\ _1 = \alpha, \lambda^* > 0 \end{cases}$	$\alpha > 0$	Example 6.33
$\{\mathbf{x} : \boldsymbol{\omega}^T \mathbf{x} \leq \beta, -\boldsymbol{\alpha} \leq \mathbf{x} \leq \boldsymbol{\alpha}\}$	$\begin{cases} \mathbf{v}_x, & \boldsymbol{\omega}^T \mathbf{v}_x \leq \beta, \\ \mathcal{S}_{\lambda^* \boldsymbol{\omega}, \boldsymbol{\alpha}}(\mathbf{x}), & \boldsymbol{\omega}^T \mathbf{v}_x > \beta, \end{cases}$ $\mathbf{v}_x = P_{\text{Box}[-\boldsymbol{\alpha}, \boldsymbol{\alpha}]}(\mathbf{x}),$ $\boldsymbol{\omega}^T \mathcal{S}_{\lambda^* \boldsymbol{\omega}, \boldsymbol{\alpha}}(\mathbf{x}) = \beta, \lambda^* > 0$	$\boldsymbol{\omega} \in \mathbb{R}_+^n, \boldsymbol{\alpha} \in [0, \infty]^n, \beta \in \mathbb{R}_{++}$	Example 6.34
$\{\mathbf{x} > \mathbf{0} : \Pi x_i \geq \alpha\}$	$\begin{cases} \mathbf{x}, & \mathbf{x} \in C, \\ \left(\frac{x_j + \sqrt{x_j^2 + 4\lambda^*}}{2} \right)_{j=1}^n, & \mathbf{x} \notin C, \end{cases}$ $\prod_{j=1}^n \left((x_j + \sqrt{x_j^2 + 4\lambda^*})/2 \right) = \alpha, \lambda^* > 0$	$\alpha > 0$	Example 6.35
$\{(\mathbf{x}, s) : \ \mathbf{x}\ _2 \leq s\}$	$\left(\frac{\ \mathbf{x}\ _2 + s}{2\ \mathbf{x}\ _2} \mathbf{x}, \frac{\ \mathbf{x}\ _2 + s}{2} \right)$ if $\ \mathbf{x}\ _2 \geq s $ $(\mathbf{0}, 0)$ if $s < \ \mathbf{x}\ _2 < -s$, (\mathbf{x}, s) if $\ \mathbf{x}\ _2 \leq s$.	—	Example 6.37
$\{(\mathbf{x}, s) : \ \mathbf{x}\ _1 \leq s\}$	$\begin{cases} (\mathbf{x}, s), & \ \mathbf{x}\ _1 \leq s, \\ (\mathcal{T}_{\lambda^*}(\mathbf{x}), s + \lambda^*), & \ \mathbf{x}\ _1 > s, \\ \ \mathcal{T}_{\lambda^*}(\mathbf{x})\ _1 - \lambda^* - s = 0, \lambda^* > 0 \end{cases}$	—	Example 6.38

6.5 The Second Prox Theorem

We can use Fermat's optimality condition (Theorem 3.63) in order to prove the second prox theorem.

Theorem 6.39 (second prox theorem). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper closed and convex function. Then for any $\mathbf{x}, \mathbf{u} \in \mathbb{E}$, the following three claims are equivalent:*

- (i) $\mathbf{u} = \text{prox}_f(\mathbf{x})$.
- (ii) $\mathbf{x} - \mathbf{u} \in \partial f(\mathbf{u})$.
- (iii) $\langle \mathbf{x} - \mathbf{u}, \mathbf{y} - \mathbf{u} \rangle \leq f(\mathbf{y}) - f(\mathbf{u})$ for any $\mathbf{y} \in \mathbb{E}$.

Proof. By definition, $\mathbf{u} = \text{prox}_f(\mathbf{x})$ if and only if \mathbf{u} is the minimizer of the problem

$$\min_{\mathbf{v}} \left\{ f(\mathbf{v}) + \frac{1}{2} \|\mathbf{v} - \mathbf{x}\|^2 \right\},$$

which, by Fermat's optimality condition (Theorem 3.63) and the sum rule of subdifferential calculus (Theorem 3.40), is equivalent to the relation

$$\mathbf{0} \in \partial f(\mathbf{u}) + \mathbf{u} - \mathbf{x}. \quad (6.31)$$

We have thus shown the equivalence between claims (i) and (ii). Finally, by the definition of the subgradient, the membership relation of claim (ii) is equivalent to (iii). \square

A direct consequence of the second prox theorem is that for a proper closed and convex function, $\mathbf{x} = \text{prox}_f(\mathbf{x})$ if and only \mathbf{x} is a minimizer of f .

Corollary 6.40. *Let f be a proper closed and convex function. Then \mathbf{x} is a minimizer of f if and only if $\mathbf{x} = \text{prox}_f(\mathbf{x})$.*

Proof. \mathbf{x} is a minimizer of f if and only if $\mathbf{0} \in \partial f(\mathbf{x})$, that is, if and only if $\mathbf{x} - \mathbf{x} \in \partial f(\mathbf{x})$, which by the second prox theorem (equivalence between (i) and (ii)) is the same as $\mathbf{x} = \text{prox}_f(\mathbf{x})$. \square

When $f = \delta_C$, with C being a nonempty closed and convex set, the equivalence between claims (i) and (iii) in the second prox theorem amounts to the second projection theorem.

Theorem 6.41 (second projection theorem). *Let $C \subseteq \mathbb{E}$ be a nonempty closed and convex set. Let $\mathbf{u} \in C$. Then $\mathbf{u} = P_C(\mathbf{x})$ if and only if*

$$\langle \mathbf{x} - \mathbf{u}, \mathbf{y} - \mathbf{u} \rangle \leq 0 \text{ for any } \mathbf{y} \in C.$$

Another rather direct result of the second prox theorem is the firm nonexpansivity of the prox operator.

Theorem 6.42 (firm nonexpansivity of the prox operator). *Let f be a proper closed and convex function. Then for any $\mathbf{x}, \mathbf{y} \in \mathbb{E}$,*

(a) **(firm nonexpansivity)**

$$\langle \mathbf{x} - \mathbf{y}, \text{prox}_f(\mathbf{x}) - \text{prox}_f(\mathbf{y}) \rangle \geq \|\text{prox}_f(\mathbf{x}) - \text{prox}_f(\mathbf{y})\|^2.$$

(b) **(nonexpansivity)**

$$\|\text{prox}_f(\mathbf{x}) - \text{prox}_f(\mathbf{y})\| \leq \|\mathbf{x} - \mathbf{y}\|.$$

Proof. (a) Denoting $\mathbf{u} = \text{prox}_f(\mathbf{x}), \mathbf{v} = \text{prox}_f(\mathbf{y})$, by the equivalence of (i) and (ii) in the second prox theorem (Theorem 6.39), it follows that

$$\mathbf{x} - \mathbf{u} \in \partial f(\mathbf{u}), \mathbf{y} - \mathbf{v} \in \partial f(\mathbf{v}).$$

Thus, by the subgradient inequality,

$$\begin{aligned} f(\mathbf{v}) &\geq f(\mathbf{u}) + \langle \mathbf{x} - \mathbf{u}, \mathbf{v} - \mathbf{u} \rangle, \\ f(\mathbf{u}) &\geq f(\mathbf{v}) + \langle \mathbf{y} - \mathbf{v}, \mathbf{u} - \mathbf{v} \rangle. \end{aligned}$$

Summing the above two inequalities, we obtain

$$0 \geq \langle \mathbf{y} - \mathbf{x} + \mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v} \rangle,$$

which is the same as

$$\langle \mathbf{x} - \mathbf{y}, \mathbf{u} - \mathbf{v} \rangle \geq \|\mathbf{u} - \mathbf{v}\|^2,$$

that is,

$$\langle \mathbf{x} - \mathbf{y}, \text{prox}_f(\mathbf{x}) - \text{prox}_f(\mathbf{y}) \rangle \geq \|\text{prox}_f(\mathbf{x}) - \text{prox}_f(\mathbf{y})\|^2.$$

(b) If $\text{prox}_f(\mathbf{x}) = \text{prox}_f(\mathbf{y})$, then the inequality is obvious. Assume that $\text{prox}_f(\mathbf{x}) \neq \text{prox}_f(\mathbf{y})$. Using (a) and the Cauchy–Schwarz inequality, it follows that

$$\begin{aligned} \|\text{prox}_f(\mathbf{x}) - \text{prox}_f(\mathbf{y})\|^2 &\leq \langle \text{prox}_h(\mathbf{x}) - \text{prox}_h(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \\ &\leq \|\text{prox}_h(\mathbf{x}) - \text{prox}_h(\mathbf{y})\| \cdot \|\mathbf{x} - \mathbf{y}\|. \end{aligned}$$

Dividing by $\|\text{prox}_h(\mathbf{x}) - \text{prox}_h(\mathbf{y})\|$, the desired result is established. \square

The following result shows how to compute the prox of the distance function to a nonempty closed and convex set. The proof is heavily based on the second prox theorem.

Lemma 6.43 (prox of the distance function). *Let $C \subseteq \mathbb{E}$ be a nonempty, closed, and convex set. Let $\lambda > 0$. Then for any $\mathbf{x} \in \mathbb{E}$,*

$$\text{prox}_{\lambda d_C}(\mathbf{x}) = \begin{cases} (1 - \theta)\mathbf{x} + \theta P_C(\mathbf{x}), & d_C(\mathbf{x}) > \lambda, \\ P_C(\mathbf{x}), & d_C(\mathbf{x}) \leq \lambda, \end{cases} \quad (6.32)$$

where³²

$$\theta = \frac{\lambda}{d_C(\mathbf{x})}. \quad (6.33)$$

Proof. Let $\mathbf{u} = \text{prox}_{\lambda d_C}(\mathbf{x})$. By the second prox theorem (Theorem 6.39),

$$\mathbf{x} - \mathbf{u} \in \lambda \partial d_C(\mathbf{u}). \quad (6.34)$$

We will split the analysis into two cases.

Case I. $\mathbf{u} \notin C$. By Example 3.49, (6.34) is the same as

$$\mathbf{x} - \mathbf{u} = \lambda \frac{\mathbf{u} - P_C(\mathbf{u})}{d_C(\mathbf{u})}.$$

Denoting $\alpha = \frac{\lambda}{d_C(\mathbf{u})}$, the last equality can be rewritten as

$$\mathbf{u} = \frac{1}{\alpha + 1} \mathbf{x} + \frac{\alpha}{\alpha + 1} P_C(\mathbf{u}) \quad (6.35)$$

or as

$$\mathbf{x} - P_C(\mathbf{u}) = (\alpha + 1)(\mathbf{u} - P_C(\mathbf{u})). \quad (6.36)$$

By the second projection theorem (Theorem 6.41), in order to show that $P_C(\mathbf{u}) = P_C(\mathbf{x})$, it is enough to show that

$$\langle \mathbf{x} - P_C(\mathbf{u}), \mathbf{y} - P_C(\mathbf{u}) \rangle \leq 0 \text{ for any } \mathbf{y} \in C. \quad (6.37)$$

Using (6.36), we can deduce that (6.37) is equivalent to

$$(\alpha + 1) \langle \mathbf{u} - P_C(\mathbf{u}), \mathbf{y} - P_C(\mathbf{u}) \rangle \leq 0 \text{ for any } \mathbf{y} \in C,$$

which is a valid inequality by the second projection theorem, and hence $P_C(\mathbf{u}) = P_C(\mathbf{x})$. Using this fact and taking the norm in both sides of (6.36), we obtain that

$$d_C(\mathbf{x}) = (\alpha + 1)d_C(\mathbf{u}) = d_C(\mathbf{u}) + \lambda,$$

which also shows that in this case $d_C(\mathbf{x}) > \lambda$ (since $d_C(\mathbf{u}) > 0$) and that

$$\frac{1}{\alpha + 1} = \frac{d_C(\mathbf{u})}{\lambda + d_C(\mathbf{u})} = \frac{d_C(\mathbf{x}) - \lambda}{d_C(\mathbf{x})} = 1 - \theta,$$

where θ is given in (6.33). Therefore, (6.35) can also be written as (recalling also that $P_C(\mathbf{u}) = P_C(\mathbf{x})$)

$$\text{prox}_{\lambda d_C}(\mathbf{x}) = (1 - \theta)\mathbf{x} + \theta P_C(\mathbf{x}). \quad (6.38)$$

Case II. If $\mathbf{u} \in C$, then $\mathbf{u} = P_C(\mathbf{x})$. To show this, let $\mathbf{v} \in C$. Since $\mathbf{u} = \text{prox}_{\lambda d_C}(\mathbf{x})$, it follows in particular that

$$\lambda d_C(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \leq \lambda d_C(\mathbf{v}) + \frac{1}{2} \|\mathbf{v} - \mathbf{x}\|^2,$$

³²Since θ is used only when $\mathbf{x} \notin C$, it follows that $d_C(\mathbf{x}) > 0$, so that θ is well defined.

and hence, since $d_C(\mathbf{u}) = d_C(\mathbf{v}) = 0$,

$$\|\mathbf{u} - \mathbf{x}\| \leq \|\mathbf{v} - \mathbf{x}\|.$$

Therefore,

$$\mathbf{u} = \operatorname{argmin}_{\mathbf{v} \in C} \|\mathbf{v} - \mathbf{x}\| = P_C(\mathbf{x}).$$

By Example 3.49, the optimality condition (6.34) becomes

$$\frac{\mathbf{x} - P_C(\mathbf{x})}{\lambda} \in N_C(\mathbf{u}) \cap B[\mathbf{0}, 1],$$

which in particular implies that

$$\left\| \frac{\mathbf{x} - P_C(\mathbf{x})}{\lambda} \right\| \leq 1,$$

that is,

$$d_C(\mathbf{x}) = \|P_C(\mathbf{x}) - \mathbf{x}\| \leq \lambda.$$

Since the first case in which (6.38) holds corresponds to vectors satisfying $d_C(\mathbf{x}) > \lambda$, while the second case in which $\operatorname{prox}_{\lambda d_C}(\mathbf{x}) = P_C(\mathbf{x})$ corresponds to vectors satisfying $d_C(\mathbf{x}) \leq \lambda$, the desired result (6.32) is established. \square

6.6 Moreau Decomposition

A key property of the prox operator is the so-called Moreau decomposition theorem, which connects the prox operator of proper closed convex functions and their conjugates.

Theorem 6.44 (Moreau decomposition). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be proper closed and convex. Then for any $\mathbf{x} \in \mathbb{E}$,*

$$\operatorname{prox}_f(\mathbf{x}) + \operatorname{prox}_{f^*}(\mathbf{x}) = \mathbf{x}.$$

Proof. Let $\mathbf{x} \in \mathbb{E}$ and denote $\mathbf{u} = \operatorname{prox}_f(\mathbf{x})$. Then by the equivalence between claims (i) and (ii) in the second prox theorem (Theorem 6.39), it follows that $\mathbf{x} - \mathbf{u} \in \partial f(\mathbf{u})$, which by the conjugate subgradient theorem (Theorem 4.20) is equivalent to $\mathbf{u} \in \partial f^*(\mathbf{x} - \mathbf{u})$. Using the second prox theorem again, we conclude that $\mathbf{x} - \mathbf{u} = \operatorname{prox}_{f^*}(\mathbf{x})$. Therefore,

$$\operatorname{prox}_f(\mathbf{x}) + \operatorname{prox}_{f^*}(\mathbf{x}) = \mathbf{u} + (\mathbf{x} - \mathbf{u}) = \mathbf{x}. \quad \square$$

The next result is a useful extension of the Moreau decomposition theorem.

Theorem 6.45 (extended Moreau decomposition). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be proper closed and convex, and let $\lambda > 0$. Then for any $\mathbf{x} \in \mathbb{E}$,*

$$\operatorname{prox}_{\lambda f}(\mathbf{x}) + \lambda \operatorname{prox}_{\lambda^{-1} f^*}(\mathbf{x}/\lambda) = \mathbf{x}. \quad (6.39)$$

Proof. Using Moreau decomposition, for any $\mathbf{x} \in \mathbb{E}$,

$$\text{prox}_{\lambda f}(\mathbf{x}) = \mathbf{x} - \text{prox}_{(\lambda f)^*}(\mathbf{x}) = \mathbf{x} - \text{prox}_{\lambda f^*(\cdot/\lambda)}(\mathbf{x}), \quad (6.40)$$

where the second equality follows by Theorem 4.14(a). By Theorem 6.12,

$$\text{prox}_{\lambda f^*(\cdot/\lambda)}(\mathbf{x}) = \lambda \text{prox}_{\lambda^{-1} f^*}(\mathbf{x}/\lambda),$$

which, combined with (6.40), yields (6.39). \square

6.6.1 Support Functions

Using Moreau decomposition, we can develop a formula for computing the prox of a support function of a given nonempty closed and convex set in terms of the orthogonal projection operator.

Theorem 6.46 (prox of support functions). *Let $C \subseteq \mathbb{E}$ be a nonempty closed and convex set, and let $\lambda > 0$. Then for any $\mathbf{x} \in \mathbb{E}$,*

$$\text{prox}_{\lambda \sigma_C}(\mathbf{x}) = \mathbf{x} - \lambda P_C(\mathbf{x}/\lambda). \quad (6.41)$$

Proof. A direct consequence of the extended Moreau decomposition formula (Theorem 6.45) along with the fact that $(\sigma_C)^* = \delta_C$ (Example 4.9). \square

Following are several examples of prox computations using formula (6.41).

Example 6.47 (prox of norms). Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}) = \lambda \|\mathbf{x}\|_\alpha$, where $\lambda > 0$ and $\|\cdot\|_\alpha$ is any norm on \mathbb{E} . Note that $\|\cdot\|_\alpha$ is not necessarily the endowed norm on \mathbb{E} , which is denoted by $\|\cdot\|$ and in this chapter is always assumed to be the Euclidean norm. We know by Example 2.31 that

$$\|\mathbf{x}\|_\alpha = \sigma_C(\mathbf{x}),$$

where

$$C = B_{\|\cdot\|_{\alpha,*}}[\mathbf{0}, 1] = \{\mathbf{x} \in \mathbb{E} : \|\mathbf{x}\|_{\alpha,*} \leq 1\}$$

with $\|\cdot\|_{\alpha,*}$ being the dual norm of $\|\cdot\|_\alpha$. Invoking Theorem 6.46, we obtain

$$\text{prox}_{\lambda \|\cdot\|_\alpha}(\mathbf{x}) = \mathbf{x} - \lambda P_{B_{\|\cdot\|_{\alpha,*}}[\mathbf{0}, 1]}(\mathbf{x}/\lambda).$$

\blacksquare

Example 6.48 (prox of l_∞ -norm). By Example 6.47 we have for all $\lambda > 0$ and $\mathbf{x} \in \mathbb{R}^n$,

$$\text{prox}_{\lambda \|\cdot\|_\infty}(\mathbf{x}) = \mathbf{x} - \lambda P_{B_{\|\cdot\|_1}[\mathbf{0}, 1]}(\mathbf{x}/\lambda).$$

The projection onto the l_1 unit ball can be easily computed by finding a root of a nonincreasing one-dimensional function; see Example 6.33. ■

Example 6.49 (prox of the max function). Consider the max function $g : \mathbb{R}^n \rightarrow \mathbb{R}$ given by $g(\mathbf{x}) = \max(\mathbf{x}) \equiv \max\{x_1, x_2, \dots, x_n\}$. It is easy to see that the max function is actually the support function of the unit simplex:

$$\max(\mathbf{x}) = \sigma_{\Delta_n}(\mathbf{x}).$$

Therefore, by Theorem 6.46, for any $\lambda > 0$ and $\mathbf{x} \in \mathbb{R}^n$,

$$\text{prox}_{\lambda \max(\cdot)}(\mathbf{x}) = \mathbf{x} - \lambda P_{\Delta_n}(\mathbf{x}/\lambda).$$

The projection onto the unit simplex can be efficiently computed by finding a root of a nonincreasing one-dimensional function; see Corollary 6.29. ■

Example 6.50 (prox of the sum-of- k -largest-values function). Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be given by

$$f(\mathbf{x}) = x_{[1]} + x_{[2]} + \dots + x_{[k]},$$

where $k \in \{1, 2, \dots, n\}$ and for any i , $x_{[i]}$ denotes i th largest value in the vector \mathbf{x} . It is not difficult to show that $f = \sigma_C$, where

$$C = \{\mathbf{y} \in \mathbb{R}^n : \mathbf{e}^T \mathbf{y} = k, \mathbf{0} \leq \mathbf{y} \leq \mathbf{e}\}.$$

Therefore, by Theorem 6.46,

$$\text{prox}_{\lambda f}(\mathbf{x}) = \mathbf{x} - \lambda P_C(\mathbf{x}/\lambda).$$

That is, for any $\mathbf{x} \in \mathbb{R}^n$,

$$\text{prox}_{\lambda f}(\mathbf{x}) = \mathbf{x} - \lambda P_{\{\mathbf{y}: \mathbf{e}^T \mathbf{y} = k, \mathbf{0} \leq \mathbf{y} \leq \mathbf{e}\}}(\mathbf{x}/\lambda).$$

As in the previous examples, computing the projection onto C amounts to finding a root of a monotone one-dimensional function; see Theorem 6.27. ■

Example 6.51 (prox of the sum-of- k -largest-absolute-values function). Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be given by

$$f(\mathbf{x}) = \sum_{i=1}^k |x_{\langle i \rangle}|,$$

where $k \in \{1, 2, \dots, n\}$ and $x_{\langle i \rangle}$ is the component of \mathbf{x} with the i th largest absolute value, meaning in particular that $|x_{\langle 1 \rangle}| \geq |x_{\langle 2 \rangle}| \geq \dots \geq |x_{\langle n \rangle}|$. Then

$$f(\mathbf{x}) = \max \left\{ \sum_{i=1}^n z_i x_i : \|\mathbf{z}\|_1 \leq k, -\mathbf{e} \leq \mathbf{z} \leq \mathbf{e} \right\}.$$

Therefore, $f = \sigma_C$, where

$$C = \{\mathbf{z} \in \mathbb{R}^n : \|\mathbf{z}\|_1 \leq k, -\mathbf{e} \leq \mathbf{z} \leq \mathbf{e}\},$$

and consequently, by Theorem 6.46,

$$\text{prox}_{\lambda f}(\mathbf{x}) = \mathbf{x} - \lambda P_C(\mathbf{x}/\lambda).$$

That is, for any $\mathbf{x} \in \mathbb{R}^n$,

$$\text{prox}_{\lambda f}(\mathbf{x}) = \mathbf{x} - \lambda P_{\{\mathbf{y}: \|\mathbf{y}\|_1 \leq k, -\mathbf{e} \leq \mathbf{y} \leq \mathbf{e}\}}(\mathbf{x}/\lambda).$$

The orthogonal projection in the above formula amounts to finding a root of a nonincreasing one-dimensional function; see Example 6.34. ■

6.7 The Moreau Envelope

6.7.1 Definition and Basic Properties

Definition 6.52 (Moreau envelope). *Given a proper closed convex function $f : \mathbb{E} \rightarrow (-\infty, \infty]$ and $\mu > 0$, the **Moreau envelope** of f is the function*

$$M_f^\mu(\mathbf{x}) = \min_{\mathbf{u} \in \mathbb{E}} \left\{ f(\mathbf{u}) + \frac{1}{2\mu} \|\mathbf{x} - \mathbf{u}\|^2 \right\}. \quad (6.42)$$

The parameter μ is called the *smoothing parameter*. The explanation for this terminology will be given in Section 6.7.2. By the first prox theorem (Theorem 6.3), the minimization problem in (6.42) has a unique solution, given by $\text{prox}_{\mu f}(\mathbf{x})$. Therefore, $M_f^\mu(\mathbf{x})$ is always a real number and

$$M_f^\mu(\mathbf{x}) = f(\text{prox}_{\mu f}(\mathbf{x})) + \frac{1}{2\mu} \|\mathbf{x} - \text{prox}_{\mu f}(\mathbf{x})\|^2.$$

Example 6.53 (Moreau envelope of indicators). Let $f = \delta_C$, where $C \subseteq \mathbb{E}$ is a nonempty closed and convex set. By Theorem 6.24, $\text{prox}_{\mu f}(\mathbf{x}) = P_C(\mathbf{x})$. Thus, for any $\mathbf{x} \in \mathbb{E}$,

$$M_f^\mu(\mathbf{x}) = \delta_C(P_C(\mathbf{x})) + \frac{1}{2\mu} \|\mathbf{x} - P_C(\mathbf{x})\|^2,$$

and hence

$$M_{\delta_C}^\mu = \frac{1}{2\mu} d_C^2.$$

The next example will show that the Moreau envelope of the (Euclidean) norm is the so-called *Huber function* defined as

$$H_\mu(\mathbf{x}) = \begin{cases} \frac{1}{2\mu} \|\mathbf{x}\|^2, & \|\mathbf{x}\| \leq \mu, \\ \|\mathbf{x}\| - \frac{\mu}{2}, & \|\mathbf{x}\| > \mu. \end{cases} \quad (6.43)$$

The one-dimensional Huber function is plotted in Figure 6.4, where it is illustrated that the function becomes smoother as μ becomes larger.

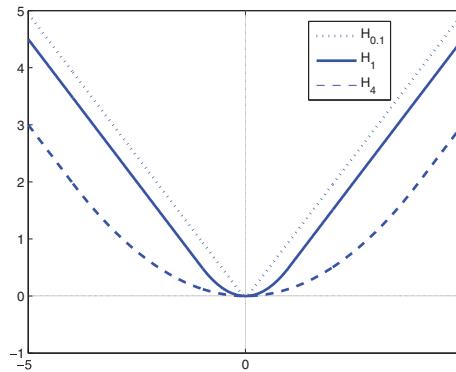


Figure 6.4. The Huber function with parameters $\mu = 0.1, 1, 4$. The function becomes smoother as μ gets larger.

Example 6.54 (Huber function). Let $f : \mathbb{E} \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}) = \|\mathbf{x}\|$. Then by Example 6.19, for any $\mathbf{x} \in \mathbb{E}$ and $\mu > 0$,

$$\text{prox}_{\mu f}(\mathbf{x}) = \left(1 - \frac{\mu}{\max\{\|\mathbf{x}\|, \mu\}}\right) \mathbf{x}.$$

Therefore,

$$M_f^\mu(\mathbf{x}) = \|\text{prox}_{\mu f}(\mathbf{x})\| + \frac{1}{2\mu} \|\mathbf{x} - \text{prox}_{\mu f}(\mathbf{x})\|^2 = \begin{cases} \frac{1}{2\mu} \|\mathbf{x}\|^2, & \|\mathbf{x}\| \leq \mu, \\ \|\mathbf{x}\| - \frac{\mu}{2}, & \|\mathbf{x}\| > \mu. \end{cases}$$

Thus, for any $\mu > 0$,

$$M_{\|\cdot\|}^\mu = H_\mu.$$

■

Note that the Moreau envelope function is actually a result of an infimal convolution operation between the function f and the function

$$\omega_\mu(\mathbf{x}) = \frac{1}{2\mu} \|\mathbf{x}\|^2. \quad (6.44)$$

That is,

$$M_f^\mu = f \square \omega_\mu.$$

One consequence of this observation is that by Theorem 2.19, if f is a proper closed³³ and convex function, then M_f^μ is convex. We summarize the above discussion in the following theorem.

³³Actually, closedness is not necessary in order to establish the convexity of the Moreau envelope.

Theorem 6.55. Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper closed and convex function, and let ω_μ be given in (6.44), where $\mu > 0$. Then

- (a) $M_f^\mu = f \square \omega_\mu$;
- (b) $M_f^\mu : \mathbb{E} \rightarrow \mathbb{R}$ is real-valued and convex.

We can immediately conclude from Theorem 6.55(a) along with the formula for the conjugate of the infimal convolution (Theorem 4.16) an expression for the conjugate of the Moreau envelope.

Corollary 6.56. Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper closed and convex function and let ω_μ be given in (6.44), where $\mu > 0$. Then

$$(M_f^\mu)^* = f^* + \omega_{\frac{1}{\mu}}.$$

Another useful algebraic property of the Moreau envelope is described in the following result.

Lemma 6.57. Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper closed and convex function, and let $\lambda, \mu > 0$. Then for any $\mathbf{x} \in \mathbb{E}$,

$$\lambda M_f^\mu(\mathbf{x}) = M_{\lambda f}^{\mu/\lambda}(\mathbf{x}). \quad (6.45)$$

Proof. For any $\mathbf{x} \in \mathbb{E}$,

$$\begin{aligned} \lambda M_f^\mu(\mathbf{x}) &= \lambda \min_{\mathbf{u}} \left\{ f(\mathbf{u}) + \frac{1}{2\mu} \|\mathbf{u} - \mathbf{x}\|^2 \right\} \\ &= \min_{\mathbf{u}} \left\{ \lambda f(\mathbf{u}) + \frac{1}{2\mu/\lambda} \|\mathbf{u} - \mathbf{x}\|^2 \right\} \\ &= M_{\lambda f}^{\mu/\lambda}(\mathbf{x}). \quad \square \end{aligned}$$

A simple calculus rule states that the Moreau envelope of a separable sum of functions is the sum of the corresponding Moreau envelopes.

Theorem 6.58 (Moreau envelope of separable functions). Suppose that $\mathbb{E} = \mathbb{E}_1 \times \mathbb{E}_2 \times \cdots \times \mathbb{E}_m$, and let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be given by

$$f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m) = \sum_{i=1}^m f_i(\mathbf{x}_i), \quad \mathbf{x}_1 \in \mathbb{E}_1, \mathbf{x}_2 \in \mathbb{E}_2, \dots, \mathbf{x}_m \in \mathbb{E}_m,$$

with $f_i : \mathbb{E}_i \rightarrow (-\infty, \infty]$ being a proper closed and convex function for any i . Then given $\mu > 0$, for any $\mathbf{x}_1 \in \mathbb{E}_1, \mathbf{x}_2 \in \mathbb{E}_2, \dots, \mathbf{x}_m \in \mathbb{E}_m$,

$$M_f^\mu(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m) = \sum_{i=1}^m M_{f_i}^\mu(\mathbf{x}_i).$$

Proof. For any $\mathbf{x}_1 \in \mathbb{E}_1, \mathbf{x}_2 \in \mathbb{E}_2, \dots, \mathbf{x}_m \in \mathbb{E}_m$, denoting $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m)$, we have

$$\begin{aligned} M_f^\mu(\mathbf{x}) &= \min_{\mathbf{u}_i \in \mathbb{E}_i, i=1,2,\dots,m} \left\{ f(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m) + \frac{1}{2\mu} \|(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m) - \mathbf{x}\|^2 \right\} \\ &= \min_{\mathbf{u}_i \in \mathbb{E}_i, i=1,2,\dots,m} \left\{ \sum_{i=1}^m f_i(\mathbf{u}_i) + \frac{1}{2\mu} \sum_{i=1}^m \|\mathbf{u}_i - \mathbf{x}_i\|^2 \right\} \\ &= \sum_{i=1}^m \min_{\mathbf{u}_i \in \mathbb{E}_i} \left\{ f_i(\mathbf{u}_i) + \frac{1}{2\mu} \|\mathbf{u}_i - \mathbf{x}_i\|^2 \right\} \\ &= \sum_{i=1}^m M_{f_i}^\mu(\mathbf{x}_i). \quad \square \end{aligned}$$

Example 6.59 (Moreau envelope of the l_1 -norm). Consider the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ given by $f(\mathbf{x}) = \|\mathbf{x}\|_1$. Note that

$$f(\mathbf{x}) = \|\mathbf{x}\|_1 = \sum_{i=1}^n g(x_i),$$

where $g(t) = |t|$. By Example 6.54, $M_g^\mu = H_\mu$. Thus, invoking Theorem 6.58, we obtain that for any $\mathbf{x} \in \mathbb{R}^n$,

$$M_f^\mu(\mathbf{x}) = \sum_{i=1}^n M_g^\mu(x_i) = \sum_{i=1}^n H_\mu(x_i). \quad \blacksquare$$

6.7.2 Differentiability of the Moreau Envelope

The main differentiability properties of the Moreau envelope function are stated in the next result.

Theorem 6.60 (smoothness of the Moreau envelope). Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper closed and convex function. Let $\mu > 0$. Then M_f^μ is $\frac{1}{\mu}$ -smooth over \mathbb{E} , and for any $\mathbf{x} \in \mathbb{E}$,

$$\nabla M_f^\mu(\mathbf{x}) = \frac{1}{\mu} (\mathbf{x} - \text{prox}_{\mu f}(\mathbf{x})).$$

Proof. By Theorem 6.55(a), $M_f^\mu = f \square \omega_\mu$, where $\omega_\mu = \frac{1}{2\mu} \|\cdot\|^2$. We can therefore invoke Theorem 5.30, whose assumptions are satisfied (taking $\omega = \omega_\mu$ and $L = \frac{1}{\mu}$), and conclude that M_f^μ is $\frac{1}{\mu}$ -smooth. In addition, since

$$\text{prox}_{\mu f}(\mathbf{x}) = \operatorname{argmin}_{\mathbf{u} \in \mathbb{E}} \left\{ f(\mathbf{u}) + \frac{1}{2\mu} \|\mathbf{u} - \mathbf{x}\|^2 \right\},$$

it follows that the vector $\mathbf{u}(\mathbf{x})$ defined in Theorem 5.30 is equal to $\text{prox}_{\mu f}(\mathbf{x})$ and that

$$\nabla M_f^\mu(\mathbf{x}) = \nabla \omega_\mu(\mathbf{x} - \mathbf{u}(\mathbf{x})) = \frac{1}{\mu} (\mathbf{x} - \text{prox}_{\mu f}(\mathbf{x})). \quad \square$$

Example 6.61 (1-smoothness of $\frac{1}{2}d_C^2$). Let $C \subseteq \mathbb{E}$ be a nonempty closed and convex set. Recall that by Example 6.53, $\frac{1}{2}d_C^2 = M_{\delta_C}^1$. Then by Theorem 6.60, $\frac{1}{2}d_C^2$ is 1-smooth and

$$\nabla \left(\frac{1}{2}d_C^2 \right) (\mathbf{x}) = \mathbf{x} - \text{prox}_{\delta_C}(\mathbf{x}) = \mathbf{x} - P_C(\mathbf{x}).$$

Note that the above expression for the gradient was already derived in Example 3.31 and that the 1-smoothness of $\frac{1}{2}d_C^2$ was already established twice in Examples 5.5 and 5.31. ■

Example 6.62 (smoothness of the Huber function). Recall that the Huber function is given by

$$H_\mu(\mathbf{x}) = \begin{cases} \frac{1}{2\mu}\|\mathbf{x}\|^2, & \|\mathbf{x}\| \leq \mu, \\ \|\mathbf{x}\| - \frac{\mu}{2}, & \|\mathbf{x}\| > \mu. \end{cases}$$

By Example 6.54, $H_\mu = M_f^\mu$, where $f(\mathbf{x}) = \|\mathbf{x}\|$. Then, by Theorem 6.60, H_μ is $\frac{1}{\mu}$ -smooth and

$$\begin{aligned} \nabla H_\mu(\mathbf{x}) &= \frac{1}{\mu} (\mathbf{x} - \text{prox}_{\mu f}(\mathbf{x})) \\ &\stackrel{(*)}{=} \frac{1}{\mu} \left(\mathbf{x} - \left(1 - \frac{\mu}{\max\{\|\mathbf{x}\|, \mu\}} \right) \mathbf{x} \right) \\ &= \begin{cases} \frac{1}{\mu}\mathbf{x}, & \|\mathbf{x}\| \leq \mu, \\ \frac{\mathbf{x}}{\|\mathbf{x}\|}, & \|\mathbf{x}\| > \mu, \end{cases} \end{aligned}$$

where the equality $(*)$ uses the expression for $\text{prox}_{\mu f}$ developed in Example 6.19. ■

6.7.3 Prox of the Moreau Envelope

An interesting and important result states that if we can compute the prox of a proper closed and convex function f , then we can also compute the prox of its Moreau envelope.

Theorem 6.63 (prox of Moreau envelope). Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper closed and convex function, and let $\mu > 0$. Then for any $\mathbf{x} \in \mathbb{E}$,

$$\text{prox}_{M_f^\mu}(\mathbf{x}) = \mathbf{x} + \frac{1}{\mu+1} (\text{prox}_{(\mu+1)f}(\mathbf{x}) - \mathbf{x}).$$

Proof. First note that

$$\min_{\mathbf{u}} \left\{ M_f^\mu(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2 \right\} = \min_{\mathbf{u}} \min_{\mathbf{y}} \left\{ f(\mathbf{y}) + \frac{1}{2\mu}\|\mathbf{u} - \mathbf{y}\|^2 + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2 \right\}. \quad (6.46)$$

Exchanging the order of minimizations, we obtain the following problem:

$$\min_{\mathbf{y}} \min_{\mathbf{u}} \left\{ f(\mathbf{y}) + \frac{1}{2\mu} \|\mathbf{u} - \mathbf{y}\|^2 + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \right\}. \quad (6.47)$$

The optimal solution of the inner minimization problem in \mathbf{u} is attained when the gradient w.r.t. \mathbf{u} vanishes:

$$\frac{1}{\mu}(\mathbf{u} - \mathbf{y}) + (\mathbf{u} - \mathbf{x}) = \mathbf{0},$$

that is, when

$$\mathbf{u} = \mathbf{u}_\mu \equiv \frac{\mu\mathbf{x} + \mathbf{y}}{\mu + 1}. \quad (6.48)$$

Therefore, the optimal value of the inner minimization problem in (6.47) is

$$\begin{aligned} f(\mathbf{y}) + \frac{1}{2\mu} \|\mathbf{u}_\mu - \mathbf{y}\|^2 + \frac{1}{2} \|\mathbf{u}_\mu - \mathbf{x}\|^2 &= f(\mathbf{y}) + \frac{1}{2\mu} \left\| \frac{\mu\mathbf{x} - \mu\mathbf{y}}{\mu + 1} \right\|^2 + \frac{1}{2} \left\| \frac{\mathbf{y} - \mathbf{x}}{\mu + 1} \right\|^2 \\ &= f(\mathbf{y}) + \frac{1}{2(\mu + 1)} \|\mathbf{x} - \mathbf{y}\|^2. \end{aligned}$$

Therefore, the optimal solution of (6.46) is given by (6.48), where \mathbf{y} is the solution of

$$\min_{\mathbf{y}} \left\{ f(\mathbf{y}) + \frac{1}{2(\mu + 1)} \|\mathbf{x} - \mathbf{y}\|^2 \right\},$$

that is, $\mathbf{y} = \text{prox}_{(\mu+1)f}(\mathbf{x})$. To summarize,

$$\text{prox}_{M_f^\mu}(\mathbf{x}) = \frac{1}{\mu + 1} \left(\mu\mathbf{x} + \text{prox}_{(\mu+1)f}(\mathbf{x}) \right). \quad \square$$

Combining Theorem 6.63 with Lemma 6.57 leads to the following corollary.

