CONVEX OPTIMIZATION THEORY

Dimitri P. Bertsekas

Athena Scientific

Convex Optimization Theory

Dimitri P. Bertsekas

Massachusetts Institute of Technology

WWW site for book information and orders

http://www.athenasc.com

Athena Scientific, Belmont, Massachusetts

Athena Scientific
Post Office Box 805
Nashua, NH 03061-0805
U.S.A.

Email: info@athenasc.com
WWW: http://www.athenasc.com

Publisher's Cataloging-in-Publication Data

Bertsekas, Dimitri P.
Convex Optimization Theory
Includes bibliographical references and index
1. Nonlinear Programming 2. Mathematical Optimization. I. Title.
T57.8.B475 2009 519.703
Library of Congress Control Number: 2002092168

ISBN-10: 1-886529-31-0, ISBN-13: 978-1-886529-31-1

Contents

ABOUT THE AUTHOR

Dimitri Bertsekas studied Mechanical and Electrical Engineering at the National Technical University of Athens, Greece, and obtained his Ph.D. in system science from the Massachusetts Institute of Technology. He has held faculty positions with the Engineering-Economic Systems Department, Stanford University, and the Electrical Engineering Department of the University of Illinois, Urbana. Since 1979 he has been teaching at the Electrical Engineering and Computer Science Department of the Massachusetts Institute of Technology (M.I.T.), where he is currently McAfee Professor of Engineering.

His teaching and research spans several fields, including deterministic optimization, dynamic programming and stochastic control, large-scale and distributed computation, and data communication networks. He has authored or coauthored numerous research papers and fourteen books, several of which are used as textbooks in MIT classes, including "Nonlinear Programming," "Dynamic Programming and Optimal Control," "Data Networks," "Introduction to Probability," as well as the present book. He often consults with private industry and has held editorial positions in several journals.

Professor Bertsekas was awarded the INFORMS 1997 Prize for Research Excellence in the Interface Between Operations Research and Computer Science for his book "Neuro-Dynamic Programming" (co-authored with John Tsitsiklis), the 2000 Greek National Award for Operations Research, and the 2001 ACC John R. Ragazzini Education Award. In 2001, he was elected to the United States National Academy of Engineering.

ATHENA SCIENTIFIC

OPTIMIZATION AND COMPUTATION SERIES

1. Convex Optimization Theory, by Dimitri P. Bertsekas, 2009, ISBN 978-1-886529-31-1, 256 pages
2. Introduction to Probability, 2nd Edition, by Dimitri P. Bertsekas and John N. Tsitsiklis, 2008, ISBN 978-1-886529-23-6, 544 pages
3. Dynamic Programming and Optimal Control, Two-Volume Set, by Dimitri P. Bertsekas, 2007, ISBN 1-886529-08-6, 1020 pages
4. Convex Analysis and Optimization, by Dimitri P. Bertsekas, with Angelia Nedić and Asuman E. Ozdaglar, 2003, ISBN 1-886529-45-0, 560 pages
5. Nonlinear Programming, 2nd Edition, by Dimitri P. Bertsekas, 1999, ISBN 1-886529-00-0, 791 pages
6. Network Optimization: Continuous and Discrete Models, by Dimitri P. Bertsekas, 1998, ISBN 1-886529-02-7, 608 pages
7. Network Flows and Monotropic Optimization, by R. Tyrrell Rockafellar, 1998, ISBN 1-886529-06-X, 634 pages
8. Introduction to Linear Optimization, by Dimitris Bertsimas and John N. Tsitsiklis, 1997, ISBN 1-886529-19-1, 608 pages
9. Parallel and Distributed Computation: Numerical Methods, by Dimitri P. Bertsekas and John N. Tsitsiklis, 1997, ISBN 1-886529-01-9, 718 pages
10. Neuro-Dynamic Programming, by Dimitri P. Bertsekas and John N. Tsitsiklis, 1996, ISBN 1-886529-10-8, 512 pages
11. Constrained Optimization and Lagrange Multiplier Methods, by Dimitri P. Bertsekas, 1996, ISBN 1-886529-04-3, 410 pages
12. Stochastic Optimal Control: The Discrete-Time Case, by Dimitri P. Bertsekas and Steven E. Shreve, 1996, ISBN 1-886529-03-5, 330 pages

Preface

This book aims at an accessible, concise, and intuitive exposition of two related subjects that find broad practical application:

(a) Convex analysis, particularly as it relates to optimization.

(b) Duality theory for optimization and minimax problems, mainly within a convexity framework.

The focus on optimization is to derive conditions for existence of primal and dual optimal solutions for constrained problems such as

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in X, \qquad g_j(x) \leq 0, \quad j = 1, \ldots, r. \end{array}$$

Other types of optimization problems, such as those arising in Fenchel duality, are also part of our scope. The focus on minimax is to derive conditions guaranteeing the equality

$$\inf_{x \in X} \sup_{z \in Z} \phi(x, z) = \sup_{z \in Z} \inf_{x \in X} \phi(x, z),$$

and the attainment of the "inf" and the "sup."

The treatment of convexity theory is fairly detailed. It touches upon nearly all major aspects of the subject, and it is sufficient for the development of the core analytical issues of convex optimization. The mathematical prerequisites are a first course in linear algebra and a first course in real analysis. A summary of the relevant material is provided in an appendix. Prior knowledge of linear and nonlinear optimization theory is not assumed, although it will undoubtedly be helpful in providing context and perspective. Other than this modest background, the development is self-contained, with rigorous proofs provided throughout.

We have aimed at a unified development of the strongest possible forms of duality with the most economic use of convexity theory. To this end, our analysis often departs from the lines of Rockafellar's classic 1970 book and other books that followed the Fenchel/Rockafellar formalism. For example, we treat differently closed set intersection theory and preservation of closure under linear transformations (Sections 1.4.2 and 1.4.3); we develop subdifferential calculus by using constrained optimization duality (Section 5.4.2); and we do not rely on concepts such as infimal convolution, image, polar sets and functions, bifunctions, and conjugate saddle functions. Perhaps our greatest departure is in duality theory itself: similar to Fenchel/Rockafellar, our development rests on Legendre/Fenchel conjugacy ideas, but is far more geometrical and visually intuitive.

Our duality framework is based on two simple geometrical problems: the *min common point problem* and the *max crossing point problem*. The salient feature of the min common/max crossing (MC/MC) framework is its highly visual geometry, through which all the core issues of duality theory become apparent and can be analyzed in a unified way. Our approach is to obtain a handful of broadly applicable theorems within the MC/MC framework, and then specialize them to particular types of problems (constrained optimization, Fenchel duality, minimax problems, etc). We address all duality questions (existence of duality gap, existence of dual optimal solutions, structure of the dual optimal solution set), and other issues (subdifferential theory, theorems of the alternative, duality gap estimates) in this way.

Fundamentally, the MC/MC framework is closely connected to the conjugacy framework, and owes its power and generality to this connection. However, the two frameworks offer complementary starting points for analysis and provide alternative views of the geometric foundation of duality: conjugacy emphasizes functional/algebraic descriptions, while MC/MC emphasizes set/epigraph descriptions. The MC/MC framework is simpler, and seems better suited for visualizing and investigating questions of strong duality and existence of dual optimal solutions. The conjugacy framework, with its emphasis on functional descriptions, is more suitable when mathematical operations on convex functions are involved, and the calculus of conjugate functions can be brought to bear for analysis or computation.

The book evolved from the earlier book of the author [BNO03] on the subject (coauthored with A. Nedić and A. Ozdaglar), but has different character and objectives. The 2003 book was quite extensive, was structured (at least in part) as a research monograph, and aimed to bridge the gap between convex and nonconvex optimization using concepts of nonsmooth analysis. By contrast, the present book is organized differently, has the character of a textbook, and concentrates exclusively on convex optimization. Despite the differences, the two books have similar style and level of mathematical sophistication, and share some material.

The chapter-by-chapter description of the book follows:

Chapter 1: This chapter develops all of the convex analysis tools that are needed for the development of duality theory in subsequent chapters. It covers basic algebraic concepts such as convex hulls and hyperplanes, and topological concepts such as relative interior, closure, preservation of closedness under linear transformations, and hyperplane separation. In addition, it develops subjects of special interest in duality and optimization, such as recession cones and conjugate functions.

Chapter 2: This chapter covers polyhedral convexity concepts: extreme points, the Farkas and Minkowski-Weyl theorems, and some of their applications in linear programming. It is not needed for the developments of subsequent chapters, and may be skipped at first reading.

Chapter 3: This chapter focuses on basic optimization concepts: types of minima, existence of solutions, and a few topics of special interest for duality theory, such as partial minimization and minimax theory.

Chapter 4: This chapter introduces the MC/MC duality framework. It discusses its connection with conjugacy theory, and it charts its applications to constrained optimization and minimax problems. It then develops broadly applicable theorems relating to strong duality and existence of dual optimal solutions.

Chapter 5: This chapter specializes the duality theorems of Chapter 4 to important contexts relating to linear programming, convex programming, and minimax theory. It also uses these theorems as an aid for the development of additional convex analysis tools, such as a powerful nonlinear version of Farkas' Lemma, subdifferential theory, and theorems of the alternative. A final section is devoted to nonconvex problems and estimates of the duality gap, with special focus on separable problems.

In aiming for brevity, I have omitted a number of topics that an instructor may wish for. One such omission is applications to specially structured problems; the book by Boyd and Vanderbergue [BoV04], as well as my book on parallel and distributed computation with John Tsitsiklis [BeT89] cover this material extensively (both books are available on line).

Another important omission is computational methods. However, I have written a long supplementary sixth chapter (over 100 pages), which covers the most popular convex optimization algorithms (and some new ones), and can be downloaded from the book's web page

http://www.athenasc.com/convexduality.html.

This chapter, together with a more comprehensive treatment of convex analysis, optimization, duality, and algorithms will be part of a more extensive textbook that I am currently writing. Until that time, the chapter will serve instructors who wish to cover convex optimization algorithms in addition to duality (as I do in my M.I.T. course). This is a "living" chapter that will be periodically updated. Its current contents are as follows:

Chapter 6 on Algorithms: 6.1. Problem Structures and Computational Approaches; 6.2. Algorithmic Descent; 6.3. Subgradient Methods; 6.4. Polyhedral Approximation Methods; 6.5. Proximal and Bundle Methods; 6.6. Dual Proximal Point Algorithms; 6.7. Interior Point Methods; 6.8. Approximate Subgradient Methods; 6.9. Optimal Algorithms and Complexity.

While I did not provide exercises in the text, I have supplied a substantial number of exercises (with detailed solutions) at the book's web page. The reader/instructor may also use the end-of-chapter problems (a total of 175) given in [BNO03], which have similar style and notation to the present book. Statements and detailed solutions of these problems can be downloaded from the book's web page and are also available on line at http://www.athenasc.com/convexity.html.

The book may be used as a text for a theoretical convex optimization course; I have taught several variants of such a course at MIT and elsewhere over the last ten years. It may also be used as a supplementary source for nonlinear programming classes, and as a theoretical foundation for classes focused on convex optimization models (rather than theory).

The book has been structured so that the reader/instructor can use the material selectively. For example, the polyhedral convexity material of Chapter 2 can be omitted in its entirety, as it is not used in Chapters 3-5. Similarly, the material on minimax theory (Sections 3.4, 4.2.5, and 5.5) may be omitted; and if this is done, Sections 3.3 and 5.3.4, which use the tools of partial minimization, may be omitted. Also, Sections 5.4-5.7 are "terminal" and may each be omitted without effect on other sections.

A "minimal" self-contained selection, which I have used in my nonlinear programming class at MIT (together with the supplementary web-based Chapter 6 on algorithms), consists of the following:

- Chapter 1, except for Sections 1.3.3 and 1.4.1.
- Section 3.1.
- Chapter 4, except for Section 4.2.5.
- Chapter 5, except for Sections 5.2, 5.3.4, and 5.5-5.7.

This selection focuses on nonlinear convex optimization, and excludes all the material relating to polyhedral convexity and minimax theory.

I would like to express my thanks to several colleagues for their contributions to the book. My collaboration with Angelia Nedić and Asuman Ozdaglar on our 2003 book was important in laying the foundations of the present book. Huizhen (Janey) Yu read carefully early drafts of portions of the book, and offered several insightful suggestions. Paul Tseng contributed substantially through our joint research on set intersection theory, given in part in Section 1.4.2 (this research was motivated by earlier collaboration with Angelia Nedić). Feedback from students and colleagues, including Dimitris Bisias, Vivek Borkar, John Tsitsiklis, Mengdi Wang, and Yunjian Xu, is highly appreciated. Finally, I wish to thank the many outstanding students in my classes, who have been a continuing source of motivation and inspiration.

1

Basic Concepts of Convex Analysis

Contents

Convex sets and functions are very useful in optimization models, and have a rich structure that is convenient for analysis and algorithms. Much of this structure can be traced to a few fundamental properties. For example, each closed convex set can be described in terms of the hyperplanes that support the set, each point on the boundary of a convex set can be approached through the relative interior of the set, and each halfline belonging to a closed convex set still belongs to the set when translated to start at any point in the set.

Yet, despite their favorable structure, convex sets and their analysis are not free of anomalies and exceptional behavior, which cause serious difficulties in theory and applications. For example, contrary to affine and compact sets, some basic operations such as linear transformation and vector sum may not preserve the closedness of closed convex sets. This in turn complicates the treatment of some fundamental optimization issues, including the existence of optimal solutions and duality.

For this reason, it is important to be rigorous in the development of convexity theory and its applications. Our aim in this first chapter is to establish the foundations for this development, with a special emphasis on issues that are relevant to optimization.

1.1 CONVEX SETS AND FUNCTIONS

We introduce in this chapter some of the basic notions relating to convex sets and functions. This material permeates all subsequent developments in this book. Appendix A provides the definitions, notational conventions, and results from linear algebra and real analysis that we will need. We first define convex sets (cf. Fig. 1.1.1).

Definition 1.1.1: A subset C of $\Re^n$ is called *convex* if

$$\alpha x + (1-\alpha)y \in C, \qquad \forall\ x, y \in C,\ \forall\ \alpha \in [0,1].$$

Note that the empty set is by convention considered to be convex. Generally, when referring to a convex set, it will usually be apparent from the context whether this set can be empty, but we will often be specific in order to minimize ambiguities. The following proposition gives some operations that preserve convexity.

Proposition 1.1.1:

(a) The intersection $\cap_{i\in I} C_i$ of any collection $\{C_i \mid i \in I\}$ of convex sets is convex.

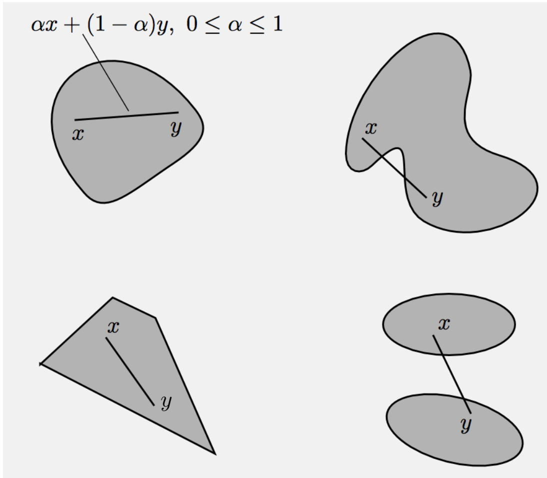

Figure 1.1.1. Illustration of the definition of a convex set. For convexity, linear interpolation between any two points of the set must yield points that lie within the set. Thus the sets on the left are convex, but the sets on the right are not.

(b) The vector sum $C_1 + C_2$ of two convex sets C_1 and C_2 is convex.

(c) The set λC is convex for any convex set C and scalar λ. Furthermore, if C is a convex set and λ_1, λ_2 are positive scalars,

$$(\lambda_1 + \lambda_2)C = \lambda_1 C + \lambda_2 C.$$

(d) The closure and the interior of a convex set are convex.

(e) The image and the inverse image of a convex set under an affine function are convex.

Proof: The proof is straightforward using the definition of convexity. To prove part (a), we take two points x and y from $\cap_{i \in I} C_i$, and we use the convexity of C_i to argue that the line segment connecting x and y belongs to all the sets C_i, and hence, to their intersection.

Similarly, to prove part (b), we take two points of $C_1 + C_2$, which we represent as $x_1 + x_2$ and $y_1 + y_2$, with $x_1, y_1 \in C_1$ and $x_2, y_2 \in C_2$. For any $\alpha \in [0, 1]$, we have

$$\alpha(x_1 + x_2) + (1-\alpha)(y_1 + y_2) = \big(\alpha x_1 + (1-\alpha) y_1\big) + \big(\alpha x_2 + (1-\alpha) y_2\big).$$

By convexity of C_1 and C_2, the vectors in the two parentheses of the right-hand side above belong to C_1 and C_2, respectively, so that their sum belongs to $C_1 + C_2$. Hence $C_1 + C_2$ is convex. The proof of part (c) is left as an exercise for the reader. The proof of part (e) is similar to the proof of part (b).

To prove part (d), let C be a convex set. Choose two points x and y from the closure of C, and sequences $\{x_k\} \subset C$ and $\{y_k\} \subset C$, such that $x_k \to x$ and $y_k \to y$. For any $\alpha \in [0,1]$, the sequence $\big\{\alpha x_k + (1-\alpha)y_k\big\}$, which belongs to C by the convexity of C, converges to $\alpha x+(1-\alpha)y$. Hence $\alpha x + (1-\alpha)y$ belongs to the closure of C, showing that the closure of C is convex. Similarly, we choose two points x and y from the interior of C, and we consider open balls that are centered at x and y, and have sufficiently small radius r so that they are contained in C. For any $\alpha \in [0,1]$, consider the open ball of radius r that is centered at $\alpha x + (1-\alpha)y$. Any point in this ball, say $\alpha x + (1-\alpha)y + z$, where $\|z\| < r$, belongs to C, because it can be expressed as the convex combination $\alpha(x+z) + (1-\alpha)(y+z)$ of the vectors $x + z$ and $y + z$, which belong to C. Hence the interior of C contains $\alpha x + (1-\alpha)y$ and is therefore convex. **Q.E.D.**

Special Convex Sets

We will often consider some special sets, which we now introduce. A *hyperplane* is a set specified by a single linear equation, i.e., a set of the form $\{x \mid a'x = b\}$, where a is a nonzero vector and b is a scalar. A *halfspace* is a set specified by a single linear inequality, i.e., a set of the form $\{x \mid a'x \le b\}$, where a is a nonzero vector and b is a scalar. It is clearly closed and convex. A set is said to be *polyhedral* if it is nonempty and it is the intersection of a finite number of halfspaces, i.e., if it has the form

$$\{x \mid a_j'x \le b_j,\ j = 1,\ldots,r\},$$

where $a_1,\ldots,a_r$ and $b_1,\ldots,b_r$ are some vectors in $\Re^n$ and scalars, respectively. A polyhedral set is convex and closed, being the intersection of halfspaces [cf. Prop. 1.1.1(a)].

A set C is said to be a *cone* if for all $x \in C$ and $\lambda > 0$, we have $\lambda x \in C$. A cone need not be convex and need not contain the origin, although the origin always lies in the closure of a nonempty cone (see Fig. 1.1.2). A *polyhedral cone* is a set of the form

$$C = \{x \mid a_j'x \le 0,\ j = 1,\ldots,r\},$$

where $a_1,\ldots,a_r$ are some vectors in $\Re^n$. A subspace is a special case of a polyhedral cone, which is in turn a special case of a polyhedral set.

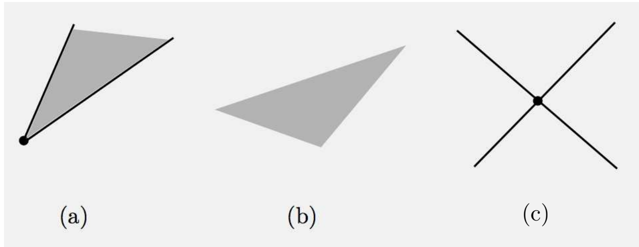

Figure 1.1.2. Illustration of convex and nonconvex cones. Cones (a) and (b) are convex, while cone (c), which consists of two lines passing through the origin, is not convex. Cone (a) is polyhedral. Cone (b) does not contain the origin.

1.1.1 Convex Functions

We now define a real-valued convex function (cf. Fig. 1.1.3).

Definition 1.1.2: Let C be a convex subset of $\Re^n$. We say that a function $f : C \mapsto \Re$ is *convex* if

$$f\big(\alpha x + (1-\alpha)y\big) \le \alpha f(x) + (1-\alpha)f(y), \qquad \forall\, x, y \in C,\ \forall\, \alpha \in [0,1]. \tag{1.1}$$

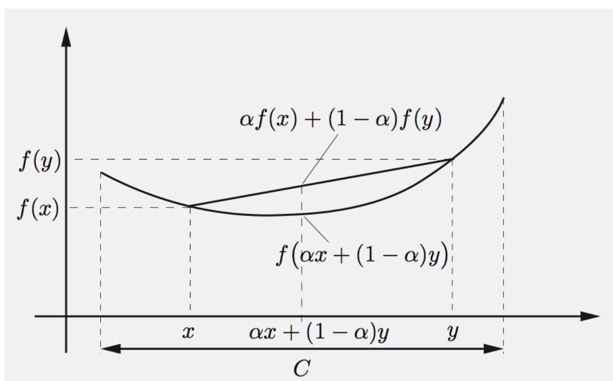

Figure 1.1.3. Illustration of the definition of a function $f : C \mapsto \Re$ that is convex. The linear interpolation $\alpha f(x) + (1-\alpha)f(y)$ overestimates the function value $f\big(\alpha x + (1-\alpha)y\big)$ for all $\alpha \in [0,1]$.

Note that, according to our definition, convexity of the domain C is a prerequisite for convexity of a function $f : C \mapsto \Re$. Thus when calling a function convex, we imply that its domain is convex.

We introduce some variants of the basic definition of convexity. A convex function $f : C \mapsto \Re$ is called *strictly convex* if the inequality (1.1) is strict for all $x, y \in C$ with $x \neq y$, and all $\alpha \in (0, 1)$. A function $f : C \mapsto \Re$, where C is a convex set, is called *concave* if the function $(-f)$ is convex.

An example of a convex function is an affine function, one of the form $f(x) = a'x + b$, where $a \in \Re^n$ and $b \in \Re$; this is straightforward to verify using the definition of convexity. Another example is a norm $\|\cdot\|$, since by the triangle inequality, we have

$$\|\alpha x + (1-\alpha)y\| \leq \|\alpha x\| + \|(1-\alpha)y\| = \alpha\|x\| + (1-\alpha)\|y\|,$$

for any $x, y \in \Re^n$, and $\alpha \in [0, 1]$.

If $f : C \mapsto \Re$ is a function and γ is a scalar, the sets $\{x \in C \mid f(x) \leq \gamma\}$ and $\{x \in C \mid f(x) < \gamma\}$, are called *level sets* of f. If f is a convex function, then all its level sets are convex. To see this, note that if $x, y \in C$ are such that $f(x) \leq \gamma$ and $f(y) \leq \gamma$, then for any $\alpha \in [0, 1]$, we have $\alpha x + (1-\alpha)y \in C$, by the convexity of C, so

$$f\big(\alpha x + (1-\alpha)y\big) \leq \alpha f(x) + (1-\alpha)f(y) \leq \gamma,$$

by the convexity of f. A similar proof also shows that the level sets $\{x \in C \mid f(x) < \gamma\}$ are convex when f is convex. Note, however, that convexity of the level sets does not imply convexity of the function; for example, the scalar function $f(x) = \sqrt{|x|}$ has convex level sets but is not convex.

Extended Real-Valued Convex Functions

We generally prefer to deal with convex functions that are real-valued and are defined over the entire space $\Re^n$ (rather than over just a convex subset), because they are mathematically simpler. However, in some situations, prominently arising in the context of optimization and duality, we will encounter operations resulting in functions that can take infinite values. For example, the function

$$f(x) = \sup_{i \in I} f_i(x),$$

where I is an infinite index set, can take the value ∞ even if the functions f_i are real-valued, and the conjugate of a real-valued function often takes infinite values (cf. Section 1.6).

Furthermore, we will encounter functions f that are convex over a convex subset C and cannot be extended to functions that are real-valued and convex over the entire space $\Re^n$ [e.g., the function $f : (0, \infty) \mapsto \Re$

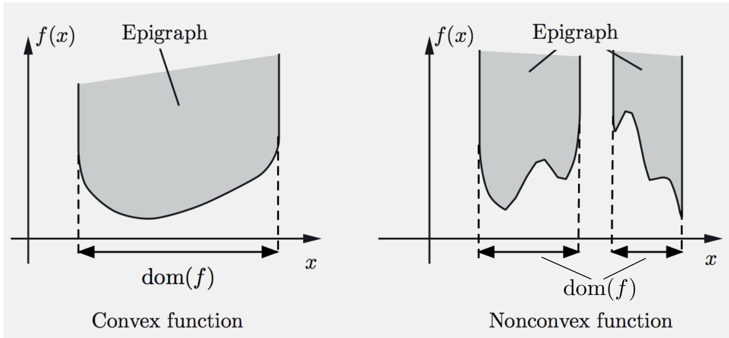

Figure 1.1.4. Illustration of the epigraphs and effective domains of extended real-valued convex and nonconvex functions.

defined by $f(x) = 1/x$]. In such situations, it may be convenient, instead of restricting the domain of f to the subset C where f takes real values, to extend the domain to all of $\Re^n$, but allow f to take infinite values.

We are thus motivated to introduce *extended real-valued* functions that can take the values $-\infty$ and ∞ at some points. Such functions can be characterized using the notion of epigraph, which we now introduce.

The *epigraph* of a function $f : X \mapsto [-\infty, \infty]$, where $X \subset \Re^n$, is defined to be the subset of $\Re^{n+1}$ given by

$$\text{epi}(f) = \big\{(x, w) \mid x \in X,\ w \in \Re,\ f(x) \le w\big\}.$$

The *effective domain* of f is defined to be the set

$$\text{dom}(f) = \big\{x \in X \mid f(x) < \infty\big\}$$

(see Fig. 1.1.4). It can be seen that

$$\text{dom}(f) = \big\{x \mid \text{there exists } w \in \Re \text{ such that } (x, w) \in \text{epi}(f)\big\},$$

i.e., dom(f) is obtained by a projection of epi(f) on $\Re^n$ (the space of x). Note that if we restrict f to its effective domain, its epigraph remains unaffected. Similarly, if we enlarge the domain of f by defining $f(x) = \infty$ for $x \notin X$, the epigraph and the effective domain remain unaffected.

It is often important to exclude the degenerate case where f is identically equal to ∞ [which is true if and only if epi(f) is empty], and the case where the function takes the value $-\infty$ at some point [which is true if and only if epi(f) contains a vertical line]. We will thus say that f is *proper* if $f(x) < \infty$ for at least one $x \in X$ and $f(x) > -\infty$ for all $x \in X$, and we will say that f *improper* if it is not proper. In words, a function is proper if and only if its epigraph is nonempty and does not contain a vertical line.

A difficulty in defining extended real-valued convex functions f that can take both values $-\infty$ and ∞ is that the term $\alpha f(x) + (1-\alpha)f(y)$ arising in our earlier definition for the real-valued case may involve the forbidden sum $-\infty + \infty$ (this, of course, may happen only if f is improper, but improper functions arise on occasion in proofs or other analyses, so we do not wish to exclude them *a priori*). The epigraph provides an effective way of dealing with this difficulty.

Definition 1.1.3: Let C be a convex subset of $\Re^n$. We say that an extended real-valued function $f : C \mapsto [-\infty, \infty]$ is *convex* if $\mathrm{epi}(f)$ is a convex subset of $\Re^{n+1}$.

It can be easily verified that, according to the above definition, convexity of f implies that its effective domain $\mathrm{dom}(f)$ and its level sets $\big\{x \in C \mid f(x) \le \gamma\big\}$ and $\big\{x \in C \mid f(x) < \gamma\big\}$ are convex sets for all scalars γ. Furthermore, if $f(x) < \infty$ for all x, or $f(x) > -\infty$ for all x, then

$$f\big(\alpha x + (1-\alpha)y\big) \le \alpha f(x) + (1-\alpha)f(y), \qquad \forall\, x, y \in C,\ \forall\, \alpha \in [0,1], \quad (1.2)$$

so the preceding definition is consistent with the earlier definition of convexity for real-valued functions.

By passing to epigraphs, we can use results about sets to infer corresponding results about functions (e.g., proving convexity). The reverse is also possible, through the notion of *indicator function* $\delta : \Re^n \mapsto (-\infty, \infty]$ of a set $X \subset \Re^n$, defined by

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

In particular, a set is convex if and only if its indicator function is convex, and it is nonempty if and only if its indicator function is proper.

A convex function $f : C \mapsto (-\infty, \infty]$ is called *strictly convex* if the inequality (1.2) is strict for all $x, y \in \mathrm{dom}(f)$ with $x \ne y$, and all $\alpha \in (0,1)$. A function $f : C \mapsto [-\infty, \infty]$, where C is a convex set, is called *concave* if the function $(-f) : C \mapsto [-\infty, \infty]$ is convex as per Definition 1.1.3.

Sometimes we will deal with functions that are defined over a (possibly nonconvex) domain C but are convex when restricted to a convex subset of their domain. The following definition formalizes this case.

Definition 1.1.4: Let C and X be subsets of $\Re^n$ such that C is nonempty and convex, and $C \subset X$. We say that an extended real-valued function $f : X \mapsto [-\infty, \infty]$ is *convex over* C if f becomes convex when the domain of f is restricted to C, i.e., if the function $\tilde{f} : C \mapsto [-\infty, \infty]$, defined by $\tilde{f}(x) = f(x)$ for all $x \in C$, is convex.

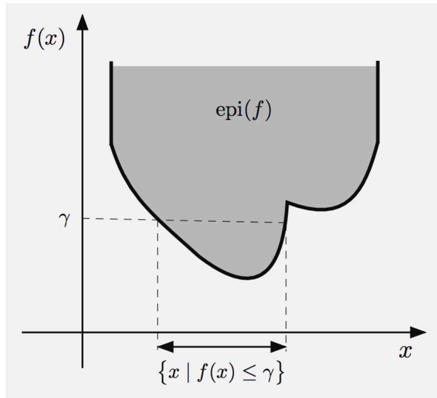

Figure 1.1.5. Visualization of the epigraph of a function in relation to its level sets. It can be seen that the level set $\{x \mid f(x) \leq \gamma\}$ can be identified with a translation of the intersection of epi(f) and the "slice" $\{(x, \gamma) \mid x \in \Re^n\}$, indicating that epi$(f)$ is closed if and only if all the level sets are closed.

By replacing the domain of an extended real-valued proper convex function with its effective domain, we can convert it to a real-valued function. In this way, we can use results stated in terms of real-valued functions, and we can also avoid calculations with ∞. Thus, nearly all the theory of convex functions can be developed without resorting to extended real-valued functions. The reverse is also true, namely that extended real-valued functions can be adopted as the norm; for example, this approach is followed by Rockafellar [Roc70]. We will adopt a flexible approach, and use both real-valued and extended real-valued functions, depending on the context.

1.1.2 Closedness and Semicontinuity

If the epigraph of a function $f : X \mapsto [-\infty, \infty]$ is a closed set, we say that f is a *closed* function. Closedness is related to the classical notion of lower semicontinuity. Recall that f is called *lower semicontinuous* at a vector $x \in X$ if

$$f(x) \leq \liminf_{k \to \infty} f(x_k)$$

for every sequence $\{x_k\} \subset X$ with $x_k \to x$. We say that f *is lower semicontinuous* if it is lower semicontinuous at each point x in its domain X. We say that f is *upper semicontinuous* if $-f$ is lower semicontinuous. These definitions are consistent with the corresponding definitions for real-valued functions [cf. Definition A.2.4(c)].

The following proposition connects closedness, lower semicontinuity, and closedness of the level sets of a function; see Fig. 1.1.5.

Proposition 1.1.2: For a function $f : \Re^n \mapsto [-\infty, \infty]$, the following are equivalent:

(i) The level set $V_\gamma = \{x \mid f(x) \leq \gamma\}$ is closed for every scalar γ.

(ii) f is lower semicontinuous.

(iii) $\mathrm{epi}(f)$ is closed.

Proof: If $f(x) = \infty$ for all x, the result trivially holds. We thus assume that $f(x) < \infty$ for at least one $x \in \Re^n$, so that $\mathrm{epi}(f)$ is nonempty and there exist level sets of f that are nonempty.

We first show that (i) implies (ii). Assume that the level set V_γ is closed for every scalar γ. Suppose, to arrive at a contradiction, that

$$f(\overline{x}) > \liminf_{k\to\infty} f(x_k)$$

for some $\overline{x}$ and sequence $\{x_k\}$ converging to $\overline{x}$, and let γ be a scalar such that

$$f(\overline{x}) > \gamma > \liminf_{k\to\infty} f(x_k).$$

Then, there exists a subsequence $\{x_k\}_\mathcal{K}$ such that $f(x_k) \leq \gamma$ for all $k \in \mathcal{K}$, so that $\{x_k\}_\mathcal{K} \subset V_\gamma$. Since V_γ is closed, $\overline{x}$ must also belong to V_γ, so $f(\overline{x}) \leq \gamma$, a contradiction.

We next show that (ii) implies (iii). Assume that f is lower semicontinuous over $\Re^n$, and let $(\overline{x}, \overline{w})$ be the limit of a sequence

$$\big\{(x_k, w_k)\big\} \subset \mathrm{epi}(f).$$

Then we have $f(x_k) \leq w_k$, and by taking the limit as $k \to \infty$ and by using the lower semicontinuity of f at $\overline{x}$, we obtain

$$f(\overline{x}) \leq \liminf_{k\to\infty} f(x_k) \leq \overline{w}.$$

Hence, $(\overline{x}, \overline{w}) \in \mathrm{epi}(f)$ and $\mathrm{epi}(f)$ is closed.

