Convex Analysis and Optimization

Dimitri P. Bertsekas

with

Angelia Nedić and Asuman E. Ozdaglar

Massachusetts Institute of Technology

WWW site for book information and orders

http://www.athenasc.com

Athena Scientific, Belmont, Massachusetts

Athena Scientific
Post Office Box 391
Belmont, Mass. 02478-9998
U.S.A.

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

Publisher's Cataloging-in-Publication Data

Bertsekas, Dimitri P., Nedić, Angelia, Ozdaglar Asuman E.
Convex Analysis and Optimization
Includes bibliographical references and index
1. Nonlinear Programming 2. Mathematical Optimization. I. Title.
T57.8.B475 2003 519.703
Library of Congress Control Number: 2002092168

ISBN 1-886529-45-0

Third printing

ATHENA SCIENTIFIC

OPTIMIZATION AND COMPUTATION SERIES

1. Convex Analysis and Optimization, by Dimitri P. Bertsekas, with Angelia Nedić and Asuman E. Ozdaglar, 2003, ISBN 1-886529-45-0, 560 pages
2. Introduction to Probability, by Dimitri P. Bertsekas and John N. Tsitsiklis, 2002, ISBN 1-886529-40-X, 430 pages
3. Dynamic Programming and Optimal Control, Two-Volume Set (2nd Edition), by Dimitri P. Bertsekas, 2001, ISBN 1-886529-08-6, 840 pages
4. Nonlinear Programming, 2nd Edition, by Dimitri P. Bertsekas, 1999, ISBN 1-886529-00-0, 791 pages
5. Network Optimization: Continuous and Discrete Models by Dimitri P. Bertsekas, 1998, ISBN 1-886529-02-7, 608 pages
6. Network Flows and Monotropic Optimization by R. Tyrrell Rockafellar, 1998, ISBN 1-886529-06-X, 634 pages
7. Introduction to Linear Optimization by Dimitris Bertsimas and John N. Tsitsiklis, 1997, ISBN 1-886529-19-1, 608 pages
8. Parallel and Distributed Computation: Numerical Methods by Dimitri P. Bertsekas and John N. Tsitsiklis, 1997, ISBN 1-886529-01-9, 718 pages
9. Neuro-Dynamic Programming, by Dimitri P. Bertsekas and John N. Tsitsiklis, 1996, ISBN 1-886529-10-8, 512 pages
10. Constrained Optimization and Lagrange Multiplier Methods, by Dimitri P. Bertsekas, 1996, ISBN 1-886529-04-3, 410 pages
11. Stochastic Optimal Control: The Discrete-Time Case by Dimitri P. Bertsekas and Steven E. Shreve, 1996, ISBN 1-886529-03-5, 330 pages

Contents

Preface

The knowledge at which geometry aims is the knowledge of the eternal

(Plato, Republic, VII, 52)

This book focuses on the theory of convex sets and functions, and its connections with a number of topics that span a broad range from continuous to discrete optimization. These topics include Lagrange multiplier theory, Lagrangian and conjugate/Fenchel duality, minimax theory, and nondifferentiable optimization.

The book evolved from a set of lecture notes for a graduate course at M.I.T. It is widely recognized that, aside from being an eminently useful subject in engineering, operations research, and economics, convexity is an excellent vehicle for assimilating some of the basic concepts of real analysis within an intuitive geometrical setting. Unfortunately, the subject's coverage in academic curricula is scant and incidental. We believe that at least part of the reason is the shortage of textbooks that are suitable for classroom instruction, particularly for nonmathematics majors. We have therefore tried to make convex analysis accessible to a broader audience by emphasizing its geometrical character, while maintaining mathematical rigor. We have included as many insightful illustrations as possible, and we have used geometric visualization as a principal tool for maintaining the students' interest in mathematical proofs.

Our treatment of convexity theory is quite comprehensive, with all major aspects of the subject receiving substantial treatment. The mathematical prerequisites are a course in linear algebra and a course in real analysis in finite dimensional spaces (which is the exclusive setting of the book). A summary of this material, without proofs, is provided in Section 1.1.

The coverage of the theory has been significantly extended in the exercises, which represent a major component of the book. Detailed solutions of all the exercises (nearly 200 pages) are internet-posted in the book's www page

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

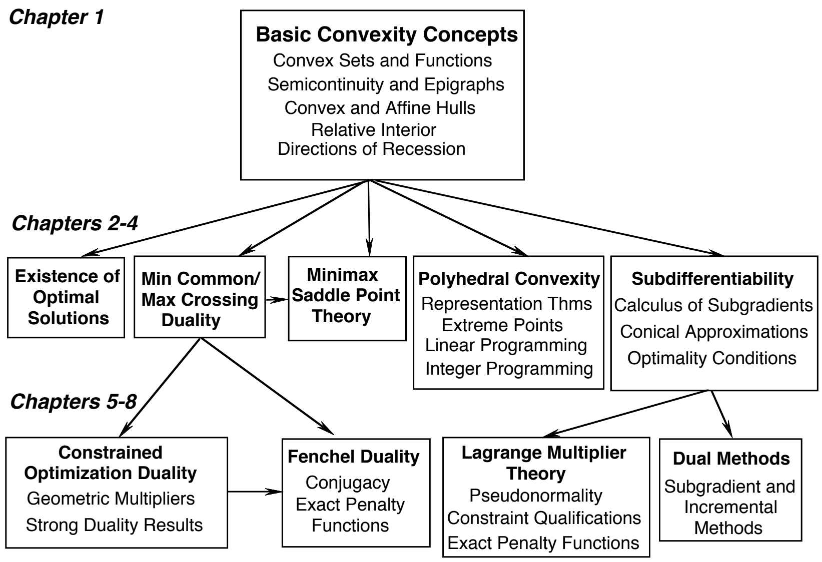
Some of the exercises may be attempted by the reader without looking at the solutions, while others are challenging but may be solved by the advanced reader with the assistance of hints. Still other exercises represent substantial theoretical results, and in some cases include new and unpublished research. Readers and instructors should decide for themselves how to make best use of the internet-posted solutions.

An important part of our approach has been to maintain a close link between the theoretical treatment of convexity and its application to optimization. For example, in Chapter 2, after the development of some of the basic facts about convexity, we discuss some of their applications to optimization and saddle point theory; in Chapter 3, after the discussion of polyhedral convexity, we discuss its application in linear and integer programming; and in Chapter 4, after the discussion of subgradients, we discuss their use in optimality conditions. We follow this style in the remaining chapters, although having developed in Chapters 1-4 most of the needed convexity theory, the discussion in the subsequent chapters is more heavily weighted towards optimization.