Corollary 6.64. *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper closed and convex function, and let $\lambda, \mu > 0$. Then for any $\mathbf{x} \in \mathbb{E}$,*

$$\text{prox}_{\lambda M_f^\mu}(\mathbf{x}) = \mathbf{x} + \frac{\lambda}{\mu + \lambda} \left(\text{prox}_{(\mu+\lambda)f}(\mathbf{x}) - \mathbf{x} \right).$$

Proof. $\text{prox}_{\lambda M_f^\mu}(\mathbf{x}) = \text{prox}_{M_{\lambda f}^{\mu/\lambda}}(\mathbf{x}) = \mathbf{x} + \frac{\lambda}{\mu + \lambda} \left(\text{prox}_{(\mu+\lambda)f}(\mathbf{x}) - \mathbf{x} \right).$ \square

Example 6.65 (prox of $\frac{\lambda}{2} d_C^2$). Let $C \subseteq \mathbb{E}$ be a nonempty closed and convex set, and let $\lambda > 0$. Consider the function $f = \frac{1}{2} d_C^2$. Then, by Example 6.53, $f = M_g^1$, where $g = \delta_C$. Recall that $\text{prox}_g = P_C$. Therefore, invoking Corollary 6.64, we obtain that for any $\mathbf{x} \in \mathbb{E}$,

$$\text{prox}_{\lambda f}(\mathbf{x}) = \text{prox}_{\lambda M_g^1}(\mathbf{x}) = \mathbf{x} + \frac{\lambda}{\lambda + 1} \left(\text{prox}_{(\lambda+1)g}(\mathbf{x}) - \mathbf{x} \right) = \mathbf{x} + \frac{\lambda}{\lambda + 1} (P_C(\mathbf{x}) - \mathbf{x}).$$

To conclude,

$$\text{prox}_{\frac{\lambda}{2}d_C^2}(\mathbf{x}) = \frac{\lambda}{\lambda+1}P_C(\mathbf{x}) + \frac{1}{\lambda+1}\mathbf{x}.$$

■

Example 6.66 (prox of the Huber function). Consider the function

$$f(\mathbf{x}) = \lambda H_\mu(\mathbf{x}),$$

where H_μ is the Huber function with a smoothing parameter $\mu > 0$ given in (6.43). By Example 6.54, $H_\mu = M_g^\mu$, where $g(\mathbf{x}) = \|\mathbf{x}\|$. Therefore, by Corollary 6.64, it follows that for any $\lambda > 0$ and $\mathbf{x} \in \mathbb{E}$ (recalling the expression for the prox of the Euclidean norm derived in Example 6.19),

$$\begin{aligned} \text{prox}_{\lambda H_\mu}(\mathbf{x}) &= \text{prox}_{\lambda M_g^\mu}(\mathbf{x}) = \mathbf{x} + \frac{\lambda}{\mu+\lambda} (\text{prox}_{(\mu+\lambda)g}(\mathbf{x}) - \mathbf{x}) \\ &= \mathbf{x} + \frac{\lambda}{\mu+\lambda} \left(\left(1 - \frac{\mu+\lambda}{\max\{\|\mathbf{x}\|, \mu+\lambda\}} \right) \mathbf{x} - \mathbf{x} \right), \end{aligned}$$

which, after some algebraic cancellations, reduces to

$$\text{prox}_{\lambda H_\mu}(\mathbf{x}) = \left(1 - \frac{\lambda}{\max\{\|\mathbf{x}\|, \mu+\lambda\}} \right) \mathbf{x}.$$

■

Similarly to the Moreau decomposition formula for the prox operator (Theorem 6.45), we can obtain a decomposition formula for the Moreau envelope function.

Theorem 6.67 (Moreau envelope decomposition). Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a proper closed and convex function, and let $\mu > 0$. Then for any $\mathbf{x} \in \mathbb{E}$,

$$M_f^\mu(\mathbf{x}) + M_{f^*}^{1/\mu}(\mathbf{x}/\mu) = \frac{1}{2\mu} \|\mathbf{x}\|^2.$$

Proof. Recall that for any $\mathbf{x} \in \mathbb{E}$,

$$M_f^\mu(\mathbf{x}) = \min_{\mathbf{u} \in \mathbb{E}} \{f(\mathbf{u}) + \psi(\mathbf{u})\},$$

where $\psi(\mathbf{u}) \equiv \frac{1}{2\mu} \|\mathbf{u} - \mathbf{x}\|^2$. By Fenchel's duality theorem (Theorem 4.15), we have

$$M_f^\mu(\mathbf{x}) = \max_{\mathbf{v} \in \mathbb{E}} \{-f^*(\mathbf{v}) - \psi^*(-\mathbf{v})\} = -\min_{\mathbf{v} \in \mathbb{E}} \{f^*(\mathbf{v}) + \psi^*(-\mathbf{v})\}.$$

Denote $\phi(\cdot) = \frac{1}{2} \|\cdot - \mathbf{x}\|^2$. Then

$$\phi^*(\mathbf{v}) = \frac{1}{2} \|\mathbf{v}\|^2 + \langle \mathbf{x}, \mathbf{v} \rangle.$$

Since $\psi = \frac{1}{\mu}\phi$, it follows by Theorem 4.14 that

$$\psi^*(\mathbf{v}) = \frac{1}{\mu}\phi^*(\mu\mathbf{v}) = \frac{\mu}{2}\|\mathbf{v}\|^2 + \langle \mathbf{x}, \mathbf{v} \rangle.$$

Therefore,

$$M_f^\mu(\mathbf{x}) = -\min_{\mathbf{v} \in \mathbb{E}} \left\{ f^*(\mathbf{v}) + \frac{\mu}{2}\|\mathbf{v}\|^2 - \langle \mathbf{x}, \mathbf{v} \rangle \right\},$$

and hence

$$M_f^\mu(\mathbf{x}) = -\min_{\mathbf{v} \in \mathbb{E}} \left\{ f^*(\mathbf{v}) + \frac{\mu}{2}\|\mathbf{v} - \mathbf{x}/\mu\|^2 - \frac{1}{2\mu}\|\mathbf{x}\|^2 \right\} = \frac{1}{2\mu}\|\mathbf{x}\|^2 - M_{f^*}^{1/\mu}(\mathbf{x}/\mu),$$

establishing the desired result. \square

6.8 Miscellaneous Prox Computations

In this section we gather several examples of prox computations that are not linked to any specific result established in this chapter.

6.8.1 Norm of a Linear Transformation over \mathbb{R}^n

Lemma 6.68. *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}) = \|\mathbf{Ax}\|_2$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is with full row rank, and let $\lambda > 0$. Then*

$$\text{prox}_{\lambda f}(\mathbf{x}) = \begin{cases} \mathbf{x} - \mathbf{A}^T(\mathbf{AA}^T)^{-1}\mathbf{Ax}, & \|(\mathbf{AA}^T)^{-1}\mathbf{Ax}\|_2 \leq \lambda, \\ \mathbf{x} - \mathbf{A}^T(\mathbf{AA}^T + \alpha^*\mathbf{I})^{-1}\mathbf{Ax}, & \|(\mathbf{AA}^T)^{-1}\mathbf{Ax}\|_2 > \lambda, \end{cases}$$

where α^* is the unique positive root of the decreasing function

$$g(\alpha) = \|(\mathbf{AA}^T + \alpha\mathbf{I})^{-1}\mathbf{Ax}\|_2^2 - \lambda^2.$$

Proof. The vector $\text{prox}_{\lambda f}(\mathbf{x})$ is the unique optimal solution to

$$\min_{\mathbf{u} \in \mathbb{R}^n} \left\{ \lambda \|\mathbf{Au}\|_2 + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|_2^2 \right\},$$

which can also be rewritten as

$$\min_{\mathbf{u} \in \mathbb{R}^n, \mathbf{z} \in \mathbb{R}^m} \left\{ \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|_2^2 + \lambda \|\mathbf{z}\|_2 : \mathbf{z} = \mathbf{Au} \right\}. \quad (6.49)$$

To construct a Lagrangian dual problem, we first form the Lagrangian:

$$\begin{aligned} L(\mathbf{u}, \mathbf{z}; \mathbf{y}) &= \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|_2^2 + \lambda \|\mathbf{z}\|_2 + \mathbf{y}^T(\mathbf{z} - \mathbf{Au}) \\ &= \left[\frac{1}{2}\|\mathbf{u} - \mathbf{x}\|_2^2 - (\mathbf{A}^T \mathbf{y})^T \mathbf{u} \right] + [\lambda \|\mathbf{z}\|_2 + \mathbf{y}^T \mathbf{z}]. \end{aligned}$$

Since the Lagrangian is separable w.r.t. \mathbf{u} and \mathbf{z} , the dual objective function can be rewritten as

$$\min_{\mathbf{u}, \mathbf{z}} L(\mathbf{u}, \mathbf{z}; \mathbf{y}) = \min_{\mathbf{u}} \left[\frac{1}{2} \|\mathbf{u} - \mathbf{x}\|_2^2 - (\mathbf{A}^T \mathbf{y})^T \mathbf{u} \right] + \min_{\mathbf{z}} [\lambda \|\mathbf{z}\|_2 + \mathbf{y}^T \mathbf{z}] . \quad (6.50)$$

The minimizer of the minimization problem in \mathbf{u} is $\tilde{\mathbf{u}} = \mathbf{x} + \mathbf{A}^T \mathbf{y}$ with a corresponding optimal value of

$$\begin{aligned} \min_{\mathbf{u}} \left[\frac{1}{2} \|\mathbf{u} - \mathbf{x}\|_2^2 - (\mathbf{A}^T \mathbf{y})^T \mathbf{u} \right] &= \frac{1}{2} \|\tilde{\mathbf{u}} - \mathbf{x}\|_2^2 - (\mathbf{A}^T \mathbf{y})^T \tilde{\mathbf{u}} \\ &= -\frac{1}{2} \mathbf{y}^T \mathbf{A} \mathbf{A}^T \mathbf{y} - (\mathbf{A} \mathbf{x})^T \mathbf{y}. \end{aligned} \quad (6.51)$$

As for the second minimization problem, note that

$$\min_{\mathbf{z}} [\lambda \|\mathbf{z}\|_2 + \mathbf{y}^T \mathbf{z}] = -\max_{\mathbf{z}} [(-\mathbf{y})^T \mathbf{z} - \lambda \|\mathbf{z}\|_2] = -g^*(-\mathbf{y}),$$

where $g(\cdot) = \lambda \|\cdot\|_2$. Since $g^*(\mathbf{w}) = \lambda \delta_{B_{\|\cdot\|_2}[\mathbf{0}, 1]}(\mathbf{w}/\lambda) = \delta_{B_{\|\cdot\|_2}[\mathbf{0}, \lambda]}$ (see Section 4.4.12 and Theorem 4.14), we can conclude that

$$\min_{\mathbf{z}} [\lambda \|\mathbf{z}\|_2 + \mathbf{y}^T \mathbf{z}] = \begin{cases} 0, & \|\mathbf{y}\|_2 \leq \lambda, \\ -\infty, & \|\mathbf{y}\|_2 > \lambda. \end{cases}$$

Combining this with (6.51), we obtain the following dual problem:

$$\max_{\mathbf{y} \in \mathbb{R}^m} \left\{ -\frac{1}{2} \mathbf{y}^T \mathbf{A} \mathbf{A}^T \mathbf{y} - (\mathbf{A} \mathbf{x})^T \mathbf{y} : \|\mathbf{y}\|_2 \leq \lambda \right\}. \quad (6.52)$$

Note that strong duality holds for the primal-dual pair of problems (6.49) and (6.52) (see Theorem A.1). To solve problem (6.52), we will first rewrite it as a minimization problem:

$$\min_{\mathbf{y} \in \mathbb{R}^m} \left\{ \frac{1}{2} \mathbf{y}^T \mathbf{A} \mathbf{A}^T \mathbf{y} + (\mathbf{A} \mathbf{x})^T \mathbf{y} : \|\mathbf{y}\|_2^2 \leq \lambda^2 \right\}. \quad (6.53)$$

So far we have shown that

$$\text{prox}_{\lambda f}(\mathbf{x}) = \mathbf{x} + \mathbf{A}^T \mathbf{y}, \quad (6.54)$$

where \mathbf{y} is an optimal solution of problem (6.53). Since problem (6.53) is convex and satisfies Slater's condition, it follows by the KKT conditions that \mathbf{y} is an optimal solution of (6.53) if and only if there exists α^* (optimal dual variable) for which

$$(\mathbf{A} \mathbf{A}^T + \alpha^* \mathbf{I}) \mathbf{y} + \mathbf{A} \mathbf{x} = \mathbf{0}, \quad (6.55)$$

$$\alpha^* (\|\mathbf{y}\|_2^2 - \lambda^2) = 0, \quad (6.56)$$

$$\|\mathbf{y}\|_2^2 \leq \lambda^2, \quad (6.57)$$

$$\alpha^* \geq 0. \quad (6.58)$$

There are two options. In the first, $\alpha^* = 0$, and then by (6.55),

$$\mathbf{y} = -(\mathbf{A} \mathbf{A}^T)^{-1} \mathbf{A} \mathbf{x}. \quad (6.59)$$

Since (6.56) and (6.58) are automatically satisfied for $\alpha^* = 0$, we can conclude that \mathbf{y} given in (6.59) is the optimal solution of (6.53) if and only if (6.57) is satisfied, meaning if and only if $\|(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{x}\|_2 \leq \lambda$. In this case, by (6.54), $\text{prox}_{\lambda f}(\mathbf{x}) = \mathbf{x} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{x}$.

On the other hand, if $\|(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{x}\|_2 > \lambda$, then $\alpha^* > 0$, and hence by the complementary slackness condition (6.56),

$$\|\mathbf{y}\|_2^2 = \lambda^2. \quad (6.60)$$

By (6.55),

$$\mathbf{y} = -(\mathbf{A}\mathbf{A}^T + \alpha^*\mathbf{I})^{-1}\mathbf{A}\mathbf{x}.$$

Using (6.60), we can conclude that α^* can be uniquely determined as the positive root of the function

$$g(\alpha) = \|(\mathbf{A}\mathbf{A}^T + \alpha\mathbf{I})^{-1}\mathbf{A}\mathbf{x}\|_2^2 - \lambda^2.$$

It is easy to see that g is strictly decreasing for $\alpha \geq 0$, and therefore g has a unique root. \square

6.8.2 Squared l_1 -Norm

The prox of the l_1 -norm has a simple formula. In this section we will show how to compute the prox of the squared l_1 -norm—a task that will prove itself to be much more complicated. We will require the following lemma that expresses $\|\mathbf{x}\|_1^2$ as the optimal value of an optimization problem written in terms of the function

$$\varphi(s, t) = \begin{cases} \frac{s^2}{t}, & t > 0, \\ 0, & s = 0, t = 0, \\ \infty & \text{else.} \end{cases} \quad (6.61)$$

By Example 2.32, φ is closed and convex (even though it is not continuous at $(s, t) = (0, 0)$).

Lemma 6.69 (variational representation of $\|\cdot\|_1^2$). *For any $\mathbf{x} \in \mathbb{R}^n$ the following holds:*

$$\min_{\lambda \in \Delta_n} \sum_{j=1}^n \varphi(x_j, \lambda_j) = \|\mathbf{x}\|_1^2, \quad (6.62)$$

where φ is defined in (6.61). An optimal solution of the minimization problem in (6.62) is given by

$$\tilde{\lambda}_j = \begin{cases} \frac{|x_j|}{\|\mathbf{x}\|_1}, & \mathbf{x} \neq \mathbf{0}, \\ \frac{1}{n}, & \mathbf{x} = \mathbf{0}, \end{cases} \quad j = 1, 2, \dots, n. \quad (6.63)$$

Proof. Since problem (6.62) consists of minimizing a closed and convex function (by Example 2.32) over a compact set, then by the Weierstrass theorem for closed

functions (Theorem 2.12), it possesses an optimal solution, which we denote by $\boldsymbol{\lambda}^* \in \Delta_n$. Define

$$\begin{aligned} I_0 &= \{i \in \{1, 2, \dots, n\} : \lambda_i^* = 0\}, \\ I_1 &= \{i \in \{1, 2, \dots, n\} : \lambda_i^* > 0\}. \end{aligned}$$

By the definitions of I_0 and I_1 , we have

$$\sum_{i \in I_1} \lambda_j^* = \sum_{i=1}^n \lambda_j^* = 1. \quad (6.64)$$

It holds that $x_i = 0$ for any $i \in I_0$, since otherwise we will have that $\varphi(x_i, \lambda_i^*) = \infty$, which is a clear contradiction to the optimality of $\boldsymbol{\lambda}^*$. Therefore, using the Cauchy–Schwarz inequality,

$$\sum_{j=1}^n |x_j| = \sum_{j \in I_1} |x_j| = \sum_{j \in I_1} \frac{|x_j|}{\sqrt{\lambda_j^*}} \sqrt{\lambda_j^*} \leq \sqrt{\sum_{j \in I_1} \frac{x_j^2}{\lambda_j^*}} \cdot \sqrt{\sum_{j \in I_1} \lambda_j^*} \stackrel{(6.64)}{=} \sqrt{\sum_{j \in I_1} \frac{x_j^2}{\lambda_j^*}}.$$

We can thus conclude that

$$\sum_{j=1}^n \varphi(x_j, \lambda_j^*) = \sum_{j \in I_1} \varphi(x_j, \lambda_j^*) = \sum_{j \in I_1} \frac{x_j^2}{\lambda_j^*} \geq \|\mathbf{x}\|_1^2. \quad (6.65)$$

On the other hand, since $\boldsymbol{\lambda}^*$ is an optimal solution of the problem in (6.62),

$$\sum_{j=1}^n \varphi(x_j, \lambda_j^*) \leq \sum_{j=1}^n \varphi(x_j, \tilde{\lambda}_j) = \|\mathbf{x}\|_1^2, \quad (6.66)$$

where $\tilde{\lambda}$ is given by (6.63). Combining (6.65) and (6.66), we finally conclude that the optimal value of the minimization problem in (6.62) is $\|\mathbf{x}\|_1^2$ and that $\tilde{\lambda}$ is an optimal solution. \square

Lemma 6.70 (prox of $\|\cdot\|_1^2$).³⁴ Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}) = \|\mathbf{x}\|_1^2$, and let $\rho > 0$. Then

$$\text{prox}_{\rho f}(\mathbf{x}) = \begin{cases} \left(\frac{\lambda_i x_i}{\lambda_i + 2\rho} \right)_{i=1}^n, & \mathbf{x} \neq \mathbf{0}, \\ \mathbf{0}, & \mathbf{x} = \mathbf{0}, \end{cases}$$

where $\lambda_i = \left[\frac{\sqrt{\rho}|x_i|}{\sqrt{\mu^*}} - 2\rho \right]_+$ with μ^* being any positive root of the nonincreasing function

$$\psi(\mu) = \sum_{i=1}^n \left[\frac{\sqrt{\rho}|x_i|}{\sqrt{\mu}} - 2\rho \right]_+ - 1.$$

³⁴The computation of the prox of the squared l_1 -norm is due to Evgeniou, Pontil, Spinellis, and Nassuphis [54].

Proof. If $\mathbf{x} = \mathbf{0}$, then obviously $\text{prox}_{\rho f}(\mathbf{x}) = \arg\min_{\mathbf{u}} \left\{ \frac{1}{2} \|\mathbf{u}\|_2^2 + \rho \|\mathbf{u}\|_1 \right\} = \mathbf{0}$. Assume that $\mathbf{x} \neq \mathbf{0}$. By Lemma 6.69, $\mathbf{u} = \text{prox}_{\rho f}(\mathbf{x})$ if and only if it is the \mathbf{u} -part of the optimal solution of

$$\min_{\mathbf{u} \in \mathbb{R}^n, \boldsymbol{\lambda} \in \Delta_n} \left\{ \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|_2^2 + \rho \sum_{i=1}^n \varphi(u_i, \lambda_i) \right\},$$

where φ is defined in (6.61). Minimizing first with respect to \mathbf{u} , we obtain that $u_i = \frac{\lambda_i x_i}{\lambda_i + 2\rho}$, and the problem thus reduces to

$$\begin{aligned} & \min_{\boldsymbol{\lambda}} \quad \sum_{i=1}^n \frac{\rho x_i^2}{\lambda_i + 2\rho} \\ & \text{s.t.} \quad \mathbf{e}^T \boldsymbol{\lambda} = 1, \\ & \quad \boldsymbol{\lambda} \geq \mathbf{0}. \end{aligned} \tag{6.67}$$

By Theorem A.1, strong duality holds for problem (6.67) (taking the underlying set as $X = \mathbb{R}_+^n$). Associating a Lagrange multiplier μ to the equality constraint, the Lagrangian is

$$L(\boldsymbol{\lambda}; \mu) = \sum_{i=1}^n \left(\frac{\rho x_i^2}{\lambda_i + 2\rho} + \lambda_i \mu \right) - \mu.$$

By Theorem A.2, $\boldsymbol{\lambda}^*$ is an optimal solution of (6.67) if and only if there exists μ^* for which

$$\boldsymbol{\lambda}^* \in \arg\min_{\boldsymbol{\lambda} \geq \mathbf{0}} L(\boldsymbol{\lambda}; \mu^*), \tag{6.68}$$

$$\mathbf{e}^T \boldsymbol{\lambda}^* = 1. \tag{6.69}$$

Since the minimum in (6.68) is finite and attained, and since $\mathbf{x} \neq \mathbf{0}$, it follows that $\mu^* > 0$ (otherwise, if $\mu^* = 0$, the minimum in (6.68) would not be attained). Exploiting the separability of the Lagrangian, it follows that (6.68) is the same as

$$\lambda_i^* = \left[\frac{\sqrt{\rho}|x_i|}{\sqrt{\mu}} - 2\rho \right]_+.$$

The dual optimal variable μ^* is chosen to satisfy (6.69):

$$\sum_{i=1}^n \left[\frac{\sqrt{\rho}|x_i|}{\sqrt{\mu}} - 2\rho \right]_+ = 1. \quad \square$$

6.8.3 Projection onto the Set of s -Sparse Vectors

Let $s \in \{1, 2, \dots, n\}$ and consider the set

$$C_s = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_0 \leq s\}.$$

The set C_s comprises all *s-sparse vectors*, meaning all vectors with at most s nonzero elements. Obviously C_s is not convex; for example, for $n = 2$, $(0, 1)^T, (1, 0)^T \in C_1$,

but $(0.5, 0.5)^T = 0.5(0, 1)^T + 0.5(1, 0)^T \notin C_1$. The set C_s is closed as a level set of the closed function $\|\cdot\|_0$ (see Example 2.11). Therefore, by Theorem 6.4, $P_{C_s} = \text{prox}_{\delta_{C_s}}$ is nonempty; however, the nonconvexity of C_s implies that $P_{C_s}(\mathbf{x})$ is not necessarily a singleton.

The set $P_{C_s}(\mathbf{x})$ is described in Lemma 6.71 below. The description requires some additional notation. For a vector $\mathbf{x} \in \mathbb{R}^n$ and a set of indices $S \subseteq \{1, 2, \dots, n\}$, the vector \mathbf{x}_S is the subvector of \mathbf{x} that corresponds to the indices in S . For example, for $n = 4$, if $\mathbf{x} = (4, 3, 5, -1)^T$, then $\mathbf{x}_{\{1,4\}} = (4, -1)^T$, $\mathbf{x}_{\{2,3\}} = (3, 5)^T$. For a given indices set $S \subseteq \{1, 2, \dots, n\}$, the matrix \mathbf{U}_S is the submatrix of the identity matrix \mathbf{I}_n comprising the columns corresponding to the indices in S . For example, for $n = 3$,

$$\mathbf{U}_{\{1,3\}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{U}_{\{2\}} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

For a given indices set $S \subseteq \{1, 2, \dots, n\}$, the complement set S^c is given by $S^c = \{1, 2, \dots, n\} \setminus S$.

Finally, we recall our notation (that was also used in Example 6.51) that for a given $\mathbf{x} \in \mathbb{R}^n$, $x_{\langle i \rangle}$ is the i th largest value among $|x_1|, |x_2|, \dots, |x_n|$. Therefore, in particular, $|x_{\langle 1 \rangle}| \geq |x_{\langle 2 \rangle}| \geq \dots \geq |x_{\langle n \rangle}|$. Lemma 6.71 shows that $P_{C_s}(\mathbf{x})$ comprises all vectors consisting of the s components of \mathbf{x} with the largest absolute values and with zeros elsewhere. There may be several choices for the s components with largest absolute values, and this is why $P_{C_s}(\mathbf{x})$ might consist of several vectors. Note that in the statement of the lemma, we characterize the property of an index set S to “comprise s indices corresponding to the s largest absolute values in \mathbf{x} ” by the relation

$$S \subseteq \{1, 2, \dots, n\}, \quad |S| = s, \quad \sum_{i \in S} |x_i| = \sum_{i=1}^s |x_{\langle i \rangle}|.$$

Lemma 6.71 (projection onto C_s). *Let $s \in \{1, 2, \dots, n\}$ and $\mathbf{x} \in \mathbb{R}^n$. Then*

$$P_{C_s}(\mathbf{x}) = \left\{ \mathbf{U}_S \mathbf{x}_S : |S| = s, S \subseteq \{1, 2, \dots, n\}, \sum_{i \in S} |x_i| = \sum_{i=1}^s |x_{\langle i \rangle}| \right\}.$$

Proof. Since C_s consists of all s -sparse vectors, it can be represented as the following union:

$$C_s = \bigcup_{S \subseteq \{1, 2, \dots, n\}, |S|=s} A_S,$$

where $A_S = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}_{S^c} = \mathbf{0}\}$. Therefore,³⁵

$$P_{C_s}(\mathbf{x}) \subseteq \bigcup_{S \subseteq \{1, 2, \dots, n\}, |S|=s} \{P_{A_S}(\mathbf{x})\}. \quad (6.70)$$

³⁵Since A_S is convex, we treat $P_{A_S}(\mathbf{x})$ as a vector and not as a singleton set. The inclusion (6.70) holds since if B_1, B_2, \dots, B_m are closed convex sets, then $P_{\bigcup_{i=1}^m B_i}(\mathbf{x}) \subseteq \bigcup_{i=1}^m \{P_{B_i}(\mathbf{x})\}$ for any \mathbf{x} .

The vectors in $P_{C_s}(\mathbf{x})$ will be the vectors $P_{A_S}(\mathbf{x})$ with the smallest possible value of $\|P_{A_S}(\mathbf{x}) - \mathbf{x}\|^2$. The vector $P_{A_S}(\mathbf{x})$ is the optimal solution of the problem

$$\min_{\mathbf{y} \in \mathbb{R}^n} \left\{ \|\mathbf{y} - \mathbf{x}\|_2^2 : \mathbf{y}_{S^c} = \mathbf{0} \right\},$$

which can be rewritten as

$$\min_{\mathbf{y} \in \mathbb{R}^n} \left\{ \|\mathbf{y}_S - \mathbf{x}_S\|_2^2 + \|\mathbf{x}_{S^c}\|_2^2 : \mathbf{y}_{S^c} = \mathbf{0} \right\}.$$

The optimal solution of the above problem is obviously given by $\mathbf{y}_S = \mathbf{x}_S, \mathbf{y}_{S^c} = \mathbf{0}$, that is, $\mathbf{y} = \mathbf{U}_S \mathbf{x}_S$, and the optimal value is $\|\mathbf{x}_{S^c}\|_2^2$. The vectors in $P_{C_s}(\mathbf{x})$ will therefore be of the form $\mathbf{U}_S \mathbf{x}_S$, with indices sets S with cardinality s and with minimal value $\|\mathbf{x}_{S^c}\|_2^2$, which is equivalent to the condition that S consists of s indices corresponding to the s largest absolute values in \mathbf{x} . \square

Example 6.72. Suppose that $n = 4$. Then

$$P_{C_2}[(2, 3, -2, 1)^T] = \{(2, 3, 0, 0)^T, (0, 3, -2, 0)^T\}. \quad \blacksquare$$

6.9 Summary of Prox Computations

$f(\mathbf{x})$	$\text{dom}(f)$	$\text{prox}_f(\mathbf{x})$	Assumptions	Reference
$\frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} + \mathbf{b}^T \mathbf{x} + c$	\mathbb{R}^n	$(\mathbf{A} + \mathbf{I})^{-1}(\mathbf{x} - \mathbf{b})$	$\mathbf{A} \in \mathbb{S}_+^n, \mathbf{b} \in \mathbb{R}^n, c \in \mathbb{R}$	Section 6.2.3
λx^3	\mathbb{R}_+	$\frac{-1 + \sqrt{1 + 12\lambda} x _+}{6\lambda}$	$\lambda > 0$	Lemma 6.5
μx	$[0, \alpha] \cap \mathbb{R}$	$\min\{\max\{x - \mu, 0\}, \alpha\}$	$\mu \in \mathbb{R}, \alpha \in [0, \infty]$	Example 6.14
$\lambda\ \mathbf{x}\ $	\mathbb{E}	$\left(1 - \frac{\lambda}{\max\{\ \mathbf{x}\ , \lambda\}}\right)\mathbf{x}$	$\ \cdot\ $ —Euclidean norm, $\lambda > 0$	Example 6.19
$-\lambda\ \mathbf{x}\ $	\mathbb{E}	$\begin{cases} \left(1 + \frac{\lambda}{\ \mathbf{x}\ }\right)\mathbf{x}, & \mathbf{x} \neq \mathbf{0}, \\ \{\mathbf{u} : \ \mathbf{u}\ = \lambda\}, & \mathbf{x} = \mathbf{0}. \end{cases}$	$\ \cdot\ $ —Euclidean norm, $\lambda > 0$	Example 6.21
$\lambda\ \mathbf{x}\ _1$	\mathbb{R}^n	$\mathcal{T}_\lambda(\mathbf{x}) = [x - \lambda e]_+ \odot \text{sgn}(\mathbf{x})$	$\lambda > 0$	Example 6.8
$\ \boldsymbol{\omega} \odot \mathbf{x}\ _1$	$\text{Box}[-\boldsymbol{\alpha}, \boldsymbol{\alpha}]$	$\mathcal{S}_{\boldsymbol{\omega}, \boldsymbol{\alpha}}(\mathbf{x})$	$\boldsymbol{\alpha} \in [0, \infty]^n, \boldsymbol{\omega} \in \mathbb{R}_+^n$	Example 6.23
$\lambda\ \mathbf{x}\ _\infty$	\mathbb{R}^n	$\mathbf{x} - \lambda P_{B_{\ \cdot\ _1}[0,1]}(\mathbf{x}/\lambda)$	$\lambda > 0$	Example 6.48
$\lambda\ \mathbf{x}\ _a$	\mathbb{E}	$\mathbf{x} - \lambda P_{B_{\ \cdot\ _a,*}[0,1]}(\mathbf{x}/\lambda)$	$\ \mathbf{x}\ _a$ —arbitrary norm, $\lambda > 0$	Example 6.47
$\lambda\ \mathbf{x}\ _0$	\mathbb{R}^n	$\mathcal{H}_{\sqrt{2\lambda}}(x_1) \times \cdots \times \mathcal{H}_{\sqrt{2\lambda}}(x_n)$	$\lambda > 0$	Example 6.10
$\lambda\ \mathbf{x}\ ^3$	\mathbb{E}	$\frac{2}{1 + \sqrt{1 + 12\lambda\ \mathbf{x}\ }}\mathbf{x}$	$\ \cdot\ $ —Euclidean norm, $\lambda > 0$,	Example 6.20
$-\lambda \sum_{j=1}^n \log x_j$	\mathbb{R}_{++}^n	$\left(\frac{x_j + \sqrt{x_j^2 + 4\lambda}}{2}\right)_{j=1}^n$	$\lambda > 0$	Example 6.9
$\delta_C(\mathbf{x})$	\mathbb{E}	$P_C(\mathbf{x})$	$\emptyset \neq C \subseteq \mathbb{E}$	Theorem 6.24
$\lambda\sigma_C(\mathbf{x})$	\mathbb{E}	$\mathbf{x} - \lambda P_C(\mathbf{x}/\lambda)$	$\lambda > 0, C \neq \emptyset$ closed convex	Theorem 6.46
$\lambda \max\{x_i\}$	\mathbb{R}^n	$\mathbf{x} - \lambda P_{\Delta_n}(\mathbf{x}/\lambda)$	$\lambda > 0$	Example 6.49
$\lambda \sum_{i=1}^k x_{[i]}$	\mathbb{R}^n	$\mathbf{x} - \lambda P_C(\mathbf{x}/\lambda), C = H_{\mathbf{e}, k} \cap \text{Box}[\mathbf{0}, \mathbf{e}]$	$\lambda > 0$	Example 6.50
$\lambda \sum_{i=1}^k x_{\langle i \rangle} $	\mathbb{R}^n	$\mathbf{x} - \lambda P_C(\mathbf{x}/\lambda), C = B_{\ \cdot\ _1}[\mathbf{0}, k] \cap \text{Box}[-\mathbf{e}, \mathbf{e}]$	$\lambda > 0$	Example 6.51
$\lambda M_f^\mu(\mathbf{x})$	\mathbb{E}	$\mathbf{x} + \frac{\lambda}{\mu + \lambda} (\text{prox}_{(\mu + \lambda)f}(\mathbf{x}) - \mathbf{x})$	$\lambda, \mu > 0, f$ proper closed convex	Corollary 6.64
$\lambda d_C(\mathbf{x})$	\mathbb{E}	$\mathbf{x} + \min\left\{\frac{\lambda}{d_C(\mathbf{x})}, 1\right\} (P_C(\mathbf{x}) - \mathbf{x})$	$\emptyset \neq C$ closed convex, $\lambda > 0$	Lemma 6.43
$\frac{\lambda}{2} d_C^2(\mathbf{x})$	\mathbb{E}	$\frac{\lambda}{\lambda + 1} P_C(\mathbf{x}) + \frac{1}{\lambda + 1} \mathbf{x}$	$\emptyset \neq C$ closed convex, $\lambda > 0$	Example 6.65
$\lambda H_\mu(\mathbf{x})$	\mathbb{E}	$\left(1 - \frac{\lambda}{\max\{\ \mathbf{x}\ , \mu + \lambda\}}\right)\mathbf{x}$	$\lambda, \mu > 0$	Example 6.66
$\rho\ \mathbf{x}\ _1^2$	\mathbb{R}^n	$\left(\frac{v_i x_i}{v_i + 2\rho}\right)_{i=1}^n, \mathbf{v} = \left[\sqrt{\frac{\rho}{\mu}} \mathbf{x} - 2\rho\right]_+, \mathbf{e}^T \mathbf{v} = 1 (\mathbf{0}$ when $\mathbf{x} = \mathbf{0})$	$\rho > 0$	Lemma 6.70
$\lambda\ \mathbf{Ax}\ _2$	\mathbb{R}^n	$\mathbf{x} - \mathbf{A}^T (\mathbf{A}\mathbf{A}^T + \alpha^* \mathbf{I})^{-1} \mathbf{Ax}, \alpha^* = 0 \text{ if } \ \mathbf{v}_0\ _2 \leq \lambda; \text{ otherwise, } \ \mathbf{v}_{\alpha^*}\ _2 = \lambda; \mathbf{v}_\alpha \equiv (\mathbf{A}\mathbf{A}^T + \alpha \mathbf{I})^{-1} \mathbf{Ax}$	$\mathbf{A} \in \mathbb{R}^{m \times n}$ with full row rank, $\lambda > 0$	Lemma 6.68

Chapter 7

Spectral Functions

In this chapter we will concentrate on spectral functions, which are functions of matrices that depend only on their singular values or on their eigenvalues. The underlying spaces in this chapter are all Euclidean. We start by defining the notion of *symmetry* w.r.t. a given set of orthogonal matrices.

7.1 Symmetric Functions

7.1.1 Definition and Examples

Definition 7.1 (symmetric functions). Let $\mathcal{A} \subseteq \mathbb{O}^n$ be a set of orthogonal matrices. A proper function $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ is called **symmetric** w.r.t. \mathcal{A} if

$$f(\mathbf{Ax}) = f(\mathbf{x}) \text{ for all } \mathbf{x} \in \mathbb{R}^n, \mathbf{A} \in \mathcal{A}.$$

The following are five types of symmetric functions, each one dictated by the choice of orthogonal matrices in \mathcal{A} .

Example 7.2 (even functions). If $\mathcal{A} = \{-\mathbf{I}\}$, then $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ is symmetric w.r.t. \mathcal{A} if

$$f(\mathbf{x}) = f(-\mathbf{x}) \text{ for all } \mathbf{x} \in \mathbb{R}^n.$$

Such functions will be called *even functions*. ■

Example 7.3 (absolutely symmetric functions). Take $\mathcal{A} = \{\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_n\} \subseteq \mathbb{R}^{n \times n}$, where \mathbf{D}_i is the diagonal matrix whose diagonal elements are all ones except for the (i, i) component which is equal to -1 . Then a proper function $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ is symmetric w.r.t. \mathcal{A} if and only if

$$f(\mathbf{x}) = f(|\mathbf{x}|) \text{ for all } \mathbf{x} \in \mathbb{R}^n.$$

We will call such a function an *absolutely symmetric function*. It is easy to show that f is absolutely symmetric if and only if there exists a function $g : \mathbb{R}_+^n \rightarrow (-\infty, \infty]$ such that $f(\mathbf{x}) = g(|\mathbf{x}|)$ for all $\mathbf{x} \in \mathbb{R}^n$. ■

Example 7.4 (norm-dependent functions). A proper function $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ is symmetric w.r.t. $\mathcal{A} = \mathbb{O}^n$ if and only if

$$f(\mathbf{x}) = f(\mathbf{U}\mathbf{x}) \text{ for all } \mathbf{x} \in \mathbb{R}^n, \mathbf{U} \in \mathbb{O}^n.$$

The above holds if and only if there exists a proper function $g : \mathbb{R} \rightarrow (-\infty, \infty]$ such that

$$f(\mathbf{x}) = g(\|\mathbf{x}\|_2) \text{ for all } \mathbf{x} \in \mathbb{R}^n.$$

A function satisfying the above is called a *norm-dependent function*. ■

We will require some additional notation before describing the next two examples. For a given vector $\mathbf{x} \in \mathbb{R}^n$, the vector \mathbf{x}^\downarrow is the vector \mathbf{x} reordered nonincreasingly. For example, if $\mathbf{x} = (2, -9, 2, 10)^T$, then $\mathbf{x}^\downarrow = (10, 2, 2, -9)^T$.

Definition 7.5 (permutation matrices). An $n \times n$ matrix is called a **permutation matrix** if all its components are either 0 or 1 and each row and each column has exactly one nonzero element. The set of all $n \times n$ permutation matrices is denoted by Λ_n .

Definition 7.6 (generalized permutation matrices). An $n \times n$ matrix is called a **generalized permutation matrix** if all its components are either 0, 1, or -1 and each row and each column has exactly one nonzero element. The set of all $n \times n$ generalized permutation matrices is denoted by Λ_n^G .

Thus, for example,

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \in \Lambda_3, \quad \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \in \Lambda_3^G.$$

By the definition of permutation and generalized permutation matrices, it is easy to see that for any $\mathbf{x} \in \mathbb{R}^n$ there exists a permutation matrix $\mathbf{P} \in \Lambda_n$ for which $\mathbf{Px} = \mathbf{x}^\downarrow$ and a generalized permutation matrix $\mathbf{Q} \in \Lambda_n^G$ for which $\mathbf{Qx} = |\mathbf{x}|^\downarrow$. It can be readily verified that permutation matrices, as well as generalized permutation matrices, are orthogonal.