We finally show that (iii) implies (i). Assume that $\mathrm{epi}(f)$ is closed, and let $\{x_k\}$ be a sequence that converges to some $\overline{x}$ and belongs to V_γ for some scalar γ. Then $(x_k, \gamma) \in \mathrm{epi}(f)$ for all k and $(x_k, \gamma) \to (\overline{x}, \gamma)$, so since $\mathrm{epi}(f)$ is closed, we have $(\overline{x}, \gamma) \in \mathrm{epi}(f)$. Hence, $\overline{x}$ belongs to V_γ, implying that this set is closed. **Q.E.D.**

For most of our development, we prefer to use the closedness notion, rather than lower semicontinuity. One reason is that contrary to closedness, lower semicontinuity is a domain-specific property. For example, the function $f : \Re \mapsto (-\infty, \infty]$ given by

$$f(x) = \begin{cases} 0 & \text{if } x \in (0,1), \\ \infty & \text{if } x \notin (0,1), \end{cases}$$

is neither closed nor lower semicontinuous; but if its domain is restricted to $(0,1)$ it becomes lower semicontinuous.

On the other hand, if a function $f : X \mapsto [-\infty, \infty]$ has a closed effective domain $\mathrm{dom}(f)$ and is lower semicontinuous at every $x \in \mathrm{dom}(f)$, then f is closed. We state this as a proposition. The proof follows from the argument we used to show that (ii) implies (iii) in Prop. 1.1.2.

Proposition 1.1.3: Let $f : X \mapsto [-\infty, \infty]$ be a function. If $\mathrm{dom}(f)$ is closed and f is lower semicontinuous at each $x \in \mathrm{dom}(f)$, then f is closed.

As an example of application of the preceding proposition, the indicator function of a set X is closed if and only if X is closed (the "if" part follows from the proposition, and the "only if" part follows using the definition of epigraph). More generally, if f_X is a function of the form

$$f_X(x) = \begin{cases} f(x) & \text{if } x \in X, \\ \infty & \text{otherwise,} \end{cases}$$

where $f : \Re^n \mapsto \Re$ is a continuous function, then it can be shown that f_X is closed if and only if X is closed.

We finally note that an improper closed convex function is very peculiar: it cannot take a finite value at any point, so it has the form

$$f(x) = \begin{cases} -\infty & \text{if } x \in \mathrm{dom}(f), \\ \infty & \text{if } x \notin \mathrm{dom}(f). \end{cases}$$

To see this, consider an improper closed convex function $f : \Re^n \mapsto [-\infty, \infty]$, and assume that there exists an x such that $f(x)$ is finite. Let $\overline{x}$ be such that $f(\overline{x}) = -\infty$ (such a point must exist since f is improper and f is not identically equal to ∞). Because f is convex, it can be seen that every point of the form

$$x_k = \frac{k-1}{k} x + \frac{1}{k} \overline{x}, \qquad \forall\ k = 1, 2, \ldots$$

satisfies $f(x_k) = -\infty$, while we have $x_k \to x$. Since f is closed, this implies that $f(x) = -\infty$, which is a contradiction. In conclusion, *a closed convex function that is improper cannot take a finite value anywhere*.

1.1.3 Operations with Convex Functions

We can verify the convexity of a given function in a number of ways. Several commonly encountered functions, such as affine functions and norms, are convex. An important type of convex function is a *polyhedral function*, which by definition is a proper convex function whose epigraph is a polyhedral set. Starting with some known convex functions, we can generate other convex functions by using some common operations that preserve convexity. Principal among these operations are the following:

(a) Composition with a linear transformation.

(b) Addition, and multiplication with a nonnegative scalar.

(c) Taking supremum.

(d) Taking partial minimum, i.e., minimizing with respect to z a function that is (jointly) convex in two vectors x and z.

The following three propositions deal with the first three cases, and Section 3.3 deals with the fourth.

Proposition 1.1.4: Let $f : \Re^m \mapsto (-\infty, \infty]$ be a given function, let A be an $m \times n$ matrix, and let $F : \Re^n \mapsto (-\infty, \infty]$ be the function

$$F(x) = f(Ax), \qquad x \in \Re^n.$$

If f is convex, then F is also convex, while if f is closed, then F is also closed.

Proof: Let f be convex. We use the definition of convexity to write for any $x, y \in \Re^n$ and $\alpha \in [0, 1]$,

$$\begin{aligned} F\big(\alpha x + (1-\alpha)y\big) &= f\big(\alpha Ax + (1-\alpha)Ay\big) \\ &\le \alpha f(Ax) + (1-\alpha) f(Ay) \\ &= \alpha F(x) + (1-\alpha) F(y). \end{aligned}$$

Hence F is convex.

Let f be closed. Then f is lower semicontinuous at every $x \in \Re^n$ (cf. Prop. 1.1.2), so for every sequence $\{x_k\}$ converging to x, we have

$$f(Ax) \le \liminf_{k \to \infty} f(Ax_k),$$

or

$$F(x) \le \liminf_{k \to \infty} F(x_k)$$

for all k. It follows that F is lower semicontinuous at every $x \in \Re^n$, and hence is closed by Prop. 1.1.2. **Q.E.D.**

The next proposition deals with sums of function and it is interesting to note that it can be viewed as a special case of the preceding proposition, which deals with compositions with linear transformations. The reason is that we may write a sum $F = f_1 + \cdots + f_m$ in the form $F(x) = f(Ax)$, where A is the matrix defined by $Ax = (x, \ldots, x)$, and $f : \Re^{mn} \mapsto (-\infty, \infty]$ is the function given by

$$f(x_1, \ldots, x_m) = f_1(x_1) + \cdots + f_m(x_m).$$

Proposition 1.1.5: Let $f_i : \Re^n \mapsto (-\infty, \infty]$, $i = 1, \ldots, m$, be given functions, let $\gamma_1, \ldots, \gamma_m$ be positive scalars, and let $F : \Re^n \mapsto (-\infty, \infty]$ be the function

$$F(x) = \gamma_1 f_1(x) + \cdots + \gamma_m f_m(x), \qquad x \in \Re^n.$$

If $f_1, \ldots, f_m$ are convex, then F is also convex, while if $f_1, \ldots, f_m$ are closed, then F is also closed.

Proof: The proof follows closely the one of Prop. 1.1.4. **Q.E.D.**

Proposition 1.1.6: Let $f_i : \Re^n \mapsto (-\infty, \infty]$ be given functions for $i \in I$, where I is an arbitrary index set, and let $f : \Re^n \mapsto (-\infty, \infty]$ be the function given by

$$f(x) = \sup_{i \in I} f_i(x).$$

If f_i, $i \in I$, are convex, then f is also convex, while if f_i, $i \in I$, are closed, then f is also closed.

Proof: A pair (x, w) belongs to $\text{epi}(f)$ if and only if $f(x) \leq w$, which is true if and only if $f_i(x) \leq w$ for all $i \in I$, or equivalently $(x, w) \in \cap_{i \in I} \text{epi}(f_i)$. Therefore,

$$\text{epi}(f) = \cap_{i \in I} \text{epi}(f_i).$$

If the functions f_i are convex, the epigraphs $\text{epi}(f_i)$ are convex, so $\text{epi}(f)$ is convex, and f is convex. If the functions f_i are closed, the epigraphs $\text{epi}(f_i)$ are closed, so $\text{epi}(f)$ is closed, and f is closed. **Q.E.D.**

1.1.4 Characterizations of Differentiable Convex Functions

For once or twice differentiable functions, there are some additional criteria for verifying convexity, as we will now discuss. A useful alternative characterization of convexity for differentiable functions is given in the following proposition and is illustrated in Fig. 1.1.6.

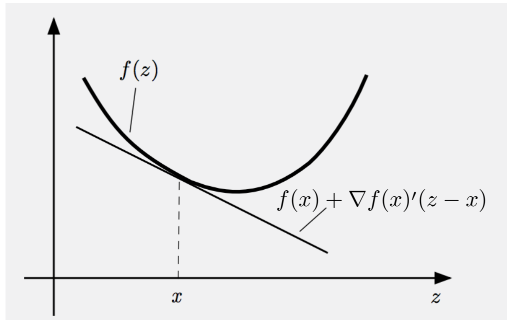

Figure 1.1.6. Characterization of convexity in terms of first derivatives. The condition $f(z) \geq f(x) + \nabla f(x)'(z - x)$ states that a linear approximation, based on the gradient, underestimates a convex function.

Proposition 1.1.7: Let C be a nonempty convex subset of $\Re^n$ and let $f : \Re^n \mapsto \Re$ be differentiable over an open set that contains C.

(a) f is convex over C if and only if

$$f(z) \geq f(x) + \nabla f(x)'(z - x), \qquad \forall\, x, z \in C. \tag{1.3}$$

(b) f is strictly convex over C if and only if the above inequality is strict whenever $x \neq z$.

Proof: The ideas of the proof are geometrically illustrated in Fig. 1.1.7. We prove (a) and (b) simultaneously. Assume that the inequality (1.3) holds. Choose any $x, y \in C$ and $\alpha \in [0, 1]$, and let $z = \alpha x + (1 - \alpha)y$. Using the inequality (1.3) twice, we obtain

$$f(x) \geq f(z) + \nabla f(z)'(x - z),$$

$$f(y) \geq f(z) + \nabla f(z)'(y - z).$$

We multiply the first inequality by α, the second by $(1 - \alpha)$, and add them to obtain

$$\alpha f(x) + (1 - \alpha) f(y) \geq f(z) + \nabla f(z)'\big(\alpha x + (1 - \alpha)y - z\big) = f(z),$$

which proves that f is convex. If the inequality (1.3) is strict as stated in part (b), then if we take $x \neq y$ and $\alpha \in (0, 1)$ above, the three preceding inequalities become strict, thus showing the strict convexity of f.

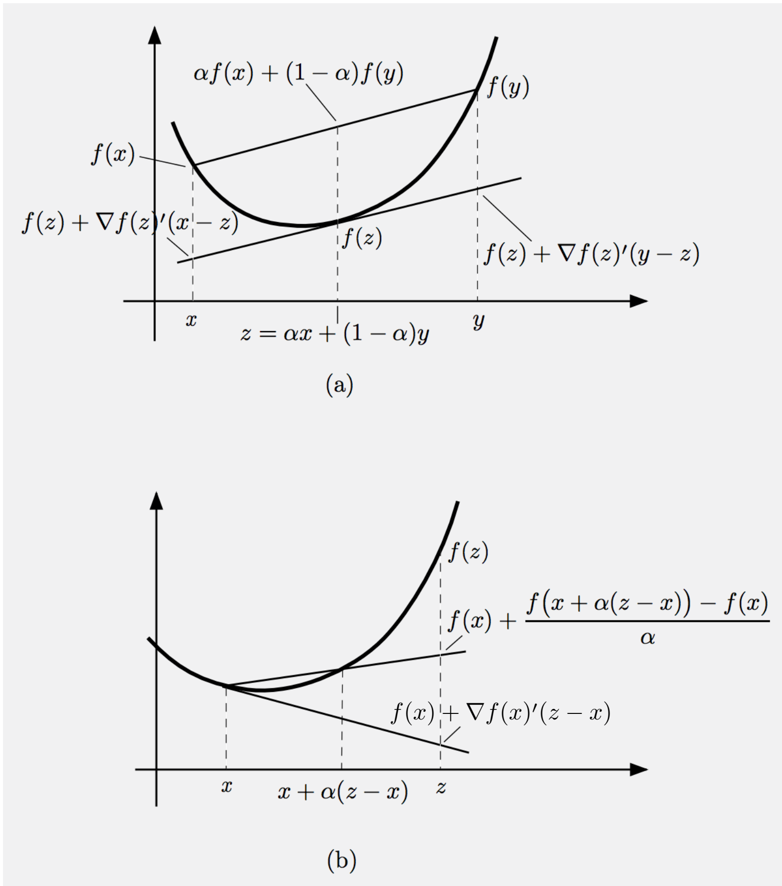

Figure 1.1.7. Geometric illustration of the ideas underlying the proof of Prop. 1.1.7. In figure (a), we linearly approximate f at $z = \alpha x + (1-\alpha)y$. The inequality (1.3) implies that

$$f(x) \geq f(z) + \nabla f(z)'(x - z),$$

$$f(y) \geq f(z) + \nabla f(z)'(y - z).$$

As can be seen from the figure, it follows that $\alpha f(x) + (1-\alpha)f(y)$ lies above $f(z)$, so f is convex.

In figure (b), we assume that f is convex, and from the figure's geometry, we note that

$$f(x) + \frac{f\big(x + \alpha(z - x)\big) - f(x)}{\alpha}$$

lies below $f(z)$, is monotonically nonincreasing as $\alpha \downarrow 0$, and converges to $f(x) + \nabla f(x)'(z - x)$. It follows that $f(z) \geq f(x) + \nabla f(x)'(z - x)$.

Conversely, assume that f is convex, let x and z be any vectors in C with $x \neq z$, and consider the function $g : (0, 1] \mapsto \Re$ given by

$$g(\alpha) = \frac{f\big(x + \alpha(z - x)\big) - f(x)}{\alpha}, \qquad \alpha \in (0, 1].$$

We will show that $g(\alpha)$ is monotonically increasing with α, and is strictly monotonically increasing if f is strictly convex. This will imply that

$$\nabla f(x)'(z - x) = \lim_{\alpha \downarrow 0} g(\alpha) \leq g(1) = f(z) - f(x),$$

with strict inequality if g is strictly monotonically increasing, thereby showing that the desired inequality (1.3) holds, and holds strictly if f is strictly convex. Indeed, consider any α_1, α_2, with $0 < \alpha_1 < \alpha_2 < 1$, and let

$$\overline{\alpha} = \frac{\alpha_1}{\alpha_2}, \qquad \overline{z} = x + \alpha_2(z - x). \tag{1.4}$$

We have

$$f\big(x + \overline{\alpha}(\overline{z} - x)\big) \leq \overline{\alpha} f(\overline{z}) + (1 - \overline{\alpha}) f(x),$$

or

$$\frac{f\big(x + \overline{\alpha}(\overline{z} - x)\big) - f(x)}{\overline{\alpha}} \leq f(\overline{z}) - f(x), \tag{1.5}$$

and the above inequalities are strict if f is strictly convex. Substituting the definitions (1.4) in Eq. (1.5), we obtain after a straightforward calculation

$$\frac{f\big(x + \alpha_1(z - x)\big) - f(x)}{\alpha_1} \leq \frac{f\big(x + \alpha_2(z - x)\big) - f(x)}{\alpha_2},$$

or

$$g(\alpha_1) \leq g(\alpha_2),$$

with strict inequality if f is strictly convex. Hence g is monotonically increasing with α, and strictly so if f is strictly convex. **Q.E.D.**

Note a simple consequence of Prop. 1.1.7(a): if $f : \Re^n \mapsto \Re$ is a differentiable convex function and $\nabla f(x^*) = 0$, then x^* minimizes f over $\Re^n$. This is a classical sufficient condition for unconstrained optimality, originally formulated (in one dimension) by Fermat in 1637. Similarly, from Prop. 1.1.7(a), we see that the condition

$$\nabla f(x^*)'(z - x^*) \geq 0, \qquad \forall\ z \in C,$$

implies that x^* minimizes a differentiable convex function f over a convex set C. This sufficient condition for optimality is also necessary. To see this, assume to arrive at a contradiction that x^* minimizes f over C and that $\nabla f(x^*)'(z - x^*) < 0$ for some $z \in C$. By differentiation, we have

$$\lim_{\alpha \downarrow 0} \frac{f\big(x^* + \alpha(z - x^*)\big) - f(x^*)}{\alpha} = \nabla f(x^*)'(z - x^*) < 0,$$

so $f\big(x^* + \alpha(z - x^*)\big)$ decreases strictly for sufficiently small $\alpha > 0$, contradicting the optimality of x^*. We state the conclusion as a proposition.

Proposition 1.1.8: Let C be a nonempty convex subset of $\Re^n$ and let $f : \Re^n \mapsto \Re$ be convex and differentiable over an open set that contains C. Then a vector $x^* \in C$ minimizes f over C if and only if

$$\nabla f(x^*)'(z - x^*) \geq 0, \qquad \forall\, z \in C.$$

Let us use the preceding optimality condition to prove a basic theorem of analysis and optimization.

Proposition 1.1.9: (Projection Theorem) Let C be a nonempty closed convex subset of $\Re^n$, and let z be a vector in $\Re^n$. There exists a unique vector that minimizes $\|z - x\|$ over $x \in C$, called the projection of z on C. Furthermore, a vector x^* is the projection of z on C if and only if

$$(z - x^*)'(x - x^*) \leq 0, \qquad \forall\, x \in C. \tag{1.6}$$

Proof: Minimizing $\|z - x\|$ is equivalent to minimizing the convex and differentiable function

$$f(x) = \tfrac{1}{2}\|z - x\|^2.$$

By Prop. 1.1.8, x^* minimizes f over C if and only if

$$\nabla f(x^*)'(x - x^*) \geq 0, \qquad \forall\, x \in C.$$

Since $\nabla f(x^*) = x^* - z$, this condition is equivalent to Eq. (1.6).

Minimizing f over C is equivalent to minimizing f over the compact set $C \cap \big\{\|z - x\| \leq \|z - w\|\big\}$, where w is any vector in C. By Weierstrass' Theorem (Prop. A.2.7), it follows that there exists a minimizing vector. To show uniqueness, let x_1^* and x_2^* be two minimizing vectors. Then by Eq. (1.6), we have

$$(z - x_1^*)'(x_2^* - x_1^*) \leq 0, \qquad (z - x_2^*)'(x_1^* - x_2^*) \leq 0.$$

Adding these two inequalities, we obtain

$$(x_2^* - x_1^*)'(x_2^* - x_1^*) = \|x_2^* - x_1^*\|^2 \le 0,$$

so $x_2^* = x_1^*$. **Q.E.D.**

For twice differentiable convex functions, there is another characterization of convexity, given by the following proposition.

Proposition 1.1.10: Let C be a nonempty convex subset of $\Re^n$ and let $f : \Re^n \mapsto \Re$ be twice continuously differentiable over an open set that contains C.

(a) If $\nabla^2 f(x)$ is positive semidefinite for all $x \in C$, then f is convex over C.

(b) If $\nabla^2 f(x)$ is positive definite for all $x \in C$, then f is strictly convex over C.

(c) If C is open and f is convex over C, then $\nabla^2 f(x)$ is positive semidefinite for all $x \in C$.

Proof: (a) Using the mean value theorem (Prop. A.3.1), we have for all $x, y \in C$,

$$f(y) = f(x) + \nabla f(x)'(y - x) + \tfrac{1}{2}(y - x)'\nabla^2 f\big(x + \alpha(y - x)\big)(y - x)$$

for some $\alpha \in [0, 1]$. Therefore, using the positive semidefiniteness of $\nabla^2 f$, we obtain

$$f(y) \ge f(x) + \nabla f(x)'(y - x), \qquad \forall\ x, y \in C.$$

From Prop. 1.1.7(a), we conclude that f is convex over C.

(b) Similar to the proof of part (a), we have $f(y) > f(x) + \nabla f(x)'(y - x)$ for all $x, y \in C$ with $x \neq y$, and the result follows from Prop. 1.1.7(b).

(c) Assume, to obtain a contradiction, that there exist some $x \in C$ and some $z \in \Re^n$ such that $z'\nabla^2 f(x)z < 0$. Since C is open and $\nabla^2 f$ is continuous, we can choose z to have small enough norm so that $x + z \in C$ and $z'\nabla^2 f(x + \alpha z)z < 0$ for every $\alpha \in [0, 1]$. Then, using again the mean value theorem, we obtain $f(x + z) < f(x) + \nabla f(x)'z$, which, in view of Prop. 1.1.7(a), contradicts the convexity of f over C. **Q.E.D.**

If f is convex over a convex set C that is not open, $\nabla^2 f(x)$ may not be positive semidefinite at any point of C [take for example $n = 2$, $C = \big\{(x_1, 0) \mid x_1 \in \Re\big\}$, and $f(x) = x_1^2 - x_2^2$]. However, it can be shown that the conclusion of Prop. 1.1.10(c) also holds if C has nonempty interior instead of being open.

1.2 CONVEX AND AFFINE HULLS

We now discuss issues relating to the convexification of nonconvex sets. Let X be a nonempty subset of $\Re^n$. The *convex hull* of a set X, denoted $\mathrm{conv}(X)$, is the intersection of all convex sets containing X, and is a convex set by Prop. 1.1.1(a). A *convex combination* of elements of X is a vector of the form $\sum_{i=1}^m \alpha_i x_i$, where m is a positive integer, $x_1, \ldots, x_m$ belong to X, and $\alpha_1, \ldots, \alpha_m$ are scalars such that

$$\alpha_i \geq 0, \quad i = 1, \ldots, m, \qquad \sum_{i=1}^m \alpha_i = 1.$$

Note that a convex combination belongs to $\mathrm{conv}(X)$ (see the construction of Fig. 1.2.1). For any convex combination and function $f : \Re^n \mapsto \Re$ that is convex over $\mathrm{conv}(X)$, we have

$$f\left(\sum_{i=1}^m \alpha_i x_i\right) \leq \sum_{i=1}^m \alpha_i f(x_i). \tag{1.7}$$

This follows by using repeatedly the definition of convexity together with the construction of Fig. 1.2.1. The preceding relation is a special case of a relation known as *Jensen's inequality*, which finds wide use in applied mathematics and probability theory.

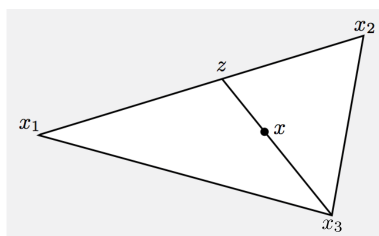

Figure 1.2.1. Construction of a convex combination of m vectors by forming a sequence of $m-1$ convex combinations of pairs of vectors (first combine two vectors, then combine the result with a third vector, etc). For example,

$$x = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 = (\alpha_1 + \alpha_2)\left(\frac{\alpha_1}{\alpha_1 + \alpha_2} x_1 + \frac{\alpha_2}{\alpha_1 + \alpha_2} x_2\right) + \alpha_3 x_3,$$

so the convex combination $\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3$ can be obtained by forming the convex combination

$$z = \frac{\alpha_1}{\alpha_1 + \alpha_2} x_1 + \frac{\alpha_2}{\alpha_1 + \alpha_2} x_2,$$

and then by forming the convex combination $x = (\alpha_1 + \alpha_2) z + \alpha_3 x_3$ as shown in the figure. This shows that a convex combination of vectors from a convex set belongs to the set, and that a convex combination of vectors from a nonconvex set belongs to the convex hull of the set.

It is straightforward to verify that the set of all convex combinations of elements of X is equal to $\mathrm{conv}(X)$. In particular, if X consists of a finite number of vectors $x_1, \ldots, x_m$, its convex hull is

$$\mathrm{conv}\big(\{x_1, \ldots, x_m\}\big) = \left\{ \sum_{i=1}^{m} \alpha_i x_i \;\middle|\; \alpha_i \geq 0,\, i = 1, \ldots, m,\, \sum_{i=1}^{m} \alpha_i = 1 \right\}.$$

Also, for any set S and linear transformation A, we have $\mathrm{conv}(AS) = A\,\mathrm{conv}(S)$. From this it follows that for any sets $S_1, \ldots, S_m$, we have $\mathrm{conv}(S_1 + \cdots + S_m) = \mathrm{conv}(S_1) + \cdots + \mathrm{conv}(S_m)$.

We recall that an affine set M in $\Re^n$ is a set of the form $x + S$, where x is some vector, and S is a subspace uniquely determined by M and called the *subspace parallel to* M. Alternatively, a set M is affine if it contains all the lines that pass through pairs of points $x, y \in M$ with $x \neq y$. If X is a subset of $\Re^n$, the *affine hull* of X, denoted $\mathrm{aff}(X)$, is the intersection of all affine sets containing X. Note that $\mathrm{aff}(X)$ is itself an affine set and that it contains $\mathrm{conv}(X)$. The dimension of $\mathrm{aff}(X)$ is defined to be the dimension of the subspace parallel to $\mathrm{aff}(X)$. It can be shown that

$$\mathrm{aff}(X) = \mathrm{aff}\big(\mathrm{conv}(X)\big) = \mathrm{aff}\big(\mathrm{cl}(X)\big).$$

For a convex set C, the *dimension* of C is defined to be the dimension of $\mathrm{aff}(C)$.

Given a nonempty subset X of $\Re^n$, a *nonnegative combination* of elements of X is a vector of the form $\sum_{i=1}^{m} \alpha_i x_i$, where m is a positive integer, $x_1, \ldots, x_m$ belong to X, and $\alpha_1, \ldots, \alpha_m$ are nonnegative scalars. If the scalars α_i are all positive, $\sum_{i=1}^{m} \alpha_i x_i$ is said to be a *positive combination*. The *cone generated by* X, denoted $\mathrm{cone}(X)$, is the set of all nonnegative combinations of elements of X. It is easily seen that $\mathrm{cone}(X)$ is a convex cone containing the origin, although it need not be closed even if X is compact, as shown in Fig. 1.2.2 [it can be proved that $\mathrm{cone}(X)$ is closed in special cases, such as when X is finite; see Section 1.4.3].

The following is a fundamental characterization of convex hulls (see Fig. 1.2.3).

Proposition 1.2.1: (Caratheodory's Theorem) Let X be a nonempty subset of $\Re^n$.

(a) Every nonzero vector from $\mathrm{cone}(X)$ can be represented as a positive combination of linearly independent vectors from X.

(b) Every vector from $\mathrm{conv}(X)$ can be represented as a convex combination of no more than $n + 1$ vectors from X.

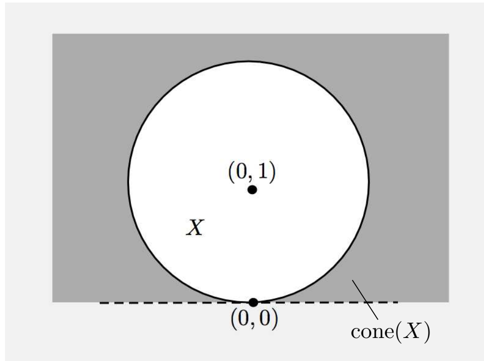

Figure 1.2.2. An example in $\Re^2$ where X is convex and compact, but cone(X) is not closed. Here

$$X = \{(x_1, x_2) \mid x_1^2 + (x_2 - 1)^2 \le 1\}, \quad \text{cone}(X) = \{(x_1, x_2) \mid x_2 > 0\} \cup \{(0, 0)\}.$$

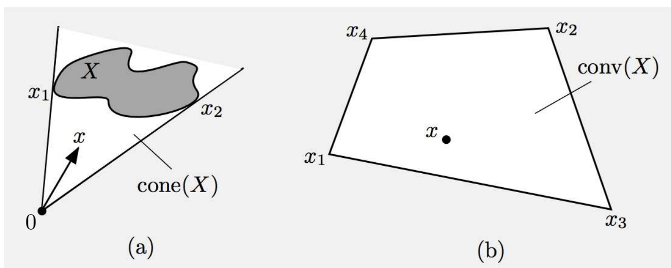

Figure 1.2.3. Illustration of Caratheodory's Theorem. In (a), X is a nonconvex set in $\Re^2$, and a point $x \in \text{cone}(X)$ is represented as a positive combination of the two linearly independent vectors $x_1, x_2 \in X$. In (b), X consists of four points x_1, x_2, x_3, x_4 in $\Re^2$, and the point $x \in \text{conv}(X)$ shown in the figure can be represented as a convex combination of the three vectors x_1, x_2, x_3. Note also that x can alternatively be represented as a convex combination of the vectors x_1, x_3, x_4, so the representation is not unique.

Proof: (a) Consider a vector $x \neq 0$ from cone(X), and let m be the smallest integer such that x has the form $\sum_{i=1}^m \alpha_i x_i$, where $\alpha_i > 0$ and $x_i \in X$ for all $i = 1, \ldots, m$. We argue by contradiction. If the vectors x_i are linearly dependent, there exist scalars $\lambda_1, \ldots, \lambda_m$, with $\sum_{i=1}^m \lambda_i x_i = 0$ and at least one λ_i is positive. Consider the linear combination $\sum_{i=1}^m (\alpha_i - \overline{\gamma}\lambda_i) x_i$, where $\overline{\gamma}$ is the largest γ such that $\alpha_i - \gamma\lambda_i \ge 0$ for all i. This combination provides a representation of x as a positive combination of fewer than m vectors of X – a contradiction. Therefore, $x_1, \ldots, x_m$ are linearly independent.

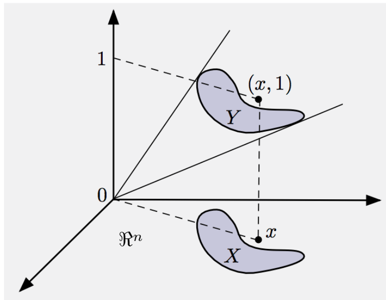

Figure 1.2.4. Illustration of the proof of Caratheodory's Theorem for convex hulls using the version of the theorem for generated cones. We consider the set $Y = \{(y,1) \mid y \in X\} \subset \Re^{n+1}$ and apply Prop. 1.2.1(a).

(b) We apply part (a) to the following subset of $\Re^{n+1}$:

$$Y = \{(y,1) \mid y \in X\}$$

(cf. Fig. 1.2.4). If $x \in \text{conv}(X)$, we have $x = \sum_{i=1}^{I} \gamma_i x_i$ for an integer $I > 0$ and scalars $\gamma_i > 0$, $i = 1, \ldots, I$, with $1 = \sum_{i=1}^{I} \gamma_i$, so that $(x,1) \in \text{cone}(Y)$. By part (a), we have $(x,1) = \sum_{i=1}^{m} \alpha_i (x_i, 1)$ for some scalars $\alpha_1, \ldots, \alpha_m > 0$ and (at most $n+1$) linearly independent vectors $(x_1, 1), \ldots, (x_m, 1)$. Thus, $x = \sum_{i=1}^{m} \alpha_i x_i$ and $1 = \sum_{i=1}^{m} \alpha_i$. **Q.E.D.**

Note that the proof of part (b) of Caratheodory's Theorem shows that if $m \geq 2$, the m vectors $x_1, \ldots, x_m \in X$ used to represent a vector in $\text{conv}(X)$ may be chosen so that $x_2 - x_1, \ldots, x_m - x_1$ are linearly independent [if $x_2 - x_1, \ldots, x_m - x_1$ were linearly dependent, there exist $\lambda_2, \ldots, \lambda_m$, not all 0, with $\sum_{i=2}^{m} \lambda_i (x_i - x_1) = 0$ so that by defining $\lambda_1 = -(\lambda_2 + \cdots + \lambda_m)$,

$$\sum_{i=1}^{m} \lambda_i (x_i, 1) = 0,$$

contradicting the linear independence of $(x_1, 1), \ldots, (x_m, 1)$].

Caratheodory's Theorem can be used to prove several other important results. An example is the following proposition.

Proposition 1.2.2: The convex hull of a compact set is compact.

Proof: Let X be a compact subset of $\Re^n$. To show that $\text{conv}(X)$ is compact, we take a sequence in $\text{conv}(X)$ and show that it has a convergent subsequence whose limit is in $\text{conv}(X)$. Indeed, by Caratheodory's Theorem, a sequence in $\text{conv}(X)$ can be expressed as $\left\{\sum_{i=1}^{n+1} \alpha_i^k x_i^k\right\}$, where for all k and i, $\alpha_i^k \geq 0$, $x_i^k \in X$, and $\sum_{i=1}^{n+1} \alpha_i^k = 1$. Since the sequence

$$\{(\alpha_1^k, \ldots, \alpha_{n+1}^k, x_1^k, \ldots, x_{n+1}^k)\}$$

is bounded, it has a limit point $\big\{(\alpha_1,\ldots,\alpha_{n+1},x_1,\ldots,x_{n+1})\big\}$, which must satisfy $\sum_{i=1}^{n+1}\alpha_i = 1$, and $\alpha_i \geq 0$, $x_i \in X$ for all i. Thus, the vector $\sum_{i=1}^{n+1}\alpha_i x_i$, which belongs to conv$(X)$, is a limit point of the sequence $\Big\{\sum_{i=1}^{n+1}\alpha_i^k x_i^k\Big\}$, showing that conv$(X)$ is compact. **Q.E.D.**

Note that the convex hull of an unbounded closed set need not be closed. As an example, for the closed subset of $\Re^2$

$$X = \big\{(0,0)\big\} \cup \big\{(x_1,x_2) \mid x_1x_2 \geq 1,\, x_1 \geq 0,\, x_2 \geq 0\big\},$$

the convex hull is

$$\text{conv}(X) = \big\{(0,0)\big\} \cup \big\{(x_1,x_2) \mid x_1 > 0,\, x_2 > 0\big\},$$

which is not closed.

We finally note that just as one can convexify nonconvex sets through the convex hull operation, one can also convexify a nonconvex function by convexification of its epigraph. In fact, this can be done in a way that the optimality of the minima of the function is maintained (see Section 1.3.3).

1.3 RELATIVE INTERIOR AND CLOSURE

We now consider some generic topological properties of convex sets and functions. Let C be a nonempty convex subset of $\Re^n$. The closure of C, denoted cl(C), is also a nonempty convex set [Prop. 1.1.1(d)]. The interior of C is also convex, but it may be empty. It turns out, however, that convexity implies the existence of interior points relative to the affine hull of C. This is an important property, which we now formalize.