The chart of the opposite page illustrates the main topics covered in the book, and their interrelations. At the top level, we have the most basic concepts of convexity theory, which are covered in Chapter 1. At the middle level, we have fundamental topics of optimization, such as existence and characterization of solutions, and minimax theory, together with some supporting convexity concepts such as hyperplane separation, polyhedral sets, and subdifferentiability (Chapters 2-4). At the lowest level, we have the core issues of convex optimization: Lagrange multipliers, Lagrange and Fenchel duality, and numerical dual optimization (Chapters 5-8).

An instructor who wishes to teach a course from the book has a choice between several different plans. One possibility is to cover in detail just the first four chapters, perhaps augmented with some selected sections from the remainder of the book, such as the first section of Chapter 7, which deals with conjugate convex functions. The idea here is to concentrate on convex analysis and illustrate its application to minimax theory through the minimax theorems of Chapters 2 and 3, and to constrained optimization theory through the Nonlinear Farkas' Lemma of Chapter 3 and the optimality conditions of Chapter 4. An alternative plan is to cover Chapters 1-4 in less detail in order to allow some time for Lagrange multiplier theory and computational methods. Other plans may also be devised, possibly including some applications or some additional theoretical topics of the instructor's choice.

While the subject of the book is classical, the treatment of several of its important topics is new and in some cases relies on new research. In



particular, our new lines of analysis include:

(a) A unified development of minimax theory and constrained optimization duality as special cases of the duality between two simple geometrical problems: the min common point problem and the max crossing point problem. Here, by minimax theory, we mean the analysis relating to the minimax 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." By constrained optimization theory, we mean the analysis of problems such as

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

and issues such as the existence of optimal solutions and Lagrange multipliers, and the absence of a duality gap [equality of the optimal value of the above problem and the optimal value of an associated dual problem, obtained by assigning multipliers to the inequality constraints $g_j(x)\le 0$].

(b) A unification of conditions for existence of solutions of convex optimization problems, conditions for the minimax equality to hold, and conditions for the absence of a duality gap in constrained optimization. This unification is based on conditions guaranteeing that a nested family of closed convex sets has a nonempty intersection.

(c) A unification of the major constraint qualifications that guarantee the existence of Lagrange multipliers for nonconvex constrained optimization. This unification is achieved through the notion of constraint pseudonormality, which is motivated by an enhanced form of the Fritz John necessary optimality conditions.

(d) The development of incremental subgradient methods for dual optimization, and the analysis of their advantages over classical subgradient methods.

We provide some orientation by informally summarizing the main ideas of each of the above topics.

Min Common/Max Crossing Duality

In this book, duality theory is captured in two easily visualized problems: the min common point problem and the max crossing point problem, introduced in Chapter 2. Fundamentally, these problems revolve around the existence of nonvertical supporting hyperplanes to convex sets that are unbounded from above along the vertical axis. When properly specialized, this turns out to be the critical issue in constrained optimization duality and saddle point/minimax theory, under standard convexity and/or concavity assumptions.

The salient feature of the min common/max crossing framework is its simple geometry, in the context of which the fundamental constraint qualifications needed for strong duality theorems are visually apparent, and admit straightforward proofs. This allows the development of duality theory in a unified way: first within the min common/max crossing framework in Chapters 2 and 3, and then by specialization, to saddle point and minimax theory in Chapters 2 and 3, and to optimization duality in Chapter 6. All of the major duality theorems discussed in this book are derived in this way, including the principal Lagrange multiplier and Fenchel duality theorems for convex programming, and the von Neuman Theorem for zero sum games.

From an instructional point of view, it is particularly desirable to unify constrained optimization duality and saddle point/minimax theory (under convexity/concavity assumptions). Their connection is well known, but it is hard to understand beyond a superficial level, because there is not enough overlap between the two theories to develop one in terms of the other. In our approach, rather than trying to build a closer connection between constrained optimization duality and saddle point/minimax theory, we show how they both stem from a common geometrical root: the min common/max crossing duality.

We note that the constructions involved in the min common and max crossing problems arise in the theories of conjugate convex functions, subgradients, and duality. As such they are implicit in several earlier analyses; in fact they have been employed for visualization purposes in the first author's nonlinear programming textbook [Ber99]. However, the two problems have not been used as a unifying theoretical framework for constrained optimization duality, saddle point theory, or other contexts, except implicitly through the theory of conjugate convex functions, and the complicated and specialized machinery of conjugate saddle functions. Pedagogically, it may be desirable to postpone the introduction of conjugacy theory until it is needed for the purposes of Fenchel duality (Chapter 7), which is what we have done.

Existence of Solutions and Strong Duality

We show that under convexity assumptions, several fundamental issues in optimization are intimately related. In particular, we give a unified analysis of conditions for optimal solutions to exist, for the minimax equality to hold, and for the absence of a duality gap in constrained optimization.

To provide a sense of the main idea, we note that given a constrained optimization problem, lower semicontinuity of the cost function and compactness of the constraint set guarantee the existence of an optimal solution (the Weierstrass Theorem). On the other hand, the same conditions plus convexity of the cost and constraint functions guarantee not only the existence of an optimal solution, but also the absence of a duality gap. This is not a coincidence, because as it turns out, the conditions for both cases critically rely on the same fundamental properties of compact sets, namely that the intersection of a nested family of nonempty compact sets is nonempty and compact, and that the projections of compact sets on any subspace are compact.

In our analysis, we extend this line of reasoning under a variety of assumptions relating to convexity, directions of recession, polyhedral sets, and special types of sets specified by quadratic and other types of inequalities. The assumptions are used to establish results asserting that the intersection of a nested family of closed convex sets is nonempty, and that the function $f(x) = \inf_z F(x, z)$, obtained by partial minimization of a convex function F, is lower semicontinuous. These results are translated in turn to a broad variety of conditions that guarantee the existence of optimal solutions, the minimax equality, and the absence of a duality gap.

Pseudonormality and Lagrange Multipliers

In Chapter 5, we discuss Lagrange multiplier theory in the context of optimization of a smooth cost function, subject to smooth equality and inequality constraints, as well as an additional set constraint. Our treatment of Lagrange multipliers is new, and aims to generalize, unify, and streamline the theory of constraint qualifications.

The starting point for our development is an enhanced set of necessary conditions of the Fritz John type, that are sharper than the classical Karush-Kuhn-Tucker conditions (they include extra conditions, which may narrow down the field of candidate local minima). They are also more general in that they apply when there is an abstract (possibly nonconvex) set constraint, in addition to the equality and inequality constraints. To achieve this level of generality, we bring to bear notions of nonsmooth analysis, and we find that the notion of regularity of the abstract constraint set provides the critical distinction between problems that do and do not admit a satisfactory theory.