Example 7.7 (permutation symmetric functions). A proper function $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ is symmetric w.r.t. Λ_n if and only if

$$f(\mathbf{x}) = f(\mathbf{Px}) \text{ for all } \mathbf{x} \in \mathbb{R}^n, \mathbf{P} \in \Lambda_n.$$

Such a function will be called a *permutation symmetric function*. It is easy to show that f is permutation symmetric if and only if

$$f(\mathbf{x}) = f(\mathbf{x}^\downarrow) \text{ for all } \mathbf{x} \in \mathbb{R}^n. \quad ■$$

Example 7.8 (absolutely permutation symmetric functions). A proper function $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ is symmetric w.r.t. Λ_n^G if and only if

$$f(\mathbf{x}) = f(\mathbf{Px}) \text{ for all } \mathbf{x} \in \mathbb{R}^n, \mathbf{P} \in \Lambda_n^G.$$

Such a function will be called an *absolutely permutation symmetric function*. It is easy to show that f is absolutely permutation symmetric if and only if

$$f(\mathbf{x}) = f(|\mathbf{x}|^\downarrow) \text{ for all } \mathbf{x} \in \mathbb{R}^n. \quad \blacksquare$$

7.1.2 The Symmetric Conjugate Theorem

We will now show that the conjugate of a symmetric function w.r.t. a set of orthogonal matrices is always symmetric w.r.t. the same set of matrices.

Theorem 7.9 (symmetric conjugate theorem).³⁶ Let $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ be a proper function which is symmetric w.r.t. a set of orthogonal matrices $\mathcal{A} \subseteq \mathbb{O}^n$. Then f^* is symmetric w.r.t. \mathcal{A} .

Proof. Let $\mathbf{A} \in \mathcal{A}$. Then by the symmetry assumption, $h = f$, where $h(\mathbf{x}) \equiv f(\mathbf{Ax})$. Thus,

$$f^*(\mathbf{y}) = h^*(\mathbf{y}) \text{ for all } \mathbf{y} \in \mathbb{R}^n. \quad (7.1)$$

By Theorem 4.13 and the orthogonality of \mathbf{A} , for any $\mathbf{y} \in \mathbb{R}^n$,

$$h^*(\mathbf{y}) = f^*((\mathbf{A}^T)^{-1}\mathbf{y}) = f^*(\mathbf{Ay}),$$

which, combined with (7.1), yields

$$f^*(\mathbf{y}) = f^*(\mathbf{Ay}) \text{ for all } \mathbf{y} \in \mathbb{R}^n.$$

Since the above holds for any $\mathbf{A} \in \mathcal{A}$, it follows that f^* is symmetric w.r.t. \mathcal{A} . \square

Example 7.10. In this example we will illustrate the symmetric conjugate theorem by verifying that the types of symmetries satisfied by the functions in the table of Section 4.4.16 also hold for their conjugates.

- even functions

$f(\mathbf{x})$	$\text{dom}(f)$	$f^*(\mathbf{y})$	Assumptions	Reference
$\frac{1}{p} \mathbf{x} ^p$	\mathbb{R}	$\frac{1}{q} \mathbf{y} ^q$	$p > 1, \frac{1}{p} + \frac{1}{q} = 1$	Section 4.4.4
$\frac{1}{2}\mathbf{x}^T \mathbf{Ax} + c$	\mathbb{R}^n	$\frac{1}{2}\mathbf{y}^T \mathbf{A}^{-1}\mathbf{y} - c$	$\mathbf{A} \in \mathbb{S}_{++}^n, c \in \mathbb{R}$	Section 4.4.6

- permutation symmetric functions

$f(\mathbf{x})$	$\text{dom}(f)$	$f^*(\mathbf{y})$	Reference
$\sum_{i=1}^n x_i \log x_i$	\mathbb{R}_+^n	$\sum_{i=1}^n e^{y_i-1}$	Section 4.4.8
$\sum_{i=1}^n x_i \log x_i$	Δ_n	$\log(\sum_{i=1}^n e^{y_i})$	Section 4.4.10
$\log(\sum_{i=1}^n e^{x_i})$	\mathbb{R}^n	$\sum_{i=1}^n y_i \log y_i$ ($\text{dom}(f^*) = \Delta_n$)	Section 4.4.11
$\max_i\{x_i\}$	\mathbb{R}^n	$\delta_{\Delta_n}(\mathbf{y})$	Example 4.10

³⁶The symmetric conjugate theorem (Theorem 7.9) is from Rockafellar [108, Corollary 12.3.1].

- **absolutely permutation symmetric functions**

$f(\mathbf{x})$	$\text{dom}(f)$	$f^*(\mathbf{y})$	Reference
$\ \mathbf{x}\ _p$	\mathbb{R}^n	$\delta_{B_{\ \cdot\ _q}[\mathbf{0}, 1]}(\mathbf{y})$	Section 4.4.12
$\frac{1}{2}\ \mathbf{x}\ _p^2$	\mathbb{E}	$\frac{1}{2}\ \mathbf{y}\ _q^2$	Section 4.4.15

where $p, q \in [1, \infty]$, $\frac{1}{p} + \frac{1}{q} = 1$ (with the convention that if $p = 1, \infty$, then $q = \infty, 1$, respectively).

- **norm-dependent functions**

f	$\text{dom}(f)$	f^*	Reference
$\ \mathbf{x}\ _2$	\mathbb{R}^n	$\delta_{B_{\ \cdot\ _2}[\mathbf{0}, 1]}(\mathbf{y})$	Section 4.4.12
$-\sqrt{\alpha^2 - \ \mathbf{x}\ _2^2}$ ($\alpha > 0$)	$B_{\ \cdot\ _2}[\mathbf{0}, \alpha]$	$\alpha\sqrt{\ \mathbf{y}\ _2^2 + 1}$	Section 4.4.13
$\sqrt{\alpha^2 + \ \mathbf{x}\ _2^2}$ ($\alpha > 0$)	\mathbb{R}^n	$-\alpha\sqrt{1 - \ \mathbf{y}\ _2^2}$ ($\text{dom } f^* = B_{\ \cdot\ _2}[\mathbf{0}, 1]$)	Section 4.4.14
$\frac{1}{2}\ \mathbf{x}\ _2^2$	\mathbb{R}^n	$\frac{1}{2}\ \mathbf{y}\ _2^2$	Section 4.4.15

■

7.2 Symmetric Spectral Functions over \mathbb{S}^n ³⁷

The main concern of this chapter are functions of matrices that are defined on either the set of symmetric matrices \mathbb{S}^n or the set of matrices $\mathbb{R}^{m \times n}$. We will deal only with functions that depend either on the eigenvalues of their argument (if the underlying space is \mathbb{S}^n) or on the singular values (if the underlying space is $\mathbb{R}^{m \times n}$). Such functions are called *spectral functions*. We first consider functions over \mathbb{S}^n . Given a matrix $\mathbf{X} \in \mathbb{S}^n$, its eigenvalues ordered nonincreasingly are denoted by

$$\lambda_1(\mathbf{X}) \geq \lambda_2(\mathbf{X}) \geq \dots \geq \lambda_n(\mathbf{X}).$$

The eigenvalues function $\boldsymbol{\lambda} : \mathbb{S}^n \rightarrow \mathbb{R}^n$ is defined as

$$\boldsymbol{\lambda}(\mathbf{X}) \equiv (\lambda_1(\mathbf{X}), \lambda_2(\mathbf{X}), \dots, \lambda_n(\mathbf{X}))^T.$$

A key fact from linear algebra is that any symmetric matrix $\mathbf{X} \in \mathbb{S}^n$ has a *spectral decomposition*, meaning an orthogonal matrix $\mathbf{U} \in \mathbb{O}^n$ for which $\mathbf{X} = \mathbf{U}\text{diag}(\boldsymbol{\lambda}(\mathbf{X}))\mathbf{U}^T$. We begin by formally defining the notion of spectral functions over \mathbb{S}^n .

Definition 7.11 (spectral functions over \mathbb{S}^n). A proper function $g : \mathbb{S}^n \rightarrow (-\infty, \infty]$ is called a **spectral function over \mathbb{S}^n** if there exists a proper function $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ for which $g = f \circ \boldsymbol{\lambda}$.

³⁷Sections 7.2 and 7.3, excluding the spectral proximal theorem, are based on the seminal papers of Lewis [80, 81] on unitarily invariant functions. The spectral proximal formulas can be found in Parikh and Boyd [102].

If $g = f \circ \boldsymbol{\lambda}$, we will refer to f (which is actually not necessarily unique) as the *associated function*. Our main interest will be to study spectral functions whose associated functions are permutation symmetric.

Definition 7.12 (symmetric spectral functions over \mathbb{S}^n). A proper function $f : \mathbb{S}^n \rightarrow (-\infty, \infty]$ is called a **symmetric spectral function over \mathbb{S}^n** if there exists a proper permutation symmetric function $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ for which $g = f \circ \boldsymbol{\lambda}$.

Example 7.13. Following is a list of permutation symmetric functions along with their associated symmetric spectral functions.

#	$f(\mathbf{x})$	$\text{dom}(f)$	$f(\boldsymbol{\lambda}(\mathbf{X}))$	$\text{dom}(f \circ \boldsymbol{\lambda})$
1	$\sum_{i=1}^n x_i$	\mathbb{R}^n	$\text{Tr}(\mathbf{X})$	\mathbb{S}^n
2	$\max_{i=1,2,\dots,n} x_i$	\mathbb{R}^n	$\lambda_{\max}(\mathbf{X})$	\mathbb{S}^n
3	$\alpha \ \mathbf{x}\ _2$ ($\alpha \in \mathbb{R}$)	\mathbb{R}^n	$\alpha \ \mathbf{X}\ _F$	\mathbb{S}^n
4	$\alpha \ \mathbf{x}\ _2^2$ ($\alpha \in \mathbb{R}$)	\mathbb{R}^n	$\alpha \ \mathbf{X}\ _F^2$	\mathbb{S}^n
5	$\alpha \ \mathbf{x}\ _\infty$ ($\alpha \in \mathbb{R}$)	\mathbb{R}^n	$\alpha \ \mathbf{X}\ _{2,2}$	\mathbb{S}^n
6	$\alpha \ \mathbf{x}\ _1$ ($\alpha \in \mathbb{R}$)	\mathbb{R}^n	$\alpha \ \mathbf{X}\ _{S_1}$	\mathbb{S}^n
7	$-\sum_{i=1}^n \log(x_i)$	\mathbb{R}_{++}^n	$-\log \det(\mathbf{X})$	\mathbb{S}_{++}^n
8	$\sum_{i=1}^n x_i \log(x_i)$	\mathbb{R}_+^n	$\sum_{i=1}^n \lambda_i(\mathbf{X}) \log(\lambda_i(\mathbf{X}))$	\mathbb{S}_+^n
9	$\sum_{i=1}^n x_i \log(x_i)$	Δ_n	$\sum_{i=1}^n \lambda_i(\mathbf{X}) \log(\lambda_i(\mathbf{X}))$	Υ_n

The domain of the last function in the above table is the *spectahedron* set given by

$$\Upsilon_n = \{\mathbf{X} \in \mathbb{S}_+^n : \text{Tr}(\mathbf{X}) = 1\}.$$

The norm used in the sixth function is the Schatten 1-norm whose expression for symmetric matrices is given by

$$\|\mathbf{X}\|_{S_1} = \sum_{i=1}^n |\lambda_i(\mathbf{X})|, \quad \mathbf{X} \in \mathbb{S}^n.$$

Schatten p -norms will be discussed in detail in Section 7.3 ■

A fundamental inequality that will be a key argument in establishing the main results of this section is the so-called Fan inequality stating that the inner product of two symmetric matrices is upper bounded by the inner product of their eigenvalues vectors and that equality holds if and only if the two matrices are simultaneously orthogonally diagonalizable.

Theorem 7.14 (Fan's Inequality [32, 119]). For any two symmetric matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{S}^n$ it holds that

$$\text{Tr}(\mathbf{XY}) \leq \langle \boldsymbol{\lambda}(\mathbf{X}), \boldsymbol{\lambda}(\mathbf{Y}) \rangle,$$

and equality holds if and only if there exists $\mathbf{V} \in \mathbb{O}^n$ for which

$$\begin{aligned}\mathbf{X} &= \mathbf{V}\text{diag}(\lambda(\mathbf{X}))\mathbf{V}^T, \\ \mathbf{Y} &= \mathbf{V}\text{diag}(\lambda(\mathbf{Y}))\mathbf{V}^T.\end{aligned}$$

7.2.1 The Spectral Conjugate Formula

A rather direct result of Fan's inequality is the spectral conjugate formula that shows how to compute the conjugate of a symmetric spectral function over \mathbb{S}^n in terms of the conjugate of its associated function.

Theorem 7.15 (spectral conjugate formula over \mathbb{S}^n). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a permutation symmetric function. Then*

$$(f \circ \boldsymbol{\lambda})^* = f^* \circ \boldsymbol{\lambda}.$$

Proof. Let $\mathbf{Y} \in \mathbb{S}^n$. Then

$$\begin{aligned}(f \circ \boldsymbol{\lambda})^*(\mathbf{Y}) &= \max_{\mathbf{X} \in \mathbb{S}^n} \{\text{Tr}(\mathbf{XY}) - f(\boldsymbol{\lambda}(\mathbf{X}))\} \\ &\leq \max_{\mathbf{X} \in \mathbb{S}^n} \{\langle \boldsymbol{\lambda}(\mathbf{X}), \boldsymbol{\lambda}(\mathbf{Y}) \rangle - f(\boldsymbol{\lambda}(\mathbf{X}))\} \\ &\leq \max_{\mathbf{x} \in \mathbb{R}^n} \{\langle \mathbf{x}, \boldsymbol{\lambda}(\mathbf{Y}) \rangle - f(\mathbf{x})\} \\ &= (f^* \circ \boldsymbol{\lambda})(\mathbf{Y}),\end{aligned}$$

where Fan's inequality (Theorem 7.14) was used in the first inequality. To show the reverse inequality, take a spectral decomposition of \mathbf{Y} :

$$\mathbf{Y} = \mathbf{U}\text{diag}(\boldsymbol{\lambda}(\mathbf{Y}))\mathbf{U}^T \quad (\mathbf{U} \in \mathbb{O}^n).$$

Then

$$\begin{aligned}(f^* \circ \boldsymbol{\lambda})(\mathbf{Y}) &= \max_{\mathbf{x} \in \mathbb{R}^n} \{\langle \mathbf{x}, \boldsymbol{\lambda}(\mathbf{Y}) \rangle - f(\mathbf{x})\} \\ &= \max_{\mathbf{x} \in \mathbb{R}^n} \{\text{Tr}(\text{diag}(\mathbf{x})\text{diag}(\boldsymbol{\lambda}(\mathbf{Y}))) - f(\mathbf{x})\} \\ &= \max_{\mathbf{x} \in \mathbb{R}^n} \{\text{Tr}(\text{diag}(\mathbf{x})\mathbf{U}^T \mathbf{Y} \mathbf{U}) - f(\mathbf{x}^\downarrow)\} \\ &= \max_{\mathbf{x} \in \mathbb{R}^n} \{\text{Tr}(\text{diag}(\mathbf{x})\mathbf{U}^T \mathbf{Y} \mathbf{U}) - f(\boldsymbol{\lambda}(\mathbf{U}\text{diag}(\mathbf{x})\mathbf{U}^T))\} \\ &= \max_{\mathbf{x} \in \mathbb{R}^n} \{\text{Tr}(\mathbf{U}\text{diag}(\mathbf{x})\mathbf{U}^T \mathbf{Y}) - f(\boldsymbol{\lambda}(\mathbf{U}\text{diag}(\mathbf{x})\mathbf{U}^T))\} \\ &\leq \max_{\mathbf{Z} \in \mathbb{S}^n} \{\text{Tr}(\mathbf{ZY}) - f(\boldsymbol{\lambda}(\mathbf{Z}))\} \\ &= (f \circ \boldsymbol{\lambda})^*(\mathbf{Y}). \quad \square\end{aligned}$$

Example 7.16. Using the spectral conjugate formula, we can compute the conjugates of the functions from the table of Example 7.13. The conjugates appear in the following table, which also includes references to the corresponding results for functions over \mathbb{R}^n . The numbering is the same as in the table of Example 7.13.

#	$g(\mathbf{X})$	$\text{dom}(g)$	$g^*(\mathbf{Y})$	$\text{dom}(g^*)$	Reference
1	$\text{Tr}(\mathbf{X})$	\mathbb{S}^n	$\delta_{\{\mathbf{I}\}}(\mathbf{Y})$	$\{\mathbf{I}\}$	Section 4.4.7
2	$\lambda_{\max}(\mathbf{X})$	\mathbb{S}^n	$\delta_{\Upsilon_n}(\mathbf{Y})$	Υ_n	Example 4.10
3	$\alpha \ \mathbf{X}\ _F (\alpha > 0)$	\mathbb{S}^n	$\delta_{B_{\ \cdot\ _F}[\mathbf{0}, \alpha]}(\mathbf{Y})$	$B_{\ \cdot\ _F}[\mathbf{0}, \alpha]$	Section 4.4.12
4	$\alpha \ \mathbf{X}\ _F^2 (\alpha > 0)$	\mathbb{S}^n	$\frac{1}{4\alpha} \ \mathbf{Y}\ _F^2$	\mathbb{S}^n	Section 4.4.6
5	$\alpha \ \mathbf{X}\ _{2,2} (\alpha > 0)$	\mathbb{S}^n	$\delta_{B_{\ \cdot\ _{S_1}}[\mathbf{0}, \alpha]}(\mathbf{Y})$	$B_{\ \cdot\ _{S_1}}[\mathbf{0}, \alpha]$	Section 4.4.12
6	$\alpha \ \mathbf{X}\ _{S_1} (\alpha > 0)$	\mathbb{S}^n	$\delta_{B_{\ \cdot\ _{2,2}}[\mathbf{0}, \alpha]}(\mathbf{Y})$	$B_{\ \cdot\ _{2,2}}[\mathbf{0}, \alpha]$	Section 4.4.12
7	$-\log \det(\mathbf{X})$	\mathbb{S}_{++}^n	$-n - \log \det(-\mathbf{Y})$	\mathbb{S}_{--}^n	Section 4.4.9
8	$\sum_{i=1}^n \lambda_i(\mathbf{X}) \log(\lambda_i(\mathbf{X}))$	\mathbb{S}_+^n	$\sum_{i=1}^n e^{\lambda_i(\mathbf{Y})-1}$	\mathbb{S}^n	Section 4.4.8
9	$\sum_{i=1}^n \lambda_i(\mathbf{X}) \log(\lambda_i(\mathbf{X}))$	Υ_n	$\log(\sum_{i=1}^n e^{\lambda_i(\mathbf{Y})})$	\mathbb{S}^n	Section 4.4.10

■

The spectral conjugate formula has several important consequences, one of which is the following theorem stating that a symmetric spectral function is closed and convex if and only if its associated function is closed and convex.

Theorem 7.17 (closedness and convexity of symmetric spectral functions over \mathbb{S}^n). *Let $F : \mathbb{S}^n \rightarrow (-\infty, \infty]$ be given by $F = f \circ \boldsymbol{\lambda}$, where $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ is a permutation symmetric proper function. Then F is closed and convex if and only if f is closed and convex.*

Proof. By the spectral conjugate formula (Theorem 7.15),

$$F^* = (f \circ \boldsymbol{\lambda})^* = f^* \circ \boldsymbol{\lambda}.$$

Since by the symmetric conjugate theorem (Theorem 7.9) f^* is permutation symmetric, we can invoke once again the spectral conjugate formula to obtain

$$F^{**} = (f^* \circ \boldsymbol{\lambda})^* = f^{**} \circ \boldsymbol{\lambda}. \quad (7.2)$$

If f is closed and convex, then by Theorem 4.8 (taking also into account the properness of f), it follows that $f^{**} = f$. Therefore, by (7.2),

$$F^{**} = f \circ \boldsymbol{\lambda} = F.$$

Thus, since F is a conjugate of another function (F^*), it follows by Theorem 4.3 that it is closed and convex. Now assume that F is closed and convex. Since F is in addition proper, it follows by Theorem 4.8 that $F^{**} = F$, which, combined with (7.2), yields the equality

$$f \circ \boldsymbol{\lambda} = F = F^{**} = f^{**} \circ \boldsymbol{\lambda}.$$

Therefore, for any $\mathbf{x} \in \mathbb{R}^n$

$$f(\mathbf{x}^\downarrow) = f(\boldsymbol{\lambda}(\text{diag}(\mathbf{x}))) = f^{**}(\boldsymbol{\lambda}(\text{diag}(\mathbf{x}))) = f^{**}(\mathbf{x}^\downarrow).$$

By the permutation symmetry property of both f and f^{**} , it follows that $f(\mathbf{x}^\downarrow) = f(\mathbf{x})$ and $f^{**}(\mathbf{x}^\downarrow) = f^{**}(\mathbf{x})$, and we thus obtained that $f(\mathbf{x}) = f^{**}(\mathbf{x})$ for any

$\mathbf{x} \in \mathbb{R}^n$, meaning that $f = f^{**}$. Therefore, f , as a conjugate of another function (f^*) is closed and convex. \square

7.2.2 The Proximal Operator of Symmetric Spectral Functions over \mathbb{S}^n

The next result shows a simple formula for computing the prox operator of a symmetric spectral function over \mathbb{S}^n which is also proper closed and convex. The prox is expressed in terms of the spectral decomposition of the argument and the prox operator of the associated function.

Theorem 7.18 (spectral prox formula over \mathbb{S}^n). *Let $F : \mathbb{S}^n \rightarrow (-\infty, \infty]$ be given by $F = f \circ \boldsymbol{\lambda}$, where $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ is a permutation symmetric proper closed and convex function. Let $\mathbf{X} \in \mathbb{S}^n$, and suppose that $\mathbf{X} = \mathbf{U}\text{diag}(\boldsymbol{\lambda}(\mathbf{X}))\mathbf{U}^T$, where $\mathbf{U} \in \mathbb{O}^n$. Then*

$$\text{prox}_F(\mathbf{X}) = \mathbf{U}\text{diag}(\text{prox}_f(\boldsymbol{\lambda}(\mathbf{X})))\mathbf{U}^T.$$

Proof. Recall that

$$\text{prox}_F(\mathbf{X}) = \underset{\mathbf{Z} \in \mathbb{S}^n}{\text{argmin}} \left\{ F(\mathbf{Z}) + \frac{1}{2} \|\mathbf{Z} - \mathbf{X}\|_F^2 \right\}. \quad (7.3)$$

Denoting $\mathbf{D} = \text{diag}(\boldsymbol{\lambda}(\mathbf{X}))$, we note that for any $\mathbf{Z} \in \mathbb{S}^n$,

$$F(\mathbf{Z}) + \frac{1}{2} \|\mathbf{Z} - \mathbf{X}\|_F^2 = F(\mathbf{Z}) + \frac{1}{2} \|\mathbf{Z} - \mathbf{U}\mathbf{D}\mathbf{U}^T\|_F^2 \stackrel{(*)}{=} F(\mathbf{U}^T\mathbf{Z}\mathbf{U}) + \frac{1}{2} \|\mathbf{U}^T\mathbf{Z}\mathbf{U} - \mathbf{D}\|_F^2,$$

where the transition $(*)$ is due to the fact that $F(\mathbf{Z}) = f(\boldsymbol{\lambda}(\mathbf{Z})) = f(\boldsymbol{\lambda}(\mathbf{U}^T\mathbf{Z}\mathbf{U})) = F(\mathbf{U}^T\mathbf{Z}\mathbf{U})$. Making the change of variables $\mathbf{W} = \mathbf{U}^T\mathbf{Z}\mathbf{U}$, we conclude that the optimal solution of (7.3) is given by

$$\mathbf{Z} = \mathbf{U}\mathbf{W}^*\mathbf{U}^T, \quad (7.4)$$

where $\mathbf{W}^* \in \mathbb{S}^n$ is the unique optimal solution of

$$\underset{\mathbf{W} \in \mathbb{S}^n}{\min} \left\{ G(\mathbf{W}) \equiv F(\mathbf{W}) + \frac{1}{2} \|\mathbf{W} - \mathbf{D}\|_F^2 \right\}. \quad (7.5)$$

We will prove that \mathbf{W}^* is diagonal. Let $i \in \{1, 2, \dots, n\}$. Take \mathbf{V}_i to be the diagonal matrix whose diagonal elements are all ones except for the (i, i) th component, which is -1 . Define $\widetilde{\mathbf{W}}_i = \mathbf{V}_i \mathbf{W}^* \mathbf{V}_i^T$. Obviously, by the fact that $\mathbf{V}_i \in \mathbb{O}^n$,

$$F(\mathbf{V}_i \mathbf{W}^* \mathbf{V}_i^T) = f(\boldsymbol{\lambda}(\mathbf{V}_i \mathbf{W}^* \mathbf{V}_i^T)) = f(\boldsymbol{\lambda}(\mathbf{W}^*)) = F(\mathbf{W}^*),$$

and we thus obtain

$$\begin{aligned} G(\widetilde{\mathbf{W}}_i) &= F(\widetilde{\mathbf{W}}_i) + \frac{1}{2} \|\widetilde{\mathbf{W}}_i - \mathbf{D}\|_F^2 \\ &= F(\mathbf{V}_i \mathbf{W}^* \mathbf{V}_i^T) + \frac{1}{2} \|\mathbf{V}_i \mathbf{W}^* \mathbf{V}_i^T - \mathbf{D}\|_F^2 \\ &= F(\mathbf{W}^*) + \frac{1}{2} \|\mathbf{W}^* - \mathbf{V}_i^T \mathbf{D} \mathbf{V}_i\|_F^2 \\ &\stackrel{(**)}{=} F(\mathbf{W}^*) + \frac{1}{2} \|\mathbf{W}^* - \mathbf{D}\|_F^2, \\ &= G(\mathbf{W}^*), \end{aligned}$$

where $(**)$ follows from the fact that \mathbf{V}_i and \mathbf{D} are both diagonal, and hence $\mathbf{V}_i^T \mathbf{D} \mathbf{V}_i = \mathbf{V}_i^T \mathbf{V}_i \mathbf{D} = \mathbf{D}$. We conclude that $\widetilde{\mathbf{W}}_i$ is also an optimal solution, but by the uniqueness of the optimal solution of problem (7.5), it follows that $\mathbf{W}^* = \mathbf{V}_i \mathbf{W}^* \mathbf{V}_i^T$. Comparing the i th rows of the two matrices, we deduce that $W_{ij}^* = 0$ for any $j \neq i$. Since this argument is valid for any $i \in \{1, 2, \dots, n\}$, it follows that \mathbf{W}^* is a diagonal matrix, and consequently the optimal solution of (7.5) is given by $\mathbf{W}^* = \text{diag}(\mathbf{w}^*)$, where \mathbf{w}^* is the optimal solution of

$$\min_{\mathbf{w}} \left\{ F(\text{diag}(\mathbf{w})) + \frac{1}{2} \|\text{diag}(\mathbf{w}) - \mathbf{D}\|_F^2 \right\}.$$

Since $F(\text{diag}(\mathbf{w})) = f(\mathbf{w}^\perp) = f(\mathbf{w})$ and $\|\text{diag}(\mathbf{w}) - \mathbf{D}\|_F^2 = \|\mathbf{w} - \boldsymbol{\lambda}(\mathbf{X})\|_2^2$, it follows that \mathbf{w}^* is given by

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmin}} \left\{ f(\mathbf{w}) + \frac{1}{2} \|\mathbf{w} - \boldsymbol{\lambda}(\mathbf{X})\|_2^2 \right\} = \text{prox}_f(\boldsymbol{\lambda}(\mathbf{X})).$$

Therefore, $\mathbf{W}^* = \text{diag}(\text{prox}_f(\boldsymbol{\lambda}(\mathbf{X})))$, which, along with (7.4), establishes the desired result. \square

Example 7.19. Using the spectral prox formula, we can compute the prox of symmetric spectral functions in terms of the prox of their associated functions. Using this observation, we present in the table below expressions of prox operators of several functions. The parameter α is always assumed to be positive, and \mathbf{U} is assumed to be an orthogonal matrix satisfying $\mathbf{X} = \mathbf{U} \text{diag}(\boldsymbol{\lambda}(\mathbf{X})) \mathbf{U}^T$. The table also includes references to the corresponding results for the associated functions, which are always defined over \mathbb{R}^n .

$F(\mathbf{X})$	$\text{dom}(F)$	$\text{prox}_F(\mathbf{X})$	Reference
$\alpha \ \mathbf{X}\ _F^2$	\mathbb{S}^n	$\frac{1}{1+2\alpha} \mathbf{X}$	Section 6.2.3
$\alpha \ \mathbf{X}\ _F$	\mathbb{S}^n	$\left(1 - \frac{\alpha}{\max\{\ \mathbf{X}\ _F, \alpha\}}\right) \mathbf{X}$	Example 6.19
$\alpha \ \mathbf{X}\ _{S_1}$	\mathbb{S}^n	$\mathbf{U} \text{diag}(\mathcal{T}_\alpha(\boldsymbol{\lambda}(\mathbf{X}))) \mathbf{U}^T$	Example 6.8
$\alpha \ \mathbf{X}\ _{2,2}$	\mathbb{S}^n	$\mathbf{U} \text{diag}(\boldsymbol{\lambda}(\mathbf{X}) - \alpha P_{B_{\ \cdot\ _1}[\mathbf{0}, \mathbf{1}]}(\boldsymbol{\lambda}(\mathbf{X})/\alpha)) \mathbf{U}^T$	Example 6.48
$-\alpha \log \det(\mathbf{X})$	\mathbb{S}_{++}^n	$\mathbf{U} \text{diag}\left(\frac{\lambda_j(\mathbf{X}) + \sqrt{\lambda_j(\mathbf{X})^2 + 4\alpha}}{2}\right) \mathbf{U}^T$	Example 6.9
$\alpha \lambda_1(\mathbf{X})$	\mathbb{S}^n	$\mathbf{U} \text{diag}(\boldsymbol{\lambda}(\mathbf{X}) - \alpha P_{\Delta_n}(\boldsymbol{\lambda}(\mathbf{X})/\alpha)) \mathbf{U}^T$	Example 6.49
$\alpha \sum_{i=1}^k \lambda_i(\mathbf{X})$	\mathbb{S}^n	$\mathbf{X} - \alpha \mathbf{U} \text{diag}(P_C(\boldsymbol{\lambda}(\mathbf{X})/\alpha)) \mathbf{U}^T, C = H_{\mathbf{e}, k} \cap \text{Box}[\mathbf{0}, \mathbf{e}]$	Example 6.50

\blacksquare

A set $T \subseteq \mathbb{S}^n$ is called a *symmetric spectral set* in \mathbb{S}^n if the indicator function δ_T is a symmetric spectral function over \mathbb{S}^n , meaning that it has the form $\delta_T = \delta_C \circ \boldsymbol{\lambda}$, where δ_C is a permutation symmetric function. The set $C \subseteq \mathbb{R}^n$ is the *associated set*. Since $\text{prox}_{\delta_T} = P_T$ and $\text{prox}_{\delta_C} = P_C$, it follows by the spectral prox formula

that if C is nonempty closed and convex, then

$$P_T(\mathbf{X}) = \mathbf{U}\text{diag}(P_C(\boldsymbol{\lambda}(\mathbf{X})))\mathbf{U}^T, \quad \mathbf{X} \in \mathbb{S}^n, \quad (7.6)$$

where \mathbf{U} is an orthogonal matrix satisfying $\mathbf{X} = \mathbf{U}\text{diag}(\boldsymbol{\lambda}(\mathbf{X}))\mathbf{U}^T$.

Example 7.20. Using formula (7.6), we present in the following table expressions for the orthogonal projection onto several symmetric spectral sets in \mathbb{S}^n . The table also includes references to the corresponding results on orthogonal projections onto the associated subsets of \mathbb{R}^n . The matrix \mathbf{U} is assumed to be an orthogonal matrix satisfying $\mathbf{X} = \mathbf{U}\text{diag}(\boldsymbol{\lambda}(\mathbf{X}))\mathbf{U}^T$.

set (T)	$P_T(\mathbf{X})$	Assumptions	Reference
\mathbb{S}_+^n	$\mathbf{U}\text{diag}([\boldsymbol{\lambda}(\mathbf{X})]_+)\mathbf{U}^T$	—	Lemma 6.26
$\{\mathbf{X} : \ell\mathbf{I} \preceq \mathbf{X} \preceq u\mathbf{I}\}$	$\mathbf{U}\text{diag}(\mathbf{v})\mathbf{U}^T, v_i = \min\{\max\{\lambda_i(\mathbf{X}), \ell\}, u\}$	$\ell \leq u$	Lemma 6.26
$B_{\ \cdot\ _F}[\mathbf{0}, r]$	$\frac{r}{\max\{\ \mathbf{X}\ _F, r\}}\mathbf{X}$	$r > 0$	Lemma 6.26
$\{\mathbf{X} : \text{Tr}(\mathbf{X}) \leq b\}$	$\mathbf{U}\text{diag}(\mathbf{v})\mathbf{U}^T, \mathbf{v} = \boldsymbol{\lambda}(\mathbf{X}) - \frac{[\mathbf{e}^T \boldsymbol{\lambda}(\mathbf{X}) - b]\mathbf{e}}{n}$	$b \in \mathbb{R}$	Lemma 6.26
Υ_n	$\mathbf{U}\text{diag}(\mathbf{v})\mathbf{U}^T, \mathbf{v} = [\boldsymbol{\lambda}(\mathbf{X}) - \mu^* \mathbf{e}]_+$ where $\mu^* \in \mathbb{R}$ satisfies $\mathbf{e}^T [\boldsymbol{\lambda}(\mathbf{X}) - \mu^* \mathbf{e}]_+ = 1$	—	Corollary 6.29
$B_{\ \cdot\ _{S_1}}[\mathbf{0}, \alpha]$	$\begin{cases} \mathbf{X}, & \ \mathbf{X}\ _{S_1} \leq \alpha, \\ \mathbf{U}\text{diag}(\mathcal{T}_{\beta^*}(\boldsymbol{\lambda}(\mathbf{X})))\mathbf{U}^T, & \ \mathbf{X}\ _{S_1} > \alpha, \\ \ \mathcal{T}_{\beta^*}(\boldsymbol{\lambda}(\mathbf{X}))\ _1 = \alpha, \beta^* > 0 & \end{cases}$	$\alpha > 0$	Example 6.33

■

7.3 Symmetric Spectral Functions over $\mathbb{R}^{m \times n}$

Let m, n be two positive integers and $r = \min\{m, n\}$. We will denote by $\sigma : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^r$ the singular values function that assigns to each matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ the vector of singular values $(\sigma_1(\mathbf{X}), \sigma_2(\mathbf{X}), \dots, \sigma_r(\mathbf{X}))^T$, where $\sigma_1(\mathbf{X}) \geq \sigma_2(\mathbf{X}) \geq \dots \geq \sigma_r(\mathbf{X}) \geq 0$. We will also require the following notation. For a vector $\mathbf{v} \in \mathbb{R}^r$, the matrix $\text{dg}(\mathbf{v})$ is the $m \times n$ matrix defined by

$$\text{dg}(\mathbf{v})_{i,j} = \begin{cases} v_i, & i = j, \\ 0 & \text{else.} \end{cases}$$

The operator $\text{dg}(\cdot)$ maps r -dimensional vectors to generalized³⁸ $m \times n$ diagonal matrices. The integers m and n (and hence also r) will be fixed throughout this section, and hence there is no need to indicate their values in the operator dg . We do not use the “diag” notation since it is reserved to square diagonal matrices.

It is well known (see Golub and Van Loan [60, Theorem 2.5.2]) that any matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ has a *singular value decomposition*, meaning matrices $\mathbf{U} \in \mathbb{O}^m, \mathbf{V} \in \mathbb{O}^n$ for which $\mathbf{X} = \mathbf{U}\text{dg}(\sigma(\mathbf{X}))\mathbf{V}^T$.

³⁸A matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ is a *generalized diagonal matrix* if $X_{ij} = 0$ for any $i \neq j$.

The analysis in this section uses very similar arguments to those used in the previous section; however, for the sake of completeness we will provide the results with their complete proofs.

We begin by formally defining the notion of spectral functions over $\mathbb{R}^{m \times n}$.

Definition 7.21 (spectral functions over $\mathbb{R}^{m \times n}$). A proper function $g : \mathbb{R}^{m \times n} \rightarrow (-\infty, \infty]$ is called a **spectral function over $\mathbb{R}^{m \times n}$** if there exists a proper function $f : \mathbb{R}^r \rightarrow (-\infty, \infty]$ for which $g = f \circ \sigma$.

Similarly to the notation in Section 7.2, if $g = f \circ \sigma$, we will refer to f (which is actually not necessarily unique) as the *associated function*. Our main interest will be with spectral functions whose associated functions are absolutely permutation symmetric.

Definition 7.22 (symmetric spectral functions over $\mathbb{R}^{m \times n}$). A proper function $f : \mathbb{R}^{m \times n} \rightarrow (-\infty, \infty]$ is called a **symmetric spectral function over $\mathbb{R}^{m \times n}$** if there exists a proper absolutely permutation symmetric function $f : \mathbb{R}^r \rightarrow (-\infty, \infty]$ for which $g = f \circ \sigma$.

Example 7.23 (Schatten p -norms). Let $p \in [1, \infty]$. Then the *Schatten p -norm* is the norm defined by

$$\|\mathbf{X}\|_{S_p} \equiv \|\sigma(\mathbf{X})\|_p, \quad \mathbf{X} \in \mathbb{R}^{m \times n}.$$

It is well known³⁹ that $\|\cdot\|_{S_p}$ is indeed a norm for any $p \in [1, \infty]$. Specific examples are the following:

- **trace-norm** (Schatten 1-norm)—also called the **nuclear norm**:

$$\|\mathbf{X}\|_{S_1} = \sum_{i=1}^r \sigma_i(\mathbf{X}).$$

- **spectral norm** (Schatten ∞ -norm):

$$\|\mathbf{X}\|_{S_\infty} = \sigma_1(\mathbf{X}) = \|\mathbf{X}\|_{2,2}.$$

- **Frobenius norm** (Schatten 2-norm):

$$\|\mathbf{X}\|_{S_2} = \sqrt{\sum_{i=1}^r \sigma_i(\mathbf{X})^2} = \sqrt{\text{Tr}(\mathbf{X}^T \mathbf{X})}.$$

The Schatten p -norm is a symmetric spectral function over $\mathbb{R}^{m \times n}$ whose associated function is the l_p -norm on \mathbb{R}^r , which is obviously an absolutely permutation symmetric function. ■

³⁹See, for example, Horn and Johnson [70].