Let C be a nonempty convex set. We say that x is a *relative interior point* of C if $x \in C$ and there exists an open sphere S centered at x such that $S \cap \text{aff}(C) \subset C$, i.e., x is an interior point of C relative to the affine hull of C. The set of all relative interior points of C is called the *relative interior of* C, and is denoted by ri(C). The set C is said to be *relatively open* if ri$(C) = C$. The vectors in cl(C) that are not relative interior points are said to be *relative boundary points* of C, and their collection is called the *relative boundary* of C.

For an example, let C be a line segment connecting two distinct points in the plane. Then ri(C) consists of all points of C except for the two end points, and the relative boundary of C consists of the two end points. For another example, let C be an affine set. Then ri$(C) = C$ and the relative boundary of C is empty.

The most fundamental fact about relative interiors is given in the following proposition.

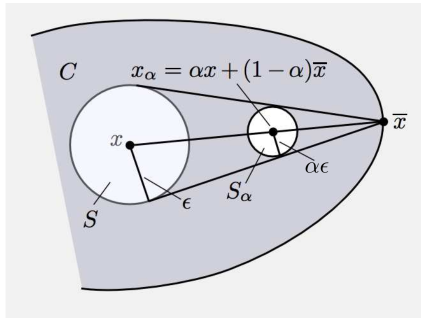

Figure 1.3.1. Proof of the Line Segment Principle for the case where $\overline{x} \in C$. Since $x \in \text{ri}(C)$, there exists an open sphere $S = \{z \mid \|z - x\| < \epsilon\}$ such that $S \cap \text{aff}(C) \subset C$. For all $\alpha \in (0, 1]$, let $x_\alpha = \alpha x + (1-\alpha)\overline{x}$ and let $S_\alpha = \{z \mid \|z - x_\alpha\| < \alpha\epsilon\}$. It can be seen that each point of $S_\alpha \cap \text{aff}(C)$ is a convex combination of $\overline{x}$ and some point of $S \cap \text{aff}(C)$. Therefore, by the convexity of C, $S_\alpha \cap \text{aff}(C) \subset C$, implying that $x_\alpha \in \text{ri}(C)$.

Proposition 1.3.1: (Line Segment Principle) Let C be a nonempty convex set. If $x \in \text{ri}(C)$ and $\overline{x} \in \text{cl}(C)$, then all points on the line segment connecting x and $\overline{x}$, except possibly $\overline{x}$, belong to $\text{ri}(C)$.

Proof: For the case where $\overline{x} \in C$, the proof is given in Fig. 1.3.1. Consider the case where $\overline{x} \notin C$. To show that for any $\alpha \in (0, 1]$ we have $x_\alpha = \alpha x + (1-\alpha)\overline{x} \in \text{ri}(C)$, consider a sequence $\{x_k\} \subset C$ that converges to $\overline{x}$, and let $x_{k,\alpha} = \alpha x + (1-\alpha)x_k$. Then as in Fig. 1.3.1, we see that $\{z \mid \|z - x_{k,\alpha}\| < \alpha\epsilon\} \cap \text{aff}(C) \subset C$ for all k, where ϵ is such that the open sphere $S = \{z \mid \|z - x\| < \epsilon\}$ satisfies $S \cap \text{aff}(C) \subset C$. Since $x_{k,\alpha} \to x_\alpha$, for large enough k, we have

$$\{z \mid \|z - x_\alpha\| < \alpha\epsilon/2\} \subset \{z \mid \|z - x_{k,\alpha}\| < \alpha\epsilon\}.$$

It follows that $\{z \mid \|z - x_\alpha\| < \alpha\epsilon/2\} \cap \text{aff}(C) \subset C$, which shows that $x_\alpha \in \text{ri}(C)$. **Q.E.D.**

A major consequence of the Line Segment Principle is given in the following proposition.

Proposition 1.3.2: (Nonemptiness of Relative Interior) Let C be a nonempty convex set. Then:

(a) $\text{ri}(C)$ is a nonempty convex set, and has the same affine hull as C.

(b) If m is the dimension of $\mathrm{aff}(C)$ and $m > 0$, there exist vectors $x_0, x_1, \ldots, x_m \in \mathrm{ri}(C)$ such that $x_1 - x_0, \ldots, x_m - x_0$ span the subspace parallel to $\mathrm{aff}(C)$.

Proof: (a) Convexity of $\mathrm{ri}(C)$ follows from the Line Segment Principle (Prop. 1.3.1). By using a translation argument if necessary, we assume without loss of generality that $0 \in C$. Then $\mathrm{aff}(C)$ is a subspace whose dimension will be denoted by m. To show that $\mathrm{ri}(C)$ is nonempty, we will use a basis for $\mathrm{aff}(C)$ to construct a relatively open set.

If the dimension m is 0, then C and $\mathrm{aff}(C)$ consist of a single point, which is a unique relative interior point. If $m > 0$, we can find m linearly independent vectors $z_1, \ldots, z_m$ in C that span $\mathrm{aff}(C)$; otherwise there would exist $r < m$ linearly independent vectors in C whose span contains C, contradicting the fact that the dimension of $\mathrm{aff}(C)$ is m. Thus $z_1, \ldots, z_m$ form a basis for $\mathrm{aff}(C)$.

Consider the set

$$X = \left\{ x \;\middle|\; x = \sum_{i=1}^{m} \alpha_i z_i,\ \sum_{i=1}^{m} \alpha_i < 1,\ \alpha_i > 0,\ i = 1, \ldots, m \right\}$$

(see Fig. 1.3.2), and note that $X \subset C$ since C is convex. We claim that this set is open relative to $\mathrm{aff}(C)$, i.e., for every vector $\overline{x} \in X$, there exists an open ball B centered at $\overline{x}$ such that $\overline{x} \in B$ and $B \cap \mathrm{aff}(C) \subset X$. To see this, fix $\overline{x} \in X$ and let x be another vector in $\mathrm{aff}(C)$. We have $\overline{x} = Z\overline{\alpha}$ and $x = Z\alpha$, where Z is the $n \times m$ matrix whose columns are the vectors $z_1, \ldots, z_m$, and $\overline{\alpha}$ and α are suitable m-dimensional vectors, which are unique since $z_1, \ldots, z_m$ form a basis for $\mathrm{aff}(C)$. Since Z has linearly independent columns, the matrix $Z'Z$ is symmetric and positive definite, so for some positive scalar γ, which is independent of x and $\overline{x}$, we have

$$\|x - \overline{x}\|^2 = (\alpha - \overline{\alpha})' Z' Z (\alpha - \overline{\alpha}) \geq \gamma \|\alpha - \overline{\alpha}\|^2. \tag{1.8}$$

Since $\overline{x} \in X$, the corresponding vector $\overline{\alpha}$ lies in the open set

$$A = \left\{ (\alpha_1, \ldots, \alpha_m) \;\middle|\; \sum_{i=1}^{m} \alpha_i < 1,\ \alpha_i > 0,\ i = 1, \ldots, m \right\}.$$

From Eq. (1.8), we see that if x lies in a suitably small ball centered at $\overline{x}$, the corresponding vector α lies in A, implying that $x \in X$. Hence X contains the intersection of $\mathrm{aff}(C)$ and an open ball centered at $\overline{x}$, so X is open relative to $\mathrm{aff}(C)$. It follows that all points of X are relative interior points of C, so that $\mathrm{ri}(C)$ is nonempty. Also, since by construction,

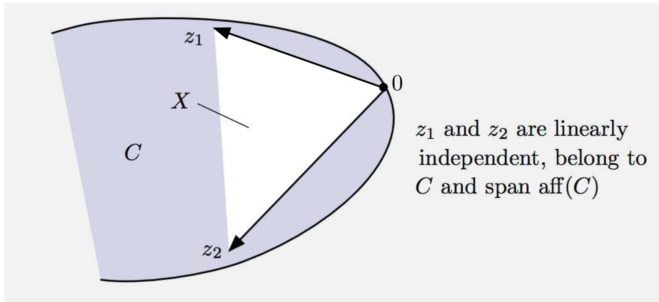

Figure 1.3.2. Construction of the relatively open set X in the proof of nonemptiness of the relative interior of a convex set C that contains the origin, assuming that $m > 0$. We choose m linearly independent vectors $z_1, \ldots, z_m \in C$, where m is the dimension of aff(C), and we let

$$X = \left\{ \sum_{i=1}^{m} \alpha_i z_i \;\middle|\; \sum_{i=1}^{m} \alpha_i < 1,\ \alpha_i > 0,\ i = 1, \ldots, m \right\}.$$

Any point in X is shown to be a relative interior point of C.

$\text{aff}(X) = \text{aff}(C)$ and $X \subset \text{ri}(C)$, we see that ri(C) and C have the same affine hull.

(b) Let x_0 be a relative interior point of C [there exists such a point by part (a)]. Translate C to $C - x_0$ (so that x_0 is translated to the origin), and consider vectors $z_1, \ldots, z_m \in C - x_0$ that span aff$(C - x_0)$, as in the proof of part (a). Let $\alpha \in (0, 1)$. Since $0 \in \text{ri}(C - x_0)$, by the Line Segment Principle (Prop. 1.3.1), we have $\alpha z_i \in \text{ri}(C - x_0)$ for all $i = 1, \ldots, m$. It follows that the vectors

$$x_i = x_0 + \alpha z_i, \qquad i = 1, \ldots, m,$$

are such that $x_1 - x_0, \ldots, x_m - x_0$ belong to ri(C) and span aff(C). **Q.E.D.**

Here is another useful consequence of the Line Segment Principle.

Proposition 1.3.3: (Prolongation Lemma) Let C be a nonempty convex set. A vector x is a relative interior point of C if and only if every line segment in C having x as one endpoint can be prolonged beyond x without leaving C [i.e., for every $\overline{x} \in C$, there exists a $\gamma > 0$ such that $x + \gamma(x - \overline{x}) \in C$].

Proof: If $x \in \text{ri}(C)$, the given condition clearly holds, using the definition of relative interior point. Conversely, let x satisfy the given condition, and let $\overline{x}$ be a point in $\mathrm{ri}(C)$ (by Prop. 1.3.2, there exists such a point). If $x = \overline{x}$, we are done, so assume that $x \neq \overline{x}$. By the given condition, there is a $\gamma > 0$ such that $y = x + \gamma(x - \overline{x}) \in C$, so that x lies strictly within the line segment connecting $\overline{x}$ and y. Since $\overline{x} \in \mathrm{ri}(C)$ and $y \in C$, by the Line Segment Principle (Prop. 1.3.1), it follows that $x \in \mathrm{ri}(C)$. **Q.E.D.**

We will see in the following chapters that the notion of relative interior is pervasive in convex optimization and duality theory. As an example, we provide an important characterization of the set of optimal solutions in the case where the cost function is concave.

Proposition 1.3.4: Let X be a nonempty convex subset of $\Re^n$, let $f : X \mapsto \Re$ be a concave function, and let X^* be the set of vectors where f attains a minimum over X, i.e.,

$$X^* = \left\{ x^* \in X \;\middle|\; f(x^*) = \inf_{x \in X} f(x) \right\}.$$

If X^* contains a relative interior point of X, then f must be constant over X, i.e., $X^* = X$.

Proof: Let x^* belong to $X^* \cap \mathrm{ri}(X)$, and let x be any vector in X. By the Prolongation Lemma (Prop. 1.3.3), there exists a $\gamma > 0$ such that

$$\hat{x} = x^* + \gamma(x^* - x)$$

belongs to X, implying that

$$x^* = \frac{1}{\gamma + 1}\hat{x} + \frac{\gamma}{\gamma + 1}x$$

(see Fig. 1.3.3). By the concavity of f, we have

$$f(x^*) \geq \frac{1}{\gamma + 1} f(\hat{x}) + \frac{\gamma}{\gamma + 1} f(x),$$

and using $f(\hat{x}) \geq f(x^*)$ and $f(x) \geq f(x^*)$, this shows that $f(x) = f(x^*)$. **Q.E.D.**

One consequence of the preceding proposition is that a linear cost function $f(x) = c'x$, with $c \neq 0$, cannot attain a minimum at some interior point of a convex constraint set, since such a function cannot be constant over an open sphere. This will be further discussed in Chapter 2, after we introduce the notion of an extreme point.

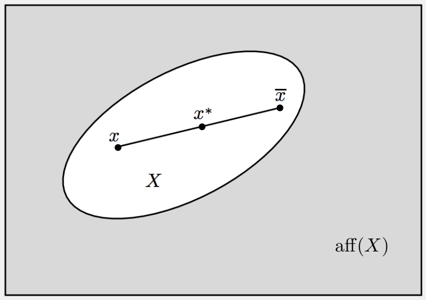

Figure 1.3.3. The idea of the proof of Prop. 1.3.4. If $x^* \in \mathrm{ri}(X)$ minimizes f over X and f is not constant over X, then there exists $x \in X$ such that $f(x) > f(x^*)$. By the Prolongation Lemma (Prop. 1.3.3), there exists $\overline{x} \in X$ such that x^* lies strictly between x and $\overline{x}$. Since f is concave and $f(x) > f(x^*)$, we must have $f(\overline{x}) < f(x^*)$ - a contradiction of the optimality of x^*.

1.3.1 Calculus of Relative Interiors and Closures

To deal with set operations such as intersection, vector sum, and linear transformation in convex analysis, we need tools for calculating the corresponding relative interiors and closures. These tools are provided in the next five propositions. Here is an informal summary of their content:

(a) Two convex sets have the same closure if and only if they have the same relative interior.

(b) Relative interior and closure commute with Cartesian product and inverse image under a linear transformation.

(c) Relative interior commutes with image under a linear transformation and vector sum, but closure does not.

(d) Neither closure nor relative interior commute with set intersection, unless the relative interiors of the sets involved have a point in common.

Proposition 1.3.5: Let C be a nonempty convex set.

(a) $\mathrm{cl}(C) = \mathrm{cl}\big(\mathrm{ri}(C)\big)$.

(b) $\mathrm{ri}(C) = \mathrm{ri}\big(\mathrm{cl}(C)\big)$.

(c) Let $\overline{C}$ be another nonempty convex set. Then the following three conditions are equivalent:

  (i) C and $\overline{C}$ have the same relative interior.

  (ii) C and $\overline{C}$ have the same closure.

  (iii) $\mathrm{ri}(C) \subset \overline{C} \subset \mathrm{cl}(C)$.

Proof: (a) Since $\mathrm{ri}(C) \subset C$, we have $\mathrm{cl}\big(\mathrm{ri}(C)\big) \subset \mathrm{cl}(C)$. Conversely, let

$\overline{x} \in \mathrm{cl}(C)$. We will show that $\overline{x} \in \mathrm{cl}\big(\mathrm{ri}(C)\big)$. Let x be any point in $\mathrm{ri}(C)$ [there exists such a point by Prop. 1.3.2(a)], and assume that $\overline{x} \neq x$ (otherwise we are done). By the Line Segment Principle (Prop. 1.3.1), we have $\alpha x + (1-\alpha)\overline{x} \in \mathrm{ri}(C)$ for all $\alpha \in (0, 1]$. Thus, $\overline{x}$ is the limit of the sequence

$$\big\{(1/k)x + (1 - 1/k)\overline{x} \mid k \geq 1\big\}$$

that lies in $\mathrm{ri}(C)$, so $\overline{x} \in \mathrm{cl}\big(\mathrm{ri}(C)\big)$.

(b) The inclusion $\mathrm{ri}(C) \subset \mathrm{ri}\big(\mathrm{cl}(C)\big)$ follows from the definition of a relative interior point and the fact $\mathrm{aff}(C) = \mathrm{aff}\big(\mathrm{cl}(C)\big)$ (the proof of this is left for the reader). To prove the reverse inclusion, let $z \in \mathrm{ri}\big(\mathrm{cl}(C)\big)$. We will show that $z \in \mathrm{ri}(C)$. By Prop. 1.3.2(a), there exists an $x \in \mathrm{ri}(C)$. We may assume that $x \neq z$ (otherwise we are done). We use the Prolongation Lemma [Prop. 1.3.3, applied within the set $\mathrm{cl}(C)$] to choose $\gamma > 0$, with γ sufficiently close to 0 so that the vector $y = z + \gamma(z - x)$ belongs to $\mathrm{cl}(C)$. Then we have $z = (1-\alpha)x + \alpha y$ where $\alpha = 1/(\gamma + 1) \in (0, 1)$, so by the Line Segment Principle (Prop. 1.3.1, applied within the set C), we obtain $z \in \mathrm{ri}(C)$.

(c) If $\mathrm{ri}(C) = \mathrm{ri}(\overline{C})$, part (a) implies that $\mathrm{cl}(C) = \mathrm{cl}(\overline{C})$. Similarly, if $\mathrm{cl}(C) = \mathrm{cl}(\overline{C})$, part (b) implies that $\mathrm{ri}(C) = \mathrm{ri}(\overline{C})$. Thus, (i) and (ii) are equivalent. Also, (i), (ii), and the relation $\mathrm{ri}(\overline{C}) \subset \overline{C} \subset \mathrm{cl}(\overline{C})$ imply condition (iii). Finally, let condition (iii) hold. Then by taking closures, we have $\mathrm{cl}\big(\mathrm{ri}(C)\big) \subset \mathrm{cl}(\overline{C}) \subset \mathrm{cl}(C)$, and by using part (a), we obtain $\mathrm{cl}(C) \subset \mathrm{cl}(\overline{C}) \subset \mathrm{cl}(C)$. Hence $\mathrm{cl}(\overline{C}) = \mathrm{cl}(C)$, i.e., (ii) holds. **Q.E.D.**

We now consider the image of a convex set C under a linear transformation A. Geometric intuition suggests that $A \cdot \mathrm{ri}(C) = \mathrm{ri}(A \cdot C)$, since spheres within C are mapped onto ellipsoids within the image $A \cdot C$ (relative to the corresponding affine hulls). This is shown in part (a) of the following proposition. However, the image of a closed convex set under a linear transformation is not closed [see part (b) of the following proposition], and this is a major source of analytical difficulty in convex optimization.

Proposition 1.3.6: Let C be a nonempty convex subset of $\Re^n$ and let A be an $m \times n$ matrix.

(a) We have $A \cdot \mathrm{ri}(C) = \mathrm{ri}(A \cdot C)$.

(b) We have $A \cdot \mathrm{cl}(C) \subset \mathrm{cl}(A \cdot C)$. Furthermore, if C is bounded, then $A \cdot \mathrm{cl}(C) = \mathrm{cl}(A \cdot C)$.

Proof: (a) For any set X, we have $A \cdot \mathrm{cl}(X) \subset \mathrm{cl}(A \cdot X)$, since if a sequence $\{x_k\} \subset X$ converges to some $x \in \mathrm{cl}(X)$ then the sequence $\{Ax_k\}$, which belongs to $A \cdot X$, converges to Ax, implying that $Ax \in \mathrm{cl}(A \cdot X)$. We use this fact and Prop. 1.3.5(a) to write

$$A \cdot \mathrm{ri}(C) \subset A \cdot C \subset A \cdot \mathrm{cl}(C) = A \cdot \mathrm{cl}\big(\mathrm{ri}(C)\big) \subset \mathrm{cl}\big(A \cdot \mathrm{ri}(C)\big).$$

Thus the convex set $A \cdot C$ lies between the convex set $A \cdot \mathrm{ri}(C)$ and the closure of that set, implying that the relative interiors of the sets $A \cdot C$ and $A \cdot \mathrm{ri}(C)$ are equal [Prop. 1.3.5(c)]. Hence $\mathrm{ri}(A \cdot C) \subset A \cdot \mathrm{ri}(C)$.

To show the reverse inclusion, we take any $z \in A \cdot \mathrm{ri}(C)$ and we show that $z \in \mathrm{ri}(A \cdot C)$. Let x be any vector in $A \cdot C$, and let $\overline{z} \in \mathrm{ri}(C)$ and $\overline{x} \in C$ be such that $A\overline{z} = z$ and $A\overline{x} = x$. By the Prolongation Lemma (Prop. 1.3.3), there exists a $\gamma > 0$ such that the vector $\overline{y} = \overline{z} + \gamma(\overline{z} - \overline{x})$ belongs to C. Thus we have $A\overline{y} \in A \cdot C$ and $A\overline{y} = z + \gamma(z - x)$, so by the Prolongation Lemma, it follows that $z \in \mathrm{ri}(A \cdot C)$.

(b) By the argument given in part (a), we have $A \cdot \mathrm{cl}(C) \subset \mathrm{cl}(A \cdot C)$. To show the converse, assuming that C is bounded, choose any $z \in \mathrm{cl}(A \cdot C)$. Then, there exists a sequence $\{x_k\} \subset C$ such that $Ax_k \to z$. Since C is bounded, $\{x_k\}$ has a subsequence that converges to some $x \in \mathrm{cl}(C)$, and we must have $Ax = z$. It follows that $z \in A \cdot \mathrm{cl}(C)$. **Q.E.D.**

Note that if C is closed and convex but unbounded, the set $A \cdot C$ need not be closed [cf. part (b) of the preceding proposition]. For example, projection on the horizontal axis of the closed convex set

$$\big\{(x_1, x_2) \mid x_1 > 0,\ x_2 > 0,\ x_1 x_2 \geq 1\big\},$$

shown in Fig. 1.3.4, is the (nonclosed) halfline $\big\{(x_1, x_2) \mid x_1 > 0,\ x_2 = 0\big\}$.

Generally, the vector sum of sets $C_1, \ldots, C_m$ can be viewed as the result of the linear transformation $(x_1, \ldots, x_m) \mapsto x_1 + \cdots + x_m$ on the Cartesian product $C_1 \times \cdots \times C_m$. Thus, results involving linear transformations, such as the one of the preceding proposition, yield corresponding results for vector sums, such as the one of the following proposition.

Proposition 1.3.7: Let C_1 and C_2 be nonempty convex sets. We have

$$\mathrm{ri}(C_1 + C_2) = \mathrm{ri}(C_1) + \mathrm{ri}(C_2), \qquad \mathrm{cl}(C_1) + \mathrm{cl}(C_2) \subset \mathrm{cl}(C_1 + C_2).$$

Furthermore, if at least one of the sets C_1 and C_2 is bounded, then

$$\mathrm{cl}(C_1) + \mathrm{cl}(C_2) = \mathrm{cl}(C_1 + C_2).$$

Proof: Consider the linear transformation $A : \Re^{2n} \mapsto \Re^n$ given by

$$A(x_1, x_2) = x_1 + x_2, \qquad x_1, x_2 \in \Re^n.$$

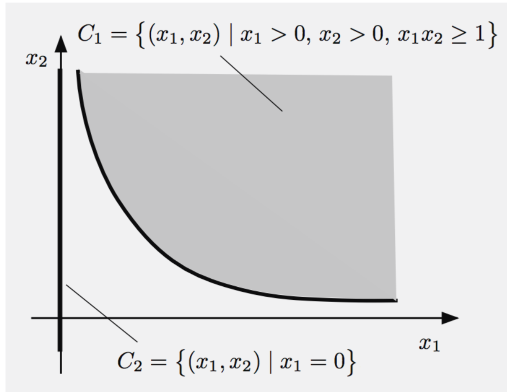

Figure 1.3.4. An example where the sum of two closed convex sets C_1 and C_2 is not closed. Here

$$C_1 = \big\{(x_1, x_2) \mid x_1 > 0,\, x_2 > 0,\, x_1x_2 \geq 1\big\}, \qquad C_2 = \big\{(x_1, x_2) \mid x_1 = 0\big\},$$

and $C_1 + C_2$ is the open halfspace $\big\{(x_1, x_2) \mid x_1 > 0\big\}$. Also the projection of the set C_1 on the horizontal axis is not closed.

The relative interior of the Cartesian product $C_1 \times C_2$ (viewed as a subset of $\Re^{2n}$) is $\mathrm{ri}(C_1) \times \mathrm{ri}(C_2)$ (the easy proof of this is left for the reader). Since

$$A(C_1 \times C_2) = C_1 + C_2,$$

from Prop. 1.3.6(a), we obtain $\mathrm{ri}(C_1 + C_2) = \mathrm{ri}(C_1) + \mathrm{ri}(C_2)$.

Similarly, the closure of $C_1 \times C_2$ is $\mathrm{cl}(C_1) \times \mathrm{cl}(C_2)$. From Prop. 1.3.6(b), we have

$$A \cdot \mathrm{cl}(C_1 \times C_2) \subset \mathrm{cl}\big(A \cdot (C_1 \times C_2)\big),$$

or equivalently, $\mathrm{cl}(C_1) + \mathrm{cl}(C_2) \subset \mathrm{cl}(C_1 + C_2)$.

To show the reverse inclusion, assuming that C_1 is bounded, let $x \in \mathrm{cl}(C_1 + C_2)$. Then there exist sequences $\{x_k^1\} \subset C_1$ and $\{x_k^2\} \subset C_2$ such that $x_k^1 + x_k^2 \to x$. Since $\{x_k^1\}$ is bounded, it follows that $\{x_k^2\}$ is also bounded. Thus, $\big\{(x_k^1, x_k^2)\big\}$ has a subsequence that converges to a vector (x^1, x^2), and we have $x^1 + x^2 = x$. Since $x^1 \in \mathrm{cl}(C_1)$ and $x^2 \in \mathrm{cl}(C_2)$, it follows that $x \in \mathrm{cl}(C_1) + \mathrm{cl}(C_2)$. Hence $\mathrm{cl}(C_1 + C_2) \subset \mathrm{cl}(C_1) + \mathrm{cl}(C_2)$. **Q.E.D.**

The requirement that at least one of the sets C_1 and C_2 be bounded is essential in the preceding proposition. This is illustrated by the example of Fig. 1.3.4.

Proposition 1.3.8: Let C_1 and C_2 be nonempty convex sets. We have

$$\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2) \subset \mathrm{ri}(C_1 \cap C_2), \qquad \mathrm{cl}(C_1 \cap C_2) \subset \mathrm{cl}(C_1) \cap \mathrm{cl}(C_2).$$

Furthermore, if the sets $\mathrm{ri}(C_1)$ and $\mathrm{ri}(C_2)$ have a nonempty intersection, then

$$\mathrm{ri}(C_1 \cap C_2) = \mathrm{ri}(C_1) \cap \mathrm{ri}(C_2), \qquad \mathrm{cl}(C_1 \cap C_2) = \mathrm{cl}(C_1) \cap \mathrm{cl}(C_2).$$

Proof: Take any $x \in \mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$ and any $y \in C_1 \cap C_2$. By the Prolongation Lemma (Prop. 1.3.3), it can be seen that the line segment connecting x and y can be prolonged beyond x by a small amount without leaving C_1 and also by another small amount without leaving C_2. Thus, by using the lemma again, it follows that $x \in \mathrm{ri}(C_1 \cap C_2)$, so that

$$\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2) \subset \mathrm{ri}(C_1 \cap C_2).$$

Also, since the set $C_1 \cap C_2$ is contained in the closed set $\mathrm{cl}(C_1) \cap \mathrm{cl}(C_2)$, we have

$$\mathrm{cl}(C_1 \cap C_2) \subset \mathrm{cl}(C_1) \cap \mathrm{cl}(C_2).$$

To show the reverse inclusions assuming that $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2) \neq \emptyset$, let $y \in \mathrm{cl}(C_1) \cap \mathrm{cl}(C_2)$, and let $x \in \mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$. By the Line Segment Principle (Prop. 1.3.1), $\alpha x + (1 - \alpha)y \in \mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$ for all $\alpha \in (0, 1]$ (see Fig. 1.3.5). Hence, y is the limit of a sequence $\alpha_k x + (1 - \alpha_k)y \subset \mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$ with $\alpha_k \to 0$, implying that $y \in \mathrm{cl}\big(\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)\big)$. Thus,

$$\mathrm{cl}(C_1) \cap \mathrm{cl}(C_2) \subset \mathrm{cl}\big(\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)\big) \subset \mathrm{cl}(C_1 \cap C_2).$$

We showed earlier that $\mathrm{cl}(C_1 \cap C_2) \subset \mathrm{cl}(C_1) \cap \mathrm{cl}(C_2)$, so equality holds throughout in the preceding relation, and therefore $\mathrm{cl}(C_1 \cap C_2) = \mathrm{cl}(C_1) \cap \mathrm{cl}(C_2)$. Furthermore, the sets $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$ and $C_1 \cap C_2$ have the same closure. Therefore, by Prop. 1.3.5(c), they have the same relative interior, so that

$$\mathrm{ri}(C_1 \cap C_2) = \mathrm{ri}\big(\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)\big) \subset \mathrm{ri}(C_1) \cap \mathrm{ri}(C_2).$$

We showed earlier the reverse inclusion, so $\mathrm{ri}(C_1 \cap C_2) = \mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$. **Q.E.D.**

The requirement that $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2) \neq \emptyset$ is essential in part (a) of the preceding proposition. As an example, consider the following subsets of the real line:

$$C_1 = \{x \mid x \geq 0\}, \qquad C_2 = \{x \mid x \leq 0\}.$$

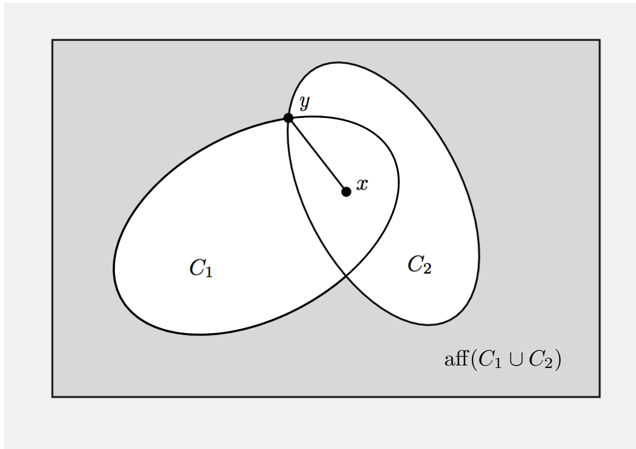

Figure 1.3.5. Construction used to show that

$$\mathrm{cl}(C_1) \cap \mathrm{cl}(C_2) \subset \mathrm{cl}(C_1 \cap C_2),$$

assuming that there exists $x \in \mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$ (cf. Prop. 1.3.8). Any $y \in \mathrm{cl}(C_1) \cap \mathrm{cl}(C_2)$ can be approached along the line segment of $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$ connecting it with x, so it belongs to the closure of $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$ and hence also to $\mathrm{cl}(C_1 \cap C_2)$.

Then we have $\mathrm{ri}(C_1 \cap C_2) = \{0\} \neq \emptyset = \mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$. Also, consider the following subsets of the real line:

$$C_1 = \{x \mid x > 0\}, \qquad C_2 = \{x \mid x < 0\}.$$

Then we have $\mathrm{cl}(C_1 \cap C_2) = \emptyset \neq \{0\} = \mathrm{cl}(C_1) \cap \mathrm{cl}(C_2)$.

Proposition 1.3.9: Let C be a nonempty convex subset of $\Re^m$, and let A be an $m \times n$ matrix. If $A^{-1} \cdot \mathrm{ri}(C)$ is nonempty, then

$$\mathrm{ri}(A^{-1} \cdot C) = A^{-1} \cdot \mathrm{ri}(C), \qquad \mathrm{cl}(A^{-1} \cdot C) = A^{-1} \cdot \mathrm{cl}(C),$$

where A^{-1} denotes inverse image of the corresponding set under A.

Proof: Define the sets

$$D = \Re^n \times C, \qquad S = \big\{(x, Ax) \mid x \in \Re^n\big\},$$

and let T be the linear transformation that maps $(x, y) \in \Re^{n+m}$ into $x \in \Re^n$. We have

$$A^{-1} \cdot C = \{x \mid Ax \in C\} = T \cdot \big\{(x, Ax) \mid Ax \in C\big\} = T \cdot (D \cap S),$$

from which

$$\mathrm{ri}(A^{-1} \cdot C) = \mathrm{ri}\big(T \cdot (D \cap S)\big). \tag{1.9}$$

Similarly, we have

$$A^{-1} \cdot \mathrm{ri}(C) = \big\{x \mid Ax \in \mathrm{ri}(C)\big\} = T \cdot \big\{(x, Ax) \mid Ax \in \mathrm{ri}(C)\big\} = T \cdot \big(\mathrm{ri}(D) \cap S\big), \tag{1.10}$$

where the last equality holds because $\mathrm{ri}(D) = \Re^n \times \mathrm{ri}(C)$ (cf. Prop. 1.3.8). Since by assumption, $A^{-1} \cdot \mathrm{ri}(C)$ is nonempty, we see that $\mathrm{ri}(D) \cap S$ is nonempty. Therefore, using the fact $\mathrm{ri}(S) = S$, and Props. 1.3.6(a) and 1.3.8, it follows that

$$\mathrm{ri}\big(T \cdot (D \cap S)\big) = T \cdot \mathrm{ri}\big(D \cap S\big) = T \cdot \big(\mathrm{ri}(D) \cap S\big). \tag{1.11}$$

Combining Eqs. (1.9)-(1.11), we obtain

$$\mathrm{ri}(A^{-1} \cdot C) = A^{-1} \cdot \mathrm{ri}(C).$$

To show the second relation, note that

$$A^{-1} \cdot \mathrm{cl}(C) = \big\{x \mid Ax \in \mathrm{cl}(C)\big\} = T \cdot \big\{(x, Ax) \mid Ax \in \mathrm{cl}(C)\big\} = T \cdot \big(\mathrm{cl}(D) \cap S\big),$$

where the last equality holds because $\mathrm{cl}(D) = \Re^n \times \mathrm{cl}(C)$. Since $\mathrm{ri}(D) \cap S$ is nonempty and $\mathrm{ri}(S) = S$, it follows from Prop. 1.3.8 that

$$\mathrm{cl}(D) \cap S = \mathrm{cl}(D \cap S).$$

Using the last two relations and the continuity of T, we obtain

$$A^{-1} \cdot \mathrm{cl}(C) = T \cdot \mathrm{cl}(D \cap S) \subset \mathrm{cl}\big(T \cdot (D \cap S)\big),$$

which combined with Eq. (1.9) yields

$$A^{-1} \cdot \mathrm{cl}(C) \subset \mathrm{cl}(A^{-1} \cdot C).$$

To show the reverse inclusion, let $\overline{x}$ be a vector in $\mathrm{cl}(A^{-1} \cdot C)$. Then there exists a sequence $\{x_k\}$ converging to $\overline{x}$ such that $Ax_k \in C$ for all k. Since $\{x_k\}$ converges to $\overline{x}$, we see that $\{Ax_k\}$ converges to $A\overline{x}$, so that $A\overline{x} \in \mathrm{cl}(C)$, or equivalently, $\overline{x} \in A^{-1} \cdot \mathrm{cl}(C)$. **Q.E.D.**

We finally show a useful characterization of the relative interior of sets involving two variables. It generalizes the Cartesian product formula

$$\mathrm{ri}(C_1 \times C_2) = \mathrm{ri}(C_1) \times \mathrm{ri}(C_2)$$

for two convex sets $C_1 \subset \Re^n$ and $C_2 \in \Re^m$.