Fundamentally, Lagrange multiplier theory should aim to identify the essential constraint structure that guarantees the existence of Lagrange multipliers. For smooth problems with equality and inequality constraints, but no abstract set constraint, this essential structure is captured by the classical notion of quasiregularity (the tangent cone at a given feasible point is equal to the cone of first order feasible variations). However, in the presence of an additional set constraint, the notion of quasiregularity breaks down as a viable unification vehicle. Our development introduces the notion of pseudonormality as a substitute for quasiregularity for the case of an abstract set constraint. Pseudonormality unifies and expands the major constraint qualifications, and simplifies the proofs of Lagrange multiplier theorems. In the case of equality constraints only, pseudonormality is implied by either one of two alternative constraint qualifications: the linear independendence of the constraint gradients and the linearity of the constraint functions. In fact, in this case, pseudonormality is not much different than the union of these two constraint qualifications. However, pseudonormality is a meaningful unifying property even in the case of an additional set constraint, where the classical proof arguments based on quasiregularity fail. Pseudonormality also provides the connecting link between constraint qualifications and the theory of exact penalty functions.

An interesting byproduct of our analysis is a taxonomy of different types of Lagrange multipliers for problems with nonunique Lagrange multipliers. Under some convexity assumptions, we show that if there exists at least one Lagrange multiplier vector, there exists at least one of a special type, called informative, which has nice sensitivity properties. The nonzero components of such a multiplier vector identify the constraints that need to be violated in order to improve the optimal cost function value. Furthermore, a particular informative Lagrange multiplier vector characterizes the direction of steepest rate of improvement of the cost function for a given level of the norm of the constraint violation. Along that direction, the equality and inequality constraints are violated consistently with the signs of the corresponding multipliers.

The theory of enhanced Fritz John conditions and pseudonormality are extended in Chapter 6 to the case of a convex programming problem, without assuming the existence of an optimal solution or the absence of a duality gap. They form the basis for a new line of analysis for asserting the existence of informative multipliers under the standard constraint qualifications.

Incremental Subgradient Methods

In Chapter 8, we discuss one of the most important uses of duality: the numerical solution of dual problems, often in the context of discrete optimization and the method of branch-and-bound. These dual problems are often nondifferentiable and have special structure. Subgradient methods have been among the most popular for the solution of these problems, but they often suffer from slow convergence.

We introduce incremental subgradient methods, which aim to accelerate the convergence by exploiting the additive structure that a dual problem often inherits from properties of its primal problem, such as separability. In particular, for the common case where the dual function is the sum of a large number of component functions, incremental methods consist of a sequence of incremental steps, each involving a single component of the dual function, rather than the sum of all components.

Our analysis aims to identify effective variants of incremental methods, and to quantify their advantages over the standard subgradient methods. An important question is the selection of the order in which the components are selected for iteration. A particularly interesting variant uses randomization of the order to resolve a worst-case complexity bottleneck associated with the natural deterministic order. According to both analysis and experiment, this randomized variant performs substantially better than the standard subgradient methods for large scale problems that typically arise in the context of duality. The randomized variant is also particularly well-suited for parallel, possibly asynchronous, implementation, and is the only available method, to our knowledge, that can be used efficiently within this context.

We are thankful to a few persons for their contributions to the book. Several colleagues contributed information, suggestions, and insights. We would like to single out Paul Tseng, who was extraordinarily helpful by proofreading portions of the book, and collaborating with us on several research topics, including the Fritz John theory of Sections 5.7 and 6.6. We would also like to thank Xin Chen and Janey Yu, who gave us valuable feedback and some specific suggestions. Finally, we wish to express our appreciation for the stimulating environment at M.I.T., which provided an excellent setting for this work.

Dimitri P. Bertsekas, dimitrib@mit.edu
Angelia Nedić, angelia.nedich@alphatech.com
Asuman E. Ozdaglar, asuman@mit.edu

Note added in the second printing (spring 2013):

The second printing of this book is identical to the first printing, except that typograpical and other minor errors have been corrected. Moreover, some relevant research references that appeared subsequent to the first printing of 2003 were noted. Also, the related book

> Convex Optimization Theory, by Dimitri P. Bertsekas, 2009,
> ISBN 978-1-886529-31-1, 256 pages

appeared in 2009. This book shares some material with the present work, but has the character of a textbook and concentrates exclusively on convex optimization. With the publication of the 2009 book, the set of exercises on convex optimization was substantially enlarged. The exercises for both books can be found (with complete solutions) at the Athena Scientific web site (http://www.athenasc.com).

1

Basic Convexity Concepts

Contents

In this chapter and the following three, we develop the theory of convex sets, which is the mathematical foundation for minimax theory, Lagrange multiplier theory, and duality. We assume no prior knowledge of the subject, and we give a detailed development. As we embark on the study of convexity, it is worth listing some of the properties of convex sets and functions that make them so special in optimization.

(a) *A convex function has no local minima that are not global*. Thus the difficulties associated with multiple disconnected local minima, whose global optimality is hard to verify in practice, are avoided (see Section 2.1).

(b) *A convex set has a nonempty relative interior*. In other words, relative to the smallest affine set containing it, a convex set has a nonempty interior (see Section 1.4). Thus convex sets avoid the analytical and computational optimization difficulties associated with "thin" and "curved" constraint surfaces.

(c) *A convex set is connected and has feasible directions at any point* (assuming it consists of more than one point). By this we mean that given any point x in a convex set X, it is possible to move from x along some directions y and stay within X for at least a nontrivial interval, i.e., $x + \alpha y \in X$ for all sufficiently small but positive stepsizes α (see Section 4.6). In fact a stronger property holds: given any two distinct points x and $\overline{x}$ in X, the direction $\overline{x} - x$ is a feasible direction at x, and all feasible directions can be characterized this way. For optimization purposes, this is important because it allows a calculus-based comparison of the cost of x with the cost of its close neighbors, and forms the basis for some important algorithms. Furthermore, much of the difficulty commonly associated with discrete constraint sets (arising for example in combinatorial optimization), is not encountered under convexity.

(d) *A nonconvex function can be "convexified" while maintaining the optimality of its global minima*, by forming the convex hull of the epigraph of the function (see Exercise 1.20).

(e) *The existence of a global minimum of a convex function over a convex set is conveniently characterized in terms of directions of recession* (see Section 2.3).