Example 7.24 (Ky Fan k -norms). Recall the notation from Example 6.51—given a vector $\mathbf{x} \in \mathbb{R}^r$, $x_{\langle i \rangle}$ is the component of \mathbf{x} with the i th largest absolute value, meaning in particular that

$$|x_{\langle 1 \rangle}| \geq |x_{\langle 2 \rangle}| \geq \cdots \geq |x_{\langle r \rangle}|.$$

The function $f_k(\mathbf{x}) = \sum_{i=1}^k |x_{\langle i \rangle}|$ is an absolutely permutation symmetric function. The corresponding symmetric spectral function is the so-called *Ky Fan k -norm* given by

$$\|\mathbf{X}\|_{\langle k \rangle} = f_k(\sigma(\mathbf{X})) = \sum_{i=1}^k \sigma_i(\mathbf{X}).$$

Obviously, $\|\cdot\|_{\langle 1 \rangle}$ is the spectral norm, which is also the Schatten ∞ -norm; the norm $\|\cdot\|_{\langle r \rangle}$ is the trace-norm, which is also the Schatten 1-norm. ■

A key inequality that is used in the analysis of spectral functions over $\mathbb{R}^{m \times n}$ is an inequality bounding the inner product of two matrices via the inner product of their singular vectors. The inequality, which is credited to von Neumann and is in a sense the “ $\mathbb{R}^{m \times n}$ -counterpart” of Fan’s inequality (Theorem 7.14).

Theorem 7.25 (von Neumann’s trace inequality [123]). *For any two matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{m \times n}$, the inequality*

$$\langle \mathbf{X}, \mathbf{Y} \rangle \leq \langle \sigma(\mathbf{X}), \sigma(\mathbf{Y}) \rangle$$

holds. Equality holds if and only if there exists a simultaneous nonincreasing singular value decomposition of \mathbf{X}, \mathbf{Y} , meaning that there exist $\mathbf{U} \in \mathbb{O}^m$ and $\mathbf{V} \in \mathbb{O}^n$ for which

$$\begin{aligned} \mathbf{X} &= \mathbf{U} \text{dg}(\sigma(\mathbf{X})) \mathbf{V}^T, \\ \mathbf{Y} &= \mathbf{U} \text{dg}(\sigma(\mathbf{Y})) \mathbf{V}^T. \end{aligned}$$

7.3.1 The Spectral Conjugate Formula

A direct result of von Neumann’s trace inequality is the spectral conjugate formula over $\mathbb{R}^{m \times n}$.

Theorem 7.26 (spectral conjugate formula over $\mathbb{R}^{m \times n}$). *Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be an absolutely permutation symmetric function. Then*

$$(f \circ \sigma)^* = f^* \circ \sigma.$$

Proof. Let $\mathbf{Y} \in \mathbb{R}^{m \times n}$. Then

$$\begin{aligned} (f \circ \sigma)^*(\mathbf{Y}) &= \max_{\mathbf{X} \in \mathbb{R}^{m \times n}} \{\text{Tr}(\mathbf{XY}) - f(\sigma(\mathbf{X}))\} \\ &\leq \max_{\mathbf{X} \in \mathbb{R}^{m \times n}} \{\langle \sigma(\mathbf{X}), \sigma(\mathbf{Y}) \rangle - f(\sigma(\mathbf{X}))\} \\ &\leq \max_{\mathbf{x} \in \mathbb{R}^r} \{\langle \mathbf{x}, \sigma(\mathbf{Y}) \rangle - f(\mathbf{x})\} \\ &= (f^* \circ \sigma)(\mathbf{Y}), \end{aligned}$$

where Von Neumann's trace inequality (Theorem 7.25) was used in the first inequality. To show the reverse inequality, take a singular value decomposition of \mathbf{Y} :

$$\mathbf{Y} = \mathbf{U}\text{dg}(\boldsymbol{\sigma}(\mathbf{Y}))\mathbf{V}^T \quad (\mathbf{U} \in \mathbb{O}^m, \mathbf{V} \in \mathbb{O}^n).$$

Then

$$\begin{aligned} (f^* \circ \boldsymbol{\sigma})(\mathbf{Y}) &= \max_{\mathbf{x} \in \mathbb{R}^r} \{ \langle \mathbf{x}, \boldsymbol{\sigma}(\mathbf{Y}) \rangle - f(\mathbf{x}) \} \\ &= \max_{\mathbf{x} \in \mathbb{R}^r} \{ \text{Tr}(\text{dg}(\mathbf{x})^T \text{dg}(\boldsymbol{\sigma}(\mathbf{Y}))) - f(\mathbf{x}) \} \\ &= \max_{\mathbf{x} \in \mathbb{R}^r} \{ \text{Tr}(\text{dg}(\mathbf{x})^T \mathbf{U}^T \mathbf{Y} \mathbf{V}) - f(\mathbf{x}^\perp) \} \\ &= \max_{\mathbf{x} \in \mathbb{R}^r} \{ \text{Tr}(\text{dg}(\mathbf{x})^T \mathbf{U}^T \mathbf{Y} \mathbf{V}) - f(\boldsymbol{\sigma}(\mathbf{U}\text{dg}(\mathbf{x})\mathbf{V}^T)) \} \\ &= \max_{\mathbf{x} \in \mathbb{R}^r} \{ \text{Tr}(\mathbf{V}\text{dg}(\mathbf{x})^T \mathbf{U}^T \mathbf{Y}) - f(\boldsymbol{\sigma}(\mathbf{U}\text{dg}(\mathbf{x})\mathbf{V}^T)) \} \\ &\leq \max_{\mathbf{Z} \in \mathbb{R}^{m \times n}} \{ \text{Tr}(\mathbf{Z}^T \mathbf{Y}) - f(\boldsymbol{\sigma}(\mathbf{Z})) \} \\ &= (f \circ \boldsymbol{\sigma})^*(\mathbf{Y}). \quad \square \end{aligned}$$

Example 7.27. Using the spectral conjugate formula over $\mathbb{R}^{m \times n}$, we present below expressions for the conjugate functions of several symmetric spectral functions over $\mathbb{R}^{m \times n}$ (all with full domain). The table also includes the references to the corresponding results on functions over \mathbb{R}^r . The constant α is assumed to be positive.

$g(\mathbf{X})$	$\text{dom}(g)$	$g^*(\mathbf{Y})$	$\text{dom}(g^*)$	Reference
$\alpha\sigma_1(\mathbf{X}) \ (\alpha > 0)$	$\mathbb{R}^{m \times n}$	$\delta_{B_{\ \cdot\ _{S_1}}[0, \alpha]}(\mathbf{Y})$	$B_{\ \cdot\ _{S_1}}[\mathbf{0}, \alpha]$	Section 4.4.12
$\alpha\ \mathbf{X}\ _F \ (\alpha > 0)$	$\mathbb{R}^{m \times n}$	$\delta_{B_{\ \cdot\ _F}[0, \alpha]}(\mathbf{Y})$	$B_{\ \cdot\ _F}[\mathbf{0}, \alpha]$	Section 4.4.12
$\alpha\ \mathbf{X}\ _F^2 \ (\alpha > 0)$	$\mathbb{R}^{m \times n}$	$\frac{1}{4\alpha}\ \mathbf{Y}\ _F^2$	$\mathbb{R}^{m \times n}$	Section 4.4.6
$\alpha\ \mathbf{X}\ _{S_1} \ (\alpha > 0)$	$\mathbb{R}^{m \times n}$	$\delta_{B_{\ \cdot\ _{S_\infty}}[0, \alpha]}(\mathbf{Y})$	$B_{\ \cdot\ _{S_\infty}}[\mathbf{0}, \alpha]$	Section 4.4.12

■

The spectral conjugate formula can be used to show that a symmetric spectral function over $\mathbb{R}^{m \times n}$ is closed and convex if and only if its associated function is closed and convex.

Theorem 7.28 (closedness and convexity of symmetric spectral functions over $\mathbb{R}^{m \times n}$). Let $F : \mathbb{R}^{m \times n} \rightarrow (-\infty, \infty]$ be given by $F = f \circ \boldsymbol{\sigma}$, where $f : \mathbb{R}^r \rightarrow (-\infty, \infty]$ is an absolutely permutation symmetric proper function. Then F is closed and convex if and only if f is closed and convex.

Proof. By the spectral conjugate formula (Theorem 7.26),

$$F^* = (f \circ \boldsymbol{\sigma})^* = f^* \circ \boldsymbol{\sigma}.$$

Since by the symmetric conjugate theorem (Theorem 7.9) f^* is absolutely permutation symmetric, we can invoke once again the spectral conjugate formula to obtain

$$F^{**} = (f^* \circ \boldsymbol{\sigma})^* = f^{**} \circ \boldsymbol{\sigma}. \quad (7.7)$$

If f is closed and convex, then by Theorem 4.8 (taking also in account the properness of f) it follows that $f^{**} = f$. Therefore, by (7.7),

$$F^{**} = f \circ \sigma = F.$$

Thus, since F is a conjugate of another function (F^*), it follows by Theorem 4.3 that it is closed and convex. Now assume that F is closed and convex. Since F is in addition proper, it follows by Theorem 4.8 that $F^{**} = F$, which, combined with (7.7), yields the equality

$$f \circ \sigma = F = F^{**} = f^{**} \circ \sigma.$$

Therefore, for any $\mathbf{x} \in \mathbb{R}^r$,

$$f(|\mathbf{x}|^\downarrow) = f(\sigma(\text{dg}(\mathbf{x}))) = f^{**}(\sigma(\text{dg}(\mathbf{x}))) = f^{**}(|\mathbf{x}|^\downarrow).$$

By the absolutely permutation symmetry property of both f and f^{**} , it follows that $f(|\mathbf{x}|^\downarrow) = f(\mathbf{x})$ and $f^{**}(|\mathbf{x}|^\downarrow) = f^{**}(\mathbf{x})$, and therefore $f(\mathbf{x}) = f^{**}(\mathbf{x})$ for any $\mathbf{x} \in \mathbb{R}^r$, meaning that $f = f^{**}$. Therefore, f , as a conjugate of another function (f^*), is closed and convex. \square

7.3.2 The Proximal Operator of Symmetric Spectral Functions over $\mathbb{R}^{m \times n}$

The next result shows a simple formula for computing the prox operator of a symmetric spectral function over $\mathbb{R}^{m \times n}$, which is also proper closed and convex. The prox is expressed in terms of the singular value decomposition of the argument and the prox operator of the associated function.

Theorem 7.29 (spectral prox formula over $\mathbb{R}^{m \times n}$). *Let $F : \mathbb{R}^{m \times n} \rightarrow (-\infty, \infty]$ be given by $F = f \circ \sigma$, where $f : \mathbb{R}^r \rightarrow (-\infty, \infty]$ is an absolutely permutation symmetric proper closed and convex function. Let $\mathbf{X} \in \mathbb{R}^{m \times n}$, and suppose that $\mathbf{X} = \mathbf{U}\text{dg}(\sigma(\mathbf{X}))\mathbf{V}^T$, where $\mathbf{U} \in \mathbb{O}^m$, $\mathbf{V} \in \mathbb{O}^n$. Then*

$$\text{prox}_F(\mathbf{X}) = \mathbf{U}\text{dg}(\text{prox}_f(\sigma(\mathbf{X})))\mathbf{V}^T.$$

Proof. Recall that

$$\text{prox}_F(\mathbf{X}) = \underset{\mathbf{Z} \in \mathbb{R}^{m \times n}}{\text{argmin}} \left\{ F(\mathbf{Z}) + \frac{1}{2} \|\mathbf{Z} - \mathbf{X}\|_F^2 \right\}. \quad (7.8)$$

Denoting $\mathbf{D} = \text{dg}(\sigma(\mathbf{X}))$, we note that for any $\mathbf{Z} \in \mathbb{R}^{m \times n}$,

$$F(\mathbf{Z}) + \frac{1}{2} \|\mathbf{Z} - \mathbf{X}\|_F^2 = F(\mathbf{Z}) + \frac{1}{2} \|\mathbf{Z} - \mathbf{UDV}^T\|_F^2 \stackrel{(*)}{=} F(\mathbf{U}^T \mathbf{Z} \mathbf{V}) + \frac{1}{2} \|\mathbf{U}^T \mathbf{Z} \mathbf{V} - \mathbf{D}\|_F^2,$$

where the transition $(*)$ is due to the fact that $F(\mathbf{Z}) = f(\sigma(\mathbf{Z})) = f(\sigma(\mathbf{U}^T \mathbf{Z} \mathbf{V})) = F(\mathbf{U}^T \mathbf{Z} \mathbf{V})$. Making the change of variables $\mathbf{W} = \mathbf{U}^T \mathbf{Z} \mathbf{V}$, we conclude that the unique optimal solution of (7.8) is given by

$$\mathbf{Z} = \mathbf{U}\mathbf{W}^*\mathbf{V}^T, \quad (7.9)$$

where \mathbf{W}^* is the unique optimal solution of

$$\min_{\mathbf{W} \in \mathbb{R}^{m \times n}} \left\{ G(\mathbf{W}) \equiv F(\mathbf{W}) + \frac{1}{2} \|\mathbf{W} - \mathbf{D}\|_F^2 \right\}. \quad (7.10)$$

We will prove that \mathbf{W}^* is a generalized diagonal matrix (meaning that all off-diagonal components are zeros). Let $i \in \{1, 2, \dots, r\}$. Take $\Sigma_i^{(1)} \in \mathbb{R}^{m \times m}$ and $\Sigma_i^{(2)} \in \mathbb{R}^{n \times n}$ to be the $m \times m$ and $n \times n$ diagonal matrices whose diagonal elements are all ones except for the (i, i) th component, which is -1 . Define $\widetilde{\mathbf{W}}_i = \Sigma_i^{(1)} \mathbf{W}^* \Sigma_i^{(2)}$. Obviously, by the fact that $\Sigma_i^{(1)} \in \mathbb{O}^m$, $\Sigma_i^{(2)} \in \mathbb{O}^n$,

$$F(\Sigma_i^{(1)} \mathbf{W}^* \Sigma_i^{(2)}) = f(\sigma(\Sigma_i^{(1)} \mathbf{W}^* \Sigma_i^{(2)})) = f(\sigma(\mathbf{W}^*)) = F(\mathbf{W}^*),$$

and we thus obtain

$$\begin{aligned} G(\widetilde{\mathbf{W}}_i) &= F(\widetilde{\mathbf{W}}_i) + \frac{1}{2} \|\widetilde{\mathbf{W}}_i - \mathbf{D}\|_F^2 \\ &= F(\Sigma_i^{(1)} \mathbf{W}^* \Sigma_i^{(2)}) + \frac{1}{2} \|\Sigma_i^{(1)} \mathbf{W}^* \Sigma_i^{(2)} - \mathbf{D}\|_F^2 \\ &= F(\mathbf{W}^*) + \frac{1}{2} \|\mathbf{W}^* - \Sigma_i^{(1)} \mathbf{D} \Sigma_i^{(2)}\|_F^2 \\ &= F(\mathbf{W}^*) + \frac{1}{2} \|\mathbf{W}^* - \mathbf{D}\|_F^2, \\ &= G(\mathbf{W}^*). \end{aligned}$$

Consequently, $\widetilde{\mathbf{W}}_i$ is also an optimal solution of (7.10), but by the uniqueness of the optimal solution of problem (7.10), we conclude that $\mathbf{W}^* = \Sigma_i^{(1)} \mathbf{W}^* \Sigma_i^{(2)}$. Comparing the i th rows and columns of the two matrices we obtain that $W_{ij}^* = 0$ and $W_{ji}^* = 0$ for any $j \neq i$. Since this argument is valid for any $i \in \{1, 2, \dots, r\}$, it follows that \mathbf{W}^* is a generalized diagonal matrix, and consequently the optimal solution of (7.10) is given by $\mathbf{W}^* = \text{dg}(\mathbf{w}^*)$, where \mathbf{w}^* is the optimal solution of

$$\min_{\mathbf{w}} \left\{ F(\text{dg}(\mathbf{w})) + \frac{1}{2} \|\text{dg}(\mathbf{w}) - \mathbf{D}\|_F^2 \right\}.$$

Since $F(\text{dg}(\mathbf{w})) = f(|\mathbf{w}|^\downarrow) = f(\mathbf{w})$ and $\|\text{dg}(\mathbf{w}) - \mathbf{D}\|_F^2 = \|\mathbf{w} - \sigma(\mathbf{X})\|_2^2$, it follows that \mathbf{w}^* is given by

$$\mathbf{w}^* = \underset{\mathbf{w}}{\text{argmin}} \left\{ f(\mathbf{w}) + \frac{1}{2} \|\mathbf{w} - \sigma(\mathbf{X})\|_2^2 \right\} = \text{prox}_f(\sigma(\mathbf{X})).$$

Therefore, $\mathbf{W}^* = \text{dg}(\text{prox}_f(\sigma(\mathbf{X})))$, which, along with (7.9), establishes the desired result. \square

Example 7.30. Using the spectral prox formula over $\mathbb{R}^{m \times n}$, we can compute the prox of symmetric spectral functions in terms of the prox of their associated functions. Using this observation, we present in the table below expressions of prox operators of several functions. The parameter α is always assumed to be positive, and $\mathbf{U} \in \mathbb{O}^m$, $\mathbf{V} \in \mathbb{O}^n$ are assumed to satisfy $\mathbf{X} = \mathbf{U} \text{dg}(\sigma(\mathbf{X})) \mathbf{V}^T$. The table also includes a reference to the corresponding results for the associated functions, which are always defined over \mathbb{R}^r .

$F(\mathbf{X})$	$\text{prox}_F(\mathbf{X})$	Reference
$\alpha \ \mathbf{X}\ _F^2$	$\frac{1}{1+2\alpha} \mathbf{X}$	Section 6.2.3
$\alpha \ \mathbf{X}\ _F$	$\left(1 - \frac{\alpha}{\max\{\ \mathbf{X}\ _F, \alpha\}}\right) \mathbf{X}$	Example 6.19
$\alpha \ \mathbf{X}\ _{S_1}$	$\mathbf{Udg}(\mathcal{T}_\alpha(\sigma(\mathbf{X}))) \mathbf{V}^T$	Example 6.8
$\alpha \ \mathbf{X}\ _{S_\infty}$	$\mathbf{X} - \alpha \mathbf{Udg}(P_{B_{\ \cdot\ _1}[\mathbf{0}, 1]}(\sigma(\mathbf{X})/\alpha)) \mathbf{V}^T$	Example 6.48
$\alpha \ \mathbf{X}\ _{\langle k \rangle}$	$\mathbf{X} - \alpha \mathbf{Udg}(P_C(\sigma(\mathbf{X})/\alpha)) \mathbf{V}^T,$ $C = B_{\ \cdot\ _1}[\mathbf{0}, k] \cap B_{\ \cdot\ _\infty}[\mathbf{0}, 1]$	Example 6.51

Note that $\|\mathbf{X}\|_{S_\infty}$ can be written as either $\sigma_1(\mathbf{X})$ or $\|\mathbf{X}\|_{2,2}$. ■

A set $T \subseteq \mathbb{R}^{m \times n}$ is called a *symmetric spectral set* in $\mathbb{R}^{m \times n}$ if the indicator function δ_T is a symmetric spectral function over $\mathbb{R}^{m \times n}$, meaning that it has the form $\delta_T = \delta_C \circ \sigma$, where δ_C is an absolutely permutation symmetric function. The set $C \subseteq \mathbb{R}^{m \times n}$ is the *associated set*. Since $\text{prox}_{\delta_T} = P_T$ and $\text{prox}_{\delta_C} = P_C$, it follows by the spectral prox formula that if C is nonempty closed and convex, then

$$P_T(\mathbf{X}) = \mathbf{Udg}(P_C(\sigma(\mathbf{X}))) \mathbf{V}^T, \quad \mathbf{X} \in \mathbb{R}^{m \times n}, \quad (7.11)$$

where $\mathbf{U} \in \mathbb{O}^m$, $\mathbf{V} \in \mathbb{O}^n$ are assumed to satisfy $\mathbf{X} = \mathbf{Udg}(\sigma(\mathbf{X})) \mathbf{V}^T$.

Example 7.31. Using formula (7.11), we present in the following table expressions for the orthogonal projection onto several symmetric spectral sets in $\mathbb{R}^{m \times n}$. The table also includes references to the corresponding results on the orthogonal projection onto the associated subset of \mathbb{R}^r . The matrices $\mathbf{U} \in \mathbb{O}^m$, $\mathbf{V} \in \mathbb{O}^n$ are assumed to satisfy $\mathbf{X} = \mathbf{Udg}(\sigma(\mathbf{X})) \mathbf{V}^T$.

set (T)	$P_T(\mathbf{X})$	Assumptions	Reference
$B_{\ \cdot\ _{S_\infty}}[\mathbf{0}, \alpha]$	$\mathbf{Udg}(\mathbf{v}) \mathbf{V}^T, v_i = \min\{\sigma_i(\mathbf{X}), \alpha\}$	$\alpha > 0$	Lemma 6.26
$B_{\ \cdot\ _F}[\mathbf{0}, r]$	$\frac{r}{\max\{\ \mathbf{X}\ _F, r\}} \mathbf{X}$	$r > 0$	Lemma 6.26
$B_{\ \cdot\ _{S_1}}[\mathbf{0}, \alpha]$	$\begin{cases} \mathbf{X}, & \ \mathbf{X}\ _{S_1} \leq \alpha, \\ \mathbf{Udg}(\mathcal{T}_{\beta^*}(\sigma(\mathbf{X}))) \mathbf{V}^T, & \ \mathbf{X}\ _{S_1} > \alpha, \\ \ \mathcal{T}_{\beta^*}(\sigma(\mathbf{X}))\ _1 = \alpha, \beta^* > 0 \end{cases}$	$\alpha > 0$	Example 6.33

■

Chapter 8

Primal and Dual Projected Subgradient Methods

Underlying Space: In this chapter \mathbb{E} is a Euclidean space, meaning a finite dimensional space endowed with an inner product $\langle \cdot, \cdot \rangle$ and the Euclidean norm $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$.

8.1 From Gradient Descent to Subgradient Descent

8.1.1 Descent Directions?

Consider the unconstrained problem

$$(P) \quad \min\{f(\mathbf{x}) : \mathbf{x} \in \mathbb{E}\}.$$

If f is differentiable over \mathbb{E} , then a well-known method for solving problem (P) is the *gradient method*, also known as *steepest descent*, which takes the form

$$\mathbf{x}^{k+1} = \mathbf{x}^k - t_k \nabla f(\mathbf{x}^k), \quad (8.1)$$

where t_k is an appropriately chosen stepsize. A key property of the direction of the negative of the gradient is that it is a *descent direction*, a notion that is now recalled.

Definition 8.1 (descent direction). Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be an extended real-valued function, and let $\mathbf{x} \in \text{int}(\text{dom}(f))$. A vector $\mathbf{0} \neq \mathbf{d} \in \mathbb{E}$ is called a **descent direction** of f at \mathbf{x} if the directional derivative $f'(\mathbf{x}; \mathbf{d})$ exists and is negative.

An important property of descent directions, which can be directly deduced from their definition, is that taking small enough steps along these directions leads to a decrease in function value.

Lemma 8.2 (descent property of descent directions [10, Lemma 4.2]). Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be an extended real-valued function. Let $\mathbf{x} \in \text{int}(\text{dom}(f))$, and assume that $\mathbf{0} \neq \mathbf{d} \in \mathbb{E}$ is a descent direction of f at \mathbf{x} . Then there exists $\varepsilon > 0$ such that $\mathbf{x} + t\mathbf{d} \in \text{dom}(f)$ and

$$f(\mathbf{x} + t\mathbf{d}) < f(\mathbf{x})$$

for any $t \in (0, \varepsilon]$.

Coming back to the gradient method, we note that the directional derivative of f at \mathbf{x}^k in the direction of $-\nabla f(\mathbf{x}^k)$ is negative as long as $\nabla f(\mathbf{x}^k) \neq \mathbf{0}$:

$$f'(\mathbf{x}^k; -\nabla f(\mathbf{x}^k)) = \langle \nabla f(\mathbf{x}^k), -\nabla f(\mathbf{x}^k) \rangle = -\|\nabla f(\mathbf{x}^k)\|^2 < 0, \quad (8.2)$$

where Theorem 3.29 was used in the first equality. We have thus shown that $-\nabla f(\mathbf{x}^k)$ is a *descent direction* of f at \mathbf{x}^k , which by Lemma 8.2 implies that there exists $\varepsilon > 0$ such that $f(\mathbf{x}^k - t\nabla f(\mathbf{x}^k)) < f(\mathbf{x}^k)$ for any $t \in (0, \varepsilon]$. In particular, this means that t_k can always be chosen in a way that guarantees a decrease in the function value from one iteration to the next. For example, one choice of stepsize that guarantees descent is the *exact line search* strategy in which t_k is chosen as

$$t_k \in \operatorname{argmin}_{t \geq 0} f(\mathbf{x}^k - t\nabla f(\mathbf{x}^k)).$$

If f is not differentiable, then scheme (8.1) is not well defined. Under our convexity assumption, a natural generalization to the nonsmooth case will consist in replacing the gradient by a subgradient (assuming that it exists):

$$\mathbf{x}^{k+1} = \mathbf{x}^k - t_k \mathbf{g}^k, \quad \mathbf{g}^k \in \partial f(\mathbf{x}^k), \quad (8.3)$$

where we assume that the choice of the subgradient from $\partial f(\mathbf{x}^k)$ is arbitrary. The scheme (8.3) is called the *subgradient method*. One substantial difference between the gradient and subgradient methods is that the direction of minus the subgradient is not necessarily a descent direction. This means that t_k cannot be chosen in a way that will guarantee a descent property in function values of the scheme (8.3).

Example 8.3 (non-descent subgradient direction).⁴⁰ Consider the function $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ given by $f(x_1, x_2) = |x_1| + 2|x_2|$. Then

$$\partial f(1, 0) = \{(1, x) : |x| \leq 2\}.$$

In particular, $(1, 2) \in \partial f(1, 0)$. However, the direction $-(1, 2)$ is not a descent direction. To show this, note that for any $t > 0$,

$$g(t) \equiv f((1, 0) - t(1, 2)) = f(1-t, -2t) = |1-t| + 4t = \begin{cases} 1 + 3t, & t \in (0, 1], \\ 5t - 1, & t \geq 1. \end{cases} \quad (8.4)$$

In particular,

$$f'((1, 0); -(1, 2)) = g'_+(0) = 3 > 0,$$

showing that $-(1, 2)$ is not a descent direction. It is also interesting to note that by (8.4), it holds that

$$f((1, 0) - t(1, 2)) \geq 1 = f(1, 0) \text{ for any } t > 0,$$

which actually shows that there is no point in the ray $\{(1, 0) - t(1, 2) : t > 0\}$ with a smaller function value than $(1, 0)$. ■

⁴⁰Example 8.3 is taken from Vandenberghe's lecture notes [122].

8.1.2 Wolfe's Example

To better understand the effect of nonsmoothness, we recall a famous example of Wolfe. The example deals with the gradient method employed on a nonsmooth convex function with stepsizes chosen by exact line search. The function is differentiable at all the iterate vectors generated by the method, which in particular means that all the directions picked by the method are descent directions, and the sequence of function values strictly decreases. However, although it seems that the nonsmoothness is “bypassed,” this is hardly the case. The sequence generated by the method converges to a nonoptimal point.

Let $\gamma > 1$, and consider the function $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ given by

$$f(x_1, x_2) = \begin{cases} \sqrt{x_1^2 + \gamma x_2^2}, & |x_2| \leq x_1, \\ \frac{x_1 + \gamma |x_2|}{\sqrt{1+\gamma}} & \text{else.} \end{cases} \quad (8.5)$$

We begin by showing in Lemma 8.5 below that the function f is closed and convex and describe its subdifferential set at any point in $\mathbb{R} \times \mathbb{R}$. For that, we will prove that f is actually a support function of a closed and convex set.⁴¹ The proof of Lemma 8.5 uses the following simple technical lemma, whose trivial proof is omitted.

Lemma 8.4. *Consider the problem*

$$(P) \quad \max\{g(\mathbf{y}) : f_1(\mathbf{y}) \leq 0, f_2(\mathbf{y}) \leq 0\},$$

where $g : \mathbb{E} \rightarrow \mathbb{R}$ is concave and $f_1, f_2 : \mathbb{E} \rightarrow \mathbb{R}$ are convex. Assume that the problem $\max\{g(\mathbf{y}) : f_1(\mathbf{y}) \leq 0\}$ has a unique solution $\tilde{\mathbf{y}}$. Let Y^* be the optimal set of problem (P). Then exactly one of the following two options holds:

- (i) $f_2(\tilde{\mathbf{y}}) \leq 0$, and in this case $Y^* = \{\tilde{\mathbf{y}}\}$.
- (ii) $f_2(\tilde{\mathbf{y}}) > 0$, and in this case $Y^* = \operatorname{argmax}\{g(\mathbf{y}) : f_1(\mathbf{y}) \leq 0, f_2(\mathbf{y}) = 0\}$.

Lemma 8.5. *Let f be given by (8.5). Then*

- (a) $f = \sigma_C$, where

$$C = \left\{ (y_1, y_2) \in \mathbb{R} \times \mathbb{R} : y_1^2 + \frac{y_2^2}{\gamma} \leq 1, y_1 \geq \frac{1}{\sqrt{1+\gamma}} \right\};$$

- (b) f is closed and convex;

- (c)

$$\partial f(x_1, x_2) = \begin{cases} C, & x_1 = x_2 = 0, \\ \frac{(x_1, \gamma x_2)}{\sqrt{x_1^2 + \gamma x_2^2}}, & |x_2| \leq x_1, x_1 \neq 0, \\ \left(\frac{1}{\sqrt{1+\gamma}}, \frac{\gamma \operatorname{sgn}(x_2)}{\sqrt{1+\gamma}} \right), & |x_2| > x_1, x_2 \neq 0, \\ \left\{ \frac{1}{\sqrt{1+\gamma}} \right\} \times \left[-\frac{\gamma}{\sqrt{1+\gamma}}, \frac{\gamma}{\sqrt{1+\gamma}} \right], & x_2 = 0, x_1 < 0. \end{cases}$$

⁴¹Recall that support functions of nonempty sets are always closed and convex (Lemma 2.23).

Proof. By the definition of support functions,

$$\sigma_C(x_1, x_2) = \max_{y_1, y_2} \left\{ x_1 y_1 + x_2 y_2 : y_1^2 + \frac{y_2^2}{\gamma} \leq 1, y_1 \geq \frac{1}{\sqrt{1+\gamma}} \right\}. \quad (8.6)$$

Note that if $(x_1, x_2) = (0, 0)$, then $\sigma_C(x_1, x_2) = 0$ and

$$\operatorname{argmax}_{y_1, y_2} \left\{ x_1 y_1 + x_2 y_2 : y_1^2 + \frac{y_2^2}{\gamma} \leq 1, y_1 \geq \frac{1}{\sqrt{1+\gamma}} \right\} = C.$$

Assume that $(x_1, x_2) \neq (0, 0)$. Denoting $g(y_1, y_2) = x_1 y_1 + x_2 y_2$, $f_1(y_1, y_2) = y_1^2 + \frac{y_2^2}{\gamma} - 1$ and $f_2(y_1, y_2) = -y_1 + \frac{1}{\sqrt{1+\gamma}}$, problem (8.6) becomes

$$\max_{y_1, y_2} \{g(y_1, y_2) : f_1(y_1, y_2) \leq 0, f_2(y_1, y_2) \leq 0\}.$$

The assumptions made in Lemma 8.4 are all met: g is concave, f_1, f_2 are convex, and the optimal solution of

$$\max_{y_1, y_2} \{g(y_1, y_2) : f_1(y_1, y_2) \leq 0\}$$

is unique and equal to $(\tilde{y}_1, \tilde{y}_2) = \frac{(x_1, \gamma x_2)}{\sqrt{x_1^2 + \gamma x_2^2}}$. Thus, by Lemma 8.4, there are two options:

Case I: $f_2(\tilde{y}_1, \tilde{y}_2) \leq 0$, meaning that $\frac{x_1}{\sqrt{x_1^2 + \gamma x_2^2}} \geq \frac{1}{\sqrt{1+\gamma}}$. It can be easily seen that the last inequality is equivalent to the condition $|x_2| \leq x_1$. Under this condition, by Lemma 8.4, $(\tilde{y}_1, \tilde{y}_2) = \frac{(x_1, \gamma x_2)}{\sqrt{x_1^2 + \gamma x_2^2}}$ is the unique optimal solution of problem (8.6) with a corresponding function value of $\sigma_C(x_1, x_2) = \sqrt{x_1^2 + \gamma x_2^2}$.

Case II: $f_2(\tilde{y}_1, \tilde{y}_2) > 0$, which is the same as $x_1 < |x_2|$. In this case, by Lemma 8.4, all the optimal solutions of problem (8.6) satisfy $y_1 = \frac{1}{\sqrt{1+\gamma}}$, and the problem thus amounts to

$$\max_{y_2} \left\{ \frac{1}{\sqrt{1+\gamma}} x_1 + x_2 y_2 : y_2^2 \leq \frac{\gamma^2}{1+\gamma} \right\}.$$

The set of maximizers of the above problem is either $\left\{ \frac{\gamma \operatorname{sgn}(x_2)}{\sqrt{1+\gamma}} \right\}$ if $x_2 \neq 0$ or $\left[-\frac{\gamma}{\sqrt{1+\gamma}}, \frac{\gamma}{\sqrt{1+\gamma}} \right]$ if $x_2 = 0$. In both options, $\sigma_C(x_1, x_2) = \frac{x_1 + \gamma |x_2|}{\sqrt{1+\gamma}}$.

To summarize, we have shown that

$$\sigma_C(x_1, x_2) = \begin{cases} \sqrt{x_1^2 + \gamma x_2^2}, & |x_2| \leq x_1, \\ \frac{x_1 + \gamma |x_2|}{\sqrt{1+\gamma}} & \text{else,} \end{cases}$$

establishing part (a), meaning that $f = \sigma_C$. Therefore, f , as a support function, is a closed and convex function, and we have thus established part (b) as well. To

prove part (c), note that we also showed that

$$\begin{aligned} & \operatorname{argmax}_{y_1, y_2} \{x_1 y_1 + x_2 y_2 : (y_1, y_2) \in C\} \\ = & \begin{cases} C, & x_1 = x_2 = 0, \\ \frac{(x_1, \gamma x_2)}{\sqrt{x_1^2 + \gamma x_2^2}}, & |x_2| \leq x_1, x_1 \neq 0, \\ \left(\frac{1}{\sqrt{1+\gamma}}, \frac{\gamma \operatorname{sgn}(x_2)}{\sqrt{1+\gamma}}\right), & |x_2| > x_1, x_2 \neq 0, \\ \left\{\frac{1}{\sqrt{1+\gamma}}\right\} \times \left[-\frac{\gamma}{\sqrt{1+\gamma}}, \frac{\gamma}{\sqrt{1+\gamma}}\right], & x_2 = 0, x_1 < 0. \end{cases} \end{aligned}$$

Combining this with the conjugate subgradient theorem (Corollary 4.21), as well as Example 4.9 and the closedness and convexity of C , implies

$$\begin{aligned} \partial f(x_1, x_2) &= \partial \sigma_C(x_1, x_2) \\ &= \operatorname{argmax}_{y_1, y_2} \{x_1 y_1 + x_2 y_2 - \sigma_C^*(y_1, y_2)\} \\ &= \operatorname{argmax}_{y_1, y_2} \{x_1 y_1 + x_2 y_2 - \delta_C(y_1, y_2)\} \\ &= \operatorname{argmax}_{y_1, y_2} \{x_1 y_1 + x_2 y_2 : (y_1, y_2) \in C\} \\ &= \begin{cases} C, & x_1 = x_2 = 0, \\ \frac{(x_1, \gamma x_2)}{\sqrt{x_1^2 + \gamma x_2^2}}, & |x_2| \leq x_1, x_1 \neq 0, \\ \left(\frac{1}{\sqrt{1+\gamma}}, \frac{\gamma \operatorname{sgn}(x_2)}{\sqrt{1+\gamma}}\right), & |x_2| > x_1, x_2 \neq 0, \\ \left\{\frac{1}{\sqrt{1+\gamma}}\right\} \times \left[-\frac{\gamma}{\sqrt{1+\gamma}}, \frac{\gamma}{\sqrt{1+\gamma}}\right], & x_2 = 0, x_1 < 0. \end{cases} \quad \square \end{aligned}$$

Note that a direct result of part (c) of Lemma 8.5 and Theorem 3.33 is that f is not differentiable only at the nonpositive part of the x_1 axis.

In the next result we will show that the gradient method with exact line search employed on f with a certain initialization converges to the nonoptimal point $(0, 0)$ even though all the points generated by the gradient method are points in which f is differentiable.

Lemma 8.6. *Let $\{(x_1^{(k)}, x_2^{(k)})\}_{k \geq 0}$ be the sequence generated by the gradient method with exact line search employed on f with initial point $(x_1^0, x_2^0) = (\gamma, 1)$, where $\gamma > 1$. Then for any $k \geq 0$,*

- (a) f is differentiable at $(x_1^{(k)}, x_2^{(k)})$;
- (b) $|x_2^{(k)}| \leq x_1^{(k)}$ and $x_1^{(k)} \neq 0$;
- (c) $(x_1^{(k)}, x_2^{(k)}) = \left(\gamma \left(\frac{\gamma-1}{\gamma+1}\right)^k, \left(-\frac{\gamma-1}{\gamma+1}\right)^k\right)$.

Proof. We only need to show part (c) since part (b) follows directly from the expression of $(x_1^{(k)}, x_2^{(k)})$ given in (c), and part (a) is then a consequence of Lemma 8.5(c).