Proposition 1.3.10: Let C be a convex subset of $\Re^{n+m}$. For $x \in \Re^n$, denote

$$C_x = \{y \mid (x, y) \in C\},$$

and let

$$D = \{x \mid C_x \neq \emptyset\}.$$

Then

$$\text{ri}(C) = \big\{(x, y) \mid x \in \text{ri}(D),\ y \in \text{ri}(C_x)\big\}.$$

Proof: Since D is the projection of C on the x-axis, from Prop. 1.3.6,

$$\text{ri}(D) = \big\{x \mid \text{there exists } y \in \Re^m \text{ with } (x, y) \in \text{ri}(C)\big\},$$

so that

$$\text{ri}(C) = \cup_{x \in \text{ri}(D)} \Big(M_x \cap \text{ri}(C)\Big),$$

where $M_x = \big\{(x, y) \mid y \in \Re^m\big\}$. For every $x \in \text{ri}(D)$, we have

$$M_x \cap \text{ri}(C) = \text{ri}(M_x \cap C) = \big\{(x, y) \mid y \in \text{ri}(C_x)\big\},$$

where the first equality follows from Prop. 1.3.8. By combining the preceding two equations, we obtain the desired result. **Q.E.D.**

1.3.2 Continuity of Convex Functions

We now derive a basic continuity property of convex functions.

Proposition 1.3.11: If $f : \Re^n \mapsto \Re$ is convex, then it is continuous. More generally, if $f : \Re^n \mapsto (-\infty, \infty]$ is a proper convex function, then f, restricted to $\text{dom}(f)$, is continuous over the relative interior of $\text{dom}(f)$.

Proof: Restricting attention to the affine hull of $\text{dom}(f)$ and using a transformation argument if necessary, we assume without loss of generality that the origin is an interior point of $\text{dom}(f)$ and that the unit cube

$$X = \{x \mid \|x\|_\infty \leq 1\}$$

is contained in $\text{dom}(f)$ (we use the norm $\|x\|_\infty = \max_{j \in \{1,\ldots,n\}} |x_j|$). It will suffice to show that f is continuous at 0, i.e., that for any sequence $\{x_k\} \subset \Re^n$ that converges to 0, we have $f(x_k) \to f(0)$.

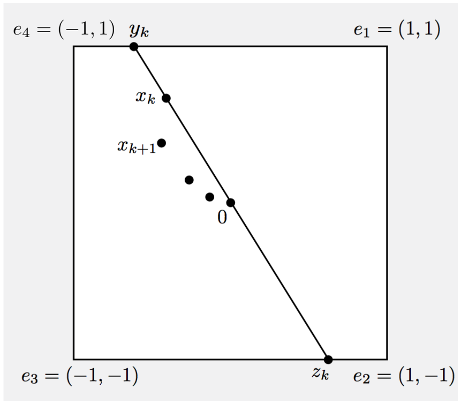

Figure 1.3.6. Construction for the proof of continuity of a real-valued convex function (cf. Prop. 1.3.11).

Let e_i, $i = 1, \ldots, 2^n$, be the corners of X, i.e., each e_i is a vector whose entries are either 1 or -1. It can be seen that any $x \in X$ can be expressed in the form $x = \sum_{i=1}^{2^n} \alpha_i e_i$, where each α_i is a nonnegative scalar and $\sum_{i=1}^{2^n} \alpha_i = 1$. Let $A = \max_i f(e_i)$. From Jensen's inequality [Eq. (1.7)], it follows that $f(x) \leq A$ for every $x \in X$.

For the purpose of proving continuity at 0, we can assume that $x_k \in X$ and $x_k \neq 0$ for all k. Consider the sequences $\{y_k\}$ and $\{z_k\}$ given by

$$y_k = \frac{x_k}{\|x_k\|_\infty}, \qquad z_k = -\frac{x_k}{\|x_k\|_\infty};$$

(cf. Fig. 1.3.6). Using the definition of a convex function for the line segment that connects y_k, x_k, and 0, we have

$$f(x_k) \leq \big(1 - \|x_k\|_\infty\big) f(0) + \|x_k\|_\infty f(y_k).$$

Since $\|x_k\|_\infty \to 0$ and $f(y_k) \leq A$ for all k, by taking the limit as $k \to \infty$, we obtain

$$\limsup_{k \to \infty} f(x_k) \leq f(0).$$

Using the definition of a convex function for the line segment that connects x_k, 0, and z_k, we have

$$f(0) \leq \frac{\|x_k\|_\infty}{\|x_k\|_\infty + 1} f(z_k) + \frac{1}{\|x_k\|_\infty + 1} f(x_k)$$

and letting $k \to \infty$, we obtain

$$f(0) \leq \liminf_{k \to \infty} f(x_k).$$

Thus, $\lim_{k \to \infty} f(x_k) = f(0)$ and f is continuous at zero. **Q.E.D.**

Among other things, the proposition implies that a real-valued convex function is continuous and hence closed (cf. Prop. 1.1.2). We also have the following stronger result for the case of a function of one variable.

Proposition 1.3.12: If C is a closed interval of the real line, and $f : C \mapsto \Re$ is closed and convex, then f is continuous over C.

Proof: By the preceding proposition, f is continuous in the relative interior of C. To show continuity at a boundary point $\overline{x}$, let $\{x_k\} \subset C$ be a sequence that converges to $\overline{x}$, and write

$$x_k = \alpha_k x_0 + (1 - \alpha_k)\overline{x}, \qquad \forall\ k,$$

where $\{\alpha_k\}$ is a nonnegative sequence with $\alpha_k \to 0$. By convexity of f, we have for all k such that $\alpha_k \leq 1$,

$$f(x_k) \leq \alpha_k f(x_0) + (1 - \alpha_k) f(\overline{x}),$$

and by taking the limit as $k \to \infty$, we obtain

$$\limsup_{k\to\infty} f(x_k) \leq f(\overline{x}).$$

Consider the function $\tilde{f} : \Re \mapsto (-\infty, \infty]$, which takes the value $f(x)$ for $x \in C$ and ∞ for $x \notin C$, and note that it is closed (since it has the same epigraph as f), and hence lower semicontinuous (cf. Prop. 1.1.2). It follows that $f(\overline{x}) \leq \liminf_{k\to\infty} f(x_k)$, thus implying that $f(x_k) \to f(\overline{x})$, and that f is continuous at $\overline{x}$. **Q.E.D.**

1.3.3 Closures of Functions

In this section, we study operations that can transform a given function to a closed and/or convex function, while preserving much of its essential character. These operations play an important role in optimization and other contexts.

A nonempty subset E of $\Re^{n+1}$ is the epigraph of some function if for every $(\overline{x}, \overline{w}) \in E$, the set $\{w \mid (\overline{x}, w) \in E\}$ is either the real line or else it is a halfline that is bounded below and contains its (lower) endpoint. Then E is the epigraph of the function $f : D \mapsto [-\infty, \infty]$, where

$$D = \{x \mid \text{there exists } w \in \Re \text{ with } (x, w) \in E\},$$

and

$$f(x) = \inf\{w \mid (x, w) \in E\}, \qquad \forall\ x \in D$$

[the infimum is actually attained if $f(x)$ is finite]. Note that E is also the epigraph of other functions with different domain than f (but the same effective domain); for example, $\tilde{f} : \Re^n \mapsto [-\infty, \infty]$, where $\tilde{f}(x) = f(x)$ for $x \in D$ and $\tilde{f}(x) = \infty$ for $x \notin D$. If E is the empty set, it is the epigraph of the function that is identically equal to ∞.

The closure of the epigraph of a function $f : X \mapsto [-\infty, \infty]$ can be seen to be a legitimate epigraph of another function. This function, called the *closure of* f and denoted $\mathrm{cl}\, f : \Re^n \mapsto [-\infty, \infty]$, is given by†

$$(\mathrm{cl}\, f)(x) = \inf\big\{w \mid (x, w) \in \mathrm{cl}\big(\mathrm{epi}(f)\big)\big\}, \qquad x \in \Re^n.$$

When f is convex, the set $\mathrm{cl}\big(\mathrm{epi}(f)\big)$ is closed and convex [since the closure of a convex set is convex by Prop. 1.1.1(d)], implying that $\mathrm{cl}\, f$ is closed and convex since $\mathrm{epi}(\mathrm{cl}\, f) = \mathrm{cl}\big(\mathrm{epi}(f)\big)$ by definition.

The closure of the convex hull of the epigraph of f is the epigraph of some function, denoted $\check{\mathrm{cl}}\, f$ called the *convex closure of* f. It can be seen that $\check{\mathrm{cl}}\, f$ is the closure of the function $F : \Re^n \mapsto [-\infty, \infty]$ given by

$$F(x) = \inf\big\{w \mid (x, w) \in \mathrm{conv}\big(\mathrm{epi}(f)\big)\big\}, \qquad x \in \Re^n. \tag{1.12}$$

It is easily shown that F is convex, but it need not be closed and its domain may be strictly contained in $\mathrm{dom}(\check{\mathrm{cl}}\, f)$ (it can be seen though that the closures of the domains of F and $\check{\mathrm{cl}}\, f$ coincide).

From the point of view of optimization, an important property is that the minimal values of f, $\mathrm{cl}\, f$, F, and $\check{\mathrm{cl}}\, f$ coincide, as stated in the following proposition:

Proposition 1.3.13: Let $f : X \mapsto [-\infty, \infty]$ be a function. Then

$$\inf_{x \in X} f(x) = \inf_{x \in X} (\mathrm{cl}\, f)(x) = \inf_{x \in \Re^n} (\mathrm{cl}\, f)(x) = \inf_{x \in \Re^n} F(x) = \inf_{x \in \Re^n} (\check{\mathrm{cl}}\, f)(x),$$

where F is given by Eq. (1.12). Furthermore, any vector that attains the infimum of f over X also attains the infimum of $\mathrm{cl}\, f$, F, and $\check{\mathrm{cl}}\, f$.

† A note regarding the definition of closure: in Rockafellar [Roc70], p. 52, what we call "closure" of f is called the "lower semi-continuous hull" of f, and "closure" of f is defined somewhat differently (but denoted $\mathrm{cl}\, f$). Our definition of "closure" of f works better for our purposes, and results in a more streamlined analysis. It coincides with the one of [Roc70] when f is proper convex. For this reason the results of this section correspond to results in [Roc70] only in the case where the functions involved are proper convex. In Rockafellar and Wets [RoW98], p. 14, our "closure" of f is called the "lsc regularization" or "lower closure" of f, and is denoted by $\mathrm{cl}\, f$. Thus our notation is consistent with the one of [RoW98].

Proof: If epi(f) is empty, i.e., $f(x) = \infty$ for all x, the results trivially hold. Assume that epi(f) is nonempty, and let $f^* = \inf_{x\in\Re^n} (\mathrm{cl}\, f)(x)$. For any sequence $\big\{(\overline{x}_k, \overline{w}_k)\big\} \subset \mathrm{cl}\big(\mathrm{epi}(f)\big)$ with $\overline{w}_k \to f^*$, we can construct a sequence $\big\{(x_k, w_k)\big\} \subset \mathrm{epi}(f)$ such that $|w_k - \overline{w}_k| \to 0$, so that $w_k \to f^*$. Since $x_k \in X$, $f(x_k) \le w_k$, we have

$$\limsup_{k\to\infty} f(x_k) \le f^* \le (\mathrm{cl}\, f)(x) \le f(x), \qquad \forall\ x \in X,$$

so that

$$\inf_{x\in X} f(x) = \inf_{x\in X} (\mathrm{cl}\, f)(x) = \inf_{x\in\Re^n} (\mathrm{cl}\, f)(x).$$

Choose $\big\{(x_k, w_k)\big\} \subset \mathrm{conv}\big(\mathrm{epi}(f)\big)$ with $w_k \to \inf_{x\in\Re^n} F(x)$. Each (x_k, w_k) is a convex combination of vectors from epi(f), so that $w_k \ge \inf_{x\in X} f(x)$. Hence $\inf_{x\in\Re^n} F(x) \ge \inf_{x\in X} f(x)$. On the other hand, we have $F(x) \le f(x)$ for all $x \in X$, so it follows that $\inf_{x\in\Re^n} F(x) = \inf_{x\in X} f(x)$. Since $\check{\mathrm{cl}}\, f$ is the closure of F, it also follows (based on what was shown in the preceding paragraph) that $\inf_{x\in\Re^n} (\check{\mathrm{cl}}\, f)(x) = \inf_{x\in\Re^n} F(x)$.

We have $f(x) \ge (\mathrm{cl}\, f)(x)$ for all x, so if x^* attains the infimum of f,

$$\inf_{x\in\Re^n} (\mathrm{cl}\, f)(x) = \inf_{x\in X} f(x) = f(x^*) \ge (\mathrm{cl}\, f)(x^*),$$

showing that x^* attains the infimum of $\mathrm{cl}\, f$. Similarly, x^* attains the infimum of F and $\check{\mathrm{cl}}\, f$. **Q.E.D.**

The following is a characterization of closures and convex closures.

Proposition 1.3.14: Let $f : \Re^n \mapsto [-\infty, \infty]$ be a function.

(a) $\mathrm{cl}\, f$ is the greatest closed function majorized by f, i.e., if $g : \Re^n \mapsto [-\infty, \infty]$ is closed and satisfies $g(x) \le f(x)$ for all $x \in \Re^n$, then $g(x) \le (\mathrm{cl}\, f)(x)$ for all $x \in \Re^n$.

(b) $\check{\mathrm{cl}}\, f$ is the greatest closed and convex function majorized by f, i.e., if $g : \Re^n \mapsto [-\infty, \infty]$ is closed and convex, and satisfies $g(x) \le f(x)$ for all $x \in \Re^n$, then $g(x) \le (\check{\mathrm{cl}}\, f)(x)$ for all $x \in \Re^n$.

Proof: (a) Let $g : \Re^n \mapsto [-\infty, \infty]$ be closed and such that $g(x) \le f(x)$ for all x. Then epi$(f) \subset$ epi(g). Since $\mathrm{epi}(\mathrm{cl}\, f) = \mathrm{cl}\big(\mathrm{epi}(f)\big)$, we have that $\mathrm{epi}(\mathrm{cl}\, f)$ is the intersection of all closed sets $E \subset \Re^{n+1}$ with $\mathrm{epi}(f) \subset E$, so that $\mathrm{epi}(\mathrm{cl}\, f) \subset \mathrm{epi}(g)$. It follows that $g(x) \le (\mathrm{cl}\, f)(x)$ for all $x \in \Re^n$.

(b) Similar to the proof of part (a). **Q.E.D.**

Working with the closure of a convex function is often useful because in some sense the closure "differs minimally" from the original. In particular, we can show that a convex function coincides with its closure on the relative interior of its domain. This and other properties of closures are derived in the following proposition.

Proposition 1.3.15: Let $f : \Re^n \mapsto [-\infty, \infty]$ be a convex function. Then:

(a) We have

$$\mathrm{cl}\big(\mathrm{dom}(f)\big) = \mathrm{cl}\big(\mathrm{dom}(\mathrm{cl}\, f)\big), \qquad \mathrm{ri}\big(\mathrm{dom}(f)\big) = \mathrm{ri}\big(\mathrm{dom}(\mathrm{cl}\, f)\big),$$

$$(\mathrm{cl}\, f)(x) = f(x), \qquad \forall\ x \in \mathrm{ri}\big(\mathrm{dom}(f)\big).$$

Furthermore, $\mathrm{cl}\, f$ is proper if and only if f is proper.

(b) If $x \in \mathrm{ri}\big(\mathrm{dom}(f)\big)$, we have

$$(\mathrm{cl}\, f)(y) = \lim_{\alpha \downarrow 0} f\big(y + \alpha(x - y)\big), \qquad \forall\ y \in \Re^n.$$

Proof: (a) From Prop. 1.3.10, we have

$$\mathrm{ri}\big(\mathrm{epi}(f)\big) = \big\{(x, w) \mid x \in \mathrm{ri}\big(\mathrm{dom}(f)\big),\ f(x) < w\big\}, \tag{1.13}$$

$$\mathrm{ri}\big(\mathrm{epi}(\mathrm{cl}\, f)\big) = \big\{(x, w) \mid x \in \mathrm{ri}\big(\mathrm{dom}(\mathrm{cl}\, f)\big),\ (\mathrm{cl}\, f)(x) < w\big\}. \tag{1.14}$$

Since $\mathrm{epi}(f)$ and $\mathrm{epi}(\mathrm{cl}\, f)$ have the same closure, they have the same relative interior [Prop. 1.3.5(c)], i.e., the sets of Eqs. (1.13) and (1.14) are equal. Hence $\mathrm{dom}(f)$ and $\mathrm{dom}(\mathrm{cl}\, f)$ have the same relative interior and therefore also the same closure. Thus, the equality of the sets (1.13) and (1.14) yields

$$\big\{(x, w) \mid x \in \mathrm{ri}\big(\mathrm{dom}(f)\big), f(x) < w\big\}$$
$$= \big\{(x, w) \mid x \in \mathrm{ri}\big(\mathrm{dom}(f)\big),\ (\mathrm{cl}\, f)(x) < w\big\},$$

from which it follows that $f(x) = (\mathrm{cl}\, f)(x)$ for all $x \in \mathrm{ri}\big(\mathrm{dom}(f)\big)$.

If $\mathrm{cl}\, f$ is proper, clearly f is proper. Conversely, if $\mathrm{cl}\, f$ is improper, then $(\mathrm{cl}\, f)(x) = -\infty$ for all $x \in \mathrm{dom}(\mathrm{cl}\, f)$ (cf. the discussion at the end of Section 1.1.2). Hence $(\mathrm{cl}\, f)(x) = -\infty$ for all $x \in \mathrm{ri}\big(\mathrm{dom}(\mathrm{cl}\, f)\big) = \mathrm{ri}\big(\mathrm{dom}(f)\big)$. Using what was just proved, it follows that $f(x) = (\mathrm{cl}\, f)(x) = -\infty$ for all $x \in \mathrm{ri}\big(\mathrm{dom}(f)\big)$, implying that f is improper.

(b) Assume first $y \notin \mathrm{dom}(\mathrm{cl}\, f)$, i.e., $(\mathrm{cl}\, f)(y) = \infty$. Then, by the lower semicontinuity of $\mathrm{cl}\, f$, we have $(\mathrm{cl}\, f)(y_k) \to \infty$ for all sequences $\{y_k\}$ with $y_k \to y$, from which $f(y_k) \to \infty$, since $(\mathrm{cl}\, f)(y_k) \le f(y_k)$. Hence $(\mathrm{cl}\, f)(y) = \lim_{\alpha \downarrow 0} f\big(y + \alpha(x - y)\big) = \infty$.

Assume next that $y \in \text{dom}(\text{cl}\, f)$, and consider the function $g : [0, 1] \mapsto \Re$ given by

$$g(\alpha) = (\text{cl}\, f)\big(y + \alpha(x - y)\big).$$

For $\alpha \in (0, 1]$, by the Line Segment Principle (Prop. 1.3.1), we have

$$y + \alpha(x - y) \in \text{ri}\big(\text{dom}(\text{cl}\, f)\big),$$

so by part (a), $y + \alpha(x - y) \in \text{ri}\big(\text{dom}(f)\big)$, and

$$g(\alpha) = (\text{cl}\, f)\big(y + \alpha(x - y)\big) = f\big(y + \alpha(x - y)\big). \tag{1.15}$$

If $(\text{cl}\, f)(y) = -\infty$, then $\text{cl}\, f$ is improper and $(\text{cl}\, f)(z) = -\infty$ for all $z \in \text{dom}(\text{cl}\, f)$, since an improper closed convex function cannot take a finite value at any point (cf. the discussion at the end of Section 1.1.2). Hence

$$f\big(y + \alpha(x - y)\big) = -\infty, \qquad \forall\ \alpha \in (0, 1],$$

and the desired equation follows. If $(\text{cl}\, f)(y) > -\infty$, then $(\text{cl}\, f)(y)$ is finite, so $\text{cl}\, f$ is proper and by part (a), f is also proper. It follows that the function g is real-valued, convex, and closed, and hence also continuous over $[0, 1]$ (Prop. 1.3.12). By taking the limit in Eq. (1.15),

$$(\text{cl}\, f)(y) = g(0) = \lim_{\alpha \downarrow 0} g(\alpha) = \lim_{\alpha \downarrow 0} f\big(y + \alpha(x - y)\big).$$

Q.E.D.

Note a corollary of part (a) of the preceding proposition: an improper convex function f takes the value $-\infty$ at all $x \in \text{ri}\big(\text{dom}(f)\big)$, since its closure does (cf. the discussion at the end of Section 1.1.2).

Calculus of Closure Operations

We now characterize the closure of functions obtained by linear composition and summation of convex functions.

Proposition 1.3.16: Let $f : \Re^m \mapsto [-\infty, \infty]$ be a convex function and A be an $m \times n$ matrix such that the range of A contains a point in $\text{ri}\big(\text{dom}(f)\big)$. The function F defined by

$$F(x) = f(Ax),$$

is convex and

$$(\text{cl}\, F)(x) = (\text{cl}\, f)(Ax), \qquad \forall\ x \in \Re^n.$$

Proof: Let z be a point in the range of A that belongs to $\mathrm{ri}\big(\mathrm{dom}(f)\big)$, and let y be such that $Ay = z$. Then, since $\mathrm{dom}(F) = A^{-1}\mathrm{dom}(f)$ and by Prop. 1.3.9, $\mathrm{ri}\big(\mathrm{dom}(F)\big) = A^{-1}\mathrm{ri}\big(\mathrm{dom}(f)\big)$, we see that $y \in \mathrm{ri}\big(\mathrm{dom}(F)\big)$. By using Prop. 1.3.15(b), we have for every $x \in \Re^n$,

$$(\mathrm{cl}\, F)(x) = \lim_{\alpha\downarrow 0} F\big(x + \alpha(y-x)\big) = \lim_{\alpha\downarrow 0} f\big(Ax + \alpha(Ay - Ax)\big) = (\mathrm{cl}\, f)(Ax).$$

Q.E.D.

The following proposition is essentially a special case of the preceding one (cf. the discussion in Section 1.1.3).

Proposition 1.3.17: Let $f_i : \Re^n \mapsto [-\infty, \infty]$, $i = 1, \ldots, m$, be convex functions such that

$$\cap_{i=1}^m \mathrm{ri}\big(\mathrm{dom}(f_i)\big) \neq \emptyset. \tag{1.16}$$

The function F defined by

$$F(x) = f_1(x) + \cdots + f_m(x),$$

is convex and

$$(\mathrm{cl}\, F)(x) = (\mathrm{cl}\, f_1)(x) + \cdots + (\mathrm{cl}\, f_m)(x), \qquad \forall\, x \in \Re^n.$$

Proof: We write F in the form $F(x) = f(Ax)$, where A is the matrix defined by $Ax = (x, \ldots, x)$, and $f : \Re^{mn} \mapsto (-\infty, \infty]$ is the function

$$f(x_1, \ldots, x_m) = f_1(x_1) + \cdots + f_m(x_m).$$

Since $\mathrm{dom}(F) = \cap_{i=1}^n \mathrm{dom}(f_i)$, Eq. (1.16) implies that

$$\cap_{i=1}^n \mathrm{ri}\big(\mathrm{dom}(f_i)\big) = \mathrm{ri}\big(\mathrm{dom}(F)\big) = \mathrm{ri}\big(A^{-1} \cdot \mathrm{dom}(f)\big) = A^{-1} \cdot \mathrm{ri}\big(\mathrm{dom}(f)\big)$$

(cf. Props. 1.3.8 and 1.3.9). Thus Eq. (1.16) is equivalent to the range of A containing a point in $\mathrm{ri}\big(\mathrm{dom}(f)\big)$, so that $(\mathrm{cl}\, F)(x) = (\mathrm{cl}\, f)(x, \ldots, x)$ (cf. Prop. 1.3.16). Let $y \in \cap_{i=1}^n \mathrm{ri}(\mathrm{dom}(f_i))$, so that $(y, \ldots, y) \in \mathrm{ri}(\mathrm{dom}(f))$. Then, from Prop. 1.3.15(b), $(\mathrm{cl}\, F)(x) = \lim_{\alpha\downarrow 0} f_1\big(x + \alpha(y-x)\big) + \cdots + \lim_{\alpha\downarrow 0} f_m\big(x + \alpha(y-x)\big) = (\mathrm{cl}\, f_1)(x) + \cdots + (\mathrm{cl}\, f_m)(x)$. **Q.E.D.**

Note that the relative interior assumption (1.16) is essential. To see this, let f_1 and f_2 be the indicator functions of two convex sets C_1 and C_2 such that $\mathrm{cl}(C_1 \cap C_2) \neq \mathrm{cl}(C_1) \cap \mathrm{cl}(C_2)$ (cf. the example following Prop. 1.3.8).

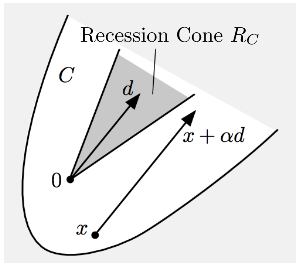

Figure 1.4.1. Illustration of the recession cone R_C of a convex set C. A direction of recession d has the property that $x + \alpha d \in C$ for all $x \in C$ and $\alpha \geq 0$.

1.4 RECESSION CONES

We will now develop some methodology to characterize the asymptotic behavior of convex sets and functions. This methodology is fundamental in several convex optimization contexts, including the issue of existence of optimal solutions, which will be discussed in Chapter 3.

Given a nonempty convex set C, we say that a vector d is a *direction of recession* of C if $x + \alpha d \in C$ for all $x \in C$ and $\alpha \geq 0$. Thus, d is a direction of recession of C if starting at any x in C and going indefinitely along d, we never cross the relative boundary of C to points outside C.

The set of all directions of recession is a cone containing the origin. It is called the *recession cone* of C and it is denoted by R_C (see Fig. 1.4.1). Thus $d \in R_C$ if $x+\alpha d \in C$ for all $x \in C$ and $\alpha \geq 0$. An important property of a *closed* convex set is that to test whether $d \in R_C$ it is enough to verify the property $x+\alpha d \in C$ for a *single* $x \in C$. This is part (b) of the following proposition.

Proposition 1.4.1: (Recession Cone Theorem) Let C be a nonempty closed convex set.

(a) The recession cone R_C is closed and convex.

(b) A vector d belongs to R_C if and only if there exists a vector $x \in C$ such that $x + \alpha d \in C$ for all $\alpha \geq 0$.

Proof: (a) If d_1, d_2 belong to R_C and γ_1, γ_2 are positive scalars such that $\gamma_1 + \gamma_2 = 1$, we have for any $x \in C$ and $\alpha \geq 0$

$$x + \alpha(\gamma_1 d_1 + \gamma_2 d_2) = \gamma_1(x + \alpha d_1) + \gamma_2(x + \alpha d_2) \in C,$$

where the last inclusion holds because C is convex, and $x+\alpha d_1$ and $x+\alpha d_2$ belong to C by the definition of R_C. Hence $\gamma_1 d_1 + \gamma_2 d_2 \in R_C$, implying that R_C is convex.

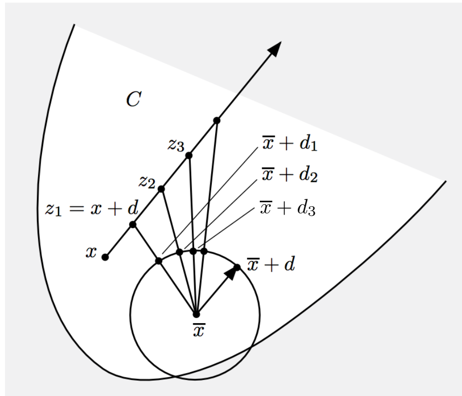

Figure 1.4.2. Construction for the proof of Prop. 1.4.1(b).

Let d be in the closure of R_C, and let $\{d_k\} \subset R_C$ be a sequence converging to d. For any $x \in C$ and $\alpha \geq 0$ we have $x + \alpha d_k \in C$ for all k, and because C is closed, $x + \alpha d \in C$. Hence $d \in R_C$, so R_C is closed.

(b) If $d \in R_C$, every vector $x \in C$ has the required property by the definition of R_C. Conversely, let d be such that there exists a vector $x \in C$ with $x + \alpha d \in C$ for all $\alpha \geq 0$. With no loss of generality, we assume that $d \neq 0$. We choose arbitrary $\overline{x} \in C$ and $\alpha > 0$, and we will show that $\overline{x} + \alpha d \in C$. In fact, it is sufficient to show that $\overline{x} + d \in C$, i.e., to assume that $\alpha = 1$, since the general case where $\alpha > 0$ can be reduced to the case where $\alpha = 1$ by replacing d with αd.

Let

$$z_k = x + kd, \qquad k = 1, 2, \ldots$$

and note that $z_k \in C$ for all k, by our choice of x and d. If $\overline{x} = z_k$ for some k, then $\overline{x} + d = x + (k+1)d$, which belongs to C and we are done. We thus assume that $\overline{x} \neq z_k$ for all k, and we define

$$d_k = \frac{z_k - \overline{x}}{\|z_k - \overline{x}\|}\|d\|, \qquad k = 1, 2, \ldots \tag{1.17}$$

so that $\overline{x} + d_k$ is the intersection of the surface of the sphere centered at $\overline{x}$ of radius $\|d\|$, and the halfline that starts at $\overline{x}$ and passes through z_k (see the construction of Fig. 1.4.2). We will now argue that $d_k \to d$, and that for large enough k, $\overline{x} + d_k \in C$, so using the closure of C, it follows that $\overline{x} + d \in C$.

Indeed, using the definition (1.17) of d_k, we have

$$\frac{d_k}{\|d\|} = \frac{\|z_k - x\|}{\|z_k - \overline{x}\|} \cdot \frac{z_k - x}{\|z_k - x\|} + \frac{x - \overline{x}}{\|z_k - \overline{x}\|} = \frac{\|z_k - x\|}{\|z_k - \overline{x}\|} \cdot \frac{d}{\|d\|} + \frac{x - \overline{x}}{\|z_k - \overline{x}\|}.$$

Because $\{z_k\}$ is an unbounded sequence,

$$\frac{\|z_k - x\|}{\|z_k - \overline{x}\|} \to 1, \qquad \frac{x - \overline{x}}{\|z_k - \overline{x}\|} \to 0,$$

so by combining the preceding relations, we have $d_k \to d$. The vector $\overline{x}+d_k$ lies between $\overline{x}$ and z_k in the line segment connecting $\overline{x}$ and z_k for all k such that $\|z_k - \overline{x}\| \geq \|d\|$, so by the convexity of C, we have $\overline{x} + d_k \in C$ for all sufficiently large k. Since $\overline{x} + d_k \to \overline{x} + d$ and C is closed, it follows that $\overline{x} + d \in C$. **Q.E.D.**

It is essential to assume that the set C is closed in the preceding proposition. For an example where part (a) fails without this assumption, consider the set

$$C = \big\{(x_1, x_2) \mid 0 < x_1,\, 0 < x_2\big\} \cup \big\{(0, 0)\big\}.$$

Its recession cone is equal to C, which is not closed. Part (b) also fails in this example, since for the direction $d = (1, 0)$ we have $x + \alpha d \in C$ for all $\alpha \geq 0$ and all $x \in C$, except for $x = (0, 0)$.

The following proposition gives some additional properties of recession cones.

Proposition 1.4.2: (Properties of Recession Cones) Let C be a nonempty closed convex set.

(a) R_C contains a nonzero direction if and only if C is unbounded.

(b) $R_C = R_{\mathrm{ri}(C)}$.

(c) For any collection of closed convex sets C_i, $i \in I$, where I is an arbitrary index set and $\cap_{i\in I} C_i \neq \emptyset$, we have

$$R_{\cap_{i\in I} C_i} = \cap_{i\in I} R_{C_i}.$$

(d) Let W be a compact and convex subset of $\Re^m$, and let A be an $m \times n$ matrix. The recession cone of the set

$$V = \{x \in C \mid Ax \in W\}$$

(assuming this set is nonempty) is $R_C \cap N(A)$, where $N(A)$ is the nullspace of A.