(f) *A polyhedral convex set (one that is specified by linear equality and inequality constraints) is characterized in terms of a finite set of extreme points and extreme directions*. This is the basis for finitely terminating methods for linear programming, including the celebrated simplex method (see Sections 3.3 and 3.4).

(g) *A convex function is continuous within the interior of its domain, and has nice differentiability properties*. In particular, a real-valued convex function is directionally differentiable at any point. Furthermore, while a convex function need not be differentiable, it possesses subgradients, which are nice and geometrically intuitive substitutes for a gradient (see Chapter 4). Just like gradients, subgradients figure prominently in optimality conditions and computational algorithms.

(h) *Convex functions are central in duality theory*. Indeed, the dual problem of a given optimization problem (discussed in Chapter 6) consists of minimization of a convex function over a convex set, even if the original problem is not convex.

(i) *Closed convex cones are self-dual with respect to polarity*. In words, we have $C = (C^*)^*$ for any closed and convex cone C, where C^* is the polar cone of C (the set of vectors that form a nonpositive inner product with all vectors in C), and $(C^*)^*$ is the polar cone of C^*. This simple and geometrically intuitive property (discussed in Section 3.1) underlies important aspects of Lagrange multiplier theory.

(j) *Convex lower semicontinuous functions are self-dual with respect to conjugacy*. It will be seen in Chapter 7 that a certain geometrically motivated conjugacy operation on a convex, lower semicontinuous function generates another convex, lower semicontinuous function, and when applied for the second time regenerates the original function. The conjugacy operation relies on a fundamental dual characterization of a closed convex set: as the union of all line segments connecting its points, and as the intersection of the closed halfspaces within which the set is contained. Conjugacy is central in duality theory, and has a nice interpretation that can be used to visualize and understand some of the most interesting aspects of convex optimization.

In this first chapter, after an introductory first section, we focus on the basic concepts of convex analysis: characterizations of convex sets and functions, convex and affine hulls, topological concepts such as closure, continuity, and relative interior, and the important notion of the recession cone.

1.1 LINEAR ALGEBRA AND REAL ANALYSIS

In this section, we list some basic definitions, notational conventions, and results from linear algebra and real analysis. We assume that the reader is familiar with this material, so no proofs are given. For related and additional material, we recommend the books by Hoffman and Kunze [HoK71], Lancaster and Tismenetsky [LaT85], and Strang [Str76] (linear algebra), and the books by Ash [Ash72], Ortega and Rheinboldt [OrR70], and Rudin [Rud76] (real analysis).

Set Notation

If X is a set and x is an element of X, we write $x \in X$. A set can be specified in the form $X = \{x \mid x \text{ satisfies } P\}$, as the set of all elements satisfying property P. The union of two sets X_1 and X_2 is denoted by $X_1 \cup X_2$ and their intersection by $X_1 \cap X_2$. The symbols $\exists$ and $\forall$ have the meanings "there exists" and "for all," respectively. The empty set is denoted by $\emptyset$.

The set of real numbers (also referred to as scalars) is denoted by $\Re$. The set $\Re$ augmented with $+\infty$ and $-\infty$ is called the *set of extended real numbers*. We write $-\infty < x < \infty$ for all real numbers x, and $-\infty \leq x \leq \infty$ for all extended real numbers x. We denote by $[a, b]$ the set of (possibly extended) real numbers x satisfying $a \leq x \leq b$. A rounded, instead of square, bracket denotes strict inequality in the definition. Thus $(a, b]$, $[a, b)$, and (a, b) denote the set of all x satisfying $a < x \leq b$, $a \leq x < b$, and $a < x < b$, respectively. Furthermore, we use the natural extensions of the rules of arithmetic: $x \cdot 0 = 0$ for every extended real number x, $x \cdot \infty = \infty$ if $x > 0$, $x \cdot \infty = -\infty$ if $x < 0$, and $x + \infty = \infty$ and $x - \infty = -\infty$ for every scalar x. The expression $\infty - \infty$ is meaningless and is never allowed to occur.

Inf and Sup Notation

The *supremum* of a nonempty set X of scalars, denoted by $\sup X$, is defined as the smallest scalar y such that $y \geq x$ for all $x \in X$. If no such scalar exists, we say that the supremum of X is ∞. Similarly, the *infimum* of X, denoted by $\inf X$, is defined as the largest scalar y such that $y \leq x$ for all $x \in X$, and is equal to $-\infty$ if no such scalar exists. For the empty set, we use the convention

$$\sup \emptyset = -\infty, \qquad \inf \emptyset = \infty.$$

If $\sup X$ is equal to a scalar $\overline{x}$ that belongs to the set X, we say that $\overline{x}$ is the *maximum point* of X and we write $\overline{x} = \max X$. Similarly, if $\inf X$ is equal to a scalar $\overline{x}$ that belongs to the set X, we say that $\overline{x}$ is the *minimum point* of X and we write $\overline{x} = \min X$. Thus, when we write $\max X$ (or $\min X$) in place of $\sup X$ (or $\inf X$, respectively), we do so just for emphasis: we indicate that it is either evident, or it is known through earlier analysis, or it is about to be shown that the maximum (or minimum, respectively) of the set X is attained at one of its points.

Function Notation

If f is a function, we use the notation $f : X \mapsto Y$ to indicate the fact that f is defined on a nonempty set X (its *domain*) and takes values in a set Y (its *range*). Thus when using the notation $f : X \mapsto Y$, we implicitly assume that X is nonempty. If $f : X \mapsto Y$ is a function, and U and V are subsets of X and Y, respectively, the set $\big\{f(x) \mid x \in U\big\}$ is called the *image* or *forward image of* U *under* f, and the set $\big\{x \in X \mid f(x) \in V\big\}$ is called the *inverse image of* V *under* f.

1.1.1 Vectors and Matrices

We denote by $\Re^n$ the set of n-dimensional real vectors. For any $x \in \Re^n$, we use x_i to indicate its ith *coordinate*, also called its ith *component*.

Vectors in $\Re^n$ will be viewed as column vectors, unless the contrary is explicitly stated. For any $x \in \Re^n$, x' denotes the transpose of x, which is an n-dimensional row vector. The *inner product* of two vectors $x, y \in \Re^n$ is defined by $x'y = \sum_{i=1}^n x_i y_i$. Two vectors $x, y \in \Re^n$ satisfying $x'y = 0$ are called *orthogonal*.