We will prove part (c) by induction. The claim is obviously correct for $k = 0$ by the choice of initial point. Assume that the claim is correct for k , that is,

$$(x_1^{(k)}, x_2^{(k)}) = \left(\gamma \left(\frac{\gamma - 1}{\gamma + 1} \right)^k, \left(-\frac{\gamma - 1}{\gamma + 1} \right)^k \right).$$

We will prove that it is correct for $k + 1$, meaning that

$$(x_1^{(k+1)}, x_2^{(k+1)}) = (\beta_k, \gamma_k), \quad (8.7)$$

where

$$\beta_k = \gamma \left(\frac{\gamma - 1}{\gamma + 1} \right)^{k+1}, \gamma_k = \left(-\frac{\gamma - 1}{\gamma + 1} \right)^{k+1}.$$

Since $|x_2^{(k)}| \leq x_1^{(k)}$ and $x_1^{(k)} \neq 0$, we have $f(x_1^{(k)}, x_2^{(k)}) = \sqrt{(x_1^{(k)})^2 + \gamma(x_2^{(k)})^2}$, and by Lemma 8.5(c), f is differentiable at $(x_1^{(k)}, x_2^{(k)})$ with

$$\nabla f(x_1^{(k)}, x_2^{(k)}) = \frac{1}{\sqrt{(x_1^{(k)})^2 + \gamma(x_2^{(k)})^2}} (x_1^{(k)}, \gamma x_2^{(k)}).$$

What is important in the above formula is that $\nabla f(x_1^{(k)}, x_2^{(k)})$ can be written in the form

$$\nabla f(x_1^{(k)}, x_2^{(k)}) = \alpha_k (x_1^{(k)}, \gamma x_2^{(k)}) \quad (8.8)$$

for some positive constant α_k . To show the validity of (8.7), we will define $g(t) \equiv f((x_1^{(k)}, x_2^{(k)}) - t(x_1^{(k)}, \gamma x_2^{(k)}))$ and prove the following two statements:

$$(A) \quad (\beta_k, \gamma_k) = (x_1^{(k)}, x_2^{(k)}) - \frac{2}{\gamma+1} (x_1^{(k)}, \gamma x_2^{(k)}).$$

$$(B) \quad g' \left(\frac{2}{\gamma+1} \right) = 0.$$

(A) and (B) are enough to show (8.7) since g is strictly convex. The proof of (A) follows by the computations below:

$$x_1^{(k)} - \frac{2}{\gamma+1} x_1^{(k)} = \frac{\gamma-1}{\gamma+1} x_1^{(k)} = \frac{\gamma-1}{\gamma+1} \gamma \left(\frac{\gamma-1}{\gamma+1} \right)^k = \gamma \left(\frac{\gamma-1}{\gamma+1} \right)^{k+1} = \beta_k,$$

$$x_2^{(k)} - \frac{2\gamma}{\gamma+1} x_2^{(k)} = \frac{-\gamma+1}{\gamma+1} x_2^{(k)} = \frac{-\gamma+1}{\gamma+1} \left(-\frac{\gamma-1}{\gamma+1} \right)^k = \left(-\frac{\gamma-1}{\gamma+1} \right)^{k+1} = \gamma_k.$$

To prove (B), note that

$$\begin{aligned} g(t) &= f \left((x_1^{(k)}, x_2^{(k)}) - t(x_1^{(k)}, \gamma x_2^{(k)}) \right) = f((1-t)x_1^{(k)}, (1-\gamma t)x_2^{(k)}) \\ &= \sqrt{(1-t)^2(x_1^{(k)})^2 + \gamma(1-\gamma t)^2(x_2^{(k)})^2}. \end{aligned}$$

Therefore,

$$g'(t) = \frac{(t-1)(x_1^{(k)})^2 + \gamma^2(\gamma t - 1)(x_2^{(k)})^2}{\sqrt{(1-t)^2(x_1^{(k)})^2 + \gamma(1-\gamma t)^2(x_2^{(k)})^2}}. \quad (8.9)$$

To prove that $g'(\frac{2}{\gamma+1}) = 0$, it is enough to show that the nominator in the last expression is equal to zero at $t = \frac{2}{\gamma+1}$. Indeed,

$$\begin{aligned} & \left(\frac{2}{\gamma+1} - 1 \right) (x_1^{(k)})^2 + \gamma^2 \left(\gamma \cdot \frac{2}{\gamma+1} - 1 \right) (x_2^{(k)})^2 \\ &= \left(-\frac{\gamma-1}{\gamma+1} \right) \gamma^2 \left(\frac{\gamma-1}{\gamma+1} \right)^{2k} + \gamma^2 \left(\frac{\gamma-1}{\gamma+1} \right) \left(-\frac{\gamma-1}{\gamma+1} \right)^{2k} \\ &= 0. \quad \square \end{aligned}$$

Obviously, by Lemma 8.6, the sequence generated by the gradient method with exact line search and initial point $(\gamma, 1)$ converges to $(0, 0)$, which is not a minimizer of f since f is not bounded below (take $x_2 = 0$ and $x_1 \rightarrow -\infty$). Actually, $(-1, 0)$ is a descent direction of f at $(0, 0)$. The contour lines of the function along with the iterates of the gradient method are described in Figure 8.1.

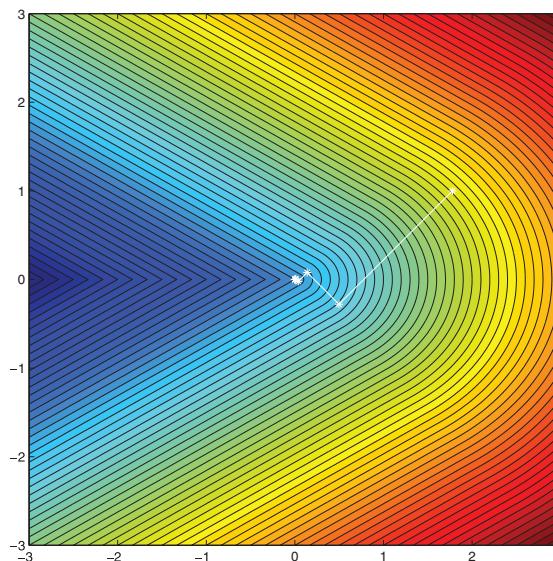


Figure 8.1. Contour lines of Wolfe's function with $\gamma = \frac{16}{9}$ along with the iterates of the gradient method with exact line search.

8.2 The Projected Subgradient Method

The main model that will be discussed in this section is

$$\min\{f(\mathbf{x}) : \mathbf{x} \in C\}, \tag{8.10}$$

where the following assumption will be made throughout this section.

Assumption 8.7.

- (A) $f : \mathbb{E} \rightarrow (-\infty, \infty]$ is proper closed and convex.
- (B) $C \subseteq \mathbb{E}$ is nonempty closed and convex.
- (C) $C \subseteq \text{int}(\text{dom}(f))$.
- (D) The optimal set of (8.10) is nonempty and denoted by X^* . The optimal value of the problem is denoted by f_{opt} .

Remark 8.8 (subdifferentiability of f and closedness of X^*). Since f is convex and $C \subseteq \text{int}(\text{dom}(f))$, it follows by Theorem 3.14 that f is subdifferentiable over C . Also, since f is closed,

$$X^* = C \cap \text{Lev}(f, f_{\text{opt}})$$

is closed. This means in particular that for any $\mathbf{x} \notin X^*$ the distance $d_{X^*}(\mathbf{x})$ is positive.

From now on, we will use the following notation: $f'(\mathbf{x})$ will denote a certain subgradient of f at \mathbf{x} , meaning a member in $\partial f(\mathbf{x})$. Thus, f' is actually a function from C to \mathbb{E}^* . The rule for choosing $f'(\mathbf{x})$ out of the members of $\partial f(\mathbf{x})$ can be arbitrary but has to be deterministic, meaning that if $f'(\mathbf{x})$ is evaluated twice, the results have to be the same.

Equipped with the observations of the previous section, we can speculate that a method which utilizes subgradients rather than gradients will not necessarily be a descent method and will not have to be based on a line search procedure for choosing its stepsizes. We will see that this is indeed the case for the projected subgradient method.

8.2.1 The Method

Each iteration of the projected subgradient method consists of a step taken toward the negative of the chosen subgradient followed by an orthogonal projection onto the underlying set C .

Projected Subgradient Method

Initialization: pick $\mathbf{x}^0 \in C$ arbitrarily.

General step: for any $k = 0, 1, 2, \dots$ execute the following steps:

- (a) pick a stepsize $t_k > 0$ and a subgradient $f'(\mathbf{x}^k) \in \partial f(\mathbf{x}^k)$;
- (b) set $\mathbf{x}^{k+1} = P_C(\mathbf{x}^k - t_k f'(\mathbf{x}^k))$.

The sequence generated by the projected subgradient method is $\{\mathbf{x}^k\}_{k \geq 0}$, while the sequence of function values generated by the method is $\{f(\mathbf{x}^k)\}_{k \geq 0}$. As

was already discussed, the sequence of function values is not necessarily monotone, and we will be also interested in the sequence of best achieved function values, which is defined by

$$f_{\text{best}}^k \equiv \min_{n=0,1,\dots,k} f(\mathbf{x}^n). \quad (8.11)$$

Obviously, the sequence $\{f_{\text{best}}^k\}_{k \geq 0}$ is nonincreasing.

Remark 8.9 (stopping criterion for the projected subgradient method). *In actual implementations of the projected subgradient method, a stopping criterion has to be incorporated, but as a rule, we will not deal in this book with stopping criteria but rather concentrate on issues of convergence.*

Remark 8.10 (zero subgradients). *In the unlikely case where $f'(\mathbf{x}^k) = \mathbf{0}$ for some k , then by Fermat's optimality condition (Theorem 3.63), \mathbf{x}^k is a minimizer of f over \mathbb{E} , and since $\mathbf{x}^k \in C$, it is also a minimizer of f over C . In this situation, the method is "stuck" at the optimal solution \mathbf{x}^k from iteration k onward, meaning that $\mathbf{x}^n = \mathbf{x}^k$ for all $n \geq k$.*

The analysis of the projected subgradient method relies on the following simple technical lemma.

Lemma 8.11 (fundamental inequality for projected subgradient). *Suppose that Assumption 8.7 holds. Let $\{\mathbf{x}^k\}_{k \geq 0}$ be the sequence generated by the projected subgradient method. Then for any $\mathbf{x}^* \in X^*$ and $k \geq 0$,*

$$\|\mathbf{x}^{k+1} - \mathbf{x}^*\|^2 \leq \|\mathbf{x}^k - \mathbf{x}^*\|^2 - 2t_k(f(\mathbf{x}^k) - f_{\text{opt}}) + t_k^2 \|f'(\mathbf{x}^k)\|^2. \quad (8.12)$$

Proof.

$$\begin{aligned} \|\mathbf{x}^{k+1} - \mathbf{x}^*\|^2 &= \|P_C(\mathbf{x}^k - t_k f'(\mathbf{x}^k)) - P_C(\mathbf{x}^*)\|^2 \\ &\stackrel{(*)}{\leq} \|\mathbf{x}^k - t_k f'(\mathbf{x}^k) - \mathbf{x}^*\|^2 \\ &= \|\mathbf{x}^k - \mathbf{x}^*\|^2 - 2t_k \langle f'(\mathbf{x}^k), \mathbf{x}^k - \mathbf{x}^* \rangle + t_k^2 \|f'(\mathbf{x}^k)\|^2 \\ &\stackrel{(**)}{\leq} \|\mathbf{x}^k - \mathbf{x}^*\|^2 - 2t_k(f(\mathbf{x}^k) - f_{\text{opt}}) + t_k^2 \|f'(\mathbf{x}^k)\|^2, \end{aligned}$$

where the inequality $(*)$ is due to the nonexpansiveness of the orthogonal projection operator (Theorem 6.42), and $(**)$ follows by the subgradient inequality. \square

8.2.2 Convergence under Polyak's Stepsize Rule

We will require an assumption in addition to Assumption 8.7 in order to prove convergence of the sequence of function values generated by the projected subgradient method.

Assumption 8.12. *There exists a constant $L_f > 0$ for which $\|\mathbf{g}\| \leq L_f$ for all $\mathbf{g} \in \partial f(\mathbf{x}), \mathbf{x} \in C$.*

Since $C \subseteq \text{int}(\text{dom}(f))$ (Assumption 8.7(C)), it follows by Theorem 3.61 that Assumption 8.12 implies that f is Lipschitz continuous over C with constant L_f :

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq L_f \|\mathbf{x} - \mathbf{y}\| \text{ for all } \mathbf{x}, \mathbf{y} \in C.$$

In addition, since (again) $C \subseteq \text{int}(\text{dom}(f))$, it follows by Theorem 3.16 that Assumption 8.12 holds if C is assumed to be compact.

One natural way to choose the stepsize t_k is by taking it as the minimizer of the right-hand side of (8.12) over $t_k \geq 0$:

$$t_k = \frac{f(\mathbf{x}^k) - f_{\text{opt}}}{\|f'(\mathbf{x}^k)\|^2}.$$

When $f'(\mathbf{x}^k) = \mathbf{0}$, the above formula is not defined, and by Remark 8.10, \mathbf{x}^k is an optimal solution of (8.10). We will artificially define $t_k = 1$ (any other positive number could also have been chosen). The complete formula is therefore

$$t_k = \begin{cases} \frac{f(\mathbf{x}^k) - f_{\text{opt}}}{\|f'(\mathbf{x}^k)\|^2}, & f'(\mathbf{x}^k) \neq \mathbf{0}, \\ 1, & f'(\mathbf{x}^k) = \mathbf{0}. \end{cases} \quad (8.13)$$

We will refer to this stepsize rule as *Polyak's stepsize rule*.⁴²

The main convergence result of the projected subgradient method with Polyak's stepsize rule is given in the next theorem.

Theorem 8.13 (convergence of projected subgradient with Polyak's stepsize). *Suppose that Assumptions 8.7 and 8.12 hold. Let $\{\mathbf{x}^k\}_{k \geq 0}$ be the sequence generated by the projected subgradient method with Polyak's stepsize rule (8.13). Then*

- (a) $\|\mathbf{x}^{k+1} - \mathbf{x}^*\|^2 \leq \|\mathbf{x}^k - \mathbf{x}^*\|^2$ for any $k \geq 0$ and $\mathbf{x}^* \in X^*$;
- (b) $f(\mathbf{x}^k) \rightarrow f_{\text{opt}}$ as $k \rightarrow \infty$;
- (c) $f_{\text{best}}^k - f_{\text{opt}} \leq \frac{L_f d_{X^*}(\mathbf{x}^0)}{\sqrt{k+1}}$ for any $k \geq 0$.

Proof. Let n be a nonnegative integer and $\mathbf{x}^* \in X^*$. By Lemma 8.11,

$$\|\mathbf{x}^{n+1} - \mathbf{x}^*\|^2 \leq \|\mathbf{x}^n - \mathbf{x}^*\|^2 - 2t_n(f(\mathbf{x}^n) - f_{\text{opt}}) + t_n^2 \|f'(\mathbf{x}^n)\|^2. \quad (8.14)$$

If $f'(\mathbf{x}^n) \neq \mathbf{0}$, then by substituting $t_n = \frac{f(\mathbf{x}^n) - f_{\text{opt}}}{\|f'(\mathbf{x}^n)\|^2}$ into (8.14), it follows that

$$\|\mathbf{x}^{n+1} - \mathbf{x}^*\|^2 \leq \|\mathbf{x}^n - \mathbf{x}^*\|^2 - \frac{(f(\mathbf{x}^n) - f_{\text{opt}})^2}{\|f'(\mathbf{x}^n)\|^2}.$$

⁴²As the name suggests, this stepsize was first suggested by Boris T. Polyak; see, for example, [104].

Using the bound $\|f'(\mathbf{x}^n)\| \leq L_f$, we thus obtain

$$\|\mathbf{x}^{n+1} - \mathbf{x}^*\|^2 \leq \|\mathbf{x}^n - \mathbf{x}^*\|^2 - \frac{(f(\mathbf{x}^n) - f_{\text{opt}})^2}{L_f^2}. \quad (8.15)$$

Inequality (8.15) also holds when $f'(\mathbf{x}^n) = \mathbf{0}$, since in this case $f(\mathbf{x}^n) = f_{\text{opt}}$ and $\mathbf{x}^{n+1} = \mathbf{x}^n$. A direct result of (8.15) is that

$$\|\mathbf{x}^{n+1} - \mathbf{x}^*\|^2 \leq \|\mathbf{x}^n - \mathbf{x}^*\|^2,$$

and part (a) is thus proved (by plugging $n = k$). Summing inequality (8.15) over $n = 0, 1, \dots, k$, we obtain that

$$\frac{1}{L_f^2} \sum_{n=0}^k (f(\mathbf{x}^n) - f_{\text{opt}})^2 \leq \|\mathbf{x}^0 - \mathbf{x}^*\|^2 - \|\mathbf{x}^{k+1} - \mathbf{x}^*\|^2,$$

and thus

$$\sum_{n=0}^k (f(\mathbf{x}^n) - f_{\text{opt}})^2 \leq L_f^2 \|\mathbf{x}^0 - \mathbf{x}^*\|^2.$$

Since the above inequality holds for any $\mathbf{x}^* \in X^*$, it follows that

$$\sum_{n=0}^k (f(\mathbf{x}^n) - f_{\text{opt}})^2 \leq L_f^2 d_{X^*}^2(\mathbf{x}^0), \quad (8.16)$$

which in particular implies that $f(\mathbf{x}^n) - f_{\text{opt}} \rightarrow 0$ as $n \rightarrow \infty$, and the validity of (b) is established. To prove part (c), note that since $f(\mathbf{x}^n) \geq f_{\text{best}}^k$ for any $n = 0, 1, \dots, k$, it follows that

$$\sum_{n=0}^k (f(\mathbf{x}^n) - f_{\text{opt}})^2 \geq (k+1)(f_{\text{best}}^k - f_{\text{opt}})^2,$$

which, combined with (8.16), yields

$$(k+1)(f_{\text{best}}^k - f_{\text{opt}})^2 \leq L_f^2 d_{X^*}^2(\mathbf{x}^0),$$

and hence

$$f_{\text{best}}^k - f_{\text{opt}} \leq \frac{L_f d_{X^*}(\mathbf{x}^0)}{\sqrt{k+1}}. \quad \square$$

Remark 8.14. Note that in the convergence result of Theorem 8.13 we can replace the constant L_f with $\max_{n=0,1,\dots,k} \|f'(\mathbf{x}^n)\|$.

The property of the sequence generated by the projected subgradient method described in part (a) of Theorem 8.13 is known as *Fejér monotonicity*.

Definition 8.15 (Fejér monotonicity). A sequence $\{\mathbf{x}^k\}_{k \geq 0} \subseteq \mathbb{E}$ is called **Fejér monotone** w.r.t. a set $S \subseteq \mathbb{E}$ if

$$\|\mathbf{x}^{k+1} - \mathbf{y}\| \leq \|\mathbf{x}^k - \mathbf{y}\| \text{ for all } k \geq 0 \text{ and } \mathbf{y} \in S.$$

Since Fejér monotonicity w.r.t. a set S implies that for all $k \geq 0$ and any $\mathbf{y} \in S$, $\|\mathbf{x}^k - \mathbf{y}\| \leq \|\mathbf{x}^0 - \mathbf{y}\|$, it follows that Fejér monotone sequences are always bounded. We will now prove that sequences which are Fejér monotone w.r.t. sets containing their limit points are convergent.

Theorem 8.16 (convergence under Fejér monotonicity). *Let $\{\mathbf{x}^k\}_{k \geq 0} \subseteq \mathbb{E}$ be a sequence, and let S be a set satisfying $D \subseteq S$, where D is the set comprising all the limit points of $\{\mathbf{x}^k\}_{k \geq 0}$. If $\{\mathbf{x}^k\}_{k \geq 0}$ is Fejér monotone w.r.t. S , then it converges to a point in D .*

Proof. Since $\{\mathbf{x}^k\}_{k \geq 0}$ is Fejér monotone, it is also bounded and hence has limit points. Let $\tilde{\mathbf{x}}$ be a limit point of the sequence $\{\mathbf{x}^k\}_{k \geq 0}$, meaning that there exists a subsequence $\{\mathbf{x}^{k_j}\}_{j \geq 0}$ such that $\mathbf{x}^{k_j} \rightarrow \tilde{\mathbf{x}}$. Since $\tilde{\mathbf{x}} \in D \subseteq S$, it follows by the Fejér monotonicity w.r.t. S that for any $k \geq 0$,

$$\|\mathbf{x}^{k+1} - \tilde{\mathbf{x}}\| \leq \|\mathbf{x}^k - \tilde{\mathbf{x}}\|.$$

Thus, $\{\|\mathbf{x}^k - \tilde{\mathbf{x}}\|\}_{k \geq 0}$ is a nonincreasing sequence which is bounded below (by zero) and hence convergent. Since $\|\mathbf{x}^{k_j} - \tilde{\mathbf{x}}\| \rightarrow 0$ as $j \rightarrow \infty$, it follows that the whole sequence $\{\|\mathbf{x}^k - \tilde{\mathbf{x}}\|\}_{k \geq 0}$ converges to zero, and consequently $\mathbf{x}^k \rightarrow \tilde{\mathbf{x}}$ as $k \rightarrow \infty$. \square

Equipped with the last theorem, we can now prove convergence of the sequence generated by the projected subgradient method with Polyak's stepsize rule.

Theorem 8.17 (convergence of the sequence generated by projected subgradient with Polyak's stepsize rule). *Suppose that Assumptions 8.7 and 8.12 hold. Let $\{\mathbf{x}^k\}_{k \geq 0}$ be the sequence generated by the projected subgradient method with Polyak's stepsize rule (8.13). Then $\{\mathbf{x}^k\}_{k \geq 0}$ converges to a point in X^* .*

Proof. By Theorem 8.13(a), the sequence is Fejér monotone w.r.t. X^* . Therefore, by Theorem 8.16, to show convergence to a point in X^* , it is enough to show that any limit point of the sequence is necessarily in X^* (that is, an optimal solution of the problem). Let then $\tilde{\mathbf{x}}$ be a limit point of the sequence. Then there exists a subsequence $\{\mathbf{x}^{k_j}\}_{j \geq 0}$ converging to $\tilde{\mathbf{x}}$. By the closedness of C , $\tilde{\mathbf{x}} \in C$. By Theorem 8.13(b),

$$f(\mathbf{x}^{k_j}) \rightarrow f_{\text{opt}} \text{ as } j \rightarrow \infty. \quad (8.17)$$

Since $\tilde{\mathbf{x}} \in C \subseteq \text{int}(\text{dom}(f))$, it follows by Theorem 2.21 that f is continuous at $\tilde{\mathbf{x}}$, which, combined with (8.17), implies that $f(\tilde{\mathbf{x}}) = f_{\text{opt}}$, meaning that $\tilde{\mathbf{x}} \in X^*$. \square

Part (c) of Theorem 8.13 provides an upper bound on the *rate of convergence* in which the sequence $\{f_{\text{best}}^k\}_{k \geq 0}$ converges to f_{opt} . Specifically, the result shows that the distance of f_{best}^k to f_{opt} is bounded above by a constant factor of $\frac{1}{\sqrt{k+1}}$ with k being the iteration index. We will sometimes refer to it as an " $O(1/\sqrt{k})$ rate of convergence result" with a slight abuse of the "big O" notation (which actually refers to asymptotic results). We can also write the rate of convergence result as a *complexity* result. For that, we first introduce the concept of an ε -optimal solution. A vector $\mathbf{x} \in C$ is called an ε -optimal solution of problem (8.10) if $f(\mathbf{x}) - f_{\text{opt}} \leq \varepsilon$. In complexity analysis, the following question is asked: *how many iterations are*

required to obtain an ε -optimal solution? That is, how many iterations are required to obtain the condition

$$f_{\text{best}}^k - f_{\text{opt}} \leq \varepsilon? \quad (8.18)$$

Using Theorem 8.13(c), it follows that a sufficient condition for (8.18) to hold is the following inequality:

$$\frac{L_f d_{X^*}(\mathbf{x}^0)}{\sqrt{k+1}} \leq \varepsilon, \quad (8.19)$$

which is the same as

$$k \geq \frac{L_f^2 d_{X^*}^2(\mathbf{x}^0)}{\varepsilon^2} - 1.$$

Therefore, an order of $\frac{1}{\varepsilon^2}$ iterations is required to obtain an ε -optimal solution. We summarize the discussion in the following theorem.

Theorem 8.18 (complexity of projected subgradient with Polyak's step-size). Suppose that Assumptions 8.7 and 8.12 hold. Let $\{\mathbf{x}^k\}_{k \geq 0}$ be the sequence generated by the projected subgradient method with Polyak's stepsize rule (8.13). Then for any nonnegative integer k satisfying

$$k \geq \frac{L_f^2 d_{X^*}^2(\mathbf{x}^0)}{\varepsilon^2} - 1,$$

it holds that

$$f_{\text{best}}^k - f_{\text{opt}} \leq \varepsilon.$$

Example 8.19. Consider the problem

$$\min_{x_1, x_2} \{f(x_1, x_2) = |x_1 + 2x_2| + |3x_1 + 4x_2|\}.$$

Since in this chapter the underlying spaces are Euclidean, it follows that the underlying space in this example is \mathbb{R}^2 endowed with the dot product and the l_2 -norm. The optimal solution of the problem is $(x_1, x_2) = (0, 0)$, and the optimal value is $f_{\text{opt}} = 0$. Clearly, both Assumptions 8.7 and 8.12 hold. Since $f(\mathbf{x}) = \|\mathbf{Ax}\|_1$, where $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, it follows that for any $\mathbf{x} \in \mathbb{R}^2$,

$$\partial f(\mathbf{x}) = \mathbf{A}^T \partial h(\mathbf{Ax}),$$

where $h(\mathbf{x}) = \|\mathbf{x}\|_1$. By Example 3.41, for any $\mathbf{w} \in \mathbb{R}^2$,

$$\partial h(\mathbf{w}) = \{\mathbf{z} \in \mathbb{R}^2 : z_i = \text{sgn}(w_i), i \in I_{\neq}(\mathbf{w}), |z_j| \leq 1, j \in I_0(\mathbf{w})\},$$

where

$$I_0(\mathbf{w}) = \{i : w_i = 0\}, I_{\neq}(\mathbf{w}) = \{i : w_i \neq 0\}.$$

Hence, if $\boldsymbol{\eta} \in \partial h(\mathbf{Ax})$, then $\boldsymbol{\eta} \in [-1, 1] \times [-1, 1]$, and, in particular, $\|\boldsymbol{\eta}\|_2 \leq \sqrt{2}$. Therefore, since any $\mathbf{g} \in \partial f(\mathbf{x})$ can be written as $\mathbf{g} = \mathbf{A}^T \boldsymbol{\eta}$ for some $\boldsymbol{\eta} \in \partial h(\mathbf{Ax})$, we have

$$\|\mathbf{g}\|_2 = \|\mathbf{A}^T \boldsymbol{\eta}\|_2 \leq \|\mathbf{A}^T\|_{2,2} \|\boldsymbol{\eta}\|_2 \leq \|\mathbf{A}^T\|_{2,2} \cdot \sqrt{2} = 7.7287.$$

We can thus choose $L_f = 7.7287$.

The subgradient method update step takes the form

$$\begin{pmatrix} x_1^{k+1} \\ x_2^{k+1} \end{pmatrix} = \begin{pmatrix} x_1^k \\ x_2^k \end{pmatrix} - \frac{|x_1^k + 2x_2^k| + |3x_1^k + 4x_2^k|}{\|\mathbf{v}(x_1^k, x_2^k)\|_2^2} \mathbf{v}(x_1^k, x_2^k),$$

where we choose

$$\mathbf{v}(x_1, x_2) = \begin{pmatrix} \operatorname{sgn}(x_1 + 2x_2) + 3\operatorname{sgn}(3x_1 + 4x_2) \\ 2\operatorname{sgn}(x_1 + 2x_2) + 4\operatorname{sgn}(3x_1 + 4x_2) \end{pmatrix} \in \partial f(x_1, x_2).$$

Note that in the terminology of this book $\operatorname{sgn}(0) = 1$ (see Section 1.7.2), which dictates the choice of the subgradient among the vectors in the subdifferential set in cases where f is not differentiable at the given point. We can immediately see that there are actually only four possible choices of directions $\mathbf{v}(x_1, x_2)$ depending on the two possible values of $\operatorname{sgn}(x_1 + 2x_2)$ and the two possible choices of $\operatorname{sgn}(3x_1 + 4x_2)$. The four possible directions are

$$\mathbf{u}_1 = \begin{pmatrix} -4 \\ -6 \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad \mathbf{u}_3 = \begin{pmatrix} -2 \\ -2 \end{pmatrix}, \quad \mathbf{u}_4 = \begin{pmatrix} 4 \\ 6 \end{pmatrix}.$$

By Remark 8.14, the constant L_f can be chosen as $\max_i \{\|\mathbf{u}_i\|_2\} = 7.2111$, which is a slightly better bound than 7.7287. The first 100 iterations of the method with a starting point $(1, 2)^T$ are described in Figure 8.2. Note that the sequence of function values is indeed not monotone (although convergence to f_{opt} is quite apparent) and that actually only two directions are being used by the method: $(-2, -2)^T$, $(4, 6)^T$. ■

8.2.3 The Convex Feasibility Problem

Let $S_1, S_2, \dots, S_m \subseteq \mathbb{E}$ be closed and convex sets. Assume that

$$S \equiv \bigcap_{i=1}^m S_i \neq \emptyset. \quad (8.20)$$

The *convex feasibility problem* is the problem of finding a point \mathbf{x} in the intersection $\bigcap_{i=1}^m S_i$. We can formulate the problem as the following minimization problem:

$$\min_{\mathbf{x}} \left\{ f(\mathbf{x}) \equiv \max_{i=1,2,\dots,m} d_{S_i}(\mathbf{x}) \right\}. \quad (8.21)$$

Since we assume that the intersection is nonempty, we have that $f_{\text{opt}} = 0$ and that the optimal set is S . Another property of f is that it is Lipschitz continuous with constant 1.

Lemma 8.20. *Let S_1, S_2, \dots, S_m be nonempty closed and convex sets. Then the function f given in (8.21) is Lipschitz continuous with constant 1.*

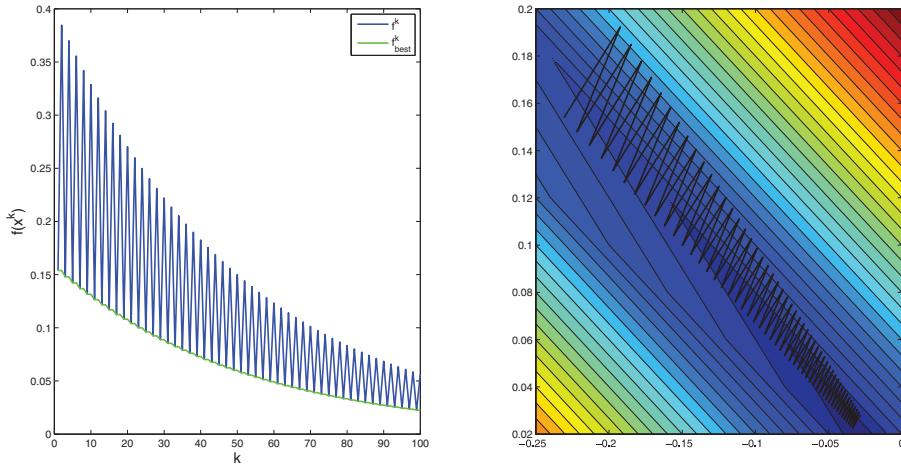


Figure 8.2. First 100 iterations of the subgradient method applied to the function $f(x_1, x_2) = |x_1 + 2x_2| + |3x_1 + 4x_2|$ with Polyak's stepsize rule and starting point $(1, 2)^T$. The left image describes the function values at each iteration, and the right image shows the contour lines along with the iterations.

Proof. Let $i \in \{1, 2, \dots, m\}$, and let $\mathbf{x}, \mathbf{y} \in \mathbb{E}$. Then

$$\begin{aligned} d_{S_i}(\mathbf{x}) &= \|\mathbf{x} - P_{S_i}(\mathbf{x})\| \\ &\leq \|\mathbf{x} - P_{S_i}(\mathbf{y})\| & [\|\mathbf{x} - P_{S_i}(\mathbf{x})\| = \operatorname{argmin}_{\mathbf{v} \in S_i} \|\mathbf{x} - \mathbf{v}\|] \\ &\leq \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{y} - P_{S_i}(\mathbf{y})\| & [\text{triangle inequality}] \\ &= \|\mathbf{x} - \mathbf{y}\| + d_{S_i}(\mathbf{y}). & [d_{S_i}(\mathbf{y}) = \|\mathbf{y} - P_{S_i}(\mathbf{y})\|] \end{aligned}$$

Thus,

$$d_{S_i}(\mathbf{x}) - d_{S_i}(\mathbf{y}) \leq \|\mathbf{x} - \mathbf{y}\|. \quad (8.22)$$

Replacing the roles of \mathbf{x} and \mathbf{y} , we obtain that

$$d_{S_i}(\mathbf{y}) - d_{S_i}(\mathbf{x}) \leq \|\mathbf{x} - \mathbf{y}\|,$$

which, combined with (8.22), yields the inequality

$$|d_{S_i}(\mathbf{x}) - d_{S_i}(\mathbf{y})| \leq \|\mathbf{x} - \mathbf{y}\|. \quad (8.23)$$

Finally, for any $\mathbf{x}, \mathbf{y} \in \mathbb{E}$,

$$|f(\mathbf{x}) - f(\mathbf{y})| = \left| \max_{i=1,2,\dots,m} d_{S_i}(\mathbf{x}) - \max_{i=1,2,\dots,m} d_{S_i}(\mathbf{y}) \right| = |\|\mathbf{v}_x\|_\infty - \|\mathbf{v}_y\|_\infty|, \quad (8.24)$$

where $\mathbf{v}_x = (d_{S_i}(\mathbf{x}))_{i=1}^m \in \mathbb{R}^m$ and $\mathbf{v}_y = (d_{S_i}(\mathbf{y}))_{i=1}^m \in \mathbb{R}^m$. Using the triangle inequality for norms, we can continue (8.24) and obtain

$$\begin{aligned} |f(\mathbf{x}) - f(\mathbf{y})| &\leq |\|\mathbf{v}_x\|_\infty - \|\mathbf{v}_y\|_\infty| \\ &\leq \|\mathbf{v}_x - \mathbf{v}_y\|_\infty \\ &= \max_{i=1,2,\dots,m} |d_{S_i}(\mathbf{x}) - d_{S_i}(\mathbf{y})| \\ (8.23) \quad &\leq \|\mathbf{x} - \mathbf{y}\|. \quad \square \end{aligned}$$

Let us write explicitly the projected subgradient method with Polyak's stepsize rule as applied to problem (8.21). The method starts with an arbitrary $\mathbf{x}^0 \in \mathbb{E}$. If the k th iteration satisfies $\mathbf{x}^k \in S$, then we can pick $f'(\mathbf{x}^k) = \mathbf{0}$ and hence $\mathbf{x}^{k+1} = \mathbf{x}^k$. Otherwise, we take a step toward minus of the subgradient with Polyak's stepsize. By Theorem 3.50, to compute a subgradient of the objective function at the k th iterate, we can use the following procedure:

- (i) compute $i_k \in \operatorname{argmax}_{i=1,2,\dots,m} d_{S_i}(\mathbf{x}^k)$;
- (ii) take any $\mathbf{g}^k \in \partial d_{S_{i_k}}(\mathbf{x}^k)$.

By Example 3.49, we can (and actually must) choose the subgradient in $\partial d_{S_{i_k}}(\mathbf{x}^k)$ as $\mathbf{g}^k = \frac{\mathbf{x}^k - P_{S_{i_k}}(\mathbf{x}^k)}{d_{S_{i_k}}(\mathbf{x}^k)}$, and in this case the update step becomes

$$\begin{aligned} \mathbf{x}^{k+1} &= \mathbf{x}^k - \frac{d_{S_{i_k}}(\mathbf{x}^k) - f_{\text{opt}}}{\|\mathbf{g}^k\|^2} \cdot \frac{\mathbf{x}^k - P_{S_{i_k}}(\mathbf{x}^k)}{d_{S_{i_k}}(\mathbf{x}^k)} \\ &= \mathbf{x}^k - d_{S_{i_k}}(\mathbf{x}^k) \frac{\mathbf{x}^k - P_{S_{i_k}}(\mathbf{x}^k)}{d_{S_{i_k}}(\mathbf{x}^k)} \\ &= P_{S_{i_k}}(\mathbf{x}^k), \end{aligned}$$

where we used in the above the facts that $f_{\text{opt}} = 0$ and $\|\mathbf{g}^k\| = 1$. What we actually obtained is the *greedy projection algorithm*, which at each iteration projects the current iterate \mathbf{x}^k onto the farthest set among S_1, S_2, \dots, S_m . The algorithm is summarized below.

Greedy Projection Algorithm

Input: m nonempty closed and convex sets S_1, S_2, \dots, S_m .

Initialization: pick $\mathbf{x}^0 \in \mathbb{E}$.

General step: for any $k = 0, 1, 2, \dots$, execute the step

$$\mathbf{x}^{k+1} = P_{S_{i_k}}(\mathbf{x}^k),$$

where $i_k \in \operatorname{argmax}_{i=1,2,\dots,m} d_{S_i}(\mathbf{x}^k)$.

We can invoke Theorems 8.13 and 8.17 to obtain the following convergence result of the algorithm.

Theorem 8.21 (convergence of the greedy projection algorithm). *Let $S_1, S_2, \dots, S_m \subseteq \mathbb{E}$ be closed and convex sets such that $S \equiv \bigcap_{i=1}^m S_i \neq \emptyset$. Let $\{\mathbf{x}^k\}_{k \geq 0}$ be the sequence generated by the greedy projection algorithm.*

(a) *For any $k \geq 0$,*

$$\min_{n=0,1,2,\dots,k} \left\{ \max_{i=1,2,\dots,m} d(\mathbf{x}^n, S_i) \right\} \leq \frac{d_S(\mathbf{x}^0)}{\sqrt{k+1}}. \quad (8.25)$$

(b) *There exists $\mathbf{x}^* \in S$ such that $\mathbf{x}^k \rightarrow \mathbf{x}^*$ as $k \rightarrow \infty$.*

Proof. To prove part (a), define $f(\mathbf{x}) \equiv \max_{i=1,2,\dots,m} d(\mathbf{x}, S_i)$ and $C = \mathbb{E}$. Then the optimal set of the problem

$$\min\{f(\mathbf{x}) : \mathbf{x} \in C\}$$

is $X^* = S$. Assumption 8.7 is satisfied since f is proper closed and convex and $C = \mathbb{E}$ is obviously nonempty closed and convex and contained in $\text{int}(\text{dom}(f)) = \mathbb{E}$. The optimal set $X^* = S$ is nonempty by the assumption in the premise of the theorem. Assumption 8.12 is satisfied with $L_f = 1$ by Lemma 8.20 and Theorem 3.61. Therefore, all the assumptions of Theorem 8.13 are satisfied, and hence, since the greedy projection algorithm is the same as the projected subgradient method with Polyak's stepsize rule, the result (8.25) holds, as it is exactly part (c) of Theorem 8.13. Part (b) follows by invoking Theorem 8.17. \square

When $m = 2$, the algorithm amounts to the *alternating projection method*, which is described below.

Alternating Projection Method

Input: two nonempty closed and convex sets S_1, S_2 .

Initialization: pick $\mathbf{x}^0 \in S_2$ arbitrarily.

General step: for any $k = 0, 1, 2, \dots$, execute the following step:

$$\mathbf{x}^{k+1} = P_{S_2}(P_{S_1}(\mathbf{x}^k)).$$

If $S_1 \cap S_2 \neq \emptyset$, by Theorem 8.21, the sequence generated by the alternating projection method converges to a point in $S_1 \cap S_2$.

Corollary 8.22 (convergence of alternating projection). *Let S_1, S_2 be closed and convex sets such that $S \equiv S_1 \cap S_2 \neq \emptyset$. Let $\{\mathbf{x}^k\}_{k \geq 0}$ be the sequence generated by the alternating projection method with initial point $\mathbf{x}^0 \in S_2$. Then*

(a) for any $k \geq 0$,

$$\min_{n=0,1,2,\dots,k} d(\mathbf{x}^n, S_1) \leq \frac{d_S(\mathbf{x}^0)}{\sqrt{k+1}};$$

(b) there exists $\mathbf{x}^* \in S$ such that $\mathbf{x}^k \rightarrow \mathbf{x}^*$ as $k \rightarrow \infty$.