Proof: (a) Assuming that C is unbounded, we will show that R_C contains a nonzero direction (the reverse implication is clear). Choose any $x \in C$

and any unbounded sequence $\{z_k\} \subset C$. Consider the sequence $\{d_k\}$, where

$$d_k = \frac{z_k - x}{\|z_k - x\|},$$

and let d be a limit point of $\{d_k\}$ (compare with the construction of Fig. 1.4.2). Without loss of generality, assume that $\|z_k - x\|$ is monotonically increasing with k. For any fixed $\alpha \geq 0$, the vector $x + \alpha d_k$ lies between x and z_k in the line segment connecting x and z_k for all k such that $\|z_k - x\| \geq \alpha$. Hence by the convexity of C, we have $x + \alpha d_k \in C$ for all sufficiently large k. Since $x + \alpha d$ is a limit point of $\{x + \alpha d_k\}$ and C is closed, we have $x + \alpha d \in C$. Hence, using also Prop. 1.4.1(b), it follows that the nonzero vector d is a direction of recession.

(b) If $d \in R_{\mathrm{ri}(C)}$, then for a fixed $x \in \mathrm{ri}(C)$ and all $\alpha \geq 0$, we have $x + \alpha d \in \mathrm{ri}(C) \subset C$. Hence, by Prop. 1.4.1(b), we have $d \in R_C$. Conversely, if $d \in R_C$, then for any $x \in \mathrm{ri}(C)$, we have $x + \alpha d \in C$ for all $\alpha \geq 0$. It follows from the Line Segment Principle (Prop. 1.3.1) that $x + \alpha d \in \mathrm{ri}(C)$ for all $\alpha \geq 0$, so that d belongs to $R_{\mathrm{ri}(C)}$.

(c) By the definition of direction of recession, $d \in R_{\cap_{i \in I} C_i}$ implies that $x + \alpha d \in \cap_{i \in I} C_i$ for all $x \in \cap_{i \in I} C_i$ and all $\alpha \geq 0$. By Prop. 1.4.1(b), this in turn implies that $d \in R_{C_i}$ for all i, so that $R_{\cap_{i \in I} C_i} \subset \cap_{i \in I} R_{C_i}$. Conversely, by the definition of direction of recession, if $d \in \cap_{i \in I} R_{C_i}$ and $x \in \cap_{i \in I} C_i$, we have $x + \alpha d \in \cap_{i \in I} C_i$ for all $\alpha \geq 0$, so $d \in R_{\cap_{i \in I} C_i}$. Thus, $\cap_{i \in I} R_{C_i} \subset R_{\cap_{i \in I} C_i}$.

(d) Consider the closed convex set $\overline{V} = \{x \mid Ax \in W\}$, and choose some $x \in \overline{V}$. Then, by Prop. 1.4.1(b), $d \in R_{\overline{V}}$ if and only if $x + \alpha d \in \overline{V}$ for all $\alpha \geq 0$, or equivalently if and only if $A(x + \alpha d) \in W$ for all $\alpha \geq 0$. Since $Ax \in W$, the last statement is equivalent to $Ad \in R_W$. Thus, $d \in R_{\overline{V}}$ if and only if $Ad \in R_W$. Since W is compact, from part (a) we have $R_W = \{0\}$, so $R_{\overline{V}}$ is equal to $\{d \mid Ad = 0\}$, which is $N(A)$. Since $V = C \cap \overline{V}$, using part (c), we have $R_V = R_C \cap N(A)$. **Q.E.D.**

For an example where part (a) of the preceding proposition fails, consider the unbounded convex set

$$C = \big\{(x_1, x_2) \mid 0 \leq x_1 < 1,\, 0 \leq x_2\big\} \cup \big\{(1, 0)\big\}.$$

By using the definition, it can be verified that C has no nonzero directions of recession. It can also be verified that $(0, 1)$ is a direction of recession of $\mathrm{ri}(C)$, so part (b) also fails. Finally, by letting

$$D = \big\{(x_1, x_2) \mid -1 \leq x_1 \leq 0,\, 0 \leq x_2\big\},$$

it can be seen that $(0, 1) \in R_D$, so $R_{C \cap D} \neq R_C \cap R_D$ and part (c) fails as well.

Note that part (c) of the preceding proposition implies that if C and D are nonempty closed and convex sets such that $C \subset D$, then $R_C \subset R_D$. This can be seen by using part (c) to write $R_C = R_{C \cap D} = R_C \cap R_D$, from which we obtain $R_C \subset R_D$. This property can fail if the sets C and D are not closed; for example, if

$$C = \big\{(x_1, x_2) \mid 0 \leq x_1 < 1,\, 0 \leq x_2\big\}, \qquad D = C \cup \big\{(1,0)\big\},$$

then the vector $(0,1)$ is a direction of recession of C but not of D.

Lineality Space

A subset of the recession cone of a convex set C that plays an important role in a number of interesting contexts is its *lineality space*, denoted by L_C. It is defined as the set of directions of recession d whose opposite, $-d$, are also directions of recession:

$$L_C = R_C \cap (-R_C).$$

Thus $d \in L_C$ if and only if the entire line $\{x + \alpha d \mid \alpha \in \Re\}$ is contained in C for every $x \in C$.

The lineality space inherits several of the properties of the recession cone that we have shown (Props. 1.4.1 and 1.4.2). We collect these properties in the following proposition.

Proposition 1.4.3: (Properties of Lineality Space) Let C be a nonempty closed convex subset of $\Re^n$.

(a) L_C is a subspace of $\Re^n$.

(b) $L_C = L_{\text{ri}(C)}$.

(c) For any collection of closed convex sets C_i, $i \in I$, where I is an arbitrary index set and $\cap_{i \in I} C_i \neq \emptyset$, we have

$$L_{\cap_{i \in I} C_i} = \cap_{i \in I} L_{C_i}.$$

(d) Let W be a compact and convex subset of $\Re^m$, and let A be an $m \times n$ matrix. The lineality space of the set

$$V = \{x \in C \mid Ax \in W\}$$

(assuming it is nonempty) is $L_C \cap N(A)$, where $N(A)$ is the nullspace of A.

Proof: (a) Let d_1 and d_2 belong to L_C, and let α_1 and α_2 be nonzero scalars. We will show that $\alpha_1 d_1 + \alpha_2 d_2$ belongs to L_C. Indeed, we have

$$\begin{aligned} \alpha_1 d_1 + \alpha_2 d_2 &= |\alpha_1|\big(\mathrm{sgn}(\alpha_1) d_1\big) + |\alpha_2|\big(\mathrm{sgn}(\alpha_2) d_2\big) \\ &= \big(|\alpha_1| + |\alpha_2|\big)\big(\alpha \overline{d}_1 + (1-\alpha)\overline{d}_2\big), \end{aligned} \tag{1.18}$$

where

$$\alpha = \frac{|\alpha_1|}{|\alpha_1| + |\alpha_2|}, \qquad \overline{d}_1 = \mathrm{sgn}(\alpha_1) d_1, \quad \overline{d}_2 = \mathrm{sgn}(\alpha_2) d_2,$$

and for a nonzero scalar s, we use the notation $\mathrm{sgn}(s) = 1$ or $\mathrm{sgn}(s) = -1$ depending on whether s is positive or negative, respectively. We now note that L_C is a convex cone, being the intersection of the convex cones R_C and $-R_C$. Hence, since $\overline{d}_1$ and $\overline{d}_2$ belong to L_C, any positive multiple of a convex combination of $\overline{d}_1$ and $\overline{d}_2$ belongs to L_C. It follows from Eq. (1.18) that $\alpha_1 d_1 + \alpha_2 d_2 \in L_C$.

(b) We have

$$L_{\mathrm{ri}(C)} = R_{\mathrm{ri}(C)} \cap \big(-R_{\mathrm{ri}(C)}\big) = R_C \cap (-R_C) = L_C,$$

where the second equality follows from Prop. 1.4.2(b).

(c) We have

$$\begin{aligned} L_{\cap_{i\in I} C_i} &= \big(R_{\cap_{i\in I} C_i}\big) \cap \big(-R_{\cap_{i\in I} C_i}\big) \\ &= (\cap_{i\in I} R_{C_i}) \cap (-\cap_{i\in I} R_{C_i}) \\ &= \cap_{i\in I}\big(R_{C_i} \cap (-R_{C_i})\big) \\ &= \cap_{i\in I} L_{C_i}, \end{aligned}$$

where the second equality follows from Prop. 1.4.2(c).

(d) We have

$$\begin{aligned} L_V &= R_V \cap (-R_V) \\ &= \big(R_C \cap N(A)\big) \cap \big((-R_C) \cap N(A)\big) \\ &= \big(R_C \cap (-R_C)\big) \cap N(A) \\ &= L_C \cap N(A), \end{aligned}$$

where the second equality follows from Prop. 1.4.2(d). **Q.E.D.**

Example 1.4.1: (Sets Specified by Linear and Convex Quadratic Inequalities)

Consider a nonempty set of the form

$$C = \{x \mid x'Qx + c'x + b \le 0\},$$

where Q is a symmetric positive semidefinite $n \times n$ matrix, c is a vector in $\Re^n$, and b is a scalar. A vector d is a direction of recession if and only if

$$(x + \alpha d)' Q (x + \alpha d) + c'(x + \alpha d) + b \leq 0, \qquad \forall\ \alpha > 0,\ x \in C,$$

or

$$x'Qx + c'x + b + \alpha(c + 2Qx)'d + \alpha^2 d'Qd \leq 0, \quad \forall\ \alpha > 0,\ x \in C. \tag{1.19}$$

Clearly, we cannot have $d'Qd > 0$, since then the left-hand side above would become arbitrarily large for a suitably large choice of α, so $d'Qd = 0$. Since Q is positive semidefinite, it can be written as $Q = M'M$ for some matrix M, so that we have $Md = 0$, implying that $Qd = 0$. It follows that Eq. (1.19) is equivalent to

$$x'Qx + c'x + b + \alpha c'd \leq 0, \qquad \forall\ \alpha > 0,\ x \in C,$$

which is true if and only if $c'd \leq 0$. Thus,

$$R_C = \{d \mid Qd = 0,\ c'd \leq 0\}.$$

Also, $L_C = R_C \cap (-R_C)$, so

$$L_C = \{d \mid Qd = 0,\ c'd = 0\}.$$

Consider now the case where C is nonempty and specified by any (possibly infinite) number of convex quadratic inequalities:

$$C = \{x \mid x'Q_j x + c_j' x + b_j \leq 0,\ j \in J\},$$

where J is some index set. Then using Props. 1.4.2(c) and 1.4.3(c), we have

$$R_C = \{d \mid Q_j d = 0,\ c_j' d \leq 0,\ \forall\ j \in J\},$$

$$L_C = \{d \mid Q_j d = 0,\ c_j' d = 0,\ \forall\ j \in J\}.$$

In particular, if C is a polyhedral set of the form

$$C = \{x \mid c_j' x + b_j \leq 0,\ j = 1, \ldots, r\},$$

we have

$$R_C = \{d \mid c_j' d \leq 0,\ j = 1, \ldots, r\}, \qquad L_C = \{d \mid c_j' d = 0,\ j = 1, \ldots, r\}.$$

Finally, let us prove a useful result that allows the decomposition of a convex set along a subspace of its lineality space (possibly the entire lineality space) and its orthogonal complement (see Fig. 1.4.3).

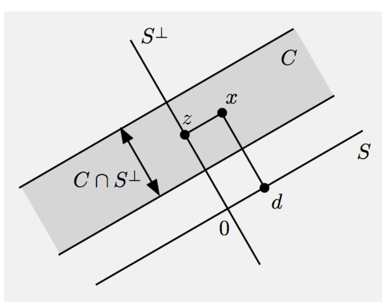

Figure 1.4.3. Illustration of the decomposition of a convex set C as

$$C = S + (C \cap S^\perp),$$

where S is a subspace contained in the lineality space L_C. A vector $x \in C$ is expressed as $x = d + z$ with $d \in S$ and $z \in C \cap S^\perp$, as shown.

Proposition 1.4.4: (Decomposition of a Convex Set) Let C be a nonempty convex subset of $\Re^n$. Then, for every subspace S that is contained in the lineality space L_C, we have

$$C = S + (C \cap S^\perp).$$

Proof: We can decompose $\Re^n$ as $S + S^\perp$, so for $x \in C$, let $x = d + z$ for some $d \in S$ and $z \in S^\perp$. Because $-d \in S \subset L_C$, the vector $-d$ is a direction of recession of C, so the vector $x - d$, which is equal to z, belongs to C, implying that $z \in C \cap S^\perp$. Thus, we have $x = d + z$ with $d \in S$ and $z \in C \cap S^\perp$ showing that $C \subset S + (C \cap S^\perp)$.

Conversely, if $x \in S + (C \cap S^\perp)$, then $x = d + z$ with $d \in S$ and $z \in C \cap S^\perp$. Thus, we have $z \in C$. Furthermore, because $S \subset L_C$, the vector d is a direction of recession of C, implying that $d + z \in C$. Hence $x \in C$, showing that $S + (C \cap S^\perp) \subset C$. **Q.E.D.**

In the special case where $S = L_C$ in Prop. 1.4.4, we obtain

$$C = L_C + (C \cap L_C^\perp). \tag{1.20}$$

Thus, C is the vector sum of two sets:

(1) The set L_C, which consists of the lines contained in C, translated to pass through the origin.

(2) The set $C \cap L_C^\perp$, which contains no lines; to see this, note that for any line $\{x + \alpha d \mid \alpha \in \Re\} \subset C \cap L_C^\perp$, we have $d \in L_C$ (since $x + \alpha d \in C$ for all $\alpha \in \Re$), so $d \perp (x + \alpha d)$ for all $\alpha \in \Re$, implying that $d = 0$.

Note that if $R_C = L_C$ and C is closed, the set $C \cap L_C^\perp$ contains no nonzero directions of recession, so it is compact [cf. Prop. 1.4.2(a)], and C can be decomposed into the sum of L_C and a compact set, as per Eq. (1.20).

1.4.1 Directions of Recession of a Convex Function

We will now develop a notion of direction of recession of a convex function. This notion is important in several contexts, including the existence of solutions of convex optimization problems, which will be discussed in Chapter 3. A key fact is that a convex function f can be described in terms of its epigraph, which is a convex set. The recession cone of epi(f) can be used to obtain the directions along which f does not increase monotonically. In particular, the "horizontal directions" in the recession cone of epi(f) correspond to the directions along which the level sets $\big\{x \mid f(x) \leq \gamma\big\}$ are unbounded. Along these directions, f is monotonically nonincreasing. This is the idea underlying the following proposition.

Proposition 1.4.5: Let $f : \Re^n \mapsto (-\infty, \infty]$ be a closed proper convex function and consider the level sets

$$V_\gamma = \big\{x \mid f(x) \leq \gamma\big\}, \qquad \gamma \in \Re.$$

Then:

(a) All the nonempty level sets V_γ have the same recession cone, denoted R_f, and given by

$$R_f = \big\{d \mid (d, 0) \in R_{\text{epi}(f)}\big\},$$

where $R_{\text{epi}(f)}$ is the recession cone of the epigraph of f.

(b) If one nonempty level set V_γ is compact, then all of these level sets are compact.

Proof: (a) Fix a γ such that V_γ is nonempty. Let S be the "γ-slice" of epi(f),

$$S = \big\{(x, \gamma) \mid f(x) \leq \gamma\big\},$$

and note that

$$S = \text{epi}(f) \cap \big\{(x, \gamma) \mid x \in \Re^n\big\}.$$

Using Prop. 1.4.2(c) [which applies since epi(f) is closed in view of the closedness of f], we have

$$R_S = R_{\text{epi}(f)} \cap \big\{(d, 0) \mid d \in \Re^n\big\} = \big\{(d, 0) \mid (d, 0) \in R_{\text{epi}(f)}\big\}.$$

From this equation and the fact $S = \big\{(x, \gamma) \mid x \in V_\gamma\big\}$, the desired formula for R_{V_γ} follows.

(b) From Prop. 1.4.2(a), a nonempty level set V_γ is compact if and only if the recession cone R_{V_γ} does not contain a nonzero direction. By part (a), all nonempty level sets V_γ have the same recession cone, so if one of them is compact, all of them are compact. **Q.E.D.**

Note that closedness of f is essential for the level sets V_γ to have a common recession cone, as per Prop. 1.4.5(a). The reader may verify this by using as an example the convex but not closed function $f : \Re^2 \mapsto (-\infty, \infty]$ given by

$$f(x_1, x_2) = \begin{cases} -x_1 & \text{if } x_1 > 0,\ x_2 \geq 0, \\ x_2 & \text{if } x_1 = 0,\ x_2 \geq 0, \\ \infty & \text{if } x_1 < 0 \text{ or } x_2 < 0. \end{cases}$$

Here, for $\gamma < 0$, we have $V_\gamma = \big\{(x_1, x_2) \mid x_1 \geq -\gamma,\ x_2 \geq 0\big\}$, so that $(0, 1) \in R_{V_\gamma}$, but $V_0 = \big\{(x_1, x_2) \mid x_1 > 0,\ x_2 \geq 0\big\} \cup \big\{(0, 0)\big\}$, so that $(0, 1) \notin R_{V_0}$.

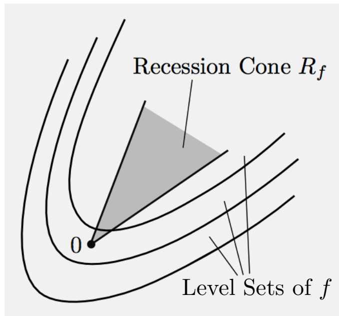

Figure 1.4.4. Illustration of the recession cone R_f of a closed proper convex function f. It is the (common) recession cone of the nonempty level sets of f.

For a closed proper convex function $f : \Re^n \mapsto (-\infty, \infty]$, the (common) recession cone R_f of the nonempty level sets is called the *recession cone of* f (cf. Fig. 1.4.4). A vector $d \in R_f$ is called a *direction of recession of* f.

The most intuitive way to look at directions of recession of f is from a descent viewpoint: if we start at any $x \in \text{dom}(f)$ and move indefinitely along a direction of recession, we must stay within each level set that contains x, or equivalently we must encounter exclusively points z with $f(z) \leq f(x)$. In words, *a direction of recession of* f *is a direction of continuous nonascent for* f. Conversely, if we start at some $x \in \text{dom}(f)$ and while moving along a direction d, we encounter a point z with $f(z) > f(x)$, then d cannot be a direction of recession. By the convexity of the level sets of f, once we cross the relative boundary of a level set, we never cross it back again, and with a little thought, it can be seen that *a direction that is not a direction of recession of* f *is a direction of eventual continuous ascent of* f [see Figs. 1.4.5(e),(f)].

Constancy Space of a Convex Function

The lineality space of the recession cone R_f of a closed proper convex function f is denoted by L_f, and is the subspace of all $d \in \Re^n$ such that both d and $-d$ are directions of recession of f, i.e.,

$$L_f = R_f \cap (-R_f).$$

Equivalently, $d \in L_f$ if and only if both d and $-d$ are directions of recession of each of the nonempty level sets $\big\{x \mid f(x) \leq \gamma\big\}$ [cf. Prop. 1.4.5(a)]. In view of the convexity of f, which implies that f is monotonically nonincreasing along a direction of recession, we see that $d \in L_f$ if and only if

$$f(x + \alpha d) = f(x), \qquad \forall\ x \in \text{dom}(f),\ \forall\ \alpha \in \Re.$$

Consequently, any $d \in L_f$ is called a *direction in which* f *is constant*, and L_f is called the *constancy space of* f.

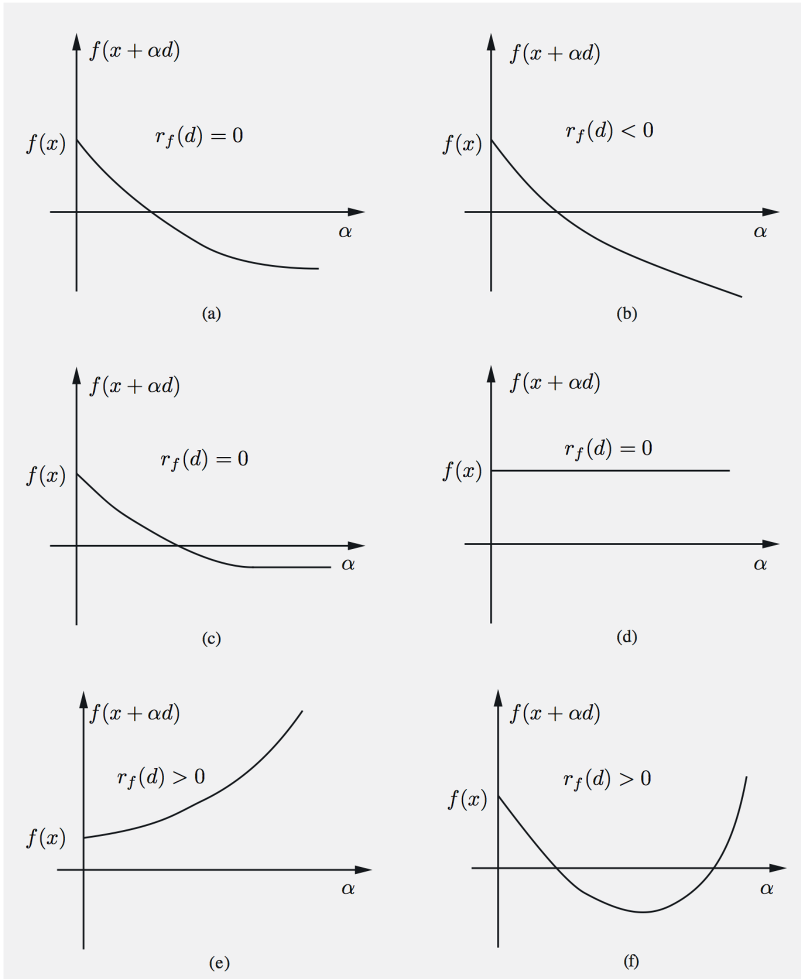

Figure 1.4.5. Ascent/descent behavior of a closed proper convex function starting at some $x \in \mathrm{dom}(f)$ and moving along a direction d. If d is a direction of recession of f, there are two possibilities: either f decreases monotonically to a finite value or $-\infty$ [figures (a) and (b), respectively], or f reaches a value that is less or equal to $f(x)$ and stays at that value [figures (c) and (d)]. If d is not a direction of recession of f, then eventually f increases monotonically to ∞ [figures (e) and (f)], i.e., for some $\overline{\alpha} \geq 0$ and all $\alpha_1, \alpha_2 \geq \overline{\alpha}$ with $\alpha_1 < \alpha_2$, we have

$$f(x + \alpha_1 d) < f(x + \alpha_2 d).$$

This behavior is determined only by d, and is independent of the choice of x within $\mathrm{dom}(f)$.

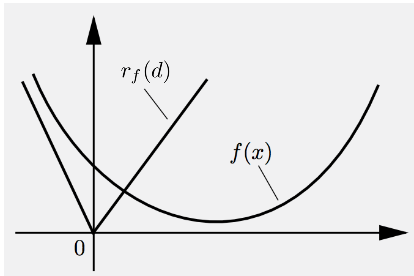

Figure 1.4.6. Illustration of the recession function r_f of a closed proper convex function f. Its epigraph is the recession cone of the epigraph of f.

As an example, if f is a quadratic function given by

$$f(x) = x'Qx + c'x + b,$$

where Q is a symmetric positive semidefinite $n \times n$ matrix, c is a vector in $\Re^n$, and b is a scalar, then its recession cone and constancy space are

$$R_f = \{d \mid Qd = 0,\ c'd \leq 0\}, \qquad L_f = \{d \mid Qd = 0,\ c'd = 0\}$$

(cf. Example 1.4.1).

Recession Function

We saw that if d is a direction of recession of f, then f is asymptotically nonincreasing along each halfline $x + \alpha d$, but in fact a stronger property holds: it turns out that *the asymptotic slope of f along d is independent of the starting point x*. The "asymptotic slope" of a closed proper convex function along a direction is expressed by a function that we now introduce.

We first note that the recession cone $R_{\text{epi}(f)}$ of the epigraph of a closed proper convex function $f : \Re^n \mapsto (-\infty, \infty]$ is itself the epigraph of another closed proper convex function. The reason is that for a given d, the set of scalars w such that $(d, w) \in R_{\text{epi}(f)}$ is either empty or it is a closed interval that is unbounded above and is bounded below (since f is proper and hence its epigraph does not contain a vertical line). Thus $R_{\text{epi}(f)}$ is the epigraph of a proper function, which must be closed and convex [since f, $\text{epi}(f)$, and $R_{\text{epi}(f)}$ are all closed and convex]. This function is called the *recession function of f* and is denoted r_f, i.e.,

$$\text{epi}(r_f) = R_{\text{epi}(f)};$$

see Fig. 1.4.6.

The recession function can be used to characterize the recession cone and constancy space of the function, as in the following proposition (cf. Fig. 1.4.5).

Proposition 1.4.6: Let $f : \Re^n \mapsto (-\infty, \infty]$ be a closed proper convex function. Then the recession cone and constancy space of f are given in terms of its recession function by

$$R_f = \big\{d \mid r_f(d) \le 0\big\}, \qquad L_f = \big\{d \mid r_f(d) = r_f(-d) = 0\big\}.$$

Proof: By Prop. 1.4.5(a) and the definition $\text{epi}(r_f) = R_{\text{epi}(f)}$ of r_f, the recession cone of f is

$$R_f = \big\{d \mid (d, 0) \in R_{\text{epi}(f)}\big\} = \big\{d \mid r_f(d) \le 0\big\}.$$

Since $L_f = R_f \cap (-R_f)$, it follows that $d \in L_f$ if and only if $r_f(d) \le 0$ and $r_f(-d) \le 0$. On the other hand, by the convexity of r_f, we have

$$r_f(d) + r_f(-d) \ge 2r_f(0) = 0, \qquad \forall\ d \in \Re^n,$$

so it follows that $d \in L_f$ if and only if $r_f(d) = r_f(-d) = 0$. **Q.E.D.**

The following proposition provides an explicit formula for the recession function.

Proposition 1.4.7: Let $f : \Re^n \mapsto (-\infty, \infty]$ be a closed proper convex function. Then, for all $x \in \text{dom}(f)$ and $d \in \Re^n$,

$$r_f(d) = \sup_{\alpha > 0} \frac{f(x + \alpha d) - f(x)}{\alpha} = \lim_{\alpha \to \infty} \frac{f(x + \alpha d) - f(x)}{\alpha}. \qquad (1.21)$$

Proof: By definition, we have that $(d, \nu) \in R_{\text{epi}(f)}$ if and only if for all $(x, w) \in \text{epi}(f)$,

$$(x + \alpha d, w + \alpha \nu) \in \text{epi}(f), \qquad \forall\ \alpha > 0,$$

or equivalently, $f(x + \alpha d) \le f(x) + \alpha \nu$ for all $\alpha > 0$, which can be written as

$$\frac{f(x + \alpha d) - f(x)}{\alpha} \le \nu, \qquad \forall\ \alpha > 0.$$

Hence

$$(d, \nu) \in R_{\text{epi}(f)} \qquad \text{if and only if} \qquad \sup_{\alpha > 0} \frac{f(x + \alpha d) - f(x)}{\alpha} \le \nu,$$

for all $x \in \mathrm{dom}(f)$. Since $R_{\mathrm{epi}(f)}$ is the epigraph of r_f, this implies the first equality in Eq. (1.21).

From the convexity of f, we see that the ratio

$$\frac{f(x+\alpha d)-f(x)}{\alpha}$$

is monotonically nondecreasing as a function of α over the range $(0,\infty)$. This implies the second equality in Eq. (1.21). **Q.E.D.**

The last expression in Eq. (1.21) leads to the interpretation of $r_f(d)$ as the "asymptotic slope" of f along the direction d. In fact, for differentiable convex functions $f : \Re^n \mapsto \Re$, this interpretation can be made more precise: we have

$$r_f(d) = \lim_{\alpha\to\infty} \nabla f(x+\alpha d)'d, \qquad \forall\ x \in \Re^n,\ d \in \Re^n. \tag{1.22}$$

Indeed, for all x, d, and $\alpha > 0$, we have using Prop. 1.1.7(a),

$$\nabla f(x)'d \le \frac{f(x+\alpha d)-f(x)}{\alpha} \le \nabla f(x+\alpha d)'d,$$

so by taking the limit as $\alpha \to \infty$ and using Eq. (1.21), it follows that

$$\nabla f(x)'d \le r_f(d) \le \lim_{\alpha\to\infty} \nabla f(x+\alpha d)'d. \tag{1.23}$$

The left-hand side above holds for all x, so replacing x with $x+\alpha d$,

$$\nabla f(x+\alpha d)'d \le r_f(d), \qquad \forall\ \alpha > 0.$$

By taking the limit as $\alpha \to \infty$, we obtain

$$\lim_{\alpha\to\infty} \nabla f(x+\alpha d)'d \le r_f(d), \tag{1.24}$$

and by combining Eqs. (1.23) and (1.24), we obtain Eq. (1.22).

The calculation of recession functions can be facilitated by nice formulas for the sum and the supremum of closed convex functions. The following proposition deals with the case of a sum.

Proposition 1.4.8: (Recession Function of a Sum) Let $f_i : \Re^n \mapsto (-\infty,\infty]$, $i = 1,\ldots,m$, be closed proper convex functions such that the function $f = f_1 + \cdots + f_m$ is proper. Then

$$r_f(d) = r_{f_1}(d) + \cdots + r_{f_m}(d), \qquad \forall\ d \in \Re^n. \tag{1.25}$$

Proof: Without loss of generality, assume that $m = 2$, and note that $f_1 + f_2$ is closed proper convex (cf. Prop. 1.1.5). By using Eq. (1.21), we have for all $x \in \text{dom}(f_1 + f_2)$ and $d \in \Re^n$,

$$\begin{aligned} r_{f_1+f_2}(d) &= \lim_{\alpha\to\infty} \left\{ \frac{f_1(x+\alpha d) - f_1(x)}{\alpha} + \frac{f_2(x+\alpha d) - f_2(x)}{\alpha} \right\} \\ &= \lim_{\alpha\to\infty} \left\{ \frac{f_1(x+\alpha d) - f_1(x)}{\alpha} \right\} + \lim_{\alpha\to\infty} \left\{ \frac{f_2(x+\alpha d) - f_2(x)}{\alpha} \right\} \\ &= r_{f_1}(d) + r_{f_2}(d), \end{aligned}$$

where the second equality holds because the limits involved exist. **Q.E.D.**

Note that for the formula (1.25) to hold, it is essential that f is proper, for otherwise its recession function is undefined. There is a similar result regarding the function

$$f(x) = \sup_{i\in I} f_i(x),$$

where I is an arbitrary index set, and $f_i : \Re^n \mapsto (-\infty, \infty]$, $i \in I$, are closed proper convex functions such that f is proper. In particular, we have

$$r_f(d) = \sup_{i\in I} r_{f_i}(d), \qquad d \in \Re^n. \tag{1.26}$$

To show this, we simply note that the epigraph of r_f is the recession cone of the epigraph of f, the intersection of the epigraphs of f_i. Thus, the epigraph of r_f is the intersection of the recession cones of the epigraphs of f_i by Prop. 1.4.2(c), which yields the formula (1.26).

1.4.2 Nonemptiness of Intersections of Closed Sets

The notions of recession cone and lineality space can be used to generalize some of the fundamental properties of compact sets to closed convex sets. One such property is that a sequence $\{C_k\}$ of nonempty and compact sets with $C_{k+1} \subset C_k$ for all k has nonempty and compact intersection [cf. Prop. A.2.4(h)]. Another property is that the image of a compact set under a linear transformation is compact [cf. Prop. A.2.6(d)]. These properties may not hold when the sets involved are closed but unbounded (cf. Fig. 1.3.4), and some additional conditions are needed for their validity. In this section we develop such conditions, using directions of recession and related notions. We focus on the case where the sets involved are convex, but the analysis generalizes to the nonconvex case (see [BeT07]).

To understand the significance of set intersection results, consider a sequence of nonempty closed sets $\{C_k\}$ in $\Re^n$ with $C_{k+1} \subset C_k$ for all k (such a sequence is said to be *nested*), and the question whether $\cap_{k=0}^{\infty} C_k$ is nonempty. Here are some of the contexts where this question arises:

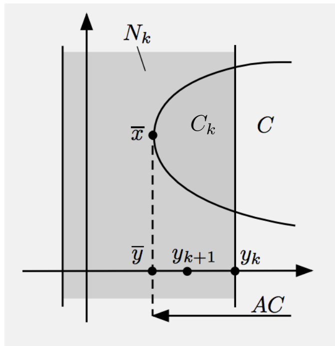

Figure 1.4.7. Set intersection argument to prove that the set AC closed when C is closed. Here A is the projection on the horizontal axis of points in the plane. For a sequence $\{y_k\} \subset AC$ that converges to some $\overline{y}$, in order to prove that $\overline{y} \in AC$, it is sufficient to prove that the intersection $\cap_{k=0}^{\infty} C_k$ is nonempty, where

$$C_k = C \cap N_k,$$

and

$$N_k = \{x \mid \|Ax - \overline{y}\| \leq \|y_k - \overline{y}\|\}.$$

(a) Does a function $f : \Re^n \mapsto \Re$ attain a minimum over a set X? This is true if and only if the intersection

$$\cap_{k=0}^{\infty} \{x \in X \mid f(x) \leq \gamma_k\}$$

is nonempty, where $\{\gamma_k\}$ is a scalar sequence with $\gamma_k \downarrow \inf_{x \in X} f(x)$.