If x is a vector in $\Re^n$, the notations $x > 0$ and $x \geq 0$ indicate that all components of x are positive and nonnegative, respectively. For any two vectors x and y, the notation $x > y$ means that $x - y > 0$. The notations $x \geq y$, $x < y$, etc., are to be interpreted accordingly.

If X is a set and λ is a scalar, we denote by λX the set $\{\lambda x \mid x \in X\}$. If X_1 and X_2 are two subsets of $\Re^n$, we denote by $X_1 + X_2$ the set

$$\{x_1 + x_2 \mid x_1 \in X_1,\, x_2 \in X_2\},$$

which is referred to as the *vector sum of* X_1 *and* X_2. We use a similar notation for the sum of any finite number of subsets. In the case where one of the subsets consists of a single vector $\overline{x}$, we simplify this notation as follows:

$$\overline{x} + X = \{\overline{x} + x \mid x \in X\}.$$

We also denote by $X_1 - X_2$ the set

$$\{x_1 - x_2 \mid x_1 \in X_1,\, x_2 \in X_2\}.$$

Given sets $X_i \subset \Re^{n_i}$, $i = 1, \ldots, m$, the *Cartesian product* of the X_i, denoted by $X_1 \times \cdots \times X_m$, is the set

$$\big\{(x_1, \ldots, x_m) \mid x_i \in X_i,\, i = 1, \ldots, m\big\},$$

which is a subset of $\Re^{n_1 + \cdots + n_m}$.

Subspaces and Linear Independence

A nonempty subset S of $\Re^n$ is called a *subspace* if $ax + by \in S$ for every $x, y \in S$ and every $a, b \in \Re$. An *affine set* in $\Re^n$ is a translated subspace, i.e., a set X of the form $X = \overline{x} + S = \{\overline{x} + x \mid x \in S\}$, where $\overline{x}$ is a vector in $\Re^n$ and S is a subspace of $\Re^n$, called the *subspace parallel to* X. Note that there can be only one subspace S associated with an affine set in this manner. [To see this, let $X = x + S$ and $X = \overline{x} + \overline{S}$ be two representations of the affine set X. Then, we must have $x = \overline{x} + \overline{s}$ for some $\overline{s} \in \overline{S}$ (since $x \in X$), so that $X = \overline{x} + \overline{s} + S$. Since we also have $X = \overline{x} + \overline{S}$, it follows that $S = \overline{S} - \overline{s} = \overline{S}$.] The *span* of a finite collection $\{x_1, \ldots, x_m\}$ of elements of $\Re^n$ is the subspace consisting of all vectors y of the form $y = \sum_{k=1}^m \alpha_k x_k$, where each α_k is a scalar.

The vectors $x_1, \ldots, x_m \in \Re^n$ are called *linearly independent* if there exists no set of scalars $\alpha_1, \ldots, \alpha_m$, at least one of which is nonzero, such that $\sum_{k=1}^m \alpha_k x_k = 0$. An equivalent definition is that $x_1 \neq 0$, and for every $k > 1$, the vector x_k does not belong to the span of $x_1, \ldots, x_{k-1}$.

If S is a subspace of $\Re^n$ containing at least one nonzero vector, a *basis* for S is a collection of vectors that are linearly independent and whose span is equal to S. Every basis of a given subspace has the same number of vectors. This number is called the *dimension* of S. By convention, the subspace $\{0\}$ is said to have dimension zero. The *dimension of an affine set* $\overline{x} + S$ is the dimension of the corresponding subspace S. Every subspace of nonzero dimension has a basis that is orthogonal (i.e., any pair of distinct vectors from the basis is orthogonal).

Given any set X, the set of vectors that are orthogonal to all elements of X is a subspace denoted by $X^\perp$:

$$X^\perp = \{y \mid y'x = 0,\ \forall\ x \in X\}.$$

If S is a subspace, $S^\perp$ is called the *orthogonal complement* of S. Any vector x can be uniquely decomposed as the sum of a vector from S and a vector from $S^\perp$. Furthermore, we have $(S^\perp)^\perp = S$.

Matrices

For any matrix A, we use A_{ij}, $[A]_{ij}$, or a_{ij} to denote its ijth element. The *transpose* of A, denoted by A', is defined by $[A']_{ij} = a_{ji}$. For any two matrices A and B of compatible dimensions, the transpose of the product matrix AB satisfies $(AB)' = B'A'$.

If X is a subset of $\Re^n$ and A is an $m \times n$ matrix, then the *image of* X *under* A is denoted by AX (or $A \cdot X$ if this enhances notational clarity):

$$AX = \{Ax \mid x \in X\}.$$

If Y is a subset of $\Re^m$, the *inverse image of* Y *under* A is denoted by $A^{-1}Y$ or $A^{-1} \cdot Y$:

$$A^{-1}Y = \{x \mid Ax \in Y\}.$$

If X and Y are subspaces, then AX and $A^{-1}Y$ are also subspaces.

Let A be a square matrix. We say that A is *symmetric* if $A' = A$. We say that A is *diagonal* if $[A]_{ij} = 0$ whenever $i \neq j$. We use I to denote the identity matrix (the diagonal matrix whose diagonal elements are equal to 1). We denote the *determinant* of A by $\det(A)$.

Let A be an $m \times n$ matrix. The *range space* of A, denoted by $R(A)$, is the set of all vectors $y \in \Re^m$ such that $y = Ax$ for some $x \in \Re^n$. The *nullspace* of A, denoted by $N(A)$, is the set of all vectors $x \in \Re^n$ such that $Ax = 0$. It is seen that the range space and the null space of A are subspaces. The *rank* of A is the dimension of the range space of A. The rank of A is equal to the maximal number of linearly independent columns of A, and is also equal to the maximal number of linearly independent rows of A. The matrix A and its transpose A' have the same rank. We say that A has *full rank*, if its rank is equal to $\min\{m, n\}$. This is true if and only if either all the rows of A are linearly independent, or all the columns of A are linearly independent.