Example 8.23 (solution of linear feasibility problems). Consider the following system of linear equalities and inequalities:

$$\mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}, \quad (8.26)$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ has full row rank and $\mathbf{b} \in \mathbb{R}^m$. The system (8.26) is one of the standard forms of feasible sets of linear programming problems. One way to solve the problem of finding a solution to (8.26) is by employing the alternating projection method. Define

$$S_1 = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} = \mathbf{b}\}, \quad S_2 = \mathbb{R}_+^n.$$

The projections on S_1 and S_2 have analytic expressions (see Lemma 6.26):

$$P_{S_1}(\mathbf{x}) = \mathbf{x} - \mathbf{A}^T (\mathbf{AA}^T)^{-1} (\mathbf{Ax} - \mathbf{b}), \quad P_{S_2}(\mathbf{x}) = [\mathbf{x}]_+.$$

The alternating projection method for finding a solution to (8.26) takes the following form:

Algorithm 1

- **Initialization:** pick $\mathbf{x}^0 \in \mathbb{R}_+^n$.
- **General step ($k \geq 0$):** $\mathbf{x}^{k+1} = [\mathbf{x}^k - \mathbf{A}^T (\mathbf{AA}^T)^{-1} (\mathbf{Ax}^k - \mathbf{b})]_+$.

The general step of the above scheme involves the computation of the expression $(\mathbf{AA}^T)^{-1}(\mathbf{Ax}^k - \mathbf{b})$, which requires the computation of the matrix \mathbf{AA}^T , as well as the solution of the linear system $(\mathbf{AA}^T)\mathbf{z} = \mathbf{Ax}^k - \mathbf{b}$. In cases when these computations are too demanding (e.g., when the dimension is large), we can employ a different projection algorithm that avoids the necessity of solving a linear system. Specifically, denoting the i th row of \mathbf{A} by \mathbf{a}_i^T and defining

$$T_i = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}_i^T \mathbf{x} = b_i\}, \quad i = 1, 2, \dots, m, \quad T_{m+1} = \mathbb{R}_+^n,$$

we obtain that finding a solution to (8.26) is the same as finding a point in the intersection $\bigcap_{i=1}^{m+1} T_i$. Note that (see Lemma 6.26)

$$P_{T_i}(\mathbf{x}) = \mathbf{x} - \frac{\mathbf{a}_i^T \mathbf{x} - b_i}{\|\mathbf{a}_i\|_2^2} \mathbf{a}_i, \quad i = 1, 2, \dots, m.$$

Hence,

$$d_{T_i}(\mathbf{x}) = \|\mathbf{x} - P_{T_i}(\mathbf{x})\| = \frac{|\mathbf{a}_i^T \mathbf{x} - b_i|}{\|\mathbf{a}_i\|_2}.$$

We can now invoke the greedy projection method that has the following form:

Algorithm 2

- **Initialization:** pick $\mathbf{x}^0 \in \mathbb{E}$.
- **General step ($k = 0, 1, \dots$):**
 - compute $i_k \in \operatorname{argmax}_{i=1,2,\dots,m} \frac{|\mathbf{a}_i^T \mathbf{x}^k - b_i|}{\|\mathbf{a}_i\|_2}$.
 - if $\frac{|\mathbf{a}_{i_k}^T \mathbf{x}^k - b_{i_k}|}{\|\mathbf{a}_{i_k}\|_2} > \|\mathbf{x}^k - [\mathbf{x}^k]_+\|_2$, then

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \frac{\mathbf{a}_{i_k}^T \mathbf{x}^k - b_{i_k}}{\|\mathbf{a}_{i_k}\|_2^2} \mathbf{a}_{i_k}.$$
 - else,

$$\mathbf{x}^{k+1} = [\mathbf{x}^k]_+.$$

Algorithm 2 is simpler than Algorithm 1 in the sense that it requires much less operations per iteration. However, simplicity has its cost. Consider, for example, the instance

$$\mathbf{A} = \begin{pmatrix} 0 & 6 & -7 & 1 \\ -1 & 2 & 10 & -1 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 0 \\ 10 \end{pmatrix}.$$

Figure 8.3 shows the constraint violation of the two sequences generated by the two algorithms initialized with the zeros vector in the first 20 iterations. Obviously, in this case, Algorithm 1 (alternating projection) reached substantially better accuracies than Algorithm 2 (greedy projection). ■

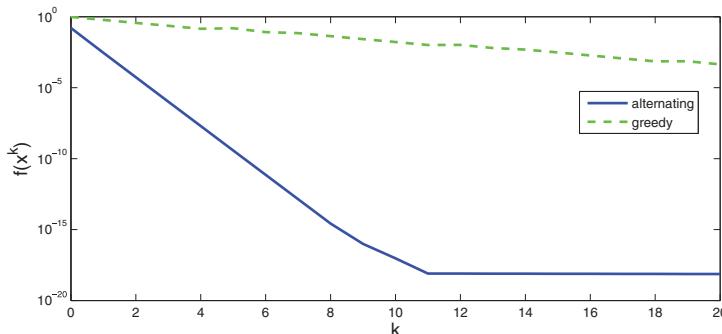


Figure 8.3. Constraints violation of alternating and greedy projection methods. Here $f(\mathbf{x}) = \max \left\{ \frac{|\mathbf{a}_1^T \mathbf{x} - b_1|}{\|\mathbf{a}_1\|_2}, \frac{|\mathbf{a}_2^T \mathbf{x} - b_2|}{\|\mathbf{a}_2\|_2}, \|\mathbf{x} - [\mathbf{x}]_+\|_2 \right\}$, where $\mathbf{a}_1^T = (0, 6, -7, 1)$, $\mathbf{a}_2^T = (-1, 2, 10, -1)$, and $\mathbf{b} = (0, 10)^T$.

8.2.4 Projected Subgradient with Dynamic Stepsizes

Polyak's stepsize is optimal in the sense that it minimizes the upper bound given in the fundamental inequality (8.12). However, a major disadvantage of this rule is that usually the optimal value f_{opt} is unknown, and in these (frequent) cases, the stepsize is incomputable. In this section we will show how to find computable

stepsize rules that still maintain the $O(1/\sqrt{k})$ rate of convergence result of the projected subgradient method. Theorem 8.25 below describes a simple condition on the stepsizes under which convergence of f_{best}^k to f_{opt} is guaranteed. The result uses the following technical lemma.

Lemma 8.24. *Suppose that Assumption 8.7 holds. Let $\{\mathbf{x}^k\}_{k \geq 0}$ be the sequence generated by the projected subgradient method with positive stepsizes $\{t_k\}_{k \geq 0}$. Then for any $\mathbf{x}^* \in X^*$ and nonnegative integer k ,*

$$\sum_{n=0}^k t_n(f(\mathbf{x}^n) - f_{\text{opt}}) \leq \frac{1}{2}\|\mathbf{x}^0 - \mathbf{x}^*\|^2 + \frac{1}{2} \sum_{n=0}^k t_n^2 \|f'(\mathbf{x}^n)\|^2. \quad (8.27)$$

Proof. By Lemma 8.11, for any $n \geq 0$ and $\mathbf{x}^* \in X^*$,

$$\frac{1}{2}\|\mathbf{x}^{n+1} - \mathbf{x}^*\|^2 \leq \frac{1}{2}\|\mathbf{x}^n - \mathbf{x}^*\|^2 - t_n(f(\mathbf{x}^n) - f_{\text{opt}}) + \frac{t_n^2}{2}\|f'(\mathbf{x}^n)\|^2.$$

Summing the above inequality over $n = 0, 1, \dots, k$ and arranging terms yields the following inequality:

$$\begin{aligned} \sum_{n=0}^k t_n(f(\mathbf{x}^n) - f_{\text{opt}}) &\leq \frac{1}{2}\|\mathbf{x}^0 - \mathbf{x}^*\|^2 - \frac{1}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^*\|^2 + \sum_{n=0}^k \frac{t_n^2}{2}\|f'(\mathbf{x}^n)\|^2 \\ &\leq \frac{1}{2}\|\mathbf{x}^0 - \mathbf{x}^*\|^2 + \frac{1}{2} \sum_{n=0}^k t_n^2 \|f'(\mathbf{x}^n)\|^2. \quad \square \end{aligned}$$

Theorem 8.25 (stepsize conditions warranting convergence of projected subgradient). *Suppose that Assumptions 8.7 and 8.12 hold. Let $\{\mathbf{x}^k\}_{k \geq 0}$ be the sequence generated by the projected subgradient method with positive stepsizes $\{t_k\}_{k \geq 0}$. If*

$$\frac{\sum_{n=0}^k t_n^2}{\sum_{n=0}^k t_n} \rightarrow 0 \text{ as } k \rightarrow \infty, \quad (8.28)$$

then

$$f_{\text{best}}^k - f_{\text{opt}} \rightarrow 0 \text{ as } k \rightarrow \infty, \quad (8.29)$$

where $\{f_{\text{best}}^k\}_{k \geq 0}$ is the sequence of best achieved values defined in (8.11).

Proof. Let L_f be a constant for which $\|\mathbf{g}\| \leq L_f$ for any $\mathbf{g} \in \partial f(\mathbf{x})$, $\mathbf{x} \in C$ whose existence is warranted by Assumption 8.12. Employing Lemma 8.24 and using the inequalities $\|f'(\mathbf{x}^n)\| \leq L_f$ and $f(\mathbf{x}^n) \geq f_{\text{best}}^k$ for $n \leq k$, we obtain

$$\left(\sum_{n=0}^k t_n \right) (f_{\text{best}}^k - f_{\text{opt}}) \leq \frac{1}{2}\|\mathbf{x}^0 - \mathbf{x}^*\|^2 + \frac{L_f^2}{2} \sum_{n=0}^k t_n^2.$$

Therefore,

$$f_{\text{best}}^k - f_{\text{opt}} \leq \frac{1}{2} \frac{\|\mathbf{x}^0 - \mathbf{x}^*\|^2}{\sum_{n=0}^k t_n} + \frac{L_f^2}{2} \frac{\sum_{n=0}^k t_n^2}{\sum_{n=0}^k t_n}.$$

The result (8.29) now follows by (8.28), and the fact that (8.28) implies the limit $\sum_{n=0}^k t_n \rightarrow \infty$ as $k \rightarrow \infty$. \square

By Theorem 8.25, we can pick, for example, the stepsizes as $t_k = \frac{1}{\sqrt{k+1}}$, and convergence of function values to f_{opt} will be guaranteed since $\sum_{n=0}^k \frac{1}{\sqrt{n+1}}$ is of the order of \sqrt{k} and $\sum_{n=0}^k \frac{1}{n+1}$ is of the order of $\log(k)$. We will analyze the convergence rate of the projected subgradient method when the stepsizes are chosen as $t_k = \frac{1}{\|f'(\mathbf{x}^k)\|\sqrt{k+1}}$ in Theorem 8.28 below. Note that in addition to proving the limit $f_{\text{best}}^k \rightarrow f_{\text{opt}}$, we will further show that the function values of a certain sequence of averages also converges to the optimal value. Such a result is called an *ergodic* convergence result.

To prove the result, we will be need to upper and lower bound sums of sequences of real numbers. For that, we will use the following technical lemma from calculus.

Lemma 8.26. *Let $f : [a - 1, b + 1] \rightarrow \mathbb{R}$ be a continuous nonincreasing function over $[a - 1, b + 1]$, where a and b are integer numbers satisfying $a \leq b$. Then*

$$\int_a^{b+1} f(t)dt \leq f(a) + f(a+1) + \cdots + f(b) \leq \int_{a-1}^b f(t)dt.$$

Using Lemma 8.26, we can prove the following lemma that will be useful in proving Theorem 8.28, as well as additional results in what follows.

Lemma 8.27. *Let $D \in \mathbb{R}$. Then*

(a) *for any $k \geq 1$,*

$$\frac{D + \sum_{n=0}^k \frac{1}{n+1}}{\sum_{n=0}^k \frac{1}{\sqrt{n+1}}} \leq \frac{D + 1 + \log(k+1)}{\sqrt{k+1}}; \quad (8.30)$$

(b) *for any $k \geq 2$,*

$$\frac{D + \sum_{n=\lceil k/2 \rceil}^k \frac{1}{n+1}}{\sum_{n=\lceil k/2 \rceil}^k \frac{1}{\sqrt{n+1}}} \leq \frac{4(D + \log(3))}{\sqrt{k+2}}. \quad (8.31)$$

Proof. (a) Using Lemma 8.26, we obtain the following inequalities:

$$\sum_{n=0}^k \frac{1}{n+1} = 1 + \sum_{n=1}^k \frac{1}{n+1} \leq 1 + \int_0^k \frac{1}{x+1} dx = 1 + \log(k+1), \quad (8.32)$$

$$\sum_{n=0}^k \frac{1}{\sqrt{n+1}} \geq \int_0^{k+1} \frac{1}{\sqrt{x+1}} dx = 2\sqrt{k+2} - 2 \geq \sqrt{k+1}, \quad (8.33)$$

where the last inequality holds for all $k \geq 1$. The result (8.30) now follows immediately from (8.32) and (8.33).

(b) Using Lemma 8.26, we obtain the following inequalities for any $k \geq 2$:

$$\begin{aligned} \sum_{n=\lceil k/2 \rceil}^k \frac{1}{n+1} &\leq \int_{\lceil k/2 \rceil - 1}^k \frac{dt}{t+1} = \log(k+1) - \log(\lceil k/2 \rceil) \\ &= \log\left(\frac{k+1}{\lceil 0.5k \rceil}\right) \leq \log\left(\frac{k+1}{0.5k}\right) = \log\left(2 + \frac{2}{k}\right) \\ &\leq \log(3) \end{aligned} \quad (8.34)$$

and

$$\begin{aligned} \sum_{n=\lceil k/2 \rceil}^k \frac{1}{\sqrt{n+1}} &\geq \int_{\lceil k/2 \rceil}^{k+1} \frac{dt}{\sqrt{t+1}} = 2\sqrt{k+2} - 2\sqrt{\lceil k/2 \rceil + 1} \\ &\geq 2\sqrt{k+2} - 2\sqrt{k/2+2} = \frac{4(k+2) - 4(0.5k+2)}{2\sqrt{k+2} + 2\sqrt{0.5k+2}} \\ &= \frac{k}{\sqrt{k+2} + \sqrt{0.5k+2}} \geq \frac{k}{2\sqrt{k+2}} \\ &\geq \frac{1}{4}\sqrt{k+2}, \end{aligned} \quad (8.35)$$

where the last inequality holds since $k \geq 2$. The result (8.31) now follows by combining (8.34) and (8.35). \square

We are now ready to prove the convergence result.

Theorem 8.28 ($O(\log(k)/\sqrt{k})$ rate of convergence of projected subgradient). Suppose that Assumptions 8.7 and 8.12 hold. Let $\{\mathbf{x}^k\}_{k \geq 0}$ be the sequence generated by the projected subgradient method with stepsizes $t_k = \frac{1}{\|f'(\mathbf{x}^k)\|\sqrt{k+1}}$ if $f'(\mathbf{x}^k) \neq 0$ and $t_k = \frac{1}{L_f}$ otherwise. Then

(a) for any $k \geq 1$,

$$f_{\text{best}}^k - f_{\text{opt}} \leq \frac{L_f}{2} \frac{\|\mathbf{x}^0 - \mathbf{x}^*\|^2 + 1 + \log(k+1)}{\sqrt{k+1}},$$

where $\{f_{\text{best}}^k\}_{k \geq 0}$ is the sequence of best achieved values defined in (8.11);

(b) for any $k \geq 1$,

$$f(\mathbf{x}^{(k)}) - f_{\text{opt}} \leq \frac{L_f}{2} \frac{\|\mathbf{x}^0 - \mathbf{x}^*\|^2 + 1 + \log(k+1)}{\sqrt{k+1}},$$

where

$$\mathbf{x}^{(k)} = \frac{1}{\sum_{n=0}^k t_n} \sum_{n=0}^k t_n \mathbf{x}^n.$$

Proof. Using (8.27) along with the inequality $f(\mathbf{x}^n) \geq f_{\text{best}}^k$ for any $n = 0, 1, 2, \dots, k$, we obtain

$$f_{\text{best}}^k - f_{\text{opt}} \leq \frac{1}{2} \frac{\|\mathbf{x}^0 - \mathbf{x}^*\|^2 + \sum_{n=0}^k t_n^2 \|f'(\mathbf{x}^n)\|^2}{\sum_{n=0}^k t_n}. \quad (8.36)$$

Alternatively, by Jensen's inequality

$$f(\mathbf{x}^{(k)}) \leq \frac{1}{\sum_{n=0}^k t_n} \sum_{n=0}^k t_n f(\mathbf{x}^n),$$

which, along with (8.27), yields

$$f(\mathbf{x}^{(k)}) - f_{\text{opt}} \leq \frac{1}{2} \frac{\|\mathbf{x}^0 - \mathbf{x}^*\|^2 + \sum_{n=0}^k t_n^2 \|f'(\mathbf{x}^n)\|^2}{\sum_{n=0}^k t_n}. \quad (8.37)$$

Therefore, combining (8.36) and (8.37), we have

$$\max\{f_{\text{best}}^k - f_{\text{opt}}, f(\mathbf{x}^{(k)}) - f_{\text{opt}}\} \leq \frac{1}{2} \frac{\|\mathbf{x}^0 - \mathbf{x}^*\|^2 + \sum_{n=0}^k t_n^2 \|f'(\mathbf{x}^n)\|^2}{\sum_{n=0}^k t_n}.$$

By the definition of t_n , $t_n^2 \|f'(\mathbf{x}^n)\|^2 \leq \frac{1}{n+1}$ (satisfied as equality when $f'(\mathbf{x}^n) \neq \mathbf{0}$ and as a strict inequality when $f'(\mathbf{x}^n) = \mathbf{0}$); in addition, since $\|f'(\mathbf{x}^n)\| \leq L_f$, we have $t_n \geq \frac{1}{L_f \sqrt{n+1}}$. Therefore,

$$\max\{f_{\text{best}}^k - f_{\text{opt}}, f(\mathbf{x}^{(k)}) - f_{\text{opt}}\} \leq \frac{L_f}{2} \frac{\|\mathbf{x}^0 - \mathbf{x}^*\|^2 + \sum_{n=0}^k \frac{1}{n+1}}{\sum_{n=0}^k \frac{1}{\sqrt{n+1}}}. \quad (8.38)$$

Invoking Lemma 8.27(a) with $D = \|\mathbf{x}^0 - \mathbf{x}^*\|^2$ implies the inequality

$$\max\{f_{\text{best}}^k - f_{\text{opt}}, f(\mathbf{x}^{(k)}) - f_{\text{opt}}\} \leq \frac{L_f}{2} \frac{\|\mathbf{x}^0 - \mathbf{x}^*\|^2 + 1 + \log(k+1)}{\sqrt{k+1}},$$

which is equivalent to the validity of the two claims (a) and (b). \square

Remark 8.29. The sequence of averages $\mathbf{x}^{(k)}$ as defined in Theorem 8.28 can be computed in an adaptive way by noting that the following simple recursion relation holds:

$$\mathbf{x}^{(k+1)} = \frac{T_k}{T_{k+1}} \mathbf{x}^{(k)} + \frac{t_{k+1}}{T_{k+1}} \mathbf{x}^{k+1},$$

where $T_k \equiv \sum_{n=0}^k t_n$ can be computed by the obvious recursion relation $T_{k+1} = T_k + t_{k+1}$.

The $O(\log(k)/\sqrt{k})$ rate of convergence proven in Theorem 8.28 is worse than the $O(1/\sqrt{k})$ rate established in Theorem 8.13 for the version of the projected subgradient method with Polyak's stepsize. It is possible to prove an $O(1/\sqrt{k})$ rate of convergence if we assume in addition that the feasible set C is compact. Note that by Theorem 3.16, the compactness of C implies the validity of Assumption 8.12, but we will nonetheless explicitly state it in the following result.

Theorem 8.30 ($O(1/\sqrt{k})$ rate of convergence of projected subgradient). Suppose that Assumptions 8.7 and 8.12 hold and assume that C is compact. Let Θ be an upper bound on the half-squared diameter of C :

$$\Theta \geq \max_{\mathbf{x}, \mathbf{y} \in C} \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2.$$

Let $\{\mathbf{x}^k\}_{k \geq 0}$ be the sequence generated by the projected subgradient method with stepsizes chosen as either

$$t_k = \frac{\sqrt{2\Theta}}{L_f \sqrt{k+1}} \quad (8.39)$$

or

$$t_k = \begin{cases} \frac{\sqrt{2\Theta}}{\|f'(\mathbf{x}^k)\| \sqrt{k+1}}, & f'(\mathbf{x}^k) \neq \mathbf{0}, \\ \frac{\sqrt{2\Theta}}{L_f \sqrt{k+1}}, & f'(\mathbf{x}^k) = \mathbf{0}. \end{cases} \quad (8.40)$$

Then for all $k \geq 2$,

$$f_{\text{best}}^k - f_{\text{opt}} \leq \frac{\delta L_f \sqrt{2\Theta}}{\sqrt{k+2}},$$

where $\delta = 2(1 + \log(3))$ and f_{best}^k is the sequence of best achieved values defined in (8.11).

Proof. By Lemma 8.11, for any $n \geq 0$,

$$\frac{1}{2} \|\mathbf{x}^{n+1} - \mathbf{x}^*\|^2 \leq \frac{1}{2} \|\mathbf{x}^n - \mathbf{x}^*\|^2 - t_n(f(\mathbf{x}^n) - f_{\text{opt}}) + \frac{t_n^2}{2} \|f'(\mathbf{x}^n)\|^2.$$

Summing the above inequality over $n = \lceil k/2 \rceil, \lceil k/2 \rceil + 1, \dots, k$, we obtain

$$\begin{aligned} \sum_{n=\lceil k/2 \rceil}^k t_n(f(\mathbf{x}^n) - f_{\text{opt}}) &\leq \frac{1}{2} \|\mathbf{x}^{\lceil k/2 \rceil} - \mathbf{x}^*\|^2 - \frac{1}{2} \|\mathbf{x}^{k+1} - \mathbf{x}^*\|^2 + \sum_{n=\lceil k/2 \rceil}^k \frac{t_n^2}{2} \|f'(\mathbf{x}^n)\|^2 \\ &\leq \Theta + \sum_{n=\lceil k/2 \rceil}^k \frac{t_n^2}{2} \|f'(\mathbf{x}^n)\|^2 \\ &\leq \Theta + \Theta \sum_{n=\lceil k/2 \rceil}^k \frac{1}{n+1}, \end{aligned} \quad (8.41)$$

where the last inequality is due to the fact that in either of the definitions of the stepsizes (8.39), (8.40), $t_n^2 \|f'(\mathbf{x}^n)\|^2 \leq \frac{2\Theta}{n+1}$.

Since $t_n \geq \frac{\sqrt{2\Theta}}{L_f \sqrt{n+1}}$ and $f(\mathbf{x}^n) \geq f_{\text{best}}^k$ for all $n \leq k$, it follows that

$$\sum_{n=\lceil k/2 \rceil}^k t_n(f(\mathbf{x}^n) - f_{\text{opt}}) \geq \left(\sum_{n=\lceil k/2 \rceil}^k \frac{\sqrt{2\Theta}}{L_f \sqrt{n+1}} \right) (f_{\text{best}}^k - f_{\text{opt}}). \quad (8.42)$$

Therefore, combining (8.41) and (8.42) yields

$$f_{\text{best}}^k - f_{\text{opt}} \leq \frac{L_f \sqrt{\Theta}}{\sqrt{2}} \frac{1 + \sum_{n=\lceil k/2 \rceil}^k \frac{1}{n+1}}{\sum_{n=\lceil k/2 \rceil}^k \frac{1}{\sqrt{n+1}}}, \quad (8.43)$$

which, combined with Lemma 8.27(b), yields the desired result. \square

8.2.5 The Strongly Convex Case⁴³

We will now show that if f is in addition strongly convex, then the $O(1/\sqrt{k})$ rate of convergence result can be improved to a rate of $O(1/k)$. The stepsizes used in order to achieve this improved rate diminish at an order of $1/k$. We will also use the growth property of strongly convex functions described in Theorem 5.25(b) in order to show a result on the rate of convergence of the sequence $\{\mathbf{x}^k\}_{k \geq 0}$ to an optimal solution.

Theorem 8.31 ($O(1/k)$ rate of convergence of projected subgradient for strongly convex functions). *Suppose that Assumptions 8.7 and 8.12 hold. Assume in addition that f is σ -strongly convex for some $\sigma > 0$, and let \mathbf{x}^* be its unique minimizer. Let $\{\mathbf{x}^k\}_{k \geq 0}$ be the sequence generated by the projected subgradient method with stepsize $t_k = \frac{2}{\sigma(k+1)}$.*

- (a) *Let $\{f_{\text{best}}^k\}_{k \geq 0}$ be the sequence of best achieved values defined in (8.11). Then for any $k \geq 0$,*

$$f_{\text{best}}^k - f_{\text{opt}} \leq \frac{2L_f^2}{\sigma(k+1)}. \quad (8.44)$$

In addition,

$$\|\mathbf{x}^{i_k} - \mathbf{x}^*\| \leq \frac{2L_f}{\sigma\sqrt{k+1}}, \quad (8.45)$$

where $i_k \in \arg\min_{i=0,1,\dots,k} f(\mathbf{x}^i)$.

- (b) *Define the sequence of averages:*

$$\mathbf{x}^{(k)} = \sum_{n=0}^k \alpha_n^k \mathbf{x}^n,$$

where $\alpha_n^k \equiv \frac{2n}{k(k+1)}$. Then for all $k \geq 0$,

$$f(\mathbf{x}^{(k)}) - f_{\text{opt}} \leq \frac{2L_f^2}{\sigma(k+1)}. \quad (8.46)$$

In addition,

$$\|\mathbf{x}^{(k)} - \mathbf{x}^*\| \leq \frac{2L_f}{\sigma\sqrt{k+1}}. \quad (8.47)$$

Proof. (a) Repeating the arguments in the proof of Lemma 8.11, we can write for any $n \geq 0$

$$\begin{aligned} \|\mathbf{x}^{n+1} - \mathbf{x}^*\|^2 &= \|P_C(\mathbf{x}^n - t_n f'(\mathbf{x}^n)) - P_C(\mathbf{x}^*)\|^2 \\ &\leq \|\mathbf{x}^n - t_n f'(\mathbf{x}^n) - \mathbf{x}^*\|^2 \\ &= \|\mathbf{x}^n - \mathbf{x}^*\|^2 - 2t_n \langle f'(\mathbf{x}^n), \mathbf{x}^n - \mathbf{x}^* \rangle + t_n^2 \|f'(\mathbf{x}^n)\|^2. \end{aligned} \quad (8.48)$$

⁴³The analysis of the stochastic and deterministic projected subgradient method in the strongly convex case is based on the work of Lacoste-Julien, Schmidt, and Bach [77].

Since f is σ -strongly convex, it follows by Theorem 5.24 that

$$f(\mathbf{x}^*) \geq f(\mathbf{x}^n) + \langle f'(\mathbf{x}^n), \mathbf{x}^* - \mathbf{x}^n \rangle + \frac{\sigma}{2} \|\mathbf{x}^n - \mathbf{x}^*\|^2.$$

That is,

$$\langle f'(\mathbf{x}^n), \mathbf{x}^n - \mathbf{x}^* \rangle \geq f(\mathbf{x}^n) - f_{\text{opt}} + \frac{\sigma}{2} \|\mathbf{x}^n - \mathbf{x}^*\|^2.$$

Plugging the above into (8.48), we obtain that

$$\|\mathbf{x}^{n+1} - \mathbf{x}^*\|^2 \leq (1 - \sigma t_n) \|\mathbf{x}^n - \mathbf{x}^*\|^2 - 2t_n(f(\mathbf{x}^n) - f_{\text{opt}}) + t_n^2 \|f'(\mathbf{x}^n)\|^2.$$

Rearranging terms, dividing by $2t_n$, and using the bound $\|f'(\mathbf{x}^n)\| \leq L_f$ leads to the following inequality:

$$f(\mathbf{x}^n) - f_{\text{opt}} \leq \frac{1}{2}(t_n^{-1} - \sigma) \|\mathbf{x}^n - \mathbf{x}^*\|^2 - \frac{1}{2}t_n^{-1} \|\mathbf{x}^{n+1} - \mathbf{x}^*\|^2 + \frac{t_n}{2} L_f^2.$$

Plugging $t_n = \frac{2}{\sigma(n+1)}$ into the latter inequality, we obtain

$$f(\mathbf{x}^n) - f_{\text{opt}} \leq \frac{\sigma(n-1)}{4} \|\mathbf{x}^n - \mathbf{x}^*\|^2 - \frac{\sigma(n+1)}{4} \|\mathbf{x}^{n+1} - \mathbf{x}^*\|^2 + \frac{1}{\sigma(n+1)} L_f^2.$$

Multiplying the above by n yields the following inequality:

$$n(f(\mathbf{x}^n) - f_{\text{opt}}) \leq \frac{\sigma n(n-1)}{4} \|\mathbf{x}^n - \mathbf{x}^*\|^2 - \frac{\sigma(n+1)n}{4} \|\mathbf{x}^{n+1} - \mathbf{x}^*\|^2 + \frac{n}{\sigma(n+1)} L_f^2.$$

Summing over $n = 0, 1, \dots, k$, we conclude that

$$\sum_{n=0}^k n(f(\mathbf{x}^n) - f_{\text{opt}}) \leq 0 - \frac{\sigma}{4} k(k+1) \|\mathbf{x}^{k+1} - \mathbf{x}^*\|^2 + \frac{L_f^2}{\sigma} \sum_{n=0}^k \frac{n}{n+1} \leq \frac{L_f^2 k}{\sigma}. \quad (8.49)$$

Therefore, using the inequality $f(\mathbf{x}^n) \geq f_{\text{best}}^k$ for all $n = 0, 1, \dots, k$, it follows that

$$\left(\sum_{n=0}^k n \right) (f_{\text{best}}^k - f_{\text{opt}}) \leq \frac{L_f^2 k}{\sigma},$$

which by the known identity $\sum_{n=0}^k n = \frac{k(k+1)}{2}$ shows that

$$f_{\text{best}}^k - f_{\text{opt}} \leq \frac{2L_f^2}{\sigma(k+1)}, \quad (8.50)$$

meaning that (8.44) holds. To prove (8.45), note that $f_{\text{best}}^k = f(\mathbf{x}^{i_k})$, and hence by Theorem 5.25(b) employed on the σ -strongly convex function $f + \delta_C$ and (8.50),

$$\frac{\sigma}{2} \|\mathbf{x}^{i_k} - \mathbf{x}^*\|^2 \leq f_{\text{best}}^k - f_{\text{opt}} \leq \frac{2L_f^2}{\sigma(k+1)},$$

which is the same as

$$\|\mathbf{x}^{i_k} - \mathbf{x}^*\| \leq \frac{2L_f}{\sigma\sqrt{k+1}}.$$

(b) To establish the ergodic convergence, we begin by dividing (8.49) by $\frac{k(k+1)}{2}$ to obtain

$$\sum_{n=0}^k \alpha_n^k (f(\mathbf{x}^n) - f_{\text{opt}}) \leq \frac{2L_f^2}{\sigma(k+1)}.$$

By Jensen's inequality (utilizing the fact that $(\alpha_n^k)_{n=0}^k \in \Delta_{k+1}$),

$$f(\mathbf{x}^{(k)}) - f_{\text{opt}} = f\left(\sum_{n=0}^k \alpha_n^k \mathbf{x}^n\right) - f_{\text{opt}} \leq \sum_{n=0}^k \alpha_n^k (f(\mathbf{x}^n) - f_{\text{opt}}) \leq \frac{2L_f^2}{\sigma(k+1)},$$

meaning that (8.46) holds. The result (8.47) now follows by the same arguments used to prove (8.45) in part (a). \square

Remark 8.32. *The sequence of averages $\mathbf{x}^{(k)}$ as defined in Theorem 8.31 can be computed in an adaptive way by noting that the following simple recursion relation holds:*

$$\mathbf{x}^{(k+1)} = \frac{k}{k+2} \mathbf{x}^{(k)} + \frac{2}{k+2} \mathbf{x}^{k+1}.$$

The $O(1/k)$ rate of convergence of the sequence of function values naturally leads to the observation that to obtain an ε -optimal solution, an order of $1/\varepsilon$ iterations is required. The proof is trivial and follows the argument of the proof of Theorem 8.18.

Theorem 8.33 (complexity of projected subgradient for strongly convex functions). *Under the setting and assumptions of Theorem 8.31, for any nonnegative integer k satisfying*

$$k \geq \frac{2L_f^2}{\sigma\varepsilon} - 1,$$

it holds that

$$f_{\text{best}}^k - f_{\text{opt}} \leq \varepsilon$$

and

$$f(\mathbf{x}^{(k)}) - f_{\text{opt}} \leq \varepsilon.$$

8.3 The Stochastic Projected Subgradient Method

8.3.1 Setting and Method

In this section we still study the model (8.10) under Assumption 8.7. The main difference will be that at each iteration we do not necessarily utilize a subgradient at the current iterate \mathbf{x}^k as the update direction vector, but rather a random estimator \mathbf{g}^k of a subgradient of f at \mathbf{x}^k (a precise characterization will be given in Assumption 8.34). The method is therefore given as follows.

The Stochastic Projected Subgradient Method

Initialization: pick $\mathbf{x}^0 \in C$ arbitrarily.

General step: for any $k = 0, 1, 2, \dots$ execute the following steps:

- (A) pick a stepsize $t_k > 0$ and a random vector $\mathbf{g}^k \in \mathbb{E}$;
- (B) set $\mathbf{x}^{k+1} = P_C(\mathbf{x}^k - t_k \mathbf{g}^k)$.

Obviously, since the vectors \mathbf{g}^k are random vectors, so are the iterate vectors \mathbf{x}^k . The exact assumptions on the random vectors \mathbf{g}^k are given below.

Assumption 8.34.

- (A) (**unbiasedness**) For any $k \geq 0$, $\mathbb{E}(\mathbf{g}^k | \mathbf{x}^k) \in \partial f(\mathbf{x}^k)$.
- (B) (**boundedness**) There exists a constant $\tilde{L}_f > 0$ such that for any $k \geq 0$, $\mathbb{E}(\|\mathbf{g}^k\|^2 | \mathbf{x}^k) \leq \tilde{L}_f^2$.

Part (A) of the assumption says that \mathbf{g}^k is an *unbiased estimator* of a subgradient at \mathbf{x}^k . This assumption can also be written as

$$f(\mathbf{z}) \geq f(\mathbf{x}^k) + \langle \mathbb{E}(\mathbf{g}^k | \mathbf{x}^k), \mathbf{z} - \mathbf{x}^k \rangle \text{ for all } \mathbf{z} \in \text{dom}(f).$$

The constant \tilde{L}_f from part (B) of Assumption 8.34 is not necessarily a Lipschitz constant of f as in the deterministic case.

8.3.2 Analysis

The analysis of the stochastic projected subgradient is almost identical to the analysis of the deterministic method. We gather the main results in the following theorem.

Theorem 8.35 (convergence of stochastic projected gradient). Suppose that Assumptions 8.7 and 8.34 hold. Let $\{\mathbf{x}^k\}_{k \geq 0}$ be the sequence generated by the stochastic projected subgradient method with positive stepsizes $\{t_k\}_{k \geq 0}$, and let $\{f_{\text{best}}^k\}_{k \geq 0}$ be the sequence of best achieved values defined in (8.11).

- (a) If $\frac{\sum_{n=0}^k t_n^2}{\sum_{n=0}^k t_n} \rightarrow 0$ as $k \rightarrow \infty$, then $\mathbb{E}(f_{\text{best}}^k) \rightarrow f_{\text{opt}}$ as $k \rightarrow \infty$.
- (b) Assume that C is compact. Let \tilde{L}_f be the positive constant defined in Assumption 8.34, and let Θ be an upper bound on the half-squared diameter of C :

$$\Theta \geq \max_{\mathbf{x}, \mathbf{y} \in C} \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2. \quad (8.51)$$

If $t_k = \frac{\sqrt{2\Theta}}{L_f \sqrt{k+1}}$, then for all $k \geq 2$,

$$\mathbb{E}(f_{\text{best}}^k) - f_{\text{opt}} \leq \frac{\delta \tilde{L}_f \sqrt{2\Theta}}{\sqrt{k+2}},$$

where $\delta = 2(1 + \log(3))$.

Proof. We have for any $n \geq 0$,

$$\begin{aligned}
\mathbb{E}(\|\mathbf{x}^{n+1} - \mathbf{x}^*\|^2 | \mathbf{x}^n) &= \mathbb{E}(\|P_C(\mathbf{x}^n - t_n \mathbf{g}^n) - P_C(\mathbf{x}^*)\|^2 | \mathbf{x}^n) \\
&\stackrel{(*)}{\leq} \mathbb{E}(\|\mathbf{x}^n - t_n \mathbf{g}^n - \mathbf{x}^*\|^2 | \mathbf{x}^n) \\
&= \|\mathbf{x}^n - \mathbf{x}^*\|^2 - 2t_n \mathbb{E}(\langle \mathbf{g}^n, \mathbf{x}^n - \mathbf{x}^* \rangle | \mathbf{x}^n) + t_n^2 \mathbb{E}(\|\mathbf{g}^n\|^2 | \mathbf{x}^n) \\
&= \|\mathbf{x}^n - \mathbf{x}^*\|^2 - 2t_n \langle \mathbb{E}(\mathbf{g}^n | \mathbf{x}^n), \mathbf{x}^n - \mathbf{x}^* \rangle + t_n^2 \mathbb{E}(\|\mathbf{g}^n\|^2 | \mathbf{x}^n) \\
&\stackrel{(**)}{\leq} \|\mathbf{x}^n - \mathbf{x}^*\|^2 - 2t_n \langle \mathbb{E}(\mathbf{g}^n | \mathbf{x}^n), \mathbf{x}^n - \mathbf{x}^* \rangle + t_n^2 \tilde{L}_f^2 \\
&\stackrel{(***)}{\leq} \|\mathbf{x}^n - \mathbf{x}^*\|^2 - 2t_n(f(\mathbf{x}^n) - f_{\text{opt}}) + t_n^2 \tilde{L}_f^2,
\end{aligned}$$

where $(*)$ follows by the nonexpansiveness property of the orthogonal projection operator (Theorem 6.42), and $(**)$ and $(***)$ follow by Assumption 8.34.