(b) If C is a closed set and A is a matrix, is AC closed? To prove this, we may let $\{y_k\}$ be a sequence in AC that converges to some $\overline{y} \in \Re^n$, and then show that $\overline{y} \in AC$. If we introduce the sets $C_k = C \cap N_k$, where

$$N_k = \{x \mid \|Ax - \overline{y}\| \leq \|y_k - \overline{y}\|\},$$

it is sufficient to show that $\cap_{k=0}^{\infty} C_k$ is nonempty (see Fig. 1.4.7).

We will next consider a nested sequence $\{C_k\}$ of nonempty closed convex sets, and in the subsequent propositions, we will derive several alternative conditions under which the intersection $\cap_{k=0}^{\infty} C_k$ is nonempty. These conditions involve a variety of assumptions about the recession cones, the lineality spaces, and the structure of the sets C_k.

Asymptotic Sequences of Convex Sets

The following line of analysis actually extends to nonconvex closed sets (see [BeT07]). However, in this book we will restrict ourselves to set intersections involving only convex sets.

Our analysis revolves around sequences $\{x_k\}$ such that $x_k \in C_k$ for each k. An important fact is that $\cap_{k=0}^{\infty} C_k$ is empty if and only if every sequence of this type is unbounded. Thus the idea is to introduce assumptions that guarantee that not all such sequences are unbounded. In fact it will be sufficient to restrict attention to unbounded sequences that escape to ∞ along common directions of recession of the sets C_k, as in the following definition.

Definition 1.4.1: Let $\{C_k\}$ be a nested sequence of nonempty closed convex sets. We say that $\{x_k\}$ is an *asymptotic sequence of* $\{C_k\}$ if $x_k \neq 0$, $x_k \in C_k$ for all k, and

$$\|x_k\| \to \infty, \qquad \frac{x_k}{\|x_k\|} \to \frac{d}{\|d\|},$$

where d is some nonzero common direction of recession of the sets C_k,

$$d \neq 0, \qquad d \in \cap_{k=0}^{\infty} R_{C_k}.$$

A special case is when all the sets C_k are equal. In particular, for a nonempty closed convex C, we say that $\{x_k\} \subset C$ is an asymptotic sequence of C if $\{x_k\}$ is asymptotic for the sequence $\{C_k\}$, where $C_k \equiv C$.

Note that given any unbounded sequence $\{x_k\}$ such that $x_k \in C_k$ for each k, there exists a subsequence $\{x_k\}_{k \in \mathcal{K}}$ that is asymptotic for the corresponding subsequence $\{C_k\}_{k \in \mathcal{K}}$. In fact, any limit point of $\big\{x_k/\|x_k\|\big\}$ is a common direction of recession of the sets C_k; this can be seen by using the proof argument of Prop. 1.4.1(b). Thus, asymptotic sequences are in a sense representative of unbounded sequences with $x_k \in C_k$ for each k.

We now introduce a special type of set sequences that have favorable properties for our purposes.

Definition 1.4.2: Let $\{C_k\}$ be a nested sequence of nonempty closed convex sets. We say that *an asymptotic sequence* $\{x_k\}$ *is retractive* if for the direction d corresponding to $\{x_k\}$ as per Definition 1.4.1, there exists an index $\overline{k}$ such that

$$x_k - d \in C_k, \qquad \forall\ k \geq \overline{k}.$$

We say that *the sequence* $\{C_k\}$ *is retractive* if all its asymptotic sequences are retractive. In the special case $C_k \equiv C$, we say that *the set* C *is retractive* if all its asymptotic sequences are retractive.

Retractive set sequences are those whose asymptotic sequences still belong to the corresponding sets C_k (for sufficiently large k) when shifted by $-d$, where d is any corresponding direction of recession. For an example, consider a nested sequence consisting of "cylindrical" sets in the plane, such as $C_k = \big\{(x^1, x^2) \mid |x^1| \leq 1/k\big\}$, whose asymptotic sequences $\{(x_k^1, x_k^2)\}$ are retractive: they satisfy $x_k^1 \to 0$, and either $x_k^2 \to \infty$ $[d = (0, 1)]$ or $x_k^2 \to -\infty$ $[d = (0, -1)]$ (see also Fig. 1.4.8). Some important types of

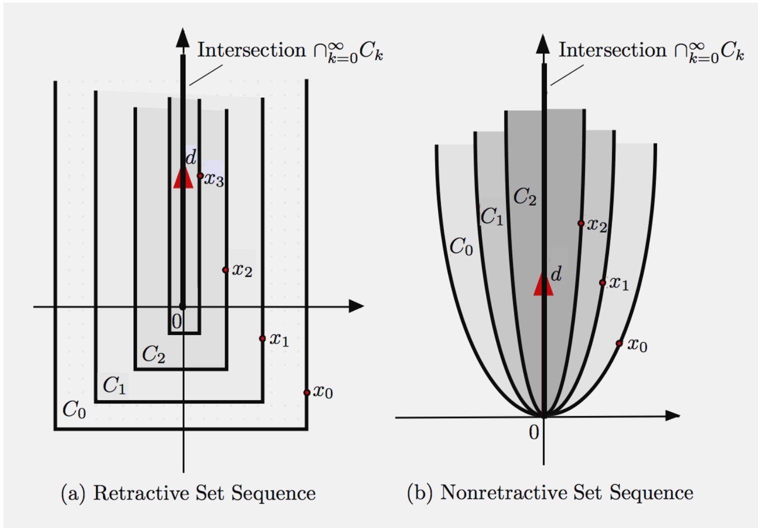

Figure 1.4.8. Illustration of retractive and nonretractive sequences in $\Re^2$. For both set sequences, the intersection is the vertical half line $\{x \mid x_2 \geq 0\}$, and the common directions of recession are of the form $(0, d_2)$ with $d_2 \geq 0$. For the example on the right, any unbounded sequence $\{x_k\}$ such that x_k is on the boundary of the set C_k is asymptotic but not retractive.

set sequences can be shown to be retractive. As an aid in this regard, we note that intersections and Cartesian products (involving a finite number of sets) preserve retractiveness, as can be easily seen from the definition. In particular, if $\{C_k^1\}, \ldots, \{C_k^r\}$ are retractive nested sequences of nonempty closed convex sets, the sequences $\{N_k\}$ and $\{T_k\}$ are retractive, where

$$N_k = C_k^1 \cap C_k^2 \cap \cdots \cap C_k^r, \qquad T_k = C_k^1 \times C_k^2 \times \cdots \times C_k^r, \qquad \forall\ k,$$

and we assume that all the sets N_k are nonempty.

The following proposition shows that a polyhedral set is retractive. Indeed, this is the most important type of retractive set for our purposes. Another retractive set of interest is the vector sum of a convex compact set and a polyhedral cone; we leave the proof of this for the reader. However, as a word of caution, we mention that a nonpolyhedral closed convex cone need not be retractive.

Proposition 1.4.9: A polyhedral set is retractive.

Proof: A closed halfspace is clearly retractive. A polyhedral set is the

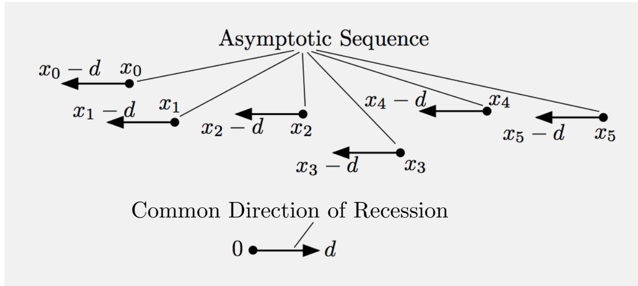

Figure 1.4.9. Geometric view of the proof idea of Prop. 1.4.10. An asymptotic sequence $\{x_k\}$ with corresponding direction of recession d eventually (for large k) gets closer to 0 when shifted by $-d$, so such a sequence cannot consist of the vectors of minimum norm from C_k without contradicting the retractiveness assumption.

nonempty intersection of a finite number of closed halfspaces, and set intersection preserves retractiveness. **Q.E.D.**

Set Intersection Theorems

The importance of retractive sequences is motivated by the following proposition.

Proposition 1.4.10: A retractive nested sequence of nonempty closed convex sets has nonempty intersection.

Proof: Let $\{C_k\}$ be the given sequence. For each k, let x_k be the vector of minimum norm in the closed set C_k (projection of the origin on C_k; cf. Prop. 1.1.9). The proof involves two ideas:

(a) The intersection $\cap_{k=0}^{\infty} C_k$ is empty if and only $\{x_k\}$ is unbounded, so there is a subsequence $\{x_k\}_{k\in\mathcal{K}}$ that is asymptotic.

(b) If a subsequence $\{x_k\}_{k\in\mathcal{K}}$ of minimum norm vectors of C_k is asymptotic with corresponding direction of recession d, then $\{x_k\}_{k\in\mathcal{K}}$ cannot be retractive, because x_k would eventually (for large k) get closer to 0 when shifted by $-d$ (see Fig. 1.4.9).

It will be sufficient to show that a subsequence $\{x_k\}_{k\in\mathcal{K}}$ is bounded. Then, since $\{C_k\}$ is nested, for each m, we have $x_k \in C_m$ for all $k \in \mathcal{K}$, $k \geq m$, and since C_m is closed, each of the limit points of $\{x_k\}_{k\in\mathcal{K}}$ will belong to each C_m and hence also to $\cap_{m=0}^{\infty} C_m$, thereby showing the result. Thus, we will prove the proposition by showing that there is no subsequence of $\{x_k\}$ that is unbounded.

Indeed, assume the contrary, let $\{x_k\}_{k\in\mathcal{K}}$ be a subsequence such that $\lim_{k\to\infty,\, k\in\mathcal{K}} \|x_k\| = \infty$, and let d be the limit of a subsequence $\big\{x_k/\|x_k\|\big\}_{k\in\overline{\mathcal{K}}}$, where $\overline{\mathcal{K}} \subset \mathcal{K}$. For each $k = 0, 1, \ldots$, define $z_k = x_m$, where m is the smallest index $m \in \overline{\mathcal{K}}$ with $k \le m$. Then, since $z_k \in C_k$ for all k and $\lim_{k\to\infty}\{z_k/\|z_k\|\} = d$, we see that d is a common direction of recession of C_k [cf. the proof of Prop. 1.4.1(b)] and $\{z_k\}$ is an asymptotic sequence corresponding to d. Using the retractiveness assumption, let $\overline{k}$ be such that $z_k - d \in C_k$ for all $k \ge \overline{k}$. We have $d'z_k \to \infty$ since

$$\frac{d'z_k}{\|z_k\|} \to \|d\|^2 = 1,$$

so for all $k \ge \overline{k}$ with $2d'z_k > 1$, we obtain

$$\|z_k - d\|^2 = \|z_k\|^2 - (2d'z_k - 1) < \|z_k\|^2.$$

This is a contradiction, since for infinitely many k, z_k is the vector of minimum norm on C_k. **Q.E.D.**

For an example, consider the sequence $\{C_k\}$ of Fig. 1.4.8(a). Here the asymptotic sequences $\{(x_k^1, x_k^2)\}$ satisfy $x_k^1 \to 0$, $x_k^2 \to \infty$ and are retractive, and indeed the intersection $\cap_{k=0}^\infty C_k$ is nonempty. On the other hand, the condition for nonemptiness of $\cap_{k=0}^\infty C_k$ of the proposition is far from necessary, e.g., the sequence $\{C_k\}$ of Fig. 1.4.8(b) has nonempty intersection but is not retractive.

A simple example where the preceding proposition applies is a "cylindrical" set sequence, where $R_{C_k} \equiv L_{C_k} \equiv L$ for some subspace L. The following proposition gives an important extension.

Proposition 1.4.11: Let $\{C_k\}$ be a nested sequence of nonempty closed convex sets. Denote

$$R = \cap_{k=0}^\infty R_{C_k}, \qquad L = \cap_{k=0}^\infty L_{C_k}.$$

(a) If $R = L$, then $\{C_k\}$ is retractive, and $\cap_{k=0}^\infty C_k$ is nonempty. Furthermore,

$$\cap_{k=0}^\infty C_k = L + \tilde{C},$$

where $\tilde{C}$ is some nonempty and compact set.

(b) Let X be a retractive closed convex set. Assume that all the sets $\overline{C}_k = X \cap C_k$ are nonempty, and that

$$R_X \cap R \ \subset \ L.$$

Then, $\{\overline{C}_k\}$ is retractive, and $\cap_{k=0}^\infty \overline{C}_k$ is nonempty.

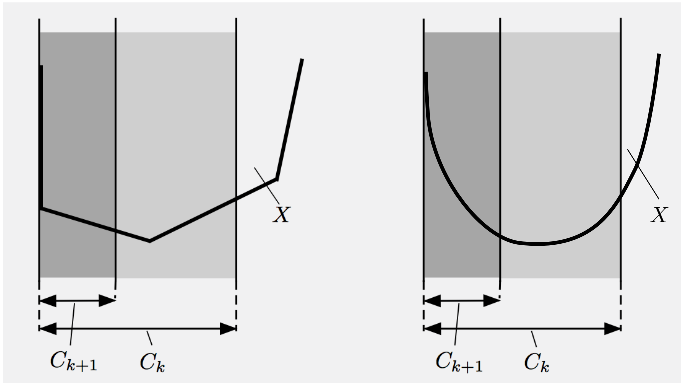

Figure 1.4.10. Illustration of the need to assume that X is retractive in Prop. 1.4.11(b). Here the intersection $\cap_{k=0}^{\infty} C_k$ is equal to the left vertical line. In the figure on the left, X is polyhedral, and the intersection $X \cap \left(\cap_{k=0}^{\infty} C_k\right)$ is nonempty. In the figure on the right, X is nonpolyhedral and nonretractive, and the intersection $X \cap \left(\cap_{k=0}^{\infty} C_k\right)$ is empty.

Proof: (a) The retractiveness of $\{C_k\}$ and consequent nonemptiness of $\cap_{k=0}^{\infty} C_k$ is the special case of part (b) where $X = \Re^n$. To show the given form of $\cap_{k=0}^{\infty} C_k$, we use the decomposition of Prop. 1.4.4, to obtain $\cap_{k=0}^{\infty} C_k = L + \tilde{C}$, where

$$\tilde{C} = (\cap_{k=0}^{\infty} C_k) \cap L^{\perp}.$$

The recession cone of $\tilde{C}$ is $R \cap L^{\perp}$, and since $R = L$, it is equal to $\{0\}$. Hence by Prop. 1.4.2(a), $\tilde{C}$ is compact.

(b) The common directions of recession of $\overline{C}_k$ are those in $R_X \cap R$, so by the hypothesis they must belong to L. Thus, for any asymptotic sequence $\{x_k\}$ of $\{\overline{C}_k\}$, corresponding to $d \in R_X \cap R$, we have $d \in L$, and hence $x_k - d \in C_k$ for all k. Since X is retractive, we also have $x_k - d \in X$ and hence $x_k - d \in \overline{C}_k$, for sufficiently large k. Hence $\{x_k\}$ is retractive, so $\{\overline{C}_k\}$ is retractive, and by Prop. 1.4.10, $\cap_{k=0}^{\infty} \overline{C}_k$ is nonempty. **Q.E.D.**

Figure 1.4.10 illustrates the need to assume that X is retractive in Prop. 1.4.11(b). The following is an important application of the preceding set intersection result.

Proposition 1.4.12: (Existence of Solutions of Convex Quadratic Programs) Let Q be a symmetric positive semidefinite $n \times n$ matrix, let c and $a_1, \ldots, a_r$ be vectors in $\Re^n$, and let $b_1, \ldots, b_r$ be scalars. Assume that the optimal value of the problem

$$\begin{array}{ll} \text{minimize} & x'Qx + c'x \\ \text{subject to} & a_j'x \le b_j, \quad j = 1, \ldots, r, \end{array}$$

is finite. Then the problem has at least one optimal solution.

Proof: Let f denote the cost function and let X be the polyhedral set of feasible solutions:

$$f(x) = x'Qx + c'x, \qquad X = \big\{x \mid a_j'x \le b_j,\ j = 1, \ldots, r\big\}.$$

Let also f^* be the optimal value, let $\{\gamma_k\}$ be a scalar sequence with $\gamma_k \downarrow f^*$, and denote

$$\overline{C}_k = X \cap \{x \mid x'Qx + c'x \le \gamma_k\}.$$

We will use Prop. 1.4.11(b) to show that the set of optimal solutions, i.e., the intersection $\cap_{k=0}^\infty \overline{C}_k$, is nonempty. Indeed, let R_X be the recession cone of X, and let R and L be the common recession cone and lineality space of the sets $\{x \in \Re^n \mid x'Qx + c'x \le \gamma_k\}$ (i.e., the recession cone and constancy space of f). By Example 1.4.1, we have

$$R = \{d \mid Qd = 0,\, c'd \le 0\}, \qquad L = \{d \mid Qd = 0,\, c'd = 0\},$$

$$R_X = \{d \mid a_j'd \le 0,\, j = 1, \ldots, r\}.$$

If d is such that $d \in R_X \cap R$ but $d \notin L$, then

$$Qd = 0, \qquad c'd < 0, \qquad a_j'd \le 0, \quad j = 1, \ldots, r,$$

which implies that for any $x \in X$, we have $x + \alpha d \in X$ for all $\alpha \ge 0$, while $f(x + \alpha d) \to -\infty$ as $\alpha \to \infty$. This contradicts the finiteness of f^*, and shows that $R_X \cap R \subset L$. The nonemptiness of $\cap_{k=0}^\infty \overline{C}_k$ now follows from Prop. 1.4.11(b). **Q.E.D.**

1.4.3 Closedness Under Linear Transformations

The conditions just obtained regarding the nonemptiness of the intersection of a sequence of closed convex sets can be translated to conditions guaranteeing the closedness of the image, AC, of a closed convex set C under a linear transformation A. This is the subject of the following proposition.

Proposition 1.4.13: Let X and C be nonempty closed convex sets in $\Re^n$, and let A be an $m \times n$ matrix with nullspace denoted by $N(A)$. If X is a retractive closed convex set and

$$R_X \cap R_C \cap N(A) \subset L_C,$$

then $A(X \cap C)$ is a closed set.

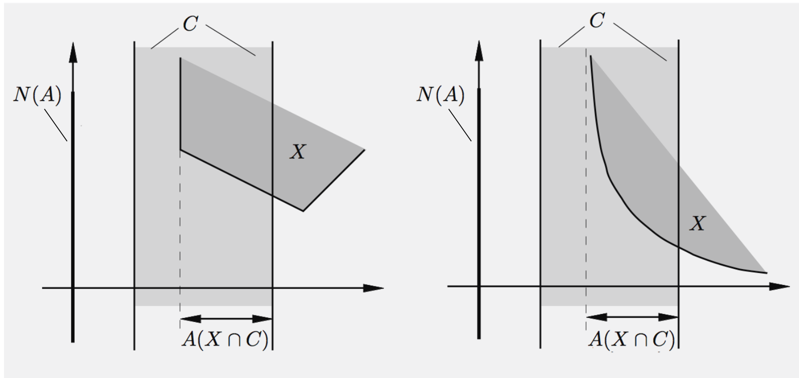

Figure 1.4.11. Illustration of the need to assume that the set X is retractive in Prop. 1.4.13. In both examples shown, the matrix A is the projection onto the horizontal axis, and its nullspace is the vertical axis. The condition $R_X \cap R_C \cap N(A) \subset L_C$ is satisfied. However, in the example on the right, X is not retractive, and the set $A(X \cap C)$ is not closed.

Proof: Let $\{y_k\}$ be a sequence in $A(X \cap C)$ converging to some $\overline{y}$. We will prove that $A(X \cap C)$ is closed by showing that $\overline{y} \in A(X \cap C)$. We introduce the sets

$$C_k = C \cap N_k,$$

where

$$N_k = \big\{x \mid \|Ax - \overline{y}\| \le \|y_k - \overline{y}\|\big\},$$

(see Fig. 1.4.7). The sets C_k are closed and convex, and their (common) recession cones and lineality spaces are $R_C \cap N(A)$ and $L_C \cap N(A)$, respectively [cf. Props. 1.4.2(d) and 1.4.3(d)]. Therefore, by Prop. 1.4.11(b), the intersection $X \cap \big(\cap_{k=0}^{\infty} C_k\big)$ is nonempty. Every point x in this intersection is such that $x \in X \cap C$ and $Ax = \overline{y}$, showing that $\overline{y} \in A(X \cap C)$. **Q.E.D.**

Figure 1.4.11 illustrates the need for the assumptions of Prop. 1.4.13. The proposition has some interesting special cases:

(a) Let $C = \Re^n$ and let X be a polyhedral set. Then, $L_C = \Re^n$ and the assumption of Prop. 1.4.13 is automatically satisfied, so it follows that $A\,X$ is closed. Thus *the image of a polyhedral set under a linear transformation is a closed set*. Simple as this result may seem, it is especially important in optimization. For example, as a special case, it yields that the cone generated by vectors $a_1, \ldots, a_r$ is a closed set, since it can be written as $A\,C$, where A is the matrix with columns $a_1, \ldots, a_r$ and C is the polyhedral set of all $(\alpha_1, \ldots, \alpha_r)$ with $\alpha_j \ge 0$ for all j. This fact is central in the proof of Farkas' Lemma, an important result given in Section 2.3.1.

(b) Let $X = \Re^n$. Then, Prop. 1.4.13 yields that AC *is closed if every direction of recession of* C *that belongs to* $N(A)$ *belongs to the lineality space of* C. This is true in particular if

$$R_C \cap N(A) = \{0\},$$

i.e., *there is no nonzero direction of recession of* C *that lies in the nullspace of* A. As a special case, this result can be used to obtain conditions that guarantee the closedness of the vector sum of closed convex sets. The idea is that the vector sum of a finite number of sets can be viewed as the image of their Cartesian product under a special type of linear transformation, as can be seen from the proof of the following proposition.

Proposition 1.4.14: Let $C_1, \ldots, C_m$ be nonempty closed convex subsets of $\Re^n$ such that the equality $d_1 + \cdots + d_m = 0$ for some vectors $d_i \in R_{C_i}$ implies that $d_i \in L_{C_i}$ for all $i = 1, \ldots, m$. Then $C_1 + \cdots + C_m$ is a closed set.

Proof: Let C be the Cartesian product $C_1 \times \cdots \times C_m$. Then, C is closed convex, and its recession cone and lineality space are given by

$$R_C = R_{C_1} \times \cdots \times R_{C_m}, \qquad L_C = L_{C_1} \times \cdots \times L_{C_m}.$$

Let A be the linear transformation that maps $(x_1, \ldots, x_m) \in \Re^{mn}$ into $x_1 + \cdots + x_m \in \Re^n$. The null space of A is the set of all $(d_1, \ldots, d_m)$ such that $d_1 + \cdots + d_m = 0$. The intersection $R_C \cap N(A)$ consists of all $(d_1, \ldots, d_m)$ such that $d_1 + \cdots + d_m = 0$ and $d_i \in R_{C_i}$ for all i. By the given condition, every $(d_1, \ldots, d_m) \in R_C \cap N(A)$ is such that $d_i \in L_{C_i}$ for all i, implying that $(d_1, \ldots, d_m) \in L_C$. Thus, $R_C \cap N(A) \subset L_C$, and by Prop. 1.4.13, the set AC is closed. Since

$$AC = C_1 + \cdots + C_m,$$

the result follows. **Q.E.D.**

When specialized to just two sets, the above proposition implies that if C_1 and $-C_2$ are closed convex sets, then $C_1 - C_2$ is closed if there is no common nonzero direction of recession of C_1 and C_2, i.e.

$$R_{C_1} \cap R_{C_2} = \{0\}.$$

This is true in particular if either C_1 or C_2 is bounded, in which case either $R_{C_1} = \{0\}$ or $R_{C_2} = \{0\}$, respectively.

Some other conditions asserting the closedness of vector sums can be derived from Prop. 1.4.13. For example, we can show that the vector sum of a finite number of polyhedral sets is closed, since it can be viewed as the image of their Cartesian product (clearly a polyhedral set) under a linear transformation (in fact this vector sum is polyhedral; see Section 2.3.2).

Another useful result is that if X is a polyhedral set, and C is a closed convex set, then $X + C$ is closed if every direction of recession of X whose opposite is a direction of recession of C lies also in the lineality space of C (replace X and C by $X \times \Re^n$ and $\Re^n \times C$, respectively, in Prop. 1.4.13, and let A map Cartesian product to sum as in the proof of Prop. 1.4.14).

1.5 HYPERPLANES

Some of the most important principles in convexity and optimization, including duality, revolve around the use of hyperplanes, i.e., $(n-1)$-dimensional affine sets, which divide $\Re^n$ into two halfspaces. For example, we will see that a closed convex set can be characterized in terms of hyperplanes: it is equal to the intersection of all the halfspaces that contain it. In the next section, we will apply this fundamental result to a convex function via its epigraph, and obtain an important dual description, encoded by another convex function, called the conjugate of the original.

A *hyperplane* in $\Re^n$ is a set of the form $\{x \mid a'x = b\}$, where a is nonzero vector in $\Re^n$ and b is a scalar. If $\overline{x}$ is any vector in a hyperplane $H = \{x \mid a'x = b\}$, then we must have $a'\overline{x} = b$, so the hyperplane can be equivalently described as

$$H = \{x \mid a'x = a'\overline{x}\},$$

or

$$H = \overline{x} + \{x \mid a'x = 0\}.$$

Thus, H is an affine set that is parallel to the subspace $\{x \mid a'x = 0\}$. The vector a is orthogonal to this subspace, and consequently, a is called the *normal* vector of H; see Fig. 1.5.1.

The sets

$$\{x \mid a'x \geq b\}, \qquad \{x \mid a'x \leq b\},$$

are called the *closed halfspaces* associated with the hyperplane (also referred to as the *positive and negative halfspaces*, respectively). The sets

$$\{x \mid a'x > b\}, \qquad \{x \mid a'x < b\},$$

are called the *open halfspaces* associated with the hyperplane.

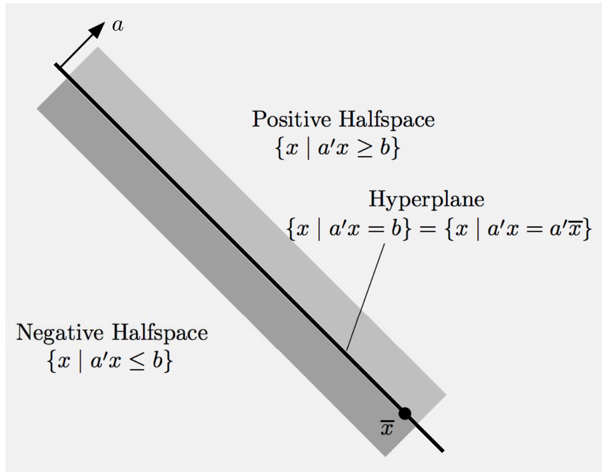

Figure 1.5.1. Illustration of the hyperplane $H = \{x \mid a'x = b\}$. If $\overline{x}$ is any vector in the hyperplane, then the hyperplane can be equivalently described as

$$H = \{x \mid a'x = a'\overline{x}\} = \overline{x} + \{x \mid a'x = 0\}.$$

The hyperplane divides the space into two halfspaces as illustrated.

1.5.1 Hyperplane Separation

We say that two sets C_1 and C_2 are *separated by a hyperplane* $H = \{x \mid a'x = b\}$ if each lies in a different closed halfspace associated with H, i.e., if either

$$a'x_1 \leq b \leq a'x_2, \qquad \forall\, x_1 \in C_1,\ \forall\, x_2 \in C_2,$$

or

$$a'x_2 \leq b \leq a'x_1, \qquad \forall\, x_1 \in C_1,\ \forall\, x_2 \in C_2.$$

We then also say that the hyperplane H *separates* C_1 and C_2, or that H is a *separating hyperplane* of C_1 and C_2. We use several different variants of this terminology. For example, the statement that two sets C_1 and C_2 *can be separated by a hyperplane* or that *there exists a hyperplane separating* C_1 and C_2, means that there exists a vector $a \neq 0$ such that

$$\sup_{x \in C_1} a'x \leq \inf_{x \in C_2} a'x;$$

[see Fig. 1.5.2(a)].

If a vector $\overline{x}$ belongs to the closure of a set C, a hyperplane that separates C and the singleton set $\{\overline{x}\}$ is said to be *supporting* C *at* $\overline{x}$. Thus

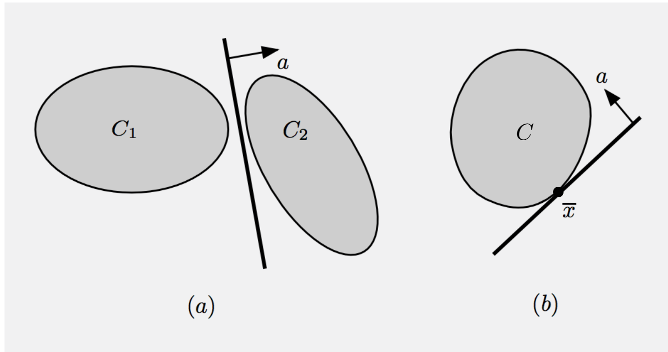

Figure 1.5.2. (a) Illustration of a hyperplane separating two sets C_1 and C_2. (b) Illustration of a hyperplane supporting a set C at a point $\overline{x}$ that belongs to the closure of C.

the statement that *there exists a supporting hyperplane of* C *at* $\overline{x}$ means that there exists a vector $a \neq 0$ such that

$$a'\overline{x} \leq a'x, \qquad \forall\ x \in C,$$

or equivalently, since $\overline{x}$ is a closure point of C,

$$a'\overline{x} = \inf_{x \in C} a'x.$$

As illustrated in Fig. 1.5.2(b), a supporting hyperplane of C is a hyperplane that "just touches" C.

We will prove several results regarding the existence of hyperplanes that separate two convex sets. Some of these results assert the existence of separating hyperplanes with special properties that will prove useful in various specialized contexts to be described later. The following proposition deals with the basic case where one of the two sets consists of a single vector. The proof is based on the Projection Theorem (Prop. 1.1.9) and is illustrated in Fig. 1.5.3.

Proposition 1.5.1: (Supporting Hyperplane Theorem) Let C be a nonempty convex subset of $\Re^n$ and let $\overline{x}$ be a vector in $\Re^n$. If $\overline{x}$ is not an interior point of C, there exists a hyperplane that passes through $\overline{x}$ and contains C in one of its closed halfspaces, i.e., there exists a vector $a \neq 0$ such that

$$a'\overline{x} \leq a'x, \qquad \forall\ x \in C. \tag{1.27}$$

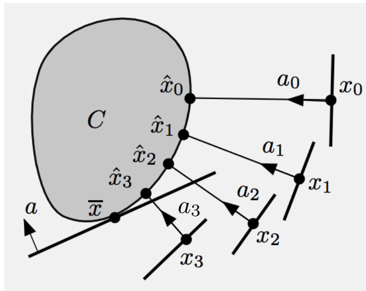

Figure 1.5.3. Illustration of the proof of the Supporting Hyperplane Theorem for the case where the vector $\overline{x}$ belongs to $\text{cl}(C)$, the closure of C. We choose a sequence $\{x_k\}$ of vectors that do not belong to $\text{cl}(C)$, with $x_k \to \overline{x}$, and we project x_k on $\text{cl}(C)$. We then consider, for each k, the hyperplane that is orthogonal to the line segment connecting x_k and its projection $\hat{x}_k$, and passes through x_k. These hyperplanes "converge" to a hyperplane that supports C at $\overline{x}$.

Proof: Consider $\text{cl}(C)$, the closure of C, which is a convex set by Prop. 1.1.1(d). Let $\{x_k\}$ be a sequence of vectors such that $x_k \to \overline{x}$ and $x_k \notin \text{cl}(C)$ for all k; such a sequence exists because $\overline{x}$ does not belong to the interior of C and hence does not belong to the interior of $\text{cl}(C)$ [cf. Prop. 1.3.5(b)]. If $\hat{x}_k$ is the projection of x_k on $\text{cl}(C)$, we have by the optimality condition of the Projection Theorem (Prop. 1.1.9)

$$(\hat{x}_k - x_k)'(x - \hat{x}_k) \geq 0, \qquad \forall\, x \in \text{cl}(C).$$

Hence we obtain for all $x \in \text{cl}(C)$ and all k,

$$(\hat{x}_k - x_k)'x \geq (\hat{x}_k - x_k)'\hat{x}_k = (\hat{x}_k - x_k)'(\hat{x}_k - x_k) + (\hat{x}_k - x_k)'x_k \geq (\hat{x}_k - x_k)'x_k.$$

We can write this inequality as

$$a_k'x \geq a_k'x_k, \qquad \forall\, x \in \text{cl}(C),\ \forall\, k, \tag{1.28}$$

where

$$a_k = \frac{\hat{x}_k - x_k}{\|\hat{x}_k - x_k\|}.$$

We have $\|a_k\| = 1$ for all k, so the sequence $\{a_k\}$ has a subsequence that converges to some $a \neq 0$. By considering Eq. (1.28) for all a_k belonging to this subsequence and by taking the limit as $k \to \infty$, we obtain Eq. (1.27). **Q.E.D.**

Note that if $\overline{x}$ is a closure point of C, then the hyperplane of the preceding proposition supports C at $\overline{x}$. Note also that if C has empty interior, then any vector $\overline{x}$ can be separated from C as in the proposition.