The range space of an $m \times n$ matrix A is equal to the orthogonal complement of the nullspace of its transpose, i.e.,

$$R(A) = N(A')^{\perp}.$$

Another way to state this result is that given vectors $a_1, \ldots, a_n \in \Re^m$ (the columns of A) and a vector $x \in \Re^m$, we have $x'y = 0$ for all y such that $a_i'y = 0$ for all i if and only if $x = \lambda_1 a_1 + \cdots + \lambda_n a_n$ for some scalars $\lambda_1, \ldots, \lambda_n$. This is a special case of Farkas' Lemma, an important result for constrained optimization, which will be discussed in Section 3.2. A useful application of this result is that if S_1 and S_2 are two subspaces of $\Re^n$, then

$$S_1^{\perp} + S_2^{\perp} = (S_1 \cap S_2)^{\perp}.$$

This follows by introducing matrices B_1 and B_2 such that $S_1 = \{x \mid B_1 x = 0\} = N(B_1)$ and $S_2 = \{x \mid B_2 x = 0\} = N(B_2)$, and writing

$$S_1^{\perp} + S_2^{\perp} = R\big([\,B_1' \quad B_2'\,]\big) = N\left(\begin{bmatrix} B_1 \\ B_2 \end{bmatrix}\right)^{\perp} = \big(N(B_1) \cap N(B_2)\big)^{\perp} = (S_1 \cap S_2)^{\perp}$$

A function $f : \Re^n \mapsto \Re$ is said to be *affine* if it has the form $f(x) = a'x + b$ for some $a \in \Re^n$ and $b \in \Re$. Similarly, a function $f : \Re^n \mapsto \Re^m$ is said to be *affine* if it has the form $f(x) = Ax + b$ for some $m \times n$ matrix A and some $b \in \Re^m$. If $b = 0$, f is said to be a *linear function* or *linear transformation*. Sometimes, with slight abuse of terminology, an equation or inequality involving a linear function, such as $a'x = b$ or $a'x \leq b$, is referred to as a *linear equation or inequality*, respectively.

1.1.2 Topological Properties

Definition 1.1.1: A *norm* $\|\cdot\|$ on $\Re^n$ is a function that assigns a scalar $\|x\|$ to every $x \in \Re^n$ and that has the following properties:

(a) $\|x\| \geq 0$ for all $x \in \Re^n$.

(b) $\|\alpha x\| = |\alpha| \cdot \|x\|$ for every scalar α and every $x \in \Re^n$.

(c) $\|x\| = 0$ if and only if $x = 0$.

(d) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in \Re^n$ (this is referred to as the *triangle inequality*).

The *Euclidean norm* of a vector $x = (x_1, \ldots, x_n)$ is defined by

$$\|x\| = (x'x)^{1/2} = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}.$$

We will use the Euclidean norm almost exclusively in this book. In particular, *in the absence of a clear indication to the contrary,* $\|\cdot\|$ *will denote the Euclidean norm.* Two important results for the Euclidean norm are:

Proposition 1.1.1: (Pythagorean Theorem) For any two vectors x and y that are orthogonal, we have

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2.$$

Proposition 1.1.2: (Schwarz Inequality) For any two vectors x and y, we have

$$|x'y| \leq \|x\| \cdot \|y\|,$$

with equality holding if and only if $x = \alpha y$ for some scalar α.

Two other important norms are the *maximum norm* $\|\cdot\|_\infty$ (also called *sup-norm* or ℓ_∞*-norm*), defined by

$$\|x\|_\infty = \max_{i=1,\ldots,n} |x_i|,$$

and the ℓ_1-*norm* $\|\cdot\|_1$, defined by

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|.$$

Sequences

We use both subscripts and superscripts in sequence notation. Generally, we prefer subscripts, but we use superscripts whenever we need to reserve the subscript notation for indexing components of vectors and functions. The meaning of the subscripts and superscripts should be clear from the context in which they are used.

A sequence $\{x_k \mid k = 1, 2, \ldots\}$ (or $\{x_k\}$ for short) of scalars is said to *converge* if there exists a scalar x such that for every $\epsilon > 0$ we have $|x_k - x| < \epsilon$ for every k greater than some integer K (that depends on ϵ). The scalar x is said to be the *limit* of $\{x_k\}$, and the sequence $\{x_k\}$ is said to *converge to* x; symbolically, $x_k \to x$ or $\lim_{k\to\infty} x_k = x$. If for every scalar b there exists some K (that depends on b) such that $x_k \geq b$ for all $k \geq K$, we write $x_k \to \infty$ and $\lim_{k\to\infty} x_k = \infty$. Similarly, if for every scalar b there exists some integer K such that $x_k \leq b$ for all $k \geq K$, we write $x_k \to -\infty$ and $\lim_{k\to\infty} x_k = -\infty$. Note, however, that implicit in any of the statements "$\{x_k\}$ converges" or "the limit of $\{x_k\}$ exists" or "$\{x_k\}$ has a limit" is that the limit of $\{x_k\}$ is a scalar.

A scalar sequence $\{x_k\}$ is said to be *bounded above* (respectively, *below*) if there exists some scalar b such that $x_k \leq b$ (respectively, $x_k \geq b$) for all k. It is said to be *bounded* if it is bounded above and bounded below. The sequence $\{x_k\}$ is said to be monotonically *nonincreasing* (respectively, *nondecreasing*) if $x_{k+1} \leq x_k$ (respectively, $x_{k+1} \geq x_k$) for all k. If $x_k \to x$ and $\{x_k\}$ is monotonically nonincreasing (nondecreasing), we also use the notation $x_k \downarrow x$ ($x_k \uparrow x$, respectively).

Proposition 1.1.3: Every bounded and monotonically nonincreasing or nondecreasing scalar sequence converges.

Note that a monotonically nondecreasing sequence $\{x_k\}$ is either bounded, in which case it converges to some scalar x by the above proposition, or else it is unbounded, in which case $x_k \to \infty$. Similarly, a monotonically nonincreasing sequence $\{x_k\}$ is either bounded and converges, or it is unbounded, in which case $x_k \to -\infty$.

Given a scalar sequence $\{x_k\}$, let

$$y_m = \sup\{x_k \mid k \geq m\}, \qquad z_m = \inf\{x_k \mid k \geq m\}.$$

The sequences $\{y_m\}$ and $\{z_m\}$ are nonincreasing and nondecreasing, respectively, and therefore have a limit whenever $\{x_k\}$ is bounded above or is bounded below, respectively (Prop. 1.1.3). The limit of y_m is denoted by $\limsup_{k\to\infty} x_k$, and is referred to as the *upper limit* of $\{x_k\}$. The limit of z_m is denoted by $\liminf_{k\to\infty} x_k$, and is referred to as the *lower limit* of $\{x_k\}$. If $\{x_k\}$ is unbounded above, we write $\limsup_{k\to\infty} x_k = \infty$, and if it is unbounded below, we write $\liminf_{k\to\infty} x_k = -\infty$.

Proposition 1.1.4: Let $\{x_k\}$ and $\{y_k\}$ be scalar sequences.