Taking expectation w.r.t. \mathbf{x}^n , we obtain

$$\mathbb{E}(\|\mathbf{x}^{n+1} - \mathbf{x}^*\|^2) \leq \mathbb{E}(\|\mathbf{x}^n - \mathbf{x}^*\|^2) - 2t_n(\mathbb{E}(f(\mathbf{x}^n)) - f_{\text{opt}}) + t_n^2 \tilde{L}_f^2.$$

Summing over $n = m, m+1, \dots, k$ (where m is an integer satisfying $m \leq k$),

$$\mathbb{E}(\|\mathbf{x}^{k+1} - \mathbf{x}^*\|^2) \leq \mathbb{E}(\|\mathbf{x}^m - \mathbf{x}^*\|^2) - 2 \sum_{n=m}^k t_n(\mathbb{E}(f(\mathbf{x}^n)) - f_{\text{opt}}) + \tilde{L}_f^2 \sum_{n=m}^k t_n^2.$$

Therefore,

$$\sum_{n=m}^k t_n(\mathbb{E}(f(\mathbf{x}^n)) - f_{\text{opt}}) \leq \frac{1}{2} \left[\mathbb{E}(\|\mathbf{x}^m - \mathbf{x}^*\|^2) + \tilde{L}_f^2 \sum_{n=m}^k t_n^2 \right],$$

which implies

$$\left(\sum_{n=m}^k t_n \right) \left(\min_{n=m, m+1, \dots, k} \mathbb{E}(f(\mathbf{x}^n)) - f_{\text{opt}} \right) \leq \frac{1}{2} \left[\mathbb{E}(\|\mathbf{x}^m - \mathbf{x}^*\|^2) + \tilde{L}_f^2 \sum_{n=m}^k t_n^2 \right].$$

Using the inequality⁴⁴

$$\mathbb{E}(f_{\text{best}}^k) \leq \mathbb{E}\left(\min_{n=m, m+1, \dots, k} f(\mathbf{x}^n)\right) \leq \min_{n=m, m+1, \dots, k} \mathbb{E}(f(\mathbf{x}^n)),$$

we can conclude that

$$\mathbb{E}(f_{\text{best}}^k) - f_{\text{opt}} \leq \frac{\mathbb{E}(\|\mathbf{x}^m - \mathbf{x}^*\|^2) + \tilde{L}_f^2 \sum_{n=m}^k t_n^2}{2 \sum_{n=m}^k t_n}. \quad (8.52)$$

Plugging $m = 0$ in (8.52), we obtain

$$\mathbb{E}(f_{\text{best}}^k) - f_{\text{opt}} \leq \frac{\|\mathbf{x}^0 - \mathbf{x}^*\|^2 + \tilde{L}_f^2 \sum_{n=0}^k t_n^2}{2 \sum_{n=0}^k t_n}.$$

⁴⁴The fact that for any p random variables $\mathbb{E}(\min\{X_1, X_2, \dots, X_p\}) \leq \min_{i=1,2,\dots,p} \mathbb{E}(X_i)$ follows by the following argument: for any $i = 1, 2, \dots, p$, the inequality $\min\{X_1, X_2, \dots, X_p\} \leq X_i$ holds. Taking expectation leads to the inequality $\mathbb{E}(\min\{X_1, X_2, \dots, X_p\}) \leq \mathbb{E}(X_i)$ for any i , from which the desired inequality $\mathbb{E}(\min\{X_1, X_2, \dots, X_p\}) \leq \min_{i=1,2,\dots,p} \mathbb{E}(X_i)$ follows.

Therefore, if $\frac{\sum_{n=0}^k t_n^2}{\sum_{n=0}^k t_n} \rightarrow 0$, then $E(f_{\text{best}}^k) \rightarrow f_{\text{opt}}$ as $k \rightarrow \infty$, proving claim (a). To show the validity of claim (b), use (8.52) with $m = \lceil k/2 \rceil$ and the bound (8.51) to obtain

$$E(f_{\text{best}}^k) - f_{\text{opt}} \leq \frac{\Theta + \frac{\tilde{L}_f^2}{2} \sum_{n=\lceil k/2 \rceil}^k t_n^2}{\sum_{n=\lceil k/2 \rceil}^k t_n}.$$

Taking $t_n = \frac{\sqrt{2\Theta}}{L_f \sqrt{n+1}}$, we get

$$E(f_{\text{best}}^k) - f_{\text{opt}} \leq \frac{\tilde{L}_f \sqrt{2\Theta}}{2} \frac{1 + \sum_{n=\lceil k/2 \rceil}^k \frac{1}{n+1}}{\sum_{n=\lceil k/2 \rceil}^k \frac{1}{\sqrt{n+1}}},$$

which, combined with Lemma 8.27(b), yields the desired result. \square

Example 8.36 (minimization of sum of convex functions). Consider the optimization model

$$(P) \quad \min \left\{ f(\mathbf{x}) \equiv \sum_{i=1}^m f_i(\mathbf{x}) : \mathbf{x} \in C \right\},$$

where $f_1, f_2, \dots, f_m : \mathbb{E} \rightarrow (-\infty, \infty]$ are proper closed and convex functions. Suppose that Assumption 8.7 holds and that C is compact, which in particular implies the validity of Assumption 8.12 with some constant L_f . By Theorem 3.61 L_f is a Lipschitz constant of f over C . Let Θ be some upper bound on the half-squared diameter of C :

$$\frac{1}{2} \max_{\mathbf{x}, \mathbf{y} \in C} \|\mathbf{x} - \mathbf{y}\|^2 \leq \Theta.$$

In addition, we will assume that for any $i = 1, 2, \dots, m$, there exists a constant L_{f_i} for which

$$\|\mathbf{g}\| \leq L_{f_i} \text{ for all } \mathbf{g} \in \partial f_i(\mathbf{x}), \mathbf{x} \in C.$$

By Theorem 3.61, L_{f_i} is a Lipschitz constant of f_i over C . We can consider two options for solving the main problem (P). The first is to employ the projected subgradient method (we assume that $f'(\mathbf{x}^k) \neq \mathbf{0}$):

Algorithm 1

- **Initialization:** pick $\mathbf{x}^0 \in C$.
- **General step ($k \geq 0$):** choose $f'_i(\mathbf{x}^k) \in \partial f_i(\mathbf{x}^k)$, $i = 1, 2, \dots, m$, and compute

$$\mathbf{x}^{k+1} = P_C \left(\mathbf{x}^k - \frac{\sqrt{2\Theta}}{\|\sum_{i=1}^m f'_i(\mathbf{x}^k)\| \sqrt{k+1}} \left(\sum_{i=1}^m f'_i(\mathbf{x}^k) \right) \right).$$

By Theorem 8.30, the following efficiency estimate holds for any $k \geq 2$:

$$f_{\text{best}}^k - f_{\text{opt}} \leq \frac{\delta L_f \sqrt{2\Theta}}{\sqrt{k+2}}, \quad (8.53)$$

where $\delta = 2(1 + \log(3))$. A direct consequence is that in order to obtain an ε -optimal solution,

$$N_1 = \max \left\{ \frac{2\delta^2 L_f^2 \Theta}{\varepsilon^2} - 2, 2 \right\}$$

iterations are sufficient. Since the computation of the subgradient of f at \mathbf{x}^k by the formula $\sum_{i=1}^m f'_i(\mathbf{x}^k)$ might be too expensive in cases where m is large, we can alternatively employ the stochastic projected subgradient method where at iteration k , we define the unbiased estimate of $f'(\mathbf{x}^k)$ as

$$\mathbf{g}^k = m f'_{i_k}(\mathbf{x}^k),$$

where i_k is randomly picked from $\{1, 2, \dots, m\}$ via a uniform distribution. Obviously,

$$\mathbb{E}(\mathbf{g}^k | \mathbf{x}^k) = \sum_{i=1}^m \frac{1}{m} m f'_i(\mathbf{x}^k) = \sum_{i=1}^m f'_i(\mathbf{x}^k) \in \partial f(\mathbf{x}^k),$$

where the inclusion in $\partial f(\mathbf{x}^k)$ follows by the sum rule of subdifferential calculus (Corollary 3.38). Also,

$$\mathbb{E}(\|\mathbf{g}^k\|^2 | \mathbf{x}^k) = \frac{1}{m} \sum_{i=1}^m m^2 \|f'_{i_k}(\mathbf{x}^k)\|^2 \leq m \sum_{i=1}^m L_{f_i}^2 \equiv \tilde{L}_f^2.$$

The stochastic projected subgradient method employed on problem (P) therefore takes the following form:

Algorithm 2

- **Initialization:** pick $\mathbf{x}^0 \in C$.
- **General step** ($k \geq 0$):
 - pick $i_k \in \{1, 2, \dots, m\}$ randomly via a uniform distribution and $f'_{i_k}(\mathbf{x}^k) \in \partial f_{i_k}(\mathbf{x}^k)$;
 - compute

$$\mathbf{x}^{k+1} = P_C \left(\mathbf{x}^k - \frac{\sqrt{2\Theta}m}{\tilde{L}_f \sqrt{k+1}} f'_{i_k}(\mathbf{x}^k) \right),$$
 where $\tilde{L}_f = \sqrt{m} \sqrt{\sum_{i=1}^m L_{f_i}^2}$.

Invoking Theorem 8.35, we obtain that

$$\mathbb{E}(f_{\text{best}}^k) - f_{\text{opt}} \leq \frac{\delta \sqrt{m} \sqrt{\sum_{i=1}^m L_{f_i}^2} \sqrt{2\Theta}}{\sqrt{k+2}}. \quad (8.54)$$

In particular,

$$N_2 = \max \left\{ \frac{2\delta^2 m \Theta \sum_{i=1}^m L_{f_i}^2}{\varepsilon^2} - 2, 2 \right\}$$

iterations are sufficient in order to ensure that an ε -optimal solution in expectation is reached. The natural question that arises is, *is it possible to compare between the two algorithms?* The answer is actually not clear. We can compare the two quantities N_2 and N_1 , but there are two major flaws in such a comparison. First, in a sense this is like comparing apples and oranges since N_1 considers a sequence of function values, while N_2 refers to a sequence of expected function values. In addition, recall that N_2 and N_1 only provide upper bounds on the amount of iterations required to obtain an ε -optimal solution (deterministically or in expectation). Comparison of *upper bounds* might be influenced dramatically by the tightness of the upper bounds. Disregarding these drawbacks, estimating the ratio between N_2 and N_1 , while neglecting the constant terms, which do not depend on ε , we get

$$\frac{N_2}{N_1} \approx \frac{\frac{2\delta^2 m \Theta \sum_{i=1}^m L_{f_i}^2}{\varepsilon^2}}{\frac{2\delta^2 L_f^2 \Theta}{\varepsilon^2}} = \frac{m \sum_{i=1}^m L_{f_i}^2}{L_f^2} \equiv \beta.$$

The value of β obviously depends on the specific problem at hand. Let us, for example, consider the instance in which $f_i(\mathbf{x}) = |\mathbf{a}_i^T \mathbf{x} + b_i|$, $i = 1, 2, \dots, m$, where $\mathbf{a}_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$, and $C = B_{\|\cdot\|_2}[\mathbf{0}, 1]$. In this case,

$$f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} + \mathbf{b}\|_1,$$

where \mathbf{A} is the $m \times n$ matrix whose rows are \mathbf{a}_i^T and $\mathbf{b} = (b_i)_{i=1}^m$. Since

$$\partial f_i(\mathbf{x}) = \begin{cases} \mathbf{a}_i, & \mathbf{a}_i^T \mathbf{x} + b_i > 0, \\ -\mathbf{a}_i, & \mathbf{a}_i^T \mathbf{x} + b_i < 0, \\ \{\xi \mathbf{a}_i : \xi \in [-1, 1]\}, & \mathbf{a}_i^T \mathbf{x} + b_i = 0, \end{cases}$$

it follows that we can choose $L_{f_i} = \|\mathbf{a}_i\|_2$. To estimate L_f , note that by Example 3.44, any $\mathbf{g} \in \partial f(\mathbf{x})$ has the form $\mathbf{g} = \mathbf{A}^T \boldsymbol{\eta}$ for some $\boldsymbol{\eta} \in [-1, 1]^m$, which in particular implies that $\|\boldsymbol{\eta}\|_2 \leq \sqrt{m}$. Thus,

$$\|\mathbf{g}\|_2 = \|\mathbf{A}^T \boldsymbol{\eta}\|_2 \leq \|\mathbf{A}^T\|_{2,2} \|\boldsymbol{\eta}\|_2 \leq \sqrt{m} \|\mathbf{A}^T\|_{2,2},$$

where $\|\cdot\|_{2,2}$ is the spectral norm. We can therefore choose $L_f = \sqrt{m} \|\mathbf{A}^T\|_{2,2}$. Thus,

$$\beta = \frac{m \sum_{i=1}^m \|\mathbf{a}_i\|_2^2}{m \|\mathbf{A}^T\|_{2,2}^2} = \frac{\|\mathbf{A}^T\|_F^2}{\|\mathbf{A}^T\|_{2,2}^2} = \frac{\sum_{i=1}^n \lambda_i(\mathbf{A}\mathbf{A}^T)}{\max_{i=1,2,\dots,n} \lambda_i(\mathbf{A}\mathbf{A}^T)},$$

where $\lambda_1(\mathbf{A}\mathbf{A}^T) \geq \lambda_2(\mathbf{A}\mathbf{A}^T) \geq \dots \geq \lambda_n(\mathbf{A}\mathbf{A}^T)$ are the eigenvalues of $\mathbf{A}\mathbf{A}^T$ ordered nonincreasingly. Using the fact that for any nonnegative numbers $\alpha_1, \alpha_2, \dots, \alpha_m$ the inequalities

$$\max_{i=1,2,\dots,m} \alpha_i \leq \sum_{i=1}^m \alpha_i \leq m \max_{i=1,2,\dots,m} \alpha_i$$

hold, we obtain that $1 \leq \beta \leq m$. The extreme case $\beta = m$ is actually quite logical in the sense that the number of subgradient computations per iteration in Algorithm 1 is m times larger than what is required in Algorithm 2, and it is thus not surprising

that the amount of iterations of Algorithm 2 might be m times larger than what is required by Algorithm 1 to obtain the same level of accuracy. What is much less intuitive is the case when β is close 1. In these instances, the two algorithms require (modulo the faults of this comparison) the same order of iterations to obtain the same order of accuracy. For example, when \mathbf{A} “close” to be of rank one, then β will be close to 1. In these cases, the two algorithms should perform similarly, although Algorithm 2 is much less computationally demanding. We can explain this result by the fact that in this instance the vectors \mathbf{a}_i are “almost” proportional to each other, and thus all the subgradient directions $f'_i(\mathbf{x}^k)$ are similar. ■

8.3.3 Stochastic Projected Subgradient—The Strongly Convex Case

The analysis of the stochastic projected subgradient method is almost identical to the one presented for the deterministic case in Theorem 8.31, but for the sake of completeness we present the result and its complete proof.

Theorem 8.37 (convergence of stochastic projected subgradient for strongly convex functions). *Suppose that Assumptions 8.7 and 8.34 hold. Let \tilde{L}_f be the positive constant defined in Assumption 8.34. Assume in addition that f is σ -strongly convex for some $\sigma > 0$. Let $\{\mathbf{x}^k\}_{k \geq 0}$ be the sequence generated by the stochastic projected subgradient method with stepsizes $t_k = \frac{2}{\sigma(k+1)}$.*

- (a) *Let $\{f_{\text{best}}^k\}_{k \geq 0}$ be the sequence of best achieved values defined in (8.11). Then for any $k \geq 0$,*

$$\mathbb{E}(f_{\text{best}}^k) - f_{\text{opt}} \leq \frac{2\tilde{L}_f^2}{\sigma(k+1)}.$$

- (b) *Define the sequence of averages*

$$\mathbf{x}^{(k)} = \sum_{n=0}^k \alpha_n^k \mathbf{x}^n,$$

where $\alpha_n^k \equiv \frac{2n}{k(k+1)}$. Then

$$\mathbb{E}(f(\mathbf{x}^{(k)})) - f_{\text{opt}} \leq \frac{2\tilde{L}_f^2}{\sigma(k+1)}.$$

Proof. (a) For any $\mathbf{x}^* \in X^*$ and $n \geq 0$,

$$\begin{aligned} \mathbb{E}(\|\mathbf{x}^{n+1} - \mathbf{x}^*\|^2 | \mathbf{x}^n) &= \mathbb{E}(\|P_C(\mathbf{x}^n - t_n \mathbf{g}^n) - P_C(\mathbf{x}^*)\|^2 | \mathbf{x}^n) \\ &\leq \mathbb{E}(\|\mathbf{x}^n - t_n \mathbf{g}^n - \mathbf{x}^*\|^2 | \mathbf{x}^n) \\ &= \|\mathbf{x}^n - \mathbf{x}^*\|^2 - 2t_n \langle \mathbb{E}(\mathbf{g}^n | \mathbf{x}^n), \mathbf{x}^n - \mathbf{x}^* \rangle \\ &\quad + t_n^2 \mathbb{E}(\|\mathbf{g}^n\|^2 | \mathbf{x}^n). \end{aligned} \tag{8.55}$$

Since f is σ -strongly convex and $\mathbb{E}(\mathbf{g}^n | \mathbf{x}^n) \in \partial f(\mathbf{x}^n)$, it follows by Theorem 5.24(ii) that

$$f(\mathbf{x}^*) \geq f(\mathbf{x}^n) + \langle \mathbb{E}(\mathbf{g}^n | \mathbf{x}^n), \mathbf{x}^* - \mathbf{x}^n \rangle + \frac{\sigma}{2} \|\mathbf{x}^n - \mathbf{x}^*\|^2.$$

That is,

$$\langle \mathbb{E}(\mathbf{g}^n | \mathbf{x}^n), \mathbf{x}^n - \mathbf{x}^* \rangle \geq f(\mathbf{x}^n) - f_{\text{opt}} + \frac{\sigma}{2} \|\mathbf{x}^n - \mathbf{x}^*\|^2.$$

Plugging the above into (8.55), we obtain that

$$\mathbb{E}(\|\mathbf{x}^{n+1} - \mathbf{x}^*\|^2 | \mathbf{x}^n) \leq (1 - \sigma t_n) \|\mathbf{x}^n - \mathbf{x}^*\|^2 - 2t_n(f(\mathbf{x}^n) - f_{\text{opt}}) + t_n^2 \mathbb{E}(\|\mathbf{g}^n\|^2 | \mathbf{x}^n).$$

Rearranging terms, dividing by $2t_n$, and using the bound $\mathbb{E}(\|\mathbf{g}^n\|^2 | \mathbf{x}^n) \leq \tilde{L}_f^2$ leads to the following inequality:

$$f(\mathbf{x}^n) - f_{\text{opt}} \leq \frac{1}{2}(t_n^{-1} - \sigma) \|\mathbf{x}^n - \mathbf{x}^*\|^2 - \frac{1}{2}t_n^{-1} \mathbb{E}(\|\mathbf{x}^{n+1} - \mathbf{x}^*\|^2 | \mathbf{x}^n) + \frac{t_n}{2} \tilde{L}_f^2.$$

Plugging $t_n = \frac{2}{\sigma(n+1)}$ into the last inequality, we obtain

$$f(\mathbf{x}^n) - f_{\text{opt}} \leq \frac{\sigma(n-1)}{4} \|\mathbf{x}^n - \mathbf{x}^*\|^2 - \frac{\sigma(n+1)}{4} \mathbb{E}(\|\mathbf{x}^{n+1} - \mathbf{x}^*\|^2 | \mathbf{x}^n) + \frac{1}{\sigma(n+1)} \tilde{L}_f^2.$$

Multiplying the above by n and taking expectation w.r.t. \mathbf{x}^n yields the following inequality:

$$\begin{aligned} n(\mathbb{E}(f(\mathbf{x}^n)) - f_{\text{opt}}) &\leq \frac{\sigma n(n-1)}{4} \mathbb{E}(\|\mathbf{x}^n - \mathbf{x}^*\|^2) - \frac{\sigma(n+1)n}{4} \mathbb{E}(\|\mathbf{x}^{n+1} - \mathbf{x}^*\|^2) \\ &\quad + \frac{n}{\sigma(n+1)} \tilde{L}_f^2. \end{aligned}$$

Summing over $n = 0, 1, \dots, k$,

$$\sum_{n=0}^k n(\mathbb{E}(f(\mathbf{x}^n)) - f_{\text{opt}}) \leq 0 - \frac{\sigma}{4} k(k+1) \mathbb{E}(\|\mathbf{x}^{k+1} - \mathbf{x}^*\|^2) + \frac{\tilde{L}_f^2}{\sigma} \sum_{n=0}^k \frac{n}{n+1} \leq \frac{\tilde{L}_f^2 k}{\sigma}. \quad (8.56)$$

Therefore, using the inequality $\mathbb{E}(f(\mathbf{x}^n)) \geq \mathbb{E}(f_{\text{best}}^k)$ for all $n = 0, 1, \dots, k$, it follows that

$$\left(\sum_{n=0}^k n \right) (\mathbb{E}(f_{\text{best}}^k) - f_{\text{opt}}) \leq \frac{\tilde{L}_f^2 k}{\sigma},$$

which, by the identity $\sum_{n=0}^k n = \frac{k(k+1)}{2}$, implies that

$$\mathbb{E}(f_{\text{best}}^k) - f_{\text{opt}} \leq \frac{2\tilde{L}_f^2}{\sigma(k+1)}.$$

(b) Divide (8.56) by $\frac{k(k+1)}{2}$ to obtain

$$\sum_{n=0}^k \alpha_n^k (\mathbb{E}(f(\mathbf{x}^n)) - f_{\text{opt}}) \leq \frac{2L_f^2}{\sigma(k+1)}.$$

By Jensen's inequality (utilizing the fact that $(\alpha_n^k)_{n=0}^k \in \Delta_{k+1}$), we finally obtain

$$\begin{aligned} \mathbb{E}(f(\mathbf{x}^{(k)})) - f_{\text{opt}} &= \mathbb{E}\left(f\left(\sum_{n=0}^k \alpha_n^k \mathbf{x}^n\right)\right) - f_{\text{opt}} \leq \sum_{n=0}^k \alpha_n^k (\mathbb{E}(f(\mathbf{x}^n)) - f_{\text{opt}}) \\ &\leq \frac{2\tilde{L}_f^2}{\sigma(k+1)}. \quad \square \end{aligned}$$

8.4 The Incremental Projected Subgradient Method

Consider the main model (8.10), where f has the form $f(\mathbf{x}) = \sum_{i=1}^m f_i(\mathbf{x})$. That is, we consider the problem

$$\min \left\{ f(\mathbf{x}) = \sum_{i=1}^m f_i(\mathbf{x}) : \mathbf{x} \in C \right\}. \quad (8.57)$$

In addition to Assumption 8.7, we make the following assumption.

Assumption 8.38.

- (a) f_i is proper closed and convex for any $i = 1, 2, \dots, m$.
- (b) There exists $L > 0$ for which $\|\mathbf{g}\| \leq L$ for any $\mathbf{g} \in \partial f_i(\mathbf{x})$, $i = 1, 2, \dots, m$, $\mathbf{x} \in C$.

In Example 8.36 the same model was also considered, and a projected subgradient method that takes a step toward a direction of the form $-f'_{i_k}(\mathbf{x}^k)$ was analyzed. The index i_k was chosen in Example 8.36 randomly by a uniform distribution over the indices $\{1, 2, \dots, m\}$, and the natural question that arises is whether we can obtain similar convergence results when i_k is chosen in a deterministic manner. We will consider the variant in which the indices are chosen in a deterministic cyclic order. The resulting method is called the *incremental projected subgradient method*. We will show that although the analysis is much more involved, it is still possible to obtain similar rates of convergence (albeit with worse constants).

An iteration of the incremental projected subgradient method is divided into subiterations. Let \mathbf{x}^k be the k th iterate vector. Then we define $\mathbf{x}^{k,0} = \mathbf{x}^k$ and produce m subiterations $\mathbf{x}^{k,1}, \mathbf{x}^{k,2}, \dots, \mathbf{x}^{k,m}$ by the rule that $\mathbf{x}^{k,i+1} = P_C(\mathbf{x}^{k,i} - t_k \mathbf{g}^{k,i})$, where $\mathbf{g}^{k,i} \in \partial f_{i+1}(\mathbf{x}^{k,i})$ and $t_k > 0$ is a positive stepsize. Finally, the next iterate is defined by $\mathbf{x}^{k+1} = \mathbf{x}^{k,m}$.

The Incremental Projected Subgradient Method

Initialization: pick $\mathbf{x}^0 \in C$ arbitrarily.

General step: for any $k = 0, 1, 2, \dots$ execute the following steps:

- (a) set $\mathbf{x}^{k,0} = \mathbf{x}^k$ and pick a stepsize $t_k > 0$;
- (b) for any $i = 0, 1, \dots, m-1$ compute

$$\mathbf{x}^{k,i+1} = P_C(\mathbf{x}^{k,i} - t_k \mathbf{g}^{k,i}),$$

where $\mathbf{g}^{k,i} \in \partial f_{i+1}(\mathbf{x}^{k,i})$;

- (c) set $\mathbf{x}^{k+1} = \mathbf{x}^{k,m}$.

The fundamental inequality from which convergence results can be deduced is proven in the following lemma. The result is similar to the result in Lemma 8.11, but the proof is considerably more complicated.

Lemma 8.39 (fundamental inequality for the incremental projected subgradient method).⁴⁵ Suppose that Assumptions 8.7 and 8.38 hold, and let $\{\mathbf{x}^k\}_{k \geq 0}$ be the sequence generated by the incremental projected subgradient method with positive stepsizes $\{t_k\}_{k \geq 0}$. Then for any $k \geq 0$,

$$\|\mathbf{x}^{k+1} - \mathbf{x}^*\|^2 \leq \|\mathbf{x}^k - \mathbf{x}^*\|^2 - 2t_k(f(\mathbf{x}^k) - f_{\text{opt}}) + t_k^2 m^2 L^2. \quad (8.58)$$

Proof. For any $\mathbf{x}^* \in X^*$, $k \geq 0$ and $i \in \{0, 1, \dots, m-1\}$,

$$\begin{aligned} \|\mathbf{x}^{k,i+1} - \mathbf{x}^*\|^2 &= \|P_C(\mathbf{x}^{k,i} - t_k \mathbf{g}^{k,i}) - \mathbf{x}^*\|^2 \\ &= \|P_C(\mathbf{x}^{k,i} - t_k \mathbf{g}^{k,i}) - P_C(\mathbf{x}^*)\|^2 \\ &\stackrel{(*)}{\leq} \|\mathbf{x}^{k,i} - t_k \mathbf{g}^{k,i} - \mathbf{x}^*\|^2 \\ &\stackrel{(**)}{\leq} \|\mathbf{x}^{k,i} - \mathbf{x}^*\|^2 - 2t_k \langle \mathbf{g}^{k,i}, \mathbf{x}^{k,i} - \mathbf{x}^* \rangle + t_k^2 L^2 \\ &\stackrel{(***)}{\leq} \|\mathbf{x}^{k,i} - \mathbf{x}^*\|^2 - 2t_k(f_{i+1}(\mathbf{x}^{k,i}) - f_{i+1}(\mathbf{x}^*)) + t_k^2 L^2, \end{aligned}$$

where $(*)$ follows by the nonexpansivity property of the orthogonal projection operator (Theorem 6.42(b)), $(**)$ by Assumption 8.38(b), and $(***)$ by the subgradient inequality. Summing the inequality over $i = 0, 1, \dots, m-1$ and using the identities $\mathbf{x}^{k,0} = \mathbf{x}^k$, $\mathbf{x}^{k,m} = \mathbf{x}^{k+1}$, we obtain that for any $\mathbf{x}^* \in X^*$,

$$\begin{aligned} \|\mathbf{x}^{k+1} - \mathbf{x}^*\|^2 &\leq \|\mathbf{x}^k - \mathbf{x}^*\|^2 - 2t_k \sum_{i=0}^{m-1} (f_{i+1}(\mathbf{x}^{k,i}) - f_{i+1}(\mathbf{x}^*)) + t_k^2 m L^2 \\ &= \|\mathbf{x}^k - \mathbf{x}^*\|^2 - 2t_k \left(f(\mathbf{x}^k) - f_{\text{opt}} + \sum_{i=0}^{m-1} (f_{i+1}(\mathbf{x}^{k,i}) - f_{i+1}(\mathbf{x}^k)) \right) + t_k^2 m L^2 \\ &\leq \|\mathbf{x}^k - \mathbf{x}^*\|^2 - 2t_k(f(\mathbf{x}^k) - f_{\text{opt}}) + 2t_k \sum_{i=0}^{m-1} L \|\mathbf{x}^{k,i} - \mathbf{x}^k\| + t_k^2 m L^2, \quad (8.59) \end{aligned}$$

where in the last inequality we used the fact that by Assumptions 8.7 and 8.38, $C \subseteq \text{int}(\text{dom}(f)) \subseteq \text{int}(\text{dom}(f_{i+1}))$ and $\|\mathbf{g}\| \leq L$ for all $\mathbf{g} \in \partial f_{i+1}(\mathbf{x})$, $\mathbf{x} \in C$, and thus, by Theorem 3.61, f_{i+1} is Lipschitz with constant L over C .

Now, using the nonexpansivity of the orthogonal projection operator,

$$\|\mathbf{x}^{k,1} - \mathbf{x}^k\| = \|P_C(\mathbf{x}^{k,0} - t_k \mathbf{g}^{k,0}) - P_C(\mathbf{x}^k)\| \leq t_k \|\mathbf{g}^{k,0}\| \leq t_k L.$$

Similarly,

$$\|\mathbf{x}^{k,2} - \mathbf{x}^k\| = \|P_C(\mathbf{x}^{k,1} - t_k \mathbf{g}^{k,1}) - P_C(\mathbf{x}^k)\| \leq \|\mathbf{x}^{k,1} - \mathbf{x}^k\| + t_k \|\mathbf{g}^{k,1}\| \leq 2t_k L.$$

In general, for any $i = 0, 1, 2, \dots, m-1$,

$$\|\mathbf{x}^{k,i} - \mathbf{x}^k\| \leq t_k i L,$$

and we can thus continue (8.59) and deduce that

$$\begin{aligned} \|\mathbf{x}^{k+1} - \mathbf{x}^*\|^2 &\leq \|\mathbf{x}^k - \mathbf{x}^*\|^2 - 2t_k(f(\mathbf{x}^k) - f_{\text{opt}}) + 2t_k^2 \sum_{i=0}^{m-1} i L^2 + t_k^2 m L^2 \\ &= \|\mathbf{x}^k - \mathbf{x}^*\|^2 - 2t_k(f(\mathbf{x}^k) - f_{\text{opt}}) + t_k^2 m^2 L^2. \quad \square \end{aligned}$$

⁴⁵The fundamental inequality for the incremental projected subgradient method is taken from Nedić and Bertsekas [89].

From this point, equipped with Lemma 8.39, we can use the same techniques used in the proofs of Theorems 8.25 and 8.30, for example, and establish the following result, whose proof is detailed here for the sake of completeness.

Theorem 8.40 (convergence of incremental projected subgradient). *Suppose that Assumptions 8.7 and 8.38 hold. Let $\{\mathbf{x}^k\}_{k \geq 0}$ be the sequence generated by the incremental stochastic projected subgradient method with positive stepsizes $\{t_k\}_{k \geq 0}$, and let $\{f_{\text{best}}^k\}_{k \geq 0}$ be the sequence of best achieved values defined in (8.11).*

- (a) *If $\frac{\sum_{n=0}^k t_n^2}{\sum_{n=0}^k t_n} \rightarrow 0$ as $k \rightarrow \infty$, then $f_{\text{best}}^k \rightarrow f_{\text{opt}}$ as $k \rightarrow \infty$.*
- (b) *Assume that C is compact. Let Θ be an upper bound on the half-squared diameter of C :*

$$\Theta \geq \max_{\mathbf{x}, \mathbf{y} \in C} \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2.$$

If $t_k = \frac{\sqrt{\Theta}}{Lm\sqrt{k+1}}$, then for all $k \geq 2$,

$$f_{\text{best}}^k - f_{\text{opt}} \leq \frac{\delta m L \sqrt{\Theta}}{\sqrt{k+2}},$$

where $\delta = 2(2 + \log(3))$.

Proof. By Lemma 8.39, for any $n \geq 0$,

$$\|\mathbf{x}^{n+1} - \mathbf{x}^*\|^2 \leq \|\mathbf{x}^n - \mathbf{x}^*\|^2 - 2t_n(f(\mathbf{x}^n) - f_{\text{opt}}) + L^2 m^2 t_n^2. \quad (8.60)$$

Summing (8.60) over $n = p, p+1, \dots, k$, we obtain

$$\|\mathbf{x}^{k+1} - \mathbf{x}^*\|^2 \leq \|\mathbf{x}^p - \mathbf{x}^*\|^2 - 2 \sum_{n=p}^k t_n(f(\mathbf{x}^n) - f_{\text{opt}}) + L^2 m^2 \sum_{n=p}^k t_n^2.$$

Therefore,

$$2 \sum_{n=p}^k t_n(f(\mathbf{x}^n) - f_{\text{opt}}) \leq \|\mathbf{x}^p - \mathbf{x}^*\|^2 + L^2 m^2 \sum_{n=p}^k t_n^2,$$

and hence

$$f_{\text{best}}^k - f_{\text{opt}} \leq \frac{\|\mathbf{x}^p - \mathbf{x}^*\|^2 + L^2 m^2 \sum_{n=p}^k t_n^2}{2 \sum_{n=p}^k t_n}. \quad (8.61)$$

Plugging $p = 0$ into (8.61), we obtain

$$f_{\text{best}}^k - f_{\text{opt}} \leq \frac{\|\mathbf{x}^0 - \mathbf{x}^*\|^2 + L^2 m^2 \sum_{n=0}^k t_n^2}{2 \sum_{n=0}^k t_n}.$$

Therefore, if $\frac{\sum_{n=0}^k t_n^2}{\sum_{n=0}^k t_n} \rightarrow 0$ as $k \rightarrow \infty$, then $f_{\text{best}}^k \rightarrow f_{\text{opt}}$ as $k \rightarrow \infty$, proving claim (a). To show the validity of claim (b), use (8.61) with $p = \lceil k/2 \rceil$ to obtain

$$f_{\text{best}}^k - f_{\text{opt}} \leq \frac{2\Theta + L^2 m^2 \sum_{n=\lceil k/2 \rceil}^k t_n^2}{2 \sum_{n=\lceil k/2 \rceil}^k t_n}.$$

Take $t_n = \frac{\sqrt{\Theta}}{Lm\sqrt{n+1}}$. Then we get

$$f_{\text{best}}^k - f_{\text{opt}} \leq \frac{Lm\sqrt{\Theta}}{2} \frac{2 + \sum_{n=\lceil k/2 \rceil}^k \frac{1}{n+1}}{\sum_{n=\lceil k/2 \rceil}^k \frac{1}{\sqrt{n+1}}},$$

which, combined with Lemma 8.27(b) (with $D = 2$), yields the desired result. \square

8.5 The Dual Projected Subgradient Method

8.5.1 The Dual Problem

Consider the problem

$$\begin{aligned} f_{\text{opt}} = \min & \quad f(\mathbf{x}) \\ \text{s.t.} & \quad \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \\ & \quad \mathbf{x} \in X, \end{aligned} \tag{8.62}$$

where the following assumptions are made.

Assumption 8.41.

- (A) $X \subseteq \mathbb{E}$ is convex.
- (B) $f : \mathbb{E} \rightarrow \mathbb{R}$ is convex.
- (C) $\mathbf{g}(\cdot) = (g_1(\cdot), g_2(\cdot), \dots, g_m(\cdot))^T$, where $g_1, g_2, \dots, g_m : \mathbb{E} \rightarrow \mathbb{R}$ are convex.
- (D) The problem has a finite optimal value denoted by f_{opt} , and the optimal set, denoted by X^* , is nonempty.
- (E) There exists $\bar{\mathbf{x}} \in X$ for which $\mathbf{g}(\bar{\mathbf{x}}) < \mathbf{0}$.
- (F) For any $\boldsymbol{\lambda} \in \mathbb{R}_+^m$, the problem $\min_{\mathbf{x} \in X} \{f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x})\}$ has an optimal solution.

The Lagrangian dual objective function of problem (8.62) is given by

$$q(\boldsymbol{\lambda}) = \min_{\mathbf{x} \in X} \left\{ L(\mathbf{x}; \boldsymbol{\lambda}) \equiv f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x}) \right\}. \tag{8.63}$$

By Assumption 8.41(F), the minimization problem in (8.63) possesses a solution, and thus, in particular, $q(\boldsymbol{\lambda})$ is finite for any $\boldsymbol{\lambda} \in \mathbb{R}_+^m$. Recall that q is concave over \mathbb{R}_+^m (as a minimum of affine and, in particular, concave functions), and hence the dual problem, which is given by

$$q_{\text{opt}} = \max \{q(\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \mathbb{R}_+^m\}, \tag{8.64}$$

is a convex problem, as it consists of maximizing a concave function over a convex set. We note that the dual problem is defined in the space \mathbb{R}^m , which we assume in this context to be endowed with the dot product and the l_2 -norm.

By Theorem A.1 and Assumption 8.41, it follows that strong duality holds for the primal-dual pair of problems (8.62) and (8.64), namely,

$$f_{\text{opt}} = q_{\text{opt}}$$

and the optimal solution of the dual problem is attained. We will denote the optimal set of the dual problem as Λ^* .

An interesting property of the dual problem under the Slater-type assumption (part (E) of Assumption 8.41) is that its superlevel sets are bounded.

Theorem 8.42 (boundedness of superlevel sets of the dual objective function).⁴⁶ Suppose that Assumption 8.41 holds. Let $\bar{\mathbf{x}} \in X$ be a point satisfying $\mathbf{g}(\bar{\mathbf{x}}) < 0$ whose existence is warranted by Assumption 8.41(E). Let $\mu \in \mathbb{R}$. Then for any $\boldsymbol{\lambda} \in S_\mu \equiv \{\boldsymbol{\lambda} \in \mathbb{R}_+^m : q(\boldsymbol{\lambda}) \geq \mu\}$,

$$\|\boldsymbol{\lambda}\|_2 \leq \frac{f(\bar{\mathbf{x}}) - \mu}{\min_{j=1,2,\dots,m} \{-g_j(\bar{\mathbf{x}})\}}.$$

Proof. Since $\boldsymbol{\lambda} \in S_\mu$, we have

$$\mu \leq q(\boldsymbol{\lambda}) \leq f(\bar{\mathbf{x}}) + \boldsymbol{\lambda}^T \mathbf{g}(\bar{\mathbf{x}}) = f(\bar{\mathbf{x}}) + \sum_{j=1}^m \lambda_j g_j(\bar{\mathbf{x}}).$$

Therefore,

$$-\sum_{j=1}^m \lambda_j g_j(\bar{\mathbf{x}}) \leq f(\bar{\mathbf{x}}) - \mu,$$

which, by the facts that $\lambda_j \geq 0$ and $g_j(\bar{\mathbf{x}}) < 0$ for all j , implies that

$$\sum_{j=1}^m \lambda_j \leq \frac{f(\bar{\mathbf{x}}) - \mu}{\min_{j=1,2,\dots,m} \{-g_j(\bar{\mathbf{x}})\}}.$$

Finally, since $\boldsymbol{\lambda} \geq \mathbf{0}$, we have that $\|\boldsymbol{\lambda}\|_2 \leq \sum_{j=1}^m \lambda_j$, and the desired result is established. \square

Taking $\mu = f_{\text{opt}} = q_{\text{opt}}$, we have $S_\mu = \Lambda^*$, and Theorem 8.42 amounts to the following corollary describing a bound on the dual optimal set.