Proposition 1.5.2: (Separating Hyperplane Theorem) Let C_1 and C_2 be two nonempty convex subsets of $\Re^n$. If C_1 and C_2 are disjoint, there exists a hyperplane that separates them, i.e., there exists a vector $a \neq 0$ such that

$$a'x_1 \leq a'x_2, \qquad \forall\, x_1 \in C_1,\ \forall\, x_2 \in C_2. \tag{1.29}$$

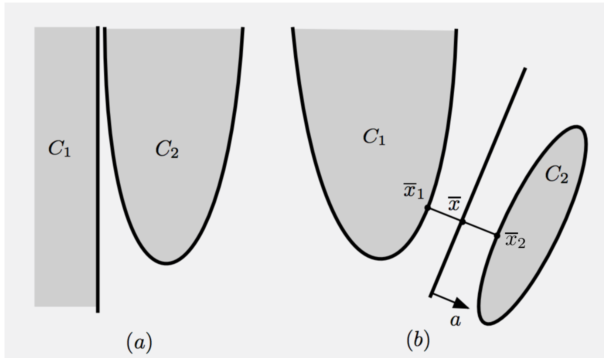

Figure 1.5.4. (a) An example of two disjoint convex sets that cannot be strictly separated. (b) Illustration of the construction of a strictly separating hyperplane.

Proof: Consider the convex set

$$C = C_2 - C_1 = \{x \mid x = x_2 - x_1,\, x_1 \in C_1,\, x_2 \in C_2\}.$$

Since C_1 and C_2 are disjoint, the origin does not belong to C, so by the Supporting Hyperplane Theorem (Prop. 1.5.1), there exists a vector $a \neq 0$ such that

$$0 \leq a'x, \qquad \forall\, x \in C,$$

which is equivalent to Eq. (1.29). **Q.E.D.**

We next consider a stronger form of separation of two sets C_1 and C_2 in $\Re^n$. We say that a hyperplane $\{x \mid a'x = b\}$ *strictly separates* C_1 *and* C_2 if it separates C_1 and C_2 while containing neither a point of C_1 nor a point of C_2, i.e.,

$$a'x_1 < b < a'x_2, \qquad \forall\, x_1 \in C_1,\ \forall\, x_2 \in C_2,$$

or

$$a'x_2 < b < a'x_1, \qquad \forall\, x_1 \in C_1,\ \forall\, x_2 \in C_2.$$

Clearly, C_1 and C_2 must be disjoint in order that they can be strictly separated. However, this is not sufficient to guarantee strict separation (see Fig. 1.5.4). The following proposition provides conditions that guarantee the existence of a strictly separating hyperplane.

Proposition 1.5.3: (Strict Separation Theorem) Let C_1 and C_2 be two disjoint nonempty convex sets. There exists a hyperplane that strictly separates C_1 and C_2 under any one of the following five conditions:

(1) $C_2 - C_1$ is closed.

(2) C_1 is closed and C_2 is compact.

(3) C_1 and C_2 are polyhedral.

(4) C_1 and C_2 are closed, and

$$R_{C_1} \cap R_{C_2} = L_{C_1} \cap L_{C_2},$$

where R_{C_i} and L_{C_i} denote the recession cone and the lineality space of C_i, $i = 1, 2$.

(5) C_1 is closed, C_2 is polyhedral, and $R_{C_1} \cap R_{C_2} \subset L_{C_1}$.

Proof: We will show the result under condition (1). The result will then follow under conditions (2)-(5), because these conditions imply condition (1) (see Prop. 1.4.14, and the discussion following its proof).

Assume that $C_2 - C_1$ is closed, and consider the vector of minimum norm (projection of the origin, cf. Prop. 1.1.9) in $C_2 - C_1$. This vector is of the form $\overline{x}_2 - \overline{x}_1$, where $\overline{x}_1 \in C_1$ and $\overline{x}_2 \in C_2$. Let

$$a = \frac{\overline{x}_2 - \overline{x}_1}{2}, \qquad \overline{x} = \frac{\overline{x}_1 + \overline{x}_2}{2}, \qquad b = a'\overline{x};$$

[cf. Fig. 1.5.4(b)]. Then, $a \neq 0$, since $\overline{x}_1 \in C_1$, $\overline{x}_2 \in C_2$, and C_1 and C_2 are disjoint. We will show that the hyperplane

$$\{x \mid a'x = b\}$$

strictly separates C_1 and C_2, i.e., that

$$a'x_1 < b < a'x_2, \qquad \forall\ x_1 \in C_1,\ \forall\ x_2 \in C_2. \tag{1.30}$$

To this end we note that $\overline{x}_1$ is the projection of $\overline{x}_2$ on $\mathrm{cl}(C_1)$ (otherwise there would exist a vector $x_1 \in C_1$ with $\|\overline{x}_2 - x_1\| < \|\overline{x}_2 - \overline{x}_1\|$ - a contradiction of the minimum norm property of $\overline{x}_2 - \overline{x}_1$). Thus, we have

$$(\overline{x}_2 - \overline{x}_1)'(x_1 - \overline{x}_1) \leq 0, \qquad \forall\ x_1 \in C_1,$$

or equivalently, since $\overline{x} - \overline{x}_1 = a$,

$$a'x_1 \leq a'\overline{x}_1 = a'\overline{x} + a'(\overline{x}_1 - \overline{x}) = b - \|a\|^2 < b, \qquad \forall\ x_1 \in C_1.$$

Thus, the left-hand side of Eq. (1.30) is proved. The right-hand side is proved similarly. **Q.E.D.**

Note that as a corollary of the preceding proposition, a closed set C can be strictly separated from a vector $\overline{x} \notin C$, i.e., from the singleton set $\{\overline{x}\}$. We will use this fact to provide the following important characterization of closed convex sets.

> **Proposition 1.5.4:** The closure of the convex hull of a set C is the intersection of the closed halfspaces that contain C. In particular, a closed convex set is the intersection of the closed halfspaces that contain it.

Proof: Let H denote the intersection of all closed halfspaces that contain C. Since every closed halfspace containing C must also contain $\text{cl}\big(\text{conv}(C)\big)$, it follows that $H \supset \text{cl}\big(\text{conv}(C)\big)$.

To show the reverse inclusion, consider a vector $x \notin \text{cl}\big(\text{conv}(C)\big)$ and a hyperplane strictly separating x and $\text{cl}\big(\text{conv}(C)\big)$. The corresponding closed halfspace that contains $\text{cl}\big(\text{conv}(C)\big)$ does not contain x, so $x \notin H$. Hence $H \subset \text{cl}\big(\text{conv}(C)\big)$. **Q.E.D.**

1.5.2 Proper Hyperplane Separation

We now discuss a form of hyperplane separation, called *proper*, which turns out to be useful in some important optimization contexts, such as the duality theorems of Chapter 4 (Props. 4.4.1 and 4.5.1).

Let C_1 and C_2 be two subsets of $\Re^n$. We say that a hyperplane *properly separates C_1 and C_2* if it separates C_1 and C_2, and does not fully contain both C_1 and C_2. Thus there exists a hyperplane that properly separates C_1 and C_2 if and only if there is a vector a such that

$$\sup_{x_1 \in C_1} a'x_1 \leq \inf_{x_2 \in C_2} a'x_2, \qquad \inf_{x_1 \in C_1} a'x_1 < \sup_{x_2 \in C_2} a'x_2;$$

(see Fig. 1.5.5). If C is a subset of $\Re^n$ and $\overline{x}$ is a vector in $\Re^n$, we say that a hyperplane *properly separates C and $\overline{x}$* if it properly separates C and the singleton set $\{\overline{x}\}$.

Note that a convex set in $\Re^n$ that has nonempty interior (and hence has dimension n) cannot be fully contained in a hyperplane (which has dimension $n-1$). Thus, in view of the Separating Hyperplane Theorem (Prop. 1.5.2), two disjoint convex sets one of which has nonempty interior can be properly separated. Similarly and more generally, two disjoint convex sets such that the affine hull of their union has dimension n can be properly separated. Figure 1.5.5(c) provides an example of two convex sets that cannot be properly separated.

The existence of a hyperplane that properly separates two convex sets is intimately tied to conditions involving the relative interiors of the sets.

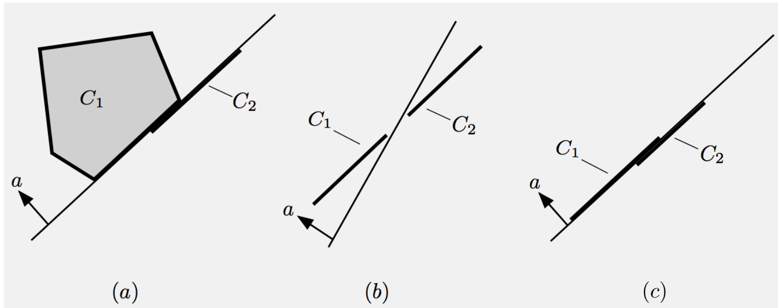

Figure 1.5.5. (a) and (b) Illustration of a properly separating hyperplanes. (c) Illustration of two convex sets that cannot be properly separated.

An important fact in this connection is that given a nonempty convex set C and a hyperplane H that contains C in one of its closed halfspaces, we have

$$C \subset H \qquad \text{if and only if} \qquad \text{ri}(C) \cap H \neq \emptyset. \tag{1.31}$$

To see this, let H be of the form $\{x \mid a'x = b\}$ with $a'x \geq b$ for all $x \in C$. Then for a vector $\overline{x} \in \text{ri}(C)$, we have $\overline{x} \in H$ if and only if $a'\overline{x} = b$, i.e., $a'x$ attains its minimum over C at $\overline{x}$. By Prop. 1.3.4, this is so if and only if $a'x = b$ for all $x \in C$, i.e., $C \subset H$.

The following propositions provide relative interior assumptions that guarantee the existence of properly separating hyperplanes.

Proposition 1.5.5: (Proper Separation Theorem) Let C be a nonempty convex subset of $\Re^n$ and let $\overline{x}$ be a vector in $\Re^n$. There exists a hyperplane that properly separates C and $\overline{x}$ if and only if $\overline{x} \notin \text{ri}(C)$.

Proof: Suppose that there exists a hyperplane H that properly separates C and $\overline{x}$. Then either $\overline{x} \notin H$, in which case $\overline{x} \notin C$ and $\overline{x} \notin \text{ri}(C)$, or else $\overline{x} \in H$ and C is not contained in H, in which case by Eq. (1.31), $\text{ri}(C) \cap H = \emptyset$, in which case again $\overline{x} \notin \text{ri}(C)$.

Conversely, assume that $\overline{x} \notin \text{ri}(C)$. To show the existence of a properly separating hyperplane, we consider two cases (see Fig. 1.5.6):

(a) $\overline{x} \notin \text{aff}(C)$. In this case, since $\text{aff}(C)$ is closed and convex, by the Strict Separation Theorem [Prop. 1.5.3 under condition (2)] there exists a hyperplane that separates $\{\overline{x}\}$ and $\text{aff}(C)$ strictly, and hence also properly separates C and $\overline{x}$.

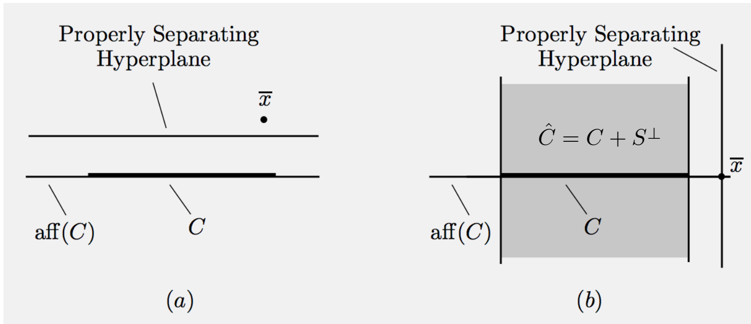

Figure 1.5.6. Illustration of the construction of a hyperplane that properly separates a convex set C and a point $\overline{x} \notin \text{ri}(C)$ (cf. the proof of Prop. 1.5.5). In case (a), where $\overline{x} \notin \text{aff}(C)$, the hyperplane is constructed as shown. In case (b), where $\overline{x} \in \text{aff}(C)$, we consider the subspace S that is parallel to $\text{aff}(C)$, we set $\hat{C} = C + S^{\perp}$, and we use the Supporting Hyperplane Theorem to separate $\overline{x}$ from $\hat{C}$ (Prop. 1.5.1).

(b) $\overline{x} \in \text{aff}(C)$. In this case, let S be the subspace that is parallel to $\text{aff}(C)$, and consider the set $\hat{C} = C + S^{\perp}$. From Prop. 1.3.7, we have $\text{ri}(\hat{C}) = \text{ri}(C) + S^{\perp}$, so that $\overline{x}$ is not an interior point of $\hat{C}$ [otherwise there must exist a vector $x \in \text{ri}(C)$ such that $x - \overline{x} \in S^{\perp}$, which, since $x \in \text{aff}(C)$, $\overline{x} \in \text{aff}(C)$, and $x - \overline{x} \in S$, implies that $x - \overline{x} = 0$, thereby contradicting the hypothesis $\overline{x} \notin \text{ri}(C)$]. By the Supporting Hyperplane Theorem (Prop. 1.5.1), it follows that there exists a vector $a \neq 0$ such that $a'x \geq a'\overline{x}$ for all $x \in \hat{C}$. Since $\hat{C}$ has nonempty interior, $a'x$ cannot be constant over $\hat{C}$, and

$$a'\overline{x} < \sup_{x \in \hat{C}} a'x = \sup_{x \in C,\, z \in S^{\perp}} a'(x + z) = \sup_{x \in C} a'x + \sup_{z \in S^{\perp}} a'z. \qquad (1.32)$$

If we had $a'\overline{z} \neq 0$ for some $\overline{z} \in S^{\perp}$, we would also have

$$\inf_{\alpha \in \Re} a'(x + \alpha\overline{z}) = -\infty,$$

which contradicts the fact $a'(x + z) \geq a'\overline{x}$ for all $x \in C$ and $z \in S^{\perp}$. It follows that

$$a'z = 0, \qquad \forall\ z \in S^{\perp},$$

which when combined with Eq. (1.32), yields

$$a'\overline{x} < \sup_{x \in C} a'x.$$

Thus the hyperplane $\{x \mid a'x = a'\overline{x}\}$ properly separates C and $\overline{x}$. **Q.E.D.**

Proposition 1.5.6: (Proper Separation of Two Convex Sets) Let C_1 and C_2 be two nonempty convex subsets of $\Re^n$. There exists a hyperplane that properly separates C_1 and C_2 if and only if

$$\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2) = \emptyset.$$

Proof: Consider the convex set $C = C_2 - C_1$. By Prop. 1.3.7, we have

$$\mathrm{ri}(C) = \mathrm{ri}(C_2) - \mathrm{ri}(C_1),$$

so the assumption $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2) = \emptyset$ is equivalent to $0 \notin \mathrm{ri}(C)$. By using Prop. 1.5.5, it follows that there exists a hyperplane properly separating C and the origin, so we have

$$0 \le \inf_{x_1 \in C_1,\, x_2 \in C_2} a'(x_2 - x_1), \qquad 0 < \sup_{x_1 \in C_1,\, x_2 \in C_2} a'(x_2 - x_1),$$

if and only if $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2) = \emptyset$. This is equivalent to the desired assertion. **Q.E.D.**

The following proposition is a variant of Prop. 1.5.6. It shows that if C_2 is polyhedral and the slightly stronger condition $\mathrm{ri}(C_1) \cap C_2 = \emptyset$ holds, then there exists a properly separating hyperplane satisfying the extra restriction that it does not contain the nonpolyhedral set C_1 (rather than just the milder requirement that it does not contain either C_1 or C_2); see Fig. 1.5.7.

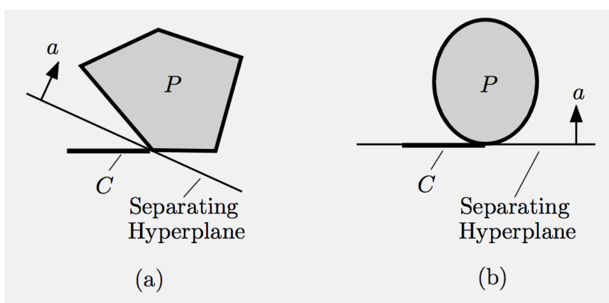

Figure 1.5.7. Illustration of the special proper separation property of a convex set C and a polyhedral set P, under the condition $\mathrm{ri}(C) \cap P = \emptyset$. In figure (a), the separating hyperplane can be chosen so that it does not contain C. If P is not polyhedral, as in figure (b), this may not be possible.

Proposition 1.5.7: (Polyhedral Proper Separation Theorem) Let C and P be two nonempty convex subsets of $\Re^n$ such that P is polyhedral. There exists a hyperplane that separates C and P, and does not contain C if and only if

$$\mathrm{ri}(C) \cap P = \emptyset.$$

Proof: First, as a general observation, we recall from our discussion of proper separation that given a convex set X and a hyperplane H that contains X in one of its closed halfspaces, we have

$$X \subset H \qquad \text{if and only if} \qquad \mathrm{ri}(X) \cap H \neq \emptyset; \tag{1.33}$$

cf. Eq. (1.31). We will use repeatedly this relation in the subsequent proof.

Assume that there exists a hyperplane H that separates C and P, and does not contain C. Then, by Eq. (1.33), H cannot contain a point in $\mathrm{ri}(C)$, and since H separates C and P, we must have $\mathrm{ri}(C) \cap P = \emptyset$.

Conversely, assume that $\mathrm{ri}(C) \cap P = \emptyset$. We will show that there exists a separating hyperplane that does not contain C. Denote

$$D = P \cap \mathrm{aff}(C).$$

If $D = \emptyset$, then since $\mathrm{aff}(C)$ and P are polyhedral, the Strict Separation Theorem [cf. Prop. 1.5.3 under condition (3)] applies and shows that there exists a hyperplane H that separates $\mathrm{aff}(C)$ and P strictly, and hence does not contain C.

We may thus assume that $D \neq \emptyset$. The idea now is to first construct a hyperplane that properly separates C and D, and then extend this hyperplane so that it suitably separates C and P. [If C had nonempty interior, the proof would be much simpler, since then $\mathrm{aff}(C) = \Re^n$ and $D = P$.]

By assumption, we have $\mathrm{ri}(C) \cap P = \emptyset$ implying that

$$\mathrm{ri}(C) \cap \mathrm{ri}(D) \subset \mathrm{ri}(C) \cap \big(P \cap \mathrm{aff}(C)\big) = \big(\mathrm{ri}(C) \cap P\big) \cap \mathrm{aff}(C) = \emptyset.$$

Hence, by Prop. 1.5.6, there exists a hyperplane H that properly separates C and D. Furthermore, H does not contain C, since if it did, H would also contain $\mathrm{aff}(C)$ and hence also D, contradicting the proper separation property. Thus, C is contained in one of the closed halfspaces of H, but not in both. Let $\overline{C}$ be the intersection of $\mathrm{aff}(C)$ and the closed halfspace of H that contains C; see Fig. 1.5.8. Note that H does not contain $\overline{C}$ (since H does not contain C), and by Eq. (1.33), we have $H \cap \mathrm{ri}(\overline{C}) = \emptyset$, implying that

$$P \cap \mathrm{ri}(\overline{C}) = \emptyset,$$

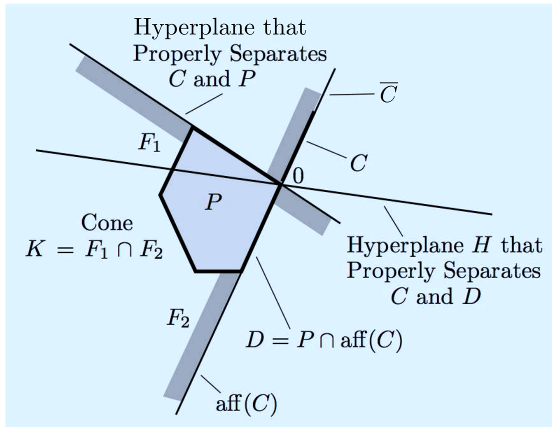

Figure 1.5.8. Illustration of the proof of Prop. 1.5.7 in the case where $D = P \cap \text{aff}(C) \neq \emptyset$. The figure shows the construction of a hyperplane that properly separates C and P, and does not contain C, starting from the hyperplane H that properly separates C and D. In this two-dimensional example we have $M = \{0\}$, so $K = \text{cone}(P) + M = \text{cone}(P)$.

[if $\overline{x} \in P \cap \text{ri}(\overline{C})$ then $\overline{x} \in D \cap \text{ri}(\overline{C})$, a contradiction since D and $\text{ri}(\overline{C})$ lie in the opposite closed halfspaces of H and $H \cap \text{ri}(\overline{C}) = \emptyset$].

If $P \cap \overline{C} = \emptyset$, then by using again the Strict Separation Theorem [cf. Prop. 1.5.3 under condition (3)], we can construct a hyperplane that strictly separates P and $\overline{C}$. This hyperplane also strictly separates P and C, and we are done. We thus assume that $P \cap \overline{C} \neq \emptyset$, and by using a translation argument if necessary, we assume that

$$0 \in P \cap \overline{C},$$

as indicated in Fig. 1.5.8. The polyhedral set P can be represented as the intersection of halfspaces $\{x \mid a_j'x \leq b_j\}$ with $b_j \geq 0$ (since $0 \in P$) and with $b_j = 0$ for at least one j [since otherwise 0 would be in the interior of P; then, by the Line Segment Principle, for any $\overline{x} \in \text{ri}(\overline{C})$ the line segment connecting 0 and $\overline{x}$ contains points in $\text{ri}(D) \cap \text{ri}(\overline{C})$, a contradiction of the fact that H properly separates D and $\overline{C}$]. Thus, we have

$$P = \{x \mid a_j'x \leq 0,\ j = 1, \ldots, m\} \cap \{x \mid a_j'x \leq b_j,\ j = m+1, \ldots, \overline{m}\},$$

for some integers $m \geq 1$ and $\overline{m} \geq m$, vectors a_j, and scalars $b_j > 0$.

Let M be the relative boundary of $\overline{C}$, i.e.,

$$M = H \cap \text{aff}(C),$$

and consider the cone

$$K = \{x \mid a_j'x \leq 0,\ j = 1, \ldots, m\} + M.$$

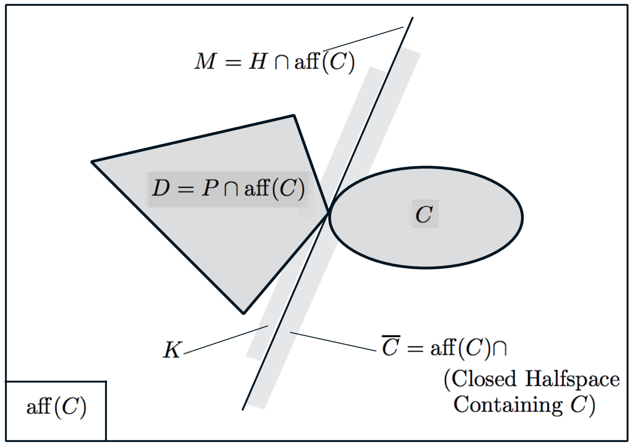

Figure 1.5.9. View of D and C within aff(C), and illustration of the construction of the cone $K = \text{cone}(P) + M$ in the proof of Prop. 1.5.7.

Note that $K = \text{cone}(P) + M$ (see Figs. 1.5.8 and 1.5.9).

We claim that $K \cap \text{ri}(\overline{C}) = \emptyset$. The proof is by contradiction. If there exists $\overline{x} \in K \cap \text{ri}(\overline{C})$, then $\overline{x}$ can be expressed as $\overline{x} = \alpha w + v$ for some $\alpha > 0$, $w \in P$, and $v \in M$ [since $K = \text{cone}(P) + M$ and $0 \in P$], so that $w = (\overline{x}/\alpha) - (v/\alpha) \in P$. On the other hand, since $\overline{x} \in \text{ri}(\overline{C})$, $0 \in \overline{C} \cap M$, and M is a subset of the lineality space of $\overline{C}$ [and hence also of the lineality space of $\text{ri}(\overline{C})$], all vectors of the form $\overline{\alpha}\,\overline{x} + \overline{v}$, with $\overline{\alpha} > 0$ and $\overline{v} \in M$, belong to $\text{ri}(\overline{C})$. In particular the vector $w = (\overline{x}/\alpha) - (v/\alpha)$ belongs to $\text{ri}(\overline{C})$, so $w \in P \cap \text{ri}(\overline{C})$. This is a contradiction since $P \cap \text{ri}(\overline{C}) = \emptyset$, and it follows that $K \cap \text{ri}(\overline{C}) = \emptyset$.

The cone K is polyhedral (since it is the vector sum of two polyhedral sets), so it is the intersection of some closed halfspaces $F_1, \ldots, F_r$ that pass through 0 (cf. Fig. 1.5.8). Since $K = \text{cone}(P) + M$, each of these closed halfspaces contains M, the relative boundary of the set $\overline{C}$, and furthermore $\overline{C}$ is the closed half of a subspace. It follows that if any of the closed halfspaces $F_1, \ldots, F_r$ contains a vector in $\text{ri}(\overline{C})$, then that closed halfspace entirely contains $\overline{C}$. Hence, since K does not contain any point in $\text{ri}(\overline{C})$, at least one of $F_1, \ldots, F_r$, say F_1, does not contain any point in $\text{ri}(\overline{C})$ (cf. Fig. 1.5.8). Therefore, the hyperplane corresponding to F_1 contains no points of $\text{ri}(\overline{C})$, and hence also no points of $\text{ri}(C)$. Thus, this hyperplane does not contain C, while separating K and C. Since K contains P, this hyperplane also separates P and C. **Q.E.D.**

Note that in the preceding proof, it is essential to introduce M, the relative boundary of the set $\overline{C}$, and to define $K = \text{cone}(P) + M$. If instead we define $K = \text{cone}(P)$, then the corresponding halfspaces $F_1, \ldots, F_r$ may all intersect $\mathrm{ri}(\overline{C})$, and the proof argument fails (see Fig. 1.5.9).

1.5.3 Nonvertical Hyperplane Separation

In the context of optimization, supporting hyperplanes are often used in conjunction with epigraphs of functions defined on $\Re^n$. Since the epigraph is a subset of $\Re^{n+1}$, we consider hyperplanes in $\Re^{n+1}$ and associate them with nonzero vectors of the form (μ, β), where $\mu \in \Re^n$ and $\beta \in \Re$. We say that such a hyperplane is *vertical* if $\beta = 0$.

Note that if a hyperplane with normal (μ, β) is nonvertical, then it crosses the $(n+1)$st axis (the axis associated with w) at a unique point. In particular, if $(\overline{u}, \overline{w})$ is any vector on the hyperplane, the crossing point has the form $(0, \xi)$, where

$$\xi = \frac{\mu'}{\beta}\overline{u} + \overline{w},$$

since from the hyperplane equation, we have $(0, \xi)'(\mu, \beta) = (\overline{u}, \overline{w})'(\mu, \beta)$. If the hyperplane is vertical, it either contains the entire $(n+1)$st axis, or else it does not cross it at all; see Fig. 1.5.10. Furthermore, a hyperplane H is vertical if and only if the recession cone of H, as well as the recession cones of the closed halfspaces associated with H, contain the $(n+1)$st axis.

Vertical lines in $\Re^{n+1}$ are sets of the form $\big\{(\overline{u}, w) \mid w \in \Re\big\}$, where $\overline{u}$ is a fixed vector in $\Re^n$. If $f : \Re^n \mapsto (-\infty, \infty]$ is a proper convex function, then $\mathrm{epi}(f)$ cannot contain a vertical line, and it appears plausible that $\mathrm{epi}(f)$ is contained in a closed halfspace corresponding to some nonvertical hyperplane. We prove this fact in greater generality in the following proposition, which will be important for the development of duality.

Proposition 1.5.8: (Nonvertical Hyperplane Theorem) Let C be a nonempty convex subset of $\Re^{n+1}$ that contains no vertical lines. Let the vectors in $\Re^{n+1}$ be denoted by (u, w), where $u \in \Re^n$ and $w \in \Re$. Then:

(a) C is contained in a closed halfspace corresponding to a nonvertical hyperplane, i.e., there exist a vector $\mu \in \Re^n$, a scalar $\beta \neq 0$, and a scalar γ such that

$$\mu' u + \beta w \geq \gamma, \qquad \forall\ (u, w) \in C.$$

(b) If $(\overline{u}, \overline{w})$ does not belong to $\mathrm{cl}(C)$, there exists a nonvertical hyperplane strictly separating $(\overline{u}, \overline{w})$ and C.

Proof: (a) Assume, to arrive at a contradiction, that every hyperplane containing C in one of its closed halfspaces is vertical. Then every hyper-

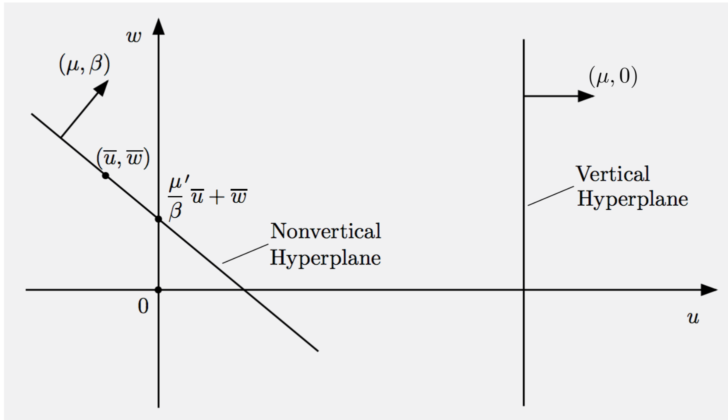

Figure 1.5.10. Illustration of vertical and nonvertical hyperplanes in $\Re^{n+1}$. A hyperplane with normal (μ,β) is nonvertical if $\beta\neq 0$, or, equivalently, if it intersects the $(n+1)$st axis at the unique point $\xi=(\mu/\beta)'\overline{u}+\overline{w}$, where $(\overline{u},\overline{w})$ is any vector on the hyperplane.

plane containing $\mathrm{cl}(C)$ in one of its closed halfspaces must also be vertical and its recession cone must contain the $(n+1)$st axis. By Prop. 1.5.4, $\mathrm{cl}(C)$ is the intersection of all closed halfspaces that contain it, so its recession cone contains the $(n+1)$st axis. Since the recession cones of $\mathrm{cl}(C)$ and $\mathrm{ri}(C)$ coincide [cf. Prop. 1.4.2(b)], for every $(\overline{u},\overline{w})\in\mathrm{ri}(C)$, the vertical line $\big\{(\overline{u},w)\mid w\in\Re\big\}$ belongs to $\mathrm{ri}(C)$ and hence to C. This contradicts the assumption that C does not contain a vertical line.

(b) If $(\overline{u},\overline{w})\notin\mathrm{cl}(C)$, then there exists a hyperplane strictly separating $(\overline{u},\overline{w})$ and $\mathrm{cl}(C)$ [cf. Prop. 1.5.3 under condition (2)]. If this hyperplane is nonvertical, since $C\subset\mathrm{cl}(C)$, we are done, so assume otherwise. Then, we have a nonzero vector $\overline{\mu}$ and a scalar $\overline{\gamma}$ such that

$$\overline{\mu}'u>\overline{\gamma}>\overline{\mu}'\,\overline{u},\qquad\forall\ (u,w)\in\mathrm{cl}(C).\tag{1.34}$$

The idea now is to combine this vertical hyperplane with a suitable nonvertical hyperplane in order to construct a nonvertical hyperplane that strictly separates $(\overline{u},\overline{w})$ from $\mathrm{cl}(C)$ (see Fig. 1.5.11).

Since, by assumption, C does not contain a vertical line, $\mathrm{ri}(C)$ also does not contain a vertical line. Since the recession cones of $\mathrm{cl}(C)$ and $\mathrm{ri}(C)$ coincide [cf. Prop. 1.4.2(b)], it follows that $\mathrm{cl}(C)$ does not contain a vertical line. Hence, by part (a), there exists a nonvertical hyperplane containing $\mathrm{cl}(C)$ in one of its closed halfspaces, so that for some (μ,β) and γ, with $\beta\neq 0$, we have

$$\mu'u+\beta w\geq\gamma,\qquad\forall\ (u,w)\in\mathrm{cl}(C).$$

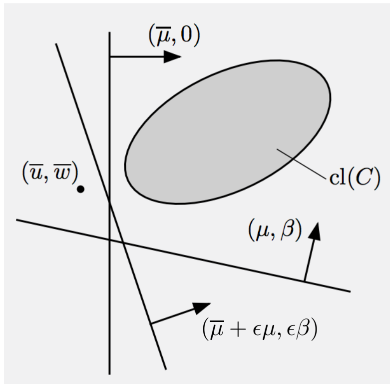

Figure 1.5.11. Construction of a strictly separating nonvertical hyperplane in the proof of Prop. 1.5.8(b).

By multiplying this relation with an $\epsilon > 0$ and combining it with Eq. (1.34), we obtain

$$(\overline{\mu} + \epsilon\mu)'u + \epsilon\beta w > \overline{\gamma} + \epsilon\gamma, \qquad \forall\ (u, w) \in \mathrm{cl}(C),\ \forall\ \epsilon > 0.$$

Since $\overline{\gamma} > \overline{\mu}'\,\overline{u}$, there is a small enough ϵ such that

$$\overline{\gamma} + \epsilon\gamma > (\overline{\mu} + \epsilon\mu)'\overline{u} + \epsilon\beta\overline{w}.$$

From the above two relations, we obtain

$$(\overline{\mu} + \epsilon\mu)'u + \epsilon\beta w > (\overline{\mu} + \epsilon\mu)'\overline{u} + \epsilon\beta\overline{w}, \qquad \forall\ (u, w) \in \mathrm{cl}(C),$$

implying that there is a nonvertical hyperplane with normal $(\overline{\mu} + \epsilon\mu, \epsilon\beta)$ that strictly separates $(\overline{u}, \overline{w})$ and cl(C). Since $C \subset \mathrm{cl}(C)$, this hyperplane also strictly separates $(\overline{u}, \overline{w})$ and C. **Q.E.D.**

1.6 CONJUGATE FUNCTIONS

We will now develop a concept that is fundamental in convex optimization. This is the conjugacy transformation, which associates with any function f, a convex function, called the conjugate of f. The idea here is to describe f in terms of the affine functions that are majorized by f. When f is closed proper convex, we will show that the description is accurate and the transformation is symmetric, i.e., f can be recovered by taking the conjugate of the conjugate of f. The conjugacy transformation thus provides an alternative view of a convex function, which often reveals interesting properties, and is useful for analysis and computation.