(a) We have

$$\inf\{x_k \mid k \geq 0\} \leq \liminf_{k\to\infty} x_k \leq \limsup_{k\to\infty} x_k \leq \sup\{x_k \mid k \geq 0\}.$$

(b) $\{x_k\}$ converges if and only if

$$-\infty < \liminf_{k\to\infty} x_k = \limsup_{k\to\infty} x_k < \infty.$$

Furthermore, if $\{x_k\}$ converges, its limit is equal to the common scalar value of $\liminf_{k\to\infty} x_k$ and $\limsup_{k\to\infty} x_k$.

(c) If $x_k \leq y_k$ for all k, then

$$\liminf_{k\to\infty} x_k \leq \liminf_{k\to\infty} y_k, \qquad \limsup_{k\to\infty} x_k \leq \limsup_{k\to\infty} y_k.$$

(d) We have

$$\liminf_{k\to\infty} x_k + \liminf_{k\to\infty} y_k \leq \liminf_{k\to\infty}(x_k + y_k),$$

$$\limsup_{k\to\infty} x_k + \limsup_{k\to\infty} y_k \geq \limsup_{k\to\infty}(x_k + y_k).$$

A sequence $\{x_k\}$ of vectors in $\Re^n$ is said to converge to some $x \in \Re^n$ if the ith component of x_k converges to the ith component of x for every i. We use the notations $x_k \to x$ and $\lim_{k\to\infty} x_k = x$ to indicate convergence for vector sequences as well. The sequence $\{x_k\}$ is called bounded if each of its corresponding component sequences is bounded. It can be seen that $\{x_k\}$ is bounded if and only if there exists a scalar c such that $\|x_k\| \leq c$ for all k. An infinite subset of a sequence $\{x_k\}$ is called a *subsequence* of $\{x_k\}$. Thus a subsequence can itself be viewed as a sequence, and can be represented as a set $\{x_k \mid k \in \mathcal{K}\}$, where $\mathcal{K}$ is an infinite subset of positive integers (the notation $\{x_k\}_\mathcal{K}$ will also be used).

A vector $x \in \Re^n$ is said to be a *limit point* of a sequence $\{x_k\}$ if there exists a subsequence of $\{x_k\}$ that converges to x.† The following is a classical result that will be used often.

Proposition 1.1.5: (Bolzano-Weierstrass Theorem) A bounded sequence in $\Re^n$ has at least one limit point.

$o(\cdot)$ Notation

For a positive integer p and a function $h : \Re^n \mapsto \Re^m$ we write

$$h(x) = o\big(\|x\|^p\big)$$

if

$$\lim_{k\to\infty} \frac{h(x_k)}{\|x_k\|^p} = 0,$$

for all sequences $\{x_k\}$ such that $x_k \to 0$ and $x_k \neq 0$ for all k.

Closed and Open Sets

We say that x is a *closure point of a subset* X of $\Re^n$ if there exists a sequence $\{x_k\} \subset X$ that converges to x. The *closure* of X, denoted $\mathrm{cl}(X)$, is the set of all closure points of X.

Definition 1.1.2: A subset X of $\Re^n$ is called *closed* if it is equal to its closure. It is called *open* if its complement, $\{x \mid x \notin X\}$, is closed. It is called *bounded* if there exists a scalar c such that $\|x\| \leq c$ for all $x \in X$. It is called *compact* if it is closed and bounded.

For any $\epsilon > 0$ and $x^* \in \Re^n$, consider the sets

$$\big\{x \mid \|x - x^*\| < \epsilon\big\}, \qquad \big\{x \mid \|x - x^*\| \leq \epsilon\big\}.$$

---

† Some authors prefer the alternative term "cluster point" of a sequence, and use the term "limit point of a set S" to indicate a point $\overline{x}$ such that $\overline{x} \notin S$ and there exists a sequence $\{x_k\} \subset S$ that converges to $\overline{x}$. With this terminology, $\overline{x}$ is a cluster point of a sequence $\{x_k \mid k = 1, 2, \ldots\}$ if and only if $(\overline{x}, 0)$ is a limit point of the set $\big\{(x_k, 1/k) \mid k = 1, 2, \ldots\big\}$. Our use of the term "limit point" of a sequence is quite popular in optimization and should not lead to any confusion.

The first set is open and is called an *open sphere* centered at x^*, while the second set is closed and is called a *closed sphere* centered at x^*. Sometimes the terms *open ball* and *closed ball* are used, respectively. A consequence of the definitions, is that a subset X of $\Re^n$ is open if and only if for every $x \in X$ there is an open sphere that is centered at x and is contained in X. A *neighborhood* of a vector x is an open set containing x.

Definition 1.1.3: We say that x is an *interior point* of a subset X of $\Re^n$ if there exists a neighborhood of x that is contained in X. The set of all interior points of X is called the *interior* of X, and is denoted by $\text{int}(X)$. A vector $x \in \text{cl}(X)$ which is not an interior point of X is said to be a *boundary point* of X. The set of all boundary points of X is called the *boundary* of X.

Proposition 1.1.6:

(a) The union of a finite collection of closed sets is closed.

(b) The intersection of any collection of closed sets is closed.

(c) The union of any collection of open sets is open.

(d) The intersection of a finite collection of open sets is open.

(e) A set is open if and only if all of its elements are interior points.

(f) Every subspace of $\Re^n$ is closed.

(g) A set X is compact if and only if every sequence of elements of X has a subsequence that converges to an element of X.

(h) If $\{X_k\}$ is a sequence of nonempty and compact sets such that $X_k \supset X_{k+1}$ for all k, then the intersection $\cap_{k=0}^{\infty} X_k$ is nonempty and compact.

The topological properties of sets in $\Re^n$, such as being open, closed, or compact, do not depend on the norm being used. This is a consequence of the following proposition, referred to as the *norm equivalence property in* $\Re^n$, which shows that if a sequence converges with respect to one norm, it converges with respect to all other norms.

Proposition 1.1.7: For any two norms $\|\cdot\|$ and $\|\cdot\|'$ on $\Re^n$, there exists a scalar c such that $\|x\| \leq c\|x\|'$ for all $x \in \Re^n$.

Using the preceding proposition, we obtain the following.

Proposition 1.1.8: If a subset of $\Re^n$ is open (respectively, closed, bounded, or compact) with respect to some norm, it is open (respectively, closed, bounded, or compact) with respect to all other norms.

Sequences of Sets

Let $\{X_k\}$ be a sequence of nonempty subsets of $\Re^n$. The *outer limit* of $\{X_k\}$, denoted $\limsup_{k\to\infty} X_k$, is the set of all $x \in \Re^n$ such that every neighborhood of x has a nonempty intersection with infinitely many of the sets X_k, $k = 1, 2, \ldots$. Equivalently, $\limsup_{k\to\infty} X_k$ is the set of all limit points of sequences $\{x_k\}$ such that $x_k \in X_k$ for all $k = 1, 2, \ldots$.