Corollary 8.43 (boundedness of the optimal dual set). Suppose that Assumption 8.41 holds, and let Λ^* be the optimal set of the dual problem (8.64). Let $\bar{\mathbf{x}} \in X$ be a point satisfying $\mathbf{g}(\bar{\mathbf{x}}) < 0$ whose existence is warranted by Assumption 8.41(E). Then for any $\boldsymbol{\lambda} \in \Lambda^*$,

$$\|\boldsymbol{\lambda}\|_2 \leq \frac{f(\bar{\mathbf{x}}) - f_{\text{opt}}}{\min_{j=1,2,\dots,m} \{-g_j(\bar{\mathbf{x}})\}}.$$

⁴⁶Theorem 8.42 is Lemma 1 from Nedić and Ozdaglar [90].

8.5.2 The Dual Projected Subgradient Method

We begin by recalling how to compute a subgradient of minus of the dual objective function. By Example 3.7, if for a given $\boldsymbol{\lambda} \in \mathbb{R}_+^m$ the minimum of the problem defining $q(\boldsymbol{\lambda})$ is attained at $\mathbf{x}_{\boldsymbol{\lambda}} \in X$, meaning if $q(\boldsymbol{\lambda}) = f(\mathbf{x}_{\boldsymbol{\lambda}}) + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x}_{\boldsymbol{\lambda}})$, then $-\mathbf{g}(\mathbf{x}_{\boldsymbol{\lambda}}) \in \partial(-q)(\boldsymbol{\lambda})$.

Using the above expression for the subgradient of $-q$, we can define the projected subgradient method employed on the dual problem.

The Dual Projected Subgradient Method

Initialization: pick $\boldsymbol{\lambda}^0 \in \mathbb{R}_+^m$ arbitrarily.

General step: for any $k = 0, 1, 2, \dots$ execute the following steps:

- (a) pick a positive number γ_k ;
- (b) compute $\mathbf{x}^k \in \operatorname{argmin}_{\mathbf{x} \in X} \left\{ f(\mathbf{x}) + (\boldsymbol{\lambda}^k)^T \mathbf{g}(\mathbf{x}) \right\}$;
- (c) if $\mathbf{g}(\mathbf{x}^k) = \mathbf{0}$, then **terminate** with an output \mathbf{x}^k ; otherwise,

$$\boldsymbol{\lambda}^{k+1} = \left[\boldsymbol{\lambda}^k + \gamma_k \frac{\mathbf{g}(\mathbf{x}^k)}{\|\mathbf{g}(\mathbf{x}^k)\|_2} \right]_+$$

The stepsize $\frac{\gamma_k}{\|\mathbf{g}(\mathbf{x}^k)\|_2}$ is similar in form to the normalized stepsizes considered in Section 8.2.4. The fact that the condition $\mathbf{g}(\mathbf{x}^k) = \mathbf{0}$ guarantees that \mathbf{x}^k is an optimal solution of problem (8.62) is established in the following lemma.

Lemma 8.44. *Suppose that Assumption 8.41 holds. Let $\bar{\boldsymbol{\lambda}} \in \mathbb{R}_+^m$, and let $\bar{\mathbf{x}} \in X$ be such that*

$$\bar{\mathbf{x}} \in \operatorname{argmin}_{\mathbf{x} \in X} \left\{ f(\mathbf{x}) + \bar{\boldsymbol{\lambda}}^T \mathbf{g}(\mathbf{x}) \right\} \quad (8.65)$$

and $\mathbf{g}(\bar{\mathbf{x}}) = \mathbf{0}$. Then $\bar{\mathbf{x}}$ is an optimal solution of problem (8.62).

Proof. Let \mathbf{x} be a feasible point of problem (8.62), meaning that $\mathbf{x} \in X$ and $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$. Then

$$\begin{aligned} f(\mathbf{x}) &\geq f(\mathbf{x}) + \bar{\boldsymbol{\lambda}}^T \mathbf{g}(\mathbf{x}) \quad [\mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \bar{\boldsymbol{\lambda}} \geq \mathbf{0}] \\ &\geq f(\bar{\mathbf{x}}) + \bar{\boldsymbol{\lambda}}^T \mathbf{g}(\bar{\mathbf{x}}) \quad [(8.65)] \\ &= f(\bar{\mathbf{x}}), \quad [\mathbf{g}(\bar{\mathbf{x}}) = \mathbf{0}] \end{aligned}$$

establishing the optimality of $\bar{\mathbf{x}}$. \square

8.5.3 Convergence Analysis

Proving convergence of the *dual* objective function sequence $\{q(\boldsymbol{\lambda}^k)\}_{k \geq 0}$ under various choices of the parameters $\{\gamma_k\}_{k \geq 0}$ is an easy task since such results were already

proven in the previous sections. The more interesting question is whether we can prove convergence in some sense of a primal sequence. The answer is yes, but perhaps quite surprisingly the sequence $\{\mathbf{x}^k\}_{k \geq 0}$ is not the “correct” primal sequence. We will consider the following two possible definitions of the primal sequence that involve averaging of the sequence $\{\mathbf{x}^k\}_{k \geq 0}$.

- **Full averaging sequence.** In this option, we perform averaging of the entire history of iterates:

$$\mathbf{x}^{(k)} = \sum_{n=0}^k \mu_n^k \mathbf{x}^n \quad (8.66)$$

with μ_n^k defined by

$$\mu_n^k = \frac{\gamma_n / \|\mathbf{g}(\mathbf{x}^n)\|_2}{\sum_{j=0}^k \frac{\gamma_j}{\|\mathbf{g}(\mathbf{x}^j)\|_2}}, \quad n = 0, 1, \dots, k. \quad (8.67)$$

- **Partial averaging sequence.** Here, at iteration k , we only perform averaging of iterations $\lceil k/2 \rceil, \lceil k/2 \rceil + 1, \dots, k$:

$$\mathbf{x}^{\langle k \rangle} = \sum_{n=\lceil k/2 \rceil}^k \eta_n^k \mathbf{x}^n \quad (8.68)$$

with η_n^k defined by

$$\eta_n^k = \frac{\gamma_n / \|\mathbf{g}(\mathbf{x}^n)\|_2}{\sum_{j=\lceil k/2 \rceil}^k \frac{\gamma_j}{\|\mathbf{g}(\mathbf{x}^j)\|_2}}, \quad n = \lceil k/2 \rceil, \dots, k. \quad (8.69)$$

Our underlying assumption will be that the method did not terminate, meaning that $\mathbf{g}(\mathbf{x}^k) \neq \mathbf{0}$ for any k .

Lemma 8.45. *Suppose that Assumption 8.41 holds, and assume further that there exists $L > 0$ such that $\|\mathbf{g}(\mathbf{x})\|_2 \leq L$ for any $\mathbf{x} \in X$. Let $\rho > 0$ be some positive number, and let $\{\mathbf{x}^k\}_{k \geq 0}$ and $\{\boldsymbol{\lambda}^k\}_{k \geq 0}$ be the sequences generated by the dual projected subgradient method. Then for any $k \geq 2$,*

$$f(\mathbf{x}^{(k)}) - f_{\text{opt}} + \rho \|\mathbf{g}(\mathbf{x}^{(k)})\|_2 \leq \frac{L}{2} \frac{(\|\boldsymbol{\lambda}^0\|_2 + \rho)^2 + \sum_{n=0}^k \gamma_n^2}{\sum_{n=0}^k \gamma_n} \quad (8.70)$$

and

$$f(\mathbf{x}^{\langle k \rangle}) - f_{\text{opt}} + \rho \|\mathbf{g}(\mathbf{x}^{\langle k \rangle})\|_2 \leq \frac{L}{2} \frac{(\|\boldsymbol{\lambda}^{\lceil k/2 \rceil}\|_2 + \rho)^2 + \sum_{n=\lceil k/2 \rceil}^k \gamma_n^2}{\sum_{n=\lceil k/2 \rceil}^k \gamma_n}, \quad (8.71)$$

where $\mathbf{x}^{(k)}$ and $\mathbf{x}^{\langle k \rangle}$ are given in (8.66) and (8.68), respectively.

Proof. Let $\bar{\lambda} \in \mathbb{R}_+^m$. Then for every $n \geq 0$,

$$\begin{aligned}\|\lambda^{n+1} - \bar{\lambda}\|_2^2 &= \left\| \left[\lambda^n + \gamma_n \frac{\mathbf{g}(\mathbf{x}^n)}{\|\mathbf{g}(\mathbf{x}^n)\|_2} \right]_+ - [\bar{\lambda}]_+ \right\|_2^2 \\ &\leq \left\| \lambda^n + \gamma_n \frac{\mathbf{g}(\mathbf{x}^n)}{\|\mathbf{g}(\mathbf{x}^n)\|_2} - \bar{\lambda} \right\|_2^2 \\ &= \|\lambda^n - \bar{\lambda}\|_2^2 + \gamma_n^2 + \frac{2\gamma_n}{\|\mathbf{g}(\mathbf{x}^n)\|_2} \mathbf{g}(\mathbf{x}^n)^T (\lambda^n - \bar{\lambda}),\end{aligned}$$

where the inequality follows by the nonexpansivity of the orthogonal projection operator (Theorem 6.42(b)). Let $p \in \{0, 1, 2, \dots, k\}$. Summing the above inequality for $n = p, p+1, \dots, k$, we obtain that

$$\|\lambda^{k+1} - \bar{\lambda}\|_2^2 \leq \|\lambda^p - \bar{\lambda}\|_2^2 + \sum_{n=p}^k \gamma_n^2 + 2 \sum_{n=p}^k \frac{\gamma_n}{\|\mathbf{g}(\mathbf{x}^n)\|_2} \mathbf{g}(\mathbf{x}^n)^T (\lambda^n - \bar{\lambda}).$$

Therefore,

$$2 \sum_{n=p}^k \frac{\gamma_n}{\|\mathbf{g}(\mathbf{x}^n)\|_2} \mathbf{g}(\mathbf{x}^n)^T (\bar{\lambda} - \lambda^n) \leq \|\lambda^p - \bar{\lambda}\|_2^2 + \sum_{n=p}^k \gamma_n^2. \quad (8.72)$$

To facilitate the proof of the lemma, we will define for any $p \in \{0, 1, \dots, k\}$

$$\mathbf{x}^{k,p} \equiv \sum_{n=p}^k \alpha_n^{k,p} \mathbf{x}^n, \quad (8.73)$$

where

$$\alpha_n^{k,p} = \frac{\frac{\gamma_n}{\|\mathbf{g}(\mathbf{x}^n)\|_2}}{\sum_{j=p}^k \frac{\gamma_j}{\|\mathbf{g}(\mathbf{x}^j)\|_2}}.$$

In particular, the sequences $\{\mathbf{x}^{k,0}\}_{k \geq 0}, \{\mathbf{x}^{k,\lceil k/2 \rceil}\}_{k \geq 0}$ are the same as the sequences $\{\mathbf{x}^{(k)}\}_{k \geq 0}$ and $\{\mathbf{x}^{\langle k \rangle}\}_{k \geq 0}$, respectively. Using the above definition of $\alpha_n^{k,p}$ and the fact that $\|\mathbf{g}(\mathbf{x}^n)\|_2 \leq L$, we conclude that (8.72) implies the following inequality:

$$\sum_{n=p}^k \alpha_n^{k,p} \mathbf{g}(\mathbf{x}^n)^T (\bar{\lambda} - \lambda^n) \leq \frac{L}{2} \frac{\|\lambda^p - \bar{\lambda}\|_2^2 + \sum_{n=p}^k \gamma_n^2}{\sum_{n=p}^k \gamma_n}. \quad (8.74)$$

By the definition of \mathbf{x}^n , we have for any $\mathbf{x}^* \in X^*$,

$$\begin{aligned}f(\mathbf{x}^*) &\geq f(\mathbf{x}^*) + (\lambda^n)^T \mathbf{g}(\mathbf{x}^*) \quad [\lambda^n \geq \mathbf{0}, \mathbf{g}(\mathbf{x}^*) \leq \mathbf{0}] \\ &\geq f(\mathbf{x}^n) + (\lambda^n)^T \mathbf{g}(\mathbf{x}^n). \quad [\mathbf{x}^n \in \operatorname{argmin}_{\mathbf{x} \in X} \{f(\mathbf{x}) + (\lambda^n)^T \mathbf{g}(\mathbf{x})\}]\end{aligned}$$

Thus,

$$-(\lambda^n)^T \mathbf{g}(\mathbf{x}^n) \geq f(\mathbf{x}^n) - f_{\text{opt}},$$

and hence

$$\begin{aligned} \sum_{n=p}^k \alpha_n^{k,p} \mathbf{g}(\mathbf{x}^n)^T (\bar{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^n) &\geq \sum_{n=p}^k \alpha_n^{k,p} \mathbf{g}(\mathbf{x}^n)^T \bar{\boldsymbol{\lambda}} + \sum_{n=p}^k \alpha_n^{k,p} f(\mathbf{x}^n) - \sum_{n=p}^k \alpha_n^{k,p} f_{\text{opt}} \\ &\geq \bar{\boldsymbol{\lambda}}^T \mathbf{g}(\mathbf{x}^{k,p}) + f(\mathbf{x}^{k,p}) - f_{\text{opt}}, \end{aligned} \quad (8.75)$$

where the last inequality follows by Jensen's inequality (recalling that f and the components of \mathbf{g} are convex) and the definition (8.73) of $\mathbf{x}^{k,p}$. Combining (8.74) and (8.75), while using the obvious inequality $\|\boldsymbol{\lambda}^p - \bar{\boldsymbol{\lambda}}\|_2 \leq \|\boldsymbol{\lambda}^p\|_2 + \|\bar{\boldsymbol{\lambda}}\|_2$, we obtain

$$f(\mathbf{x}^{k,p}) - f_{\text{opt}} + \bar{\boldsymbol{\lambda}}^T \mathbf{g}(\mathbf{x}^{k,p}) \leq \frac{L}{2} \frac{(\|\boldsymbol{\lambda}^p\|_2 + \|\bar{\boldsymbol{\lambda}}\|_2)^2 + \sum_{n=p}^k \gamma_n^2}{\sum_{n=p}^k \gamma_n}. \quad (8.76)$$

Plugging

$$\bar{\boldsymbol{\lambda}} = \begin{cases} \rho \frac{[\mathbf{g}(\mathbf{x}^{k,p})]_+}{\|[\mathbf{g}(\mathbf{x}^{k,p})]_+\|_2}, & [\mathbf{g}(\mathbf{x}^{k,p})]_+ \neq \mathbf{0}, \\ \mathbf{0}, & [\mathbf{g}(\mathbf{x}^{k,p})]_+ = \mathbf{0} \end{cases}$$

into (8.76), we obtain the inequality

$$f(\mathbf{x}^{k,p}) - f_{\text{opt}} + \rho \|[\mathbf{g}(\mathbf{x}^{k,p})]_+\|_2 \leq \frac{L}{2} \frac{(\|\boldsymbol{\lambda}^p\|_2 + \rho)^2 + \sum_{n=p}^k \gamma_n^2}{\sum_{n=p}^k \gamma_n}. \quad (8.77)$$

Substituting $p = 0$ and $p = \lceil k/2 \rceil$ in (8.77) yields the inequalities (8.70) and (8.71), respectively. \square

Analysis of the Full Averaging Scheme

We begin by developing a convergence rate related to the sequence $\{\mathbf{x}^{(k)}\}_{k \geq 0}$ given by (8.66). Similarly to the analysis for the primal projected subgradient, choosing $\gamma_k = \frac{1}{\sqrt{k+1}}$ will imply that the right-hand side of (8.70) will converge to zero. In principle, the fact that the left-hand side of (8.70) converges to zero does not necessarily imply that both the expression for the distance to optimality in function values $f(\mathbf{x}^{(k)}) - f_{\text{opt}}$ and the expression for the constraints violation $\|[\mathbf{g}(\mathbf{x}^{(k)})]_+\|_2$ converge to zero. However, using Theorem 3.60, we can show the convergence of these terms as long as ρ is chosen appropriately.

Theorem 8.46 ($O(\log(k)/\sqrt{k})$ rate of convergence of the full averaging sequence). Suppose that Assumption 8.41 holds, and assume further that there exists $L > 0$ for which $\|\mathbf{g}(\mathbf{x})\|_2 \leq L$ for any $\mathbf{x} \in X$. Let $\{\mathbf{x}^k\}_{k \geq 0}$, and let $\{\boldsymbol{\lambda}^k\}_{k \geq 0}$ be the sequences generated by the dual projected subgradient method with $\gamma_k = \frac{1}{\sqrt{k+1}}$. Then for any $k \geq 1$,

$$f(\mathbf{x}^{(k)}) - f_{\text{opt}} \leq \frac{L}{2} \frac{(\|\boldsymbol{\lambda}^0\|_2 + 2\alpha)^2 + 1 + \log(k+1)}{\sqrt{k+1}}, \quad (8.78)$$

$$\|[\mathbf{g}(\mathbf{x}^{(k)})]_+\|_2 \leq \frac{L}{2\alpha} \frac{(\|\boldsymbol{\lambda}^0\|_2 + 2\alpha)^2 + 1 + \log(k+1)}{\sqrt{k+1}}, \quad (8.79)$$

where $\{\mathbf{x}^{(k)}\}_{k \geq 0}$ is given in (8.66) and

$$\alpha = \frac{f(\bar{\mathbf{x}}) - f_{\text{opt}}}{\min_{j=1,2,\dots,m}\{-g_j(\bar{\mathbf{x}})\}},$$

with $\bar{\mathbf{x}}$ being a Slater point whose existence is guaranteed by Assumption 8.41(E).

Proof. Employing Lemma 8.45 with $\rho = 2\alpha$, and substituting $\gamma_n = \frac{1}{\sqrt{n+1}}$, we have

$$f(\mathbf{x}^{(k)}) - f_{\text{opt}} + 2\alpha \|[\mathbf{g}(\mathbf{x}^{(k)})]_+\|_2 \leq \frac{L}{2} \frac{(\|\boldsymbol{\lambda}^0\|_2 + 2\alpha)^2 + \sum_{n=0}^k \frac{1}{n+1}}{\sum_{n=0}^k \frac{1}{\sqrt{n+1}}}. \quad (8.80)$$

Using Lemma 8.27(a), we have

$$\frac{(\|\boldsymbol{\lambda}^0\|_2 + 2\alpha)^2 + \sum_{n=0}^k \frac{1}{n+1}}{\sum_{n=0}^k \frac{1}{\sqrt{n+1}}} \leq \frac{(\|\boldsymbol{\lambda}^0\|_2 + 2\alpha)^2 + 1 + \log(k+1)}{\sqrt{k+1}},$$

which, combined with (8.80), yields the inequality

$$f(\mathbf{x}^{(k)}) - f_{\text{opt}} + 2\alpha \|[\mathbf{g}(\mathbf{x}^{(k)})]_+\|_2 \leq \frac{L}{2} \frac{(\|\boldsymbol{\lambda}^0\|_2 + 2\alpha)^2 + 1 + \log(k+1)}{\sqrt{k+1}}. \quad (8.81)$$

Since by Corollary 8.43 2α is an upper bound on twice the l_2 -norm of any dual optimal solution, it follows by Theorem 3.60 that the inequality (8.81) implies the two inequalities (8.78) and (8.79). \square

Analysis of the Partial Averaging Scheme

We will now show an $O(1/\sqrt{k})$ rate of convergence in terms of function values as well as constraint violation of the partial averaging sequence given in (8.68). The proof is similar to the proof of Theorem 8.46 and utilizes inequality (8.71) but in addition utilizes the boundedness of the sequence of dual variables—a fact established in the next lemma.

Lemma 8.47.⁴⁷ Suppose that Assumption 8.41 holds and assume further that there exists $L > 0$ for which $\|\mathbf{g}(\mathbf{x})\|_2 \leq L$ for any $\mathbf{x} \in X$. Let $\{\mathbf{x}^k\}_{k \geq 0}$ and $\{\boldsymbol{\lambda}^k\}_{k \geq 0}$ be the sequences generated by the dual projected subgradient method with positive stepsizes γ_k satisfying $\gamma_k \leq \gamma_0$ for all $k \geq 0$. Then

$$\|\boldsymbol{\lambda}^k\|_2 \leq M, \quad (8.82)$$

where⁴⁸

$$M = \left\{ \|\boldsymbol{\lambda}^0\|_2 + 2\alpha, \frac{f(\bar{\mathbf{x}}) - q_{\text{opt}}}{\beta} + \frac{\gamma_0 L}{2\beta} + 2\alpha + \gamma_0 \right\}, \quad (8.83)$$

⁴⁷Lemma 8.47 is Lemma 3 from Nedić and Ozdaglar [90].

⁴⁸Recall that in our setting $q_{\text{opt}} = f_{\text{opt}}$.

with

$$\alpha = \frac{f(\bar{\mathbf{x}}) - f_{\text{opt}}}{\min_{j=1,2,\dots,m} \{-g_j(\bar{\mathbf{x}})\}}, \quad \beta = \min_{j=1,2,\dots,m} \{-g_j(\bar{\mathbf{x}})\},$$

where $\bar{\mathbf{x}}$ is a Slater point of problem (8.62) whose existence is guaranteed by Assumption 8.41(E).

Proof. Let $\boldsymbol{\lambda}^*$ be an optimal solution of the dual problem (8.64). We begin by showing by induction on k that for any $k \geq 0$,

$$\|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\|_2 \leq \max \left\{ \|\boldsymbol{\lambda}^0 - \boldsymbol{\lambda}^*\|_2, \frac{f(\bar{\mathbf{x}}) - q_{\text{opt}}}{\beta} + \frac{\gamma_0 L}{2\beta} + \|\boldsymbol{\lambda}^*\|_2 + \gamma_0 \right\}. \quad (8.84)$$

The inequality holds trivially for $k = 0$. Assume that it holds for k , and we will show that it holds for $k + 1$. We will consider two cases.

Case I. Assume that $q(\boldsymbol{\lambda}^k) \geq q_{\text{opt}} - \frac{\gamma_k L}{2}$. Then, by Theorem 8.42,

$$\|\boldsymbol{\lambda}^k\|_2 \leq \frac{f(\bar{\mathbf{x}}) - q_{\text{opt}} + \frac{\gamma_k L}{2}}{\beta},$$

where $\beta = \min_{i=1,2,\dots,m} \{-g_i(\bar{\mathbf{x}})\}$. Therefore,

$$\begin{aligned} \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^*\|_2 &\leq \left\| \boldsymbol{\lambda}^k + \frac{\gamma_k}{\|\mathbf{g}(\mathbf{x}^k)\|_2} \mathbf{g}(\mathbf{x}^k) - \boldsymbol{\lambda}^* \right\|_2 \\ &\leq \|\boldsymbol{\lambda}^k\|_2 + \|\boldsymbol{\lambda}^*\|_2 + \gamma_k \\ &\leq \frac{f(\bar{\mathbf{x}}) - q_{\text{opt}}}{\beta} + \frac{\gamma_k L}{2\beta} + \|\boldsymbol{\lambda}^*\|_2 + \gamma_k \\ &\leq \frac{f(\bar{\mathbf{x}}) - q_{\text{opt}}}{\beta} + \frac{\gamma_0 L}{2\beta} + \|\boldsymbol{\lambda}^*\|_2 + \gamma_0. \end{aligned}$$

Case II. Now assume that $q(\boldsymbol{\lambda}^k) < q_{\text{opt}} - \frac{\gamma_k L}{2}$. In this case we can write

$$\begin{aligned} \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^*\|_2^2 &= \left\| \left[\boldsymbol{\lambda}^k + \frac{\gamma_k}{\|\mathbf{g}(\mathbf{x}^k)\|_2} \mathbf{g}(\mathbf{x}^k) \right]_+ - \boldsymbol{\lambda}^* \right\|_2^2 \\ &\leq \left\| \boldsymbol{\lambda}^k - \boldsymbol{\lambda}^* + \frac{\gamma_k}{\|\mathbf{g}(\mathbf{x}^k)\|_2} \mathbf{g}(\mathbf{x}^k) \right\|_2^2 \\ &= \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\|_2^2 + 2 \frac{\gamma_k}{\|\mathbf{g}(\mathbf{x}^k)\|_2} (\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*)^T \mathbf{g}(\mathbf{x}^k) + \gamma_k^2. \quad (8.85) \end{aligned}$$

Since $-\mathbf{g}(\mathbf{x}^k) \in \partial(-q)(\boldsymbol{\lambda}^k)$ (Example 3.7), it follows by the subgradient inequality that

$$-q_{\text{opt}} \geq -q(\boldsymbol{\lambda}^k) - \mathbf{g}(\mathbf{x}^k)^T (\boldsymbol{\lambda}^* - \boldsymbol{\lambda}^k).$$

Therefore, continuing (8.85),

$$\begin{aligned}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^*\|_2^2 &\leq \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\|_2^2 + 2\frac{\gamma_k}{\|\mathbf{g}(\mathbf{x}^k)\|_2}(q(\boldsymbol{\lambda}^k) - q_{\text{opt}}) + \gamma_k^2 \\ &\leq \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\|_2^2 + 2\frac{\gamma_k}{L}(q(\boldsymbol{\lambda}^k) - q_{\text{opt}}) + \gamma_k^2 \\ &= \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\|_2^2 + 2\frac{\gamma_k}{L} \left(q(\boldsymbol{\lambda}^k) - q_{\text{opt}} + \frac{\gamma_k L}{2} \right) \\ &< \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\|_2^2,\end{aligned}$$

where in the last inequality we used our assumption that $q(\boldsymbol{\lambda}^k) < q_{\text{opt}} - \frac{\gamma_k L}{2}$. We can now use the induction hypothesis and conclude that

$$\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^*\|_2 \leq \max \left\{ \|\boldsymbol{\lambda}^0 - \boldsymbol{\lambda}^*\|_2, \frac{f(\bar{\mathbf{x}}) - q_{\text{opt}}}{\beta} + \frac{\gamma_0 L}{2\beta} + \|\boldsymbol{\lambda}^*\|_2 + \gamma_0 \right\}.$$

We have thus established the validity of (8.84) for all $k \geq 0$. The result (8.82) now follows by recalling that by Corollary 8.43, $\|\boldsymbol{\lambda}^*\|_2 \leq \alpha$, and hence

$$\begin{aligned}\|\boldsymbol{\lambda}^k\|_2 &\leq \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\|_2 + \|\boldsymbol{\lambda}^*\|_2 \\ &\leq \max \left\{ \|\boldsymbol{\lambda}^0 - \boldsymbol{\lambda}^*\|_2, \frac{f(\bar{\mathbf{x}}) - q_{\text{opt}}}{\beta} + \frac{\gamma_0 L}{2\beta} + \|\boldsymbol{\lambda}^*\|_2 + \gamma_0 \right\} + \|\boldsymbol{\lambda}^*\|_2 \\ &\leq \max \left\{ \|\boldsymbol{\lambda}^0\|_2 + 2\alpha, \frac{f(\bar{\mathbf{x}}) - q_{\text{opt}}}{\beta} + \frac{\gamma_0 L}{2\beta} + 2\alpha + \gamma_0 \right\}.\quad \square\end{aligned}$$

Equipped with the upper bound on the sequence of dual variables, we can prove, using a similar argument to the one used in the proof of Theorem 8.46, an $O(1/\sqrt{k})$ rate of convergence related to the partial averaging sequence generated by the dual projected subgradient method.

Theorem 8.48 ($O(1/\sqrt{k})$ rate of convergence of the partial averaging sequence). Suppose that Assumption 8.41 holds, and assume further that there exists $L > 0$ for which $\|\mathbf{g}(\mathbf{x})\|_2 \leq L$ for any $\mathbf{x} \in X$. Let $\{\mathbf{x}^k\}_{k \geq 0}$, and let $\{\boldsymbol{\lambda}^k\}_{k \geq 0}$ be the sequences generated by the dual projected subgradient method with $\gamma_k = \frac{1}{\sqrt{k+1}}$. Then for any $k \geq 2$,

$$f(\mathbf{x}^{(k)}) - f_{\text{opt}} \leq \frac{2L((M+2\alpha)^2 + \log(3))}{\sqrt{k+2}}, \quad (8.86)$$

$$\|\mathbf{g}(\mathbf{x}^{(k)})\|_2 \leq \frac{2L((M+2\alpha)^2 + \log(3))}{\alpha\sqrt{k+2}}, \quad (8.87)$$

where $\{\mathbf{x}^{(k)}\}_{k \geq 0}$ is given in (8.68), M in (8.83), and

$$\alpha = \frac{f(\bar{\mathbf{x}}) - f_{\text{opt}}}{\min_{j=1,2,\dots,m} \{-g_j(\bar{\mathbf{x}})\}}$$

with $\bar{\mathbf{x}}$ being a Slater point whose existence is guaranteed by Assumption 8.41(E).

Proof. Employing Lemma 8.45 with $\rho = 2\alpha$, and substituting $\gamma_n = \frac{1}{\sqrt{n+1}}$, we have

$$\begin{aligned} f(\mathbf{x}^{(k)}) - f_{\text{opt}} + 2\alpha \|[\mathbf{g}(\mathbf{x}^{(k)})]_+\|_2 &\leq \frac{L}{2} \frac{(\|\boldsymbol{\lambda}^{\lceil k/2 \rceil}\|_2 + 2\alpha)^2 + \sum_{n=\lceil k/2 \rceil}^k \frac{1}{n+1}}{\sum_{n=\lceil k/2 \rceil}^k \frac{1}{\sqrt{n+1}}} \\ &\leq \frac{L}{2} \frac{(M + 2\alpha)^2 + \sum_{n=\lceil k/2 \rceil}^k \frac{1}{n+1}}{\sum_{n=\lceil k/2 \rceil}^k \frac{1}{\sqrt{n+1}}}, \end{aligned} \quad (8.88)$$

where in the last inequality we used the bound on the dual iterates given in Lemma 8.47. Now, using Lemma 8.27(b), we have

$$\frac{(M + 2\alpha)^2 + \sum_{n=\lceil k/2 \rceil}^k \frac{1}{n+1}}{\sum_{n=\lceil k/2 \rceil}^k \frac{1}{\sqrt{n+1}}} \leq \frac{4((M + 2\alpha)^2 + \log(3))}{\sqrt{k+2}},$$

which, combined with (8.88), yields the inequality

$$f(\mathbf{x}^{(k)}) - f_{\text{opt}} + 2\alpha \|[\mathbf{g}(\mathbf{x}^{(k)})]_+\|_2 \leq \frac{2L((M + 2\alpha)^2 + \log(3))}{\sqrt{k+2}}. \quad (8.89)$$

Since, by Corollary 8.43, 2α is an upper bound on twice the l_2 -norm of any dual optimal solution, it follows by Theorem 3.60 that the inequality (8.89) implies the two inequalities (8.86) and (8.87). \square

To derive a complexity result for the dual projected subgradient method, we should first note that the primal sequence is not feasible, as it does not necessarily satisfy the inequality constraints $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$. Therefore, there is no point in asking how many iterations are required to obtain an ε -optimal solution. Instead, we will consider the related concept of an ε -optimal and feasible solution. A vector $\mathbf{x} \in X$ is called an ε -optimal and feasible solution of problem (8.62) if $f(\mathbf{x}) - f_{\text{opt}} \leq \varepsilon$ and $\|[\mathbf{g}(\mathbf{x})]_+\|_2 \leq \varepsilon$. Theorem 8.48 immediately implies a complexity result stating that an order of $\frac{1}{\varepsilon^2}$ iterations are required to obtain an ε -optimal and feasible solution.

Corollary 8.49 ($O(1/\varepsilon^2)$ complexity result for the dual projected subgradient method). *Under the setting of Theorem 8.48, if $k \geq 2$ satisfies*

$$k \geq \frac{4L^2((M + 2\alpha)^2 + \log(3))^2}{\min\{\alpha^2, 1\}\varepsilon^2} - 2,$$

then

$$\begin{aligned} f(\mathbf{x}^{(k)}) - f_{\text{opt}} &\leq \varepsilon, \\ \|[\mathbf{g}(\mathbf{x}^{(k)})]_+\|_2 &\leq \varepsilon. \end{aligned}$$

Example 8.50 (linear programming example). Consider the linear programming problem

$$\begin{aligned} \min \quad & \mathbf{c}^T \mathbf{x} \\ (\text{LP}) \quad \text{s.t.} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b}, \\ & \mathbf{x} \in \Delta_n, \end{aligned}$$

where $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{m \times n}$, and $\mathbf{b} \in \mathbb{R}^m$. We will consider the dual projected subgradient method when the underlying set X is Δ_n and $\mathbf{g}(\mathbf{x}) \equiv \mathbf{Ax} - \mathbf{b}$. The vector \mathbf{x}^k is calculated by the update rule $\mathbf{x}^k \in \operatorname{argmin}_{\mathbf{x} \in \Delta_n} (\mathbf{c} + \mathbf{A}^T \boldsymbol{\lambda}^k)^T \mathbf{x}$. It is easy to see that an optimal solution of this subproblem is given by \mathbf{e}_i , where i is an index for which $(\mathbf{c} + \mathbf{A}^T \boldsymbol{\lambda}^k)_i$ is minimal. Therefore, the algorithm (with $\gamma_k = \frac{1}{\sqrt{k+1}}$) takes the following form:

Dual Projected Subgradient for solving (LP)

- **Initialization:** pick $\boldsymbol{\lambda}^0 \in \mathbb{R}_+^m$.
- **General step ($k \geq 0$):**

$$\begin{aligned} i_k &\in \operatorname{argmin}_{i=1,2,\dots,n} v_i; \quad \mathbf{v} = \mathbf{c} + \mathbf{A}^T \boldsymbol{\lambda}^k, \\ \mathbf{x}^k &= \mathbf{e}_{i_k}, \\ \boldsymbol{\lambda}^{k+1} &= \left[\boldsymbol{\lambda}^k + \frac{1}{\sqrt{k+1}} \frac{\mathbf{Ax}^k - \mathbf{b}}{\|\mathbf{Ax}^k - \mathbf{b}\|_2} \right]_+. \end{aligned}$$

Note that we make the implicit assumption that $\mathbf{Ax}^k \neq \mathbf{b}$. The above description of the dual projected subgradient method illustrates the fact that the sequence $\{\mathbf{x}^k\}_{k \geq 0}$ is not the “correct” primal sequence. Indeed, in this case, the vectors \mathbf{x}^k are always unit vectors, and there is no particular reason why the solution of (LP) should be attained at a unit vector. As a specific example, consider the problem

$$\begin{aligned} \min \quad & x_1 + 3x_2 + 2x_3 \\ \text{s.t.} \quad & 3x_1 + 2x_2 - x_3 \leq 1, \\ & -2x_3 \leq 2, \\ & x_1 + x_2 + x_3 = 1, \\ & x_1, x_2, x_3 \geq 0, \end{aligned} \tag{8.90}$$

which fits problem (LP) with

$$\mathbf{A} = \begin{pmatrix} 3 & 2 & -1 \\ 0 & 0 & -2 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}.$$

The optimal solution of problem (8.90) is $(\frac{1}{2}, 0, \frac{1}{2})$. We employed the dual projected subgradient method as described above with $\boldsymbol{\lambda}^0 = \mathbf{0}$ and compared the behavior of the full and partial averaging schemes during the first 100 iterations. The results are described in Figure 8.4. Obviously, the partial averaging scheme exhibits superior behavior compared to the full averaging scheme. ■

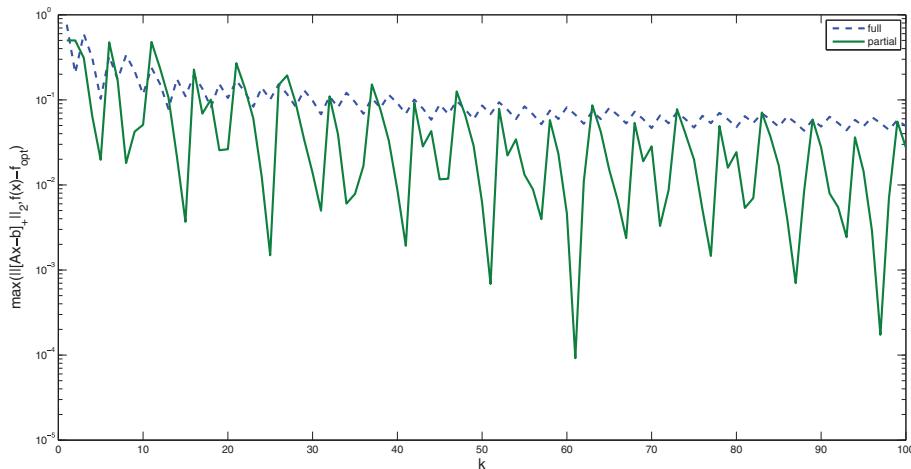


Figure 8.4. First 100 iterations of the dual projected subgradient method employed on problem (8.90). The y-axis describes (in log scale) the quantities $\max\{f(\mathbf{x}^{(k)}) - f_{\text{opt}}, \|\mathbf{A}\mathbf{x}^{(k)} - \mathbf{b}\|_2\}$ and $\max\{f(\mathbf{x}^{(k)}) - f_{\text{opt}}, \|\mathbf{A}\mathbf{x}^{(k)} - \mathbf{b}\|_2\}$.

8.5.4 Example—Network Utility Maximization

Consider a network that consists of a set $\mathcal{S} = \{1, 2, \dots, S\}$ of sources and a set $\mathcal{L} = \{1, 2, \dots, L\}$ of links, where a link ℓ has a capacity c_ℓ . For each source $s \in \mathcal{S}$, we denote by $\mathcal{L}(s) \subseteq \mathcal{L}$ the set of all links used by source s . Similarly, for a given link $\ell \in \mathcal{L}$, the set $\mathcal{S}(\ell) \subseteq \mathcal{S}$ comprises all sources that use link ℓ . In particular, for a pair $\ell \in \mathcal{L}$ and $s \in \mathcal{S}$, the relation $s \in \mathcal{S}(\ell)$ holds if and only if $\ell \in \mathcal{L}(s)$. Each source $s \in \mathcal{S}$ is associated with a concave utility function $u_s : \mathbb{R} \rightarrow \mathbb{R}$, meaning that if source s sends data at a rate x_s , it gains a utility $u_s(x_s)$. We also assume that the rate of source s is constrained to be in the interval $I_s = [0, M_s]$, where $M_s \in \mathbb{R}_{++}$. The goal of the network utility maximization problem (abbreviated NUM) is to allocate the source rates as the optimal solution of the following convex problem:

$$\begin{aligned} \max \quad & \sum_{s \in \mathcal{S}} u_s(x_s) \\ \text{s.t.} \quad & \sum_{s \in \mathcal{S}(\ell)} x_s \leq c_\ell, \quad \ell \in \mathcal{L}, \\ & x_s \in I_s, \quad s \in \mathcal{S}. \end{aligned} \tag{8.91}$$

Problem (8.91) in its minimization form is a convex problem and fits the main model (8.62) with

$$\begin{aligned} \mathbf{g}(\mathbf{x}) &= \left(\sum_{s \in \mathcal{S}(\ell)} x_s - c_\ell \right)_{\ell=1,2,\dots,L}, \\ X &= I_1 \times I_2 \times \cdots \times I_S, \\ f(\mathbf{x}) &= - \sum_{s=1}^S u_s(x_s). \end{aligned}$$