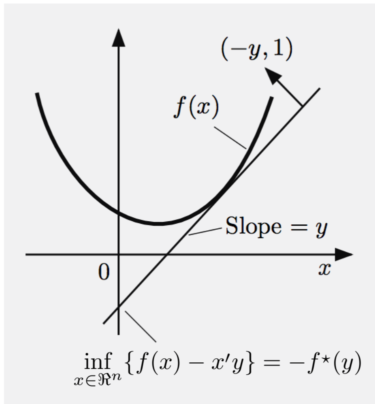

Figure 1.6.1. Visualization of the conjugate function

$$f^\star(y) = \sup_{x\in\Re^n}\big\{x'y - f(x)\big\}$$

of a function f. The crossing point of the vertical axis with the hyperplane that has normal $(-y,1)$ and passes through a point $\big(\overline{x}, f(\overline{x})\big)$ on the graph of f is

$$f(\overline{x}) - \overline{x}'y.$$

Thus, the crossing point corresponding to the hyperplane that supports the epigraph of f is

$$\inf_{x\in\Re^n}\{f(x) - x'y\},$$

which by definition is equal to $-f^\star(y)$.

Consider an extended real-valued function $f : \Re^n \mapsto [-\infty,\infty]$. The *conjugate function* of f is the function $f^\star : \Re^n \mapsto [-\infty,\infty]$ defined by

$$f^\star(y) = \sup_{x\in\Re^n}\big\{x'y - f(x)\big\}, \qquad y \in \Re^n. \tag{1.35}$$

Figure 1.6.1 provides a geometrical interpretation of the definition.

Note that regardless of the structure of f, the conjugate $f^\star$ is a closed convex function, since it is the pointwise supremum of the collection of affine functions

$$x'y - f(x), \qquad \forall\ x \text{ such that } f(x) \text{ is finite},$$

(Prop. 1.1.6). Note also that $f^\star$ need not be proper, even if f is. We will show, however, that in the case where f is convex, $f^\star$ is proper if and only if f is.

Figure 1.6.2 shows some examples of conjugate functions. In this figure, all the functions are closed proper convex, and it can be verified that the conjugate of the conjugate yields the original function. This is a manifestation of a result that we will show shortly.

For a function $f : \Re^n \mapsto [-\infty,\infty]$, consider the conjugate of the conjugate function $f^\star$ (or *double conjugate*). It is denoted by $f^{\star\star}$, it is given by

$$f^{\star\star}(x) = \sup_{y\in\Re^n}\big\{y'x - f^\star(y)\big\}, \qquad x \in \Re^n,$$

and it can be constructed as shown in Fig. 1.6.3. As this figure suggests, and part (d) of the following proposition shows, by constructing $f^{\star\star}$, we

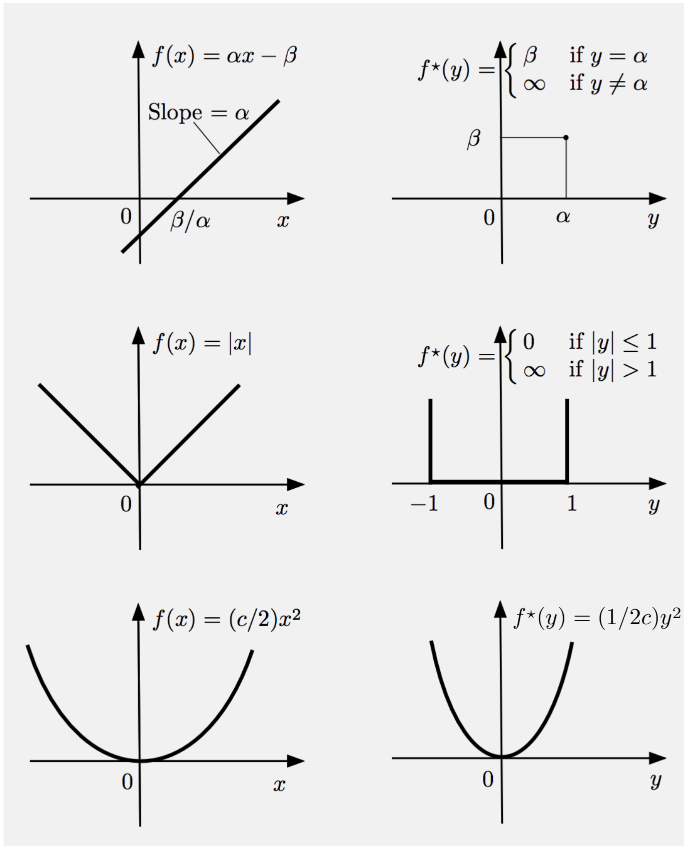

Figure 1.6.2. Some examples of conjugate functions. It can be verified that in each case, the conjugate of the conjugate is the original, i.e., the conjugates of the functions on the right are the corresponding functions on the left.

typically obtain the convex closure of f [the function that has as epigraph the closure of the convex hull of epi(f); cf. Section 1.3.3]. In particular, part (c) of the proposition shows that if f is closed proper convex, then $f^{\star\star} = f$.

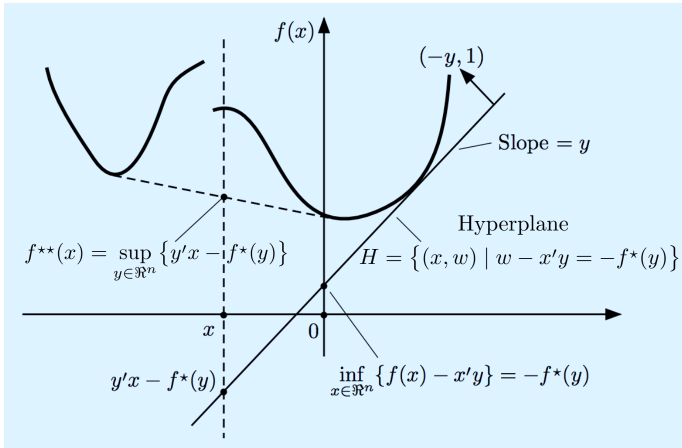

Figure 1.6.3. Visualization of the double conjugate (conjugate of the conjugate)

$$f^{\star\star}(x) = \sup_{y\in\Re^n}\big\{y'x - f^\star(y)\big\}$$

of a function f, where $f^\star$ is the conjugate of f,

$$f^\star(y) = \sup_{x\in\Re^n}\big\{x'y - f(x)\big\}.$$

For each $x \in \Re^n$, we consider the vertical line in $\Re^{n+1}$ that passes through the point $(x,0)$, and for each y in the effective domain of $f^\star$, we consider the crossing point of this line with the hyperplane with normal $(-y,1)$ that supports the graph of f [and therefore passes through the point $(0,-f^\star(y))$]. This crossing point is $y'x - f^\star(y)$, so $f^{\star\star}(x)$ is equal to the highest crossing level. As the figure indicates (and Prop. 1.6.1 shows), the double conjugate $f^{\star\star}$ is the convex closure of f (barring the exceptional case where the convex closure takes the value $-\infty$ at some point, in which case the figure above is not valid).

Proposition 1.6.1: (Conjugacy Theorem) Let $f : \Re^n \mapsto [-\infty,\infty]$ be a function, let $f^\star$ be its conjugate, and consider the double conjugate $f^{\star\star}$. Then:

(a) We have

$$f(x) \geq f^{\star\star}(x), \qquad \forall\ x \in \Re^n.$$

(b) If f is convex, then properness of any one of the functions f, $f^\star$, and $f^{\star\star}$ implies properness of the other two.

(c) If f is closed proper convex, then

$$f(x) = f^{\star\star}(x), \qquad \forall\, x \in \Re^n.$$

(d) The conjugates of f and its convex closure $\check{\text{cl}}\, f$ are equal. Furthermore, if $\check{\text{cl}}\, f$ is proper, then

$$(\check{\text{cl}}\, f)(x) = f^{\star\star}(x), \qquad \forall\, x \in \Re^n.$$

Proof: (a) For all x and y, we have

$$f^\star(y) \geq x'y - f(x),$$

so that

$$f(x) \geq x'y - f^\star(y), \qquad \forall\, x, y \in \Re^n.$$

Hence

$$f(x) \geq \sup_{y \in \Re^n} \big\{ x'y - f^\star(y) \big\} = f^{\star\star}(x), \qquad \forall\, x \in \Re^n.$$

(b) Assume that f is proper, in addition to being convex. Then its epigraph is nonempty and convex, and contains no vertical line. By applying the Nonvertical Hyperplane Theorem [Prop. 1.5.8(a)], with C being the set $\text{epi}(f)$, it follows that there exists a nonvertical hyperplane with normal $(y, 1)$ that contains $\text{epi}(f)$ in its positive halfspace. In particular, this implies the existence of a vector y and a scalar c such that

$$y'x + f(x) \geq c, \qquad \forall\, x \in \Re^n.$$

We thus obtain

$$f^\star(-y) = \sup_{x \in \Re^n} \big\{ -y'x - f(x) \big\} \leq -c,$$

so that $f^\star$ is not identically equal to ∞. Also, by the properness of f, there exists a vector $\overline{x}$ such that $f(\overline{x})$ is finite. For every $y \in \Re^n$, we have

$$f^\star(y) \geq y'\overline{x} - f(\overline{x}),$$

so $f^\star(y) > -\infty$ for all $y \in \Re^n$. Thus, $f^\star$ is proper.

Conversely, assume that $f^\star$ is proper. The preceding argument shows that properness of $f^\star$ implies properness of its conjugate, $f^{\star\star}$, so that $f^{\star\star}(x) > -\infty$ for all $x \in \Re^n$. By part (a), $f(x) \geq f^{\star\star}(x)$, so $f(x) > -\infty$

for all $x \in \Re^n$. Also, f cannot be identically equal to ∞, since then by its definition, $f^\star$ would be identically equal to $-\infty$. Thus f is proper.

We have thus shown that a convex function is proper if and only if its conjugate is proper, and the result follows in view of the conjugacy relations between f, $f^\star$, and $f^{\star\star}$.

(c) We will apply the Nonvertical Hyperplane Theorem (Prop. 1.5.8), with C being the closed and convex set $\text{epi}(f)$, which contains no vertical line since f is proper. Let (x, γ) belong to $\text{epi}(f^{\star\star})$, i.e., $x \in \text{dom}(f^{\star\star})$, $\gamma \geq f^{\star\star}(x)$, and suppose, to arrive at a contradiction, that (x, γ) does not belong to $\text{epi}(f)$. Then by Prop. 1.5.8(b), there exists a nonvertical hyperplane with normal (y, ζ), where $\zeta \neq 0$, and a scalar c such that

$$y'z + \zeta w < c < y'x + \zeta\gamma, \qquad \forall\ (z, w) \in \text{epi}(f).$$

Since w can be made arbitrarily large, we have $\zeta < 0$, and without loss of generality, we can take $\zeta = -1$, so that

$$y'z - w < c < y'x - \gamma, \qquad \forall\ (z, w) \in \text{epi}(f).$$

Since $\gamma \geq f^{\star\star}(x)$ and $\big(z, f(z)\big) \in \text{epi}(f)$ for all $z \in \text{dom}(f)$, we obtain

$$y'z - f(z) < c < y'x - f^{\star\star}(x), \qquad \forall\ z \in \text{dom}(f).$$

Hence

$$\sup_{z \in \Re^n} \big\{y'z - f(z)\big\} \leq c < y'x - f^{\star\star}(x),$$

or

$$f^\star(y) < y'x - f^{\star\star}(x),$$

which contradicts the definition $f^{\star\star}(x) = \sup_{y \in \Re^n}\big\{y'x - f^\star(y)\big\}$. Thus, we have $\text{epi}(f^{\star\star}) \subset \text{epi}(f)$, which implies that $f(x) \leq f^{\star\star}(x)$ for all $x \in \Re^n$. This, together with part (a), shows that $f^{\star\star}(x) = f(x)$ for all x.

(d) Let $\check{f}^\star$ be the conjugate of $\check{\text{cl}}\, f$. For any y, $-f^\star(y)$ and $-\check{f}^\star(y)$ are the supremum crossing levels of the vertical axis with the hyperplanes with normal $(-y, 1)$ that contain the sets $\text{epi}(f)$ and $\text{cl}\big(\text{conv}\big(\text{epi}(f)\big)\big)$, respectively, in their positive closed halfspaces (cf. Fig. 1.6.1). Since the hyperplanes of this type are the same for the two sets, we have $f^\star(y) = \check{f}^\star(y)$ for all y. Thus, $f^{\star\star}$ is equal to the conjugate of $\check{f}^\star$, which is $\check{\text{cl}}\, f$ by part (c) when $\check{\text{cl}}\, f$ is proper. **Q.E.D.**

The properness assumptions on f and $\check{\text{cl}}\, f$ are essential for the validity of parts (c) and (d), respectively, of the preceding proposition. For an illustrative example, consider the closed convex (but improper) function

$$f(x) = \begin{cases} \infty & \text{if } x > 0, \\ -\infty & \text{if } x \leq 0. \end{cases} \tag{1.36}$$

We have $f = \check{\text{cl}}\, f$, and it can be verified that $f^\star(y) = \infty$ for all y and $f^{\star\star}(x) = -\infty$ for all x, so that $f \neq f^{\star\star}$ and $\check{\text{cl}}\, f \neq f^{\star\star}$.

For an example where f is proper (but not closed convex), while $\check{\text{cl}}\, f$ is improper and we have $\check{\text{cl}}\, f \neq f^{\star\star}$, let

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

Then $\check{\text{cl}}\, f$ is equal to the function (1.36), and $\check{\text{cl}}\, f \neq f^{\star\star}$.

The exceptional behavior in the preceding example can be attributed to a subtle difference in the constructions of the conjugate function and the convex closure: while the conjugate functions $f^\star$ and $f^{\star\star}$ are defined exclusively in terms of nonvertical hyperplanes via the construction of Fig. 1.6.3, the epigraph of the convex closure $\check{\text{cl}}\, f$ is defined in terms of nonvertical *and* vertical hyperplanes. This difference is inconsequential when there exists at least one nonvertical hyperplane containing the epigraph of f in one of its closed halfspaces [this is true in particular if $\check{\text{cl}}\, f$ is proper; see Props. 1.5.8(a) and 1.6.1(d)]. The reason is that, in this case, the epigraph of $\check{\text{cl}}\, f$ can equivalently be defined by using just nonvertical hyperplanes [this can be seen using Prop. 1.5.8(b)].

Example 1.6.1: (Indicator/Support Function Conjugacy)

Given a nonempty set X, consider the *indicator function* of X, defined by

$$\delta_X(x) = \begin{cases} 0 & \text{if } x \in X, \\ \infty & \text{if } x \notin X. \end{cases}$$

The conjugate of δ_X is given by

$$\sigma_X(y) = \sup_{x \in X} y'x$$

and is called the *support function of* X (see Fig. 1.6.4). By the generic closedness and convexity properties of conjugate functions, σ_X is closed and convex. It is also proper since X is nonempty, so that $\sigma_X(0) = 0$ (an improper closed convex function cannot take finite values; cf. the discussion at the end of Section 1.1.2). Furthermore, the sets X, $\text{cl}(X)$, $\text{conv}(X)$, and $\text{cl}\big(\text{conv}(X)\big)$ all have the same support function [cf. Prop. 1.6.1(d)].

Example 1.6.2: (Support Function of a Cone - Polar Cones)

Let C be a convex cone. By the preceding example, the conjugate of its indicator function δ_C is its support function,

$$\sigma_C(y) = \sup_{x \in C} y'x.$$

Since C is a cone, we see that

$$\sigma_C(y) = \begin{cases} 0 & \text{if } y'x \leq 0, \ \forall\, x \in C, \\ \infty & \text{otherwise.} \end{cases}$$

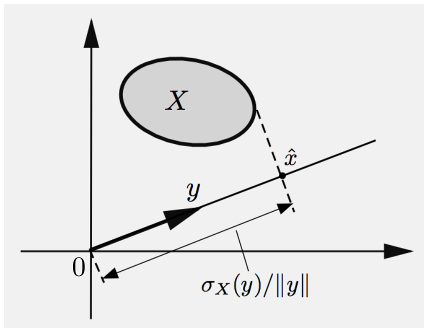

Figure 1.6.4. Visualization of the support function

$$\sigma_X(y) = \sup_{x \in X} y'x$$

of a set X. To determine the value $\sigma_X(y)$ for a given vector y, we project the set X on the line determined by y, and we find $\hat{x}$, the extreme point of projection in the direction y. Then

$$\sigma_X(y) = \|\hat{x}\| \cdot \|y\|.$$

Thus the support/conjugate function σ_C is the indicator function δ_{C^*} of the cone

$$C^* = \{y \mid y'x \leq 0, \ \forall \ x \in C\}, \tag{1.37}$$

called the *polar cone of* C. It follows that the conjugate of σ_C is the indicator function of the polar cone of C^*, and by the Conjugacy Theorem [Prop. 1.6.1(c)] it is also $\mathrm{cl}\,\delta_C$. Thus the polar cone of C^* is $\mathrm{cl}(C)$. In particular, if C is closed, the polar of its polar is equal to the original. This is a special case of the *Polar Cone Theorem*, which will be discussed in more detail in Section 2.2.

A special case of particular interest in when $C = \mathrm{cone}\big(\{a_1, \ldots, a_r\}\big)$, the cone generated by a finite set of vectors $a_1, \ldots, a_r$. Then it can be seen that

$$C^* = \{x \mid a_j'x \leq 0, \ j = 1, \ldots, r\}.$$

From this it follows that $(C^*)^* = C$ because C is a closed set [we have essentially shown this: $\mathrm{cone}\big(\{a_1, \ldots, a_r\}\big)$ is the image of the positive orthant $\{\alpha \mid \alpha \geq 0\}$ under the linear transformation that maps α to $\sum_{j=1}^r \alpha_j a_j$, and the image of any polyhedral set under a linear transformation is a closed set (see the discussion following the proof of Prop. 1.4.13)]. The assertion $(C^*)^* = C$ for the case $C = \mathrm{cone}\big(\{a_1, \ldots, a_r\}\big)$ is known as Farkas' Lemma and will be discussed further in what follows (Sections 2.3 and 5.1).

Let us finally note that we may define the polar cone of any set C via Eq. (1.37). However, unless C is a cone, the support function of C will not be an indicator function, nor will the relation $(C^*)^* = C$ hold. Instead, we will show in Section 2.2 that in general we have $(C^*)^* = \mathrm{cl}\big(\mathrm{cone}(C)\big)$.

1.7 SUMMARY

In this section, we discuss how the material of this chapter is used in subsequent chapters. First, let us note that we have aimed to develop in this chapter the part of convexity theory that is needed for the optimization and duality analysis of Chapters 3-5, and no more. We have developed the basic principles of polyhedral convexity in Chapter 2 for completeness and a broader perspective, but this material is not needed for the mathematical development of Chapters 3-5. Despite the fact that we focus only on the essential, we still cover considerable ground, and it may help the reader to know how the various topics of this chapter connect with each other, and how they are used later. We thus provide a summary and guide on a section-by-section basis:

Section 1.1: The definitions and results on convexity and closure (Sections 1.1.1-1.1.3) are very basic and should be read carefully. The material of Section 1.1.4 on differentiable convex functions, optimality conditions, and the projection theorem, is also basic. The reader may skip Prop. 1.1.10 on twice differentiable functions, which is used in nonessential ways later.

Section 1.2: The definitions of convex and affine hulls, and generated cones, as well as Caratheodory's theorem should also be read carefully. Proposition 1.2.2 on the compactness of the convex hull of a compact set is used only in the proof of Prop. 4.3.2, which is in turn used only in the specialized MC/MC framework of Section 5.7 on estimates of duality gap.

Section 1.3: The definitions and results on relative interior and closure up to and including Section 1.3.2 are pervasive in what follows. However, the somewhat tedious proofs of Props. 1.3.5-1.3.10 may be skipped at first reading. Similarly, the proof of continuity of a real-valued convex function (Prop. 1.3.11) is specialized and may be skipped. Regarding Section 1.3.3, it is important to understand the definitions, and gain intuition on closures and convex closures of functions, as they arise in the context of conjugacy. However, Props. 1.3.13-1.3.17 are used substantively later only for the development of minimax theory (Sections 3.4, 4.2.5, and 5.5) and the theory of directional derivatives (Section 5.4.4), which are themselves "terminal" and do not affect other sections.

Section 1.4: The material on directions of recession, up to and including Section 1.4.1, is very important for later developments, although the use of Props. 1.4.5-1.4.6 on recession functions is somewhat focused. In particular, Prop. 1.4.7 is used only to prove Prop. 1.4.8, and Props. 1.4.6-1.4.8 are used only for the development of existence of solutions criteria in Section 3.2. The set intersection analysis of Section 1.4.2, which may challenge some readers at first, is used for the development of some important theory: the existence of optimal solutions for linear and quadratic programs (Prop. 1.4.12), the criteria for preservations of closedness under linear transformations and vector sum (Props. 1.4.13 and 1.4.14), and the existence of solutions analysis of Chapter 3.

Section 1.5: The material on hyperplane separation is used extensively and should be read in its entirety. However, the long proof of the polyhedral proper separation theorem (Prop. 1.5.7, due to Rockafellar [Roc70], Th. 20.2) may be postponed for later. This theorem is important for our purposes. For example it is used (through Props. 4.5.1 and 4.5.2) to establish part of the Nonlinear Farkas' Lemma (Prop. 5.1.1), on which the constrained optimization and Fenchel duality theories rest.

Section 1.6: The material on conjugate functions and the Conjugacy Theorem (Prop. 1.6.1) is very basic. In our duality development, we use the theorem somewhat sparingly because we argue mostly in terms of the MC/MC framework of Chapter 4 and the Strong Duality Theorem (Prop. 4.3.1), which serves as an effective substitute.

Notes and Sources

There is a very extensive literature on convex analysis and optimization and it is beyond our scope to give a complete bibliography. We are providing instead a brief historical account and list some of the main textbooks in the field.

Among early classical works on convexity, we mention Caratheodory [Car11], Minkowski [Min11], and Steinitz [Ste13], [Ste14], [Ste16]. In particular, Caratheodory gave the theorem on convex hulls that carries his name, while Steinitz developed the theory of relative interiors and recession cones. Minkowski is credited with initiating the theory of hyperplane separation of convex sets and the theory of support functions (a precursor to conjugate convex functions). Furthermore, Minkowski and Farkas (whose work, published in Hungarian, spans a 30-year period starting around 1894), are credited with laying the foundations of polyhedral convexity.

The work of Fenchel was instrumental in launching the modern era of convex analysis, when the subject came to a sharp focus thanks to its rich applications in optimization and game theory. In his 1951 lecture notes [Fen51], Fenchel laid the foundations of convex duality theory, and together with related works by von Neumann on saddle points and game theory, and Kuhn and Tucker on nonlinear programming [KuT51], inspired much subsequent work on convexity and its connections with optimization. Furthermore, Fenchel developed most of the topics that are fundamental in our exposition, such as the theory of conjugate convex functions (introduced earlier in a more limited form by Legendre), and the theory of subgradients.

There are several books that relate to both convex analysis and optimization. The book by Rockafellar [Roc70] has been an important influence to subsequent convex optimization books, including the present work, but unfortunately it is written in a style that does not facilitate intuition through visualization (it does not contain a single figure). The book by Rockafellar and Wets [RoW98] is an extensive treatment of "variational analysis," a broad spectrum of topics that integrate classical analysis, convexity, and optimization of both convex and nonconvex (possibly nonsmooth) functions. Stoer and Witzgall [StW70] discuss similar topics as Rockafellar [Roc70] but less comprehensively. Ekeland and Temam

[EkT76], and Zalinescu [Zal02] develop the subject in infinite dimensional spaces. Hiriart-Urruty and Lemarechal [HiL93] emphasize algorithms for dual and nondifferentiable optimization. Rockafellar [Roc84] focuses on convexity and duality in network optimization, and an important generalization, called monotropic programming. Bertsekas [Ber98] also gives a detailed coverage of this material, which owes much to the early work of Minty [Min60] on network optimization. Schrijver [Sch86] provides an extensive account of polyhedral convexity with applications to integer programming and combinatorial optimization, and gives many historical references. Bonnans and Shapiro [BoS00] emphasize sensitivity analysis and discuss infinite dimensional problems as well. Borwein and Lewis [BoL00] develop many of the concepts in Rockafellar and Wets [RoW98], but more succinctly. The author's earlier book with Nedić and Ozdaglar [BNO03] also straddles the boundary between convex and variational analysis. Ben-Tal and Nemirovski [BeN01] focus on conic and semidefinite programming [see also the 2005 class notes by Nemirovski (on line)]. Auslender and Teboulle [AuT03] emphasize the question of existence of solutions for convex as well as nonconvex optimization problems, and related issues in duality theory and variational inequalities. Boyd and Vanderbergue [BoV04] discuss many applications of convex optimization.

We also note a few books that focus on the geometry and other properties of convex sets, but have limited connection with duality and optimization: Bonnesen and Fenchel [BoF34], Eggleston [Egg58], Valentine [Val64], Grunbaum [Gru67], Webster [Web94], and Barvinok [Bar02].

The MC/MC framework differentiates the present book from alternative treatments of convex optimization. It was initially developed by the author in joint research with A. Nedić and A. Ozdaglar, which is described in the book [BNO03]. The present account is improved and more comprehensive. In particular, it contains more streamlined proofs and some new results, particularly in connection with minimax problems (Sections 4.2.5 and 5.7.2), and nonconvex problems (Section 5.7, which generalizes the work on duality gap estimates in [Ber82], Section 5.6.1).

In a nutshell, the MC/MC framework aims to reduce duality theory to its bare essentials. The starting point is the two dual descriptions of a closed convex set: as the union of its points and as the intersection of the halfspaces that contain it. Thus a closed convex function f admits two dual descriptions: as the union of the points of its epigraph $\text{epi}(f)$ and as the intersection of the halfspaces that contain $\text{epi}(f)$ (the latter description defines the conjugate f^*).

Fenchel duality provides dual descriptions of optimization problems through the use of conjugate functions, but in the author's view, this is an indirect and far too complicated starting point for optimization duality. A simpler and more fundamental approach is to start from a single set and the two dual problems that it defines: the min common point problem and max crossing point problem. Developing duality theory around these two problems leads to a simpler, unified, and visually transparent analysis of the central frameworks of convex optimization: Lagrange duality, Fenchel duality, min-max theory, and theorems of the alternative.

REFERENCES

[Ash72] Ash, R. B., 1972. Real Analysis and Probability, Academic, N. Y.

[AuT03] Auslender, A., and Teboulle, M., 2003. Asymptotic Cones and Functions in Optimization and Variational Inequalities, Springer, N. Y.

[BNO03] Bertsekas, D. P., with Nedić, A., and Ozdaglar, A. E., 2003. Convex Analysis and Optimization, Athena Scientific, Belmont, MA.

[Bar02] Barvinok, A., 2002. A Course in Convexity, American Math. Society, Providence, RI.

[BeN01] Ben-Tal, A., and Nemirovski, A., 2001. Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications, SIAM, Phila.

[BeT89] Bertsekas, D. P., and Tsitsiklis, J. N., 1989. Parallel and Distributed Computation: Numerical Methods, Prentice-Hall, Englewood Cliffs, N. J; republished by Athena Scientific, Belmont, MA, 1997.

[BeT97] Bertsimas, D., and Tsitsiklis, J. N., 1997. Introduction to Linear Optimization, Athena Scientific, Belmont, MA.

[BeT07] Bertsekas, D. P., and Tseng, P., 2007. "Set Intersection Theorems and Existence of Optimal Solutions," Math. Programming J., Vol. 110, pp. 287-314.

[Ber72] Bertsekas, D. P., 1972. "Stochastic Optimization Problems with Nondifferentiable Cost Functionals with an Application in Stochastic Programming," Proc. 1972 IEEE Conf. Decision and Control, pp. 555-559.

[Ber73] Bertsekas, D. P., 1973. "Stochastic Optimization Problems with Nondifferentiable Cost Functionals," J. of Optimization Theory and Applications, Vol. 12, pp. 218-231.

[Ber82] Bertsekas, D. P., 1982. Constrained Optimization and Lagrange Multiplier Methods, Academic Press, N. Y.; republished in 1996 by Athena Scientific, Belmont, MA.

[Ber98] Bertsekas, D. P., 1998. Network Optimization: Continuous and Discrete Models, Athena Scientific, Belmont, MA.

[Ber99] Bertsekas, D. P., 1999. Nonlinear Programming: 2nd Edition, Athena Scientific, Belmont, MA.

[BoF34] Bonnesen, T., and Fenchel, W., 1934. Theorie der Konvexen Korper, Springer, Berlin; republished by Chelsea, N. Y., 1948.

[BoL00] Borwein, J. M., and Lewis, A. S., 2000. Convex Analysis and Nonlinear Optimization, Springer-Verlag, N. Y.

[BoS00] Bonnans, J. F., and Shapiro, A., 2000. Perturbation Analysis of Optimization Problems, Springer-Verlag, N. Y.

[BoV04] Boyd, S., and Vanderbergue, L., 2004. Convex Optimization, Cambridge Univ. Press, Cambridge, UK.

[Car11] Caratheodory, C., 1911. "Uber den Variabilitatsbereich der Fourierschen Konstanten von Positiven Harmonischen Funktionen," Rendiconto del Circolo Matematico di Palermo, Vol. 32, pp. 193-217; reprinted in Constantin Caratheodory, Gesammelte Mathematische Schriften, Band III (H. Tietze, ed.), C. H. Beck'sche Verlagsbuchhandlung, Munchen, 1955, pp. 78-110.

[Chv83] Chvatal, V., 1983. Linear Programming, W. H. Freeman and Co., N. Y.

[Dan63] Dantzig, G. B., 1963. Linear Programming and Extensions, Princeton Univ. Press, Princeton, N. J.

[Egg58] Eggleston, H. G., 1958. Convexity, Cambridge Univ. Press, Cambridge.

[EkT76] Ekeland, I., and Temam, R., 1976. Convex Analysis and Variational Problems, North-Holland Publ., Amsterdam.

[Fen51] Fenchel, W., 1951. Convex Cones, Sets, and Functions, Mimeographed Notes, Princeton Univ.

[Gru67] Grunbaum, B., 1967. Convex Polytopes, Wiley, N. Y.

[HiL93] Hiriart-Urruty, J.-B., and Lemarechal, C., 1993. Convex Analysis and Minimization Algorithms, Vols. I and II, Springer-Verlag, Berlin and N. Y.

[HoK71] Hoffman, K., and Kunze, R., 1971. Linear Algebra, Prentice-Hall, Englewood Cliffs, N. J.

[KuT51] Kuhn, H. W., and Tucker, A. W., 1951. "Nonlinear Programming," in Proc. of the Second Berkeley Symposium on Math. Statistics and Probability, Neyman, J., (Ed.), Univ. of California Press, Berkeley, CA, pp. 481-492.

[LaT85] Lancaster, P., and Tismenetsky, M., 1985. The Theory of Matrices, Academic Press, N. Y.

[Lue69] Luenberger, D. G., 1969. Optimization by Vector Space Methods, Wiley, N. Y.

[Min11] Minkowski, H., 1911. "Theorie der Konvexen Korper, Insbesondere Begrundung Ihres Ober Flachenbegriffs," Gesammelte Abhandlungen, II, Teubner, Leipsig.

[Min60] Minty, G. J., 1960. "Monotone Networks," Proc. Roy. Soc. London, A, Vol. 257, pp. 194-212.

[OrR70] Ortega, J. M., and Rheinboldt, W. C., 1970. Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, N. Y.

[RoW98] Rockafellar, R. T., and Wets, R. J.-B., 1998. Variational Analysis, Springer-Verlag, Berlin.

[Roc70] Rockafellar, R. T., 1970. Convex Analysis, Princeton Univ. Press, Princeton, N. J.

[Roc84] Rockafellar, R. T., 1984. Network Flows and Monotropic Optimization, Wiley, N. Y.; republished by Athena Scientific, Belmont, MA, 1998.

[Rud76] Rudin, W., 1976. Principles of Mathematical Analysis, McGraw, N. Y.

[Sch86] Schrijver, A., 1986. Theory of Linear and Integer Programming, Wiley, N. Y.

[Ste13] Steinitz, H., 1913. "Bedingt Konvergente Reihen und Konvexe System, I," J. of Math., Vol. 143, pp. 128-175.

[Ste14] Steinitz, H., 1914. "Bedingt Konvergente Reihen und Konvexe System, II," J. of Math., Vol. 144, pp. 1-40.

[Ste16] Steinitz, H., 1916. "Bedingt Konvergente Reihen und Konvexe System, III," J. of Math., Vol. 146, pp. 1-52.

[StW70] Stoer, J., and Witzgall, C., 1970. Convexity and Optimization in Finite Dimensions, Springer-Verlag, Berlin.

[Str76] Strang, G., 1976. Linear Algebra and its Applications, Academic, N. Y.

[Val64] Valentine, F. A., 1964. Convex Sets, McGraw-Hill, N. Y.

[Web94] Webster, R., 1994. Convexity, Oxford Univ. Press, Oxford, UK.

[Zal02] Zalinescu, C., 2002. Convex Analysis in General Vector Spaces, World Scientific, Singapore.