The *inner limit* of $\{X_k\}$, denoted $\liminf_{k\to\infty} X_k$, is the set of all $x \in \Re^n$ such that every neighborhood of x has a nonempty intersection with all except finitely many of the sets X_k, $k = 1, 2, \ldots$. Equivalently, $\liminf_{k\to\infty} X_k$ is the set of all limits of convergent sequences $\{x_k\}$ such that $x_k \in X_k$ for all $k = 1, 2, \ldots$.

The sequence $\{X_k\}$ is said to converge to a set X if

$$X = \liminf_{k\to\infty} X_k = \limsup_{k\to\infty} X_k.$$

In this case, X is called the *limit of* $\{X_k\}$, and is denoted by $\lim_{k\to\infty} X_k$.

The inner and outer limits are closed (possibly empty) sets. If each set X_k consists of a single point x_k, $\limsup_{k\to\infty} X_k$ is the set of limit points of $\{x_k\}$, while $\liminf_{k\to\infty} X_k$ is just the limit of $\{x_k\}$ if $\{x_k\}$ converges, and otherwise it is empty.

Continuity

Let $f : X \mapsto \Re^m$ be a function, where X is a subset of $\Re^n$, and let x be a vector in X. If there exists a vector $y \in \Re^m$ such that the sequence $\big\{f(x_k)\big\}$ converges to y for every sequence $\{x_k\} \subset X$ such that $\lim_{k\to\infty} x_k = x$, we write $\lim_{z\to x} f(z) = y$. If there exists a vector $y \in \Re^m$ such that the sequence $\big\{f(x_k)\big\}$ converges to y for every sequence $\{x_k\} \subset X$ such that $\lim_{k\to\infty} x_k = x$ and $x_k \leq x$ (respectively, $x_k \geq x$) for all k, we write $\lim_{z\uparrow x} f(z) = y$ [respectively, $\lim_{z\downarrow x} f(z)$].

Definition 1.1.4: Let X be a subset of $\Re^n$.

(a) A function $f : X \mapsto \Re^m$ is called *continuous* at a vector $x \in X$ if $\lim_{z \to x} f(z) = f(x)$.

(b) A function $f : X \mapsto \Re^m$ is called *right-continuous* (respectively, *left-continuous*) at a vector $x \in X$ if $\lim_{z \downarrow x} f(z) = f(x)$ [respectively, $\lim_{z \uparrow x} f(z) = f(x)$].

(c) A real-valued function $f : X \mapsto \Re$ is called *upper semicontinuous* (respectively, *lower semicontinuous*) at a vector $x \in X$ if $f(x) \geq \limsup_{k \to \infty} f(x_k)$ [respectively, $f(x) \leq \liminf_{k \to \infty} f(x_k)$] for every sequence $\{x_k\} \subset X$ that converges to x.

If $f : X \mapsto \Re^m$ is continuous at every vector in a subset of its domain X, we say that f *is continuous over that subset*. If $f : X \mapsto \Re^m$ is continuous at every vector in its domain X, we say that f *is continuous*. We use similar terminology for right-continuous, left-continuous, upper semicontinuous, and lower semicontinuous functions.

Proposition 1.1.9:

(a) Any vector norm on $\Re^n$ is a continuous function.

(b) Let $f : \Re^m \mapsto \Re^p$ and $g : \Re^n \mapsto \Re^m$ be continuous functions. The composition $f \cdot g : \Re^n \mapsto \Re^p$, defined by $(f \cdot g)(x) = f\big(g(x)\big)$, is a continuous function.

(c) Let $f : \Re^n \mapsto \Re^m$ be continuous, and let Y be an open (respectively, closed) subset of $\Re^m$. Then the inverse image of Y, $\big\{x \in \Re^n \mid f(x) \in Y\big\}$, is open (respectively, closed).

(d) Let $f : \Re^n \mapsto \Re^m$ be continuous, and let X be a compact subset of $\Re^n$. Then the image of X, $\big\{f(x) \mid x \in X\big\}$, is compact.

Matrix Norms

A norm $\|\cdot\|$ on the set of $n \times n$ matrices is a real-valued function that has the same properties as vector norms do when the matrix is viewed as a vector in $\Re^{n^2}$. The norm of an $n \times n$ matrix A is denoted by $\|A\|$.

An important class of matrix norms are *induced norms*, which are constructed as follows. Given any vector norm $\|\cdot\|$, the corresponding induced matrix norm, also denoted by $\|\cdot\|$, is defined by

$$\|A\| = \sup_{\|x\|=1} \|Ax\|.$$

It is easily verified that for any vector norm, the above equation defines a matrix norm.

Let $\|\cdot\|$ denote the Euclidean norm. Then by the Schwarz inequality (Prop. 1.1.2), we have

$$\|A\| = \sup_{\|x\|=1} \|Ax\| = \sup_{\|y\|=\|x\|=1} |y'Ax|.$$

By reversing the roles of x and y in the above relation and by using the equality $y'Ax = x'A'y$, it follows that $\|A\| = \|A'\|$.

1.1.3 Square Matrices

Definition 1.1.5: A square matrix A is called *singular* if its determinant is zero. Otherwise it is called *nonsingular* or *invertible*.

Proposition 1.1.10:

(a) Let A be an $n \times n$ matrix. The following are equivalent:

(i) The matrix A is nonsingular.

(ii) The matrix A' is nonsingular.

(iii) For every nonzero $x \in \Re^n$, we have $Ax \neq 0$.

(iv) For every $y \in \Re^n$, there is a unique $x \in \Re^n$ such that $Ax = y$.

(v) There is an $n \times n$ matrix B such that $AB = I = BA$.

(vi) The columns of A are linearly independent.

(vii) The rows of A are linearly independent.

(b) Assuming that A is nonsingular, the matrix B of statement (v) (called the *inverse* of A and denoted by A^{-1}) is unique.

(c) For any two square invertible matrices A and B of the same dimensions, we have $(AB)^{-1} = B^{-1}A^{-1}$.

Definition 1.1.6: A symmetric $n \times n$ matrix A is called *positive definite* if $x'Ax > 0$ for all $x \in \Re^n$, $x \neq 0$. It is called *positive semidefinite* if $x'Ax \geq 0$ for all $x \in \Re^n$.

Throughout this book, the notion of positive definiteness applies exclusively to symmetric matrices. Thus *whenever we say that a matrix is positive (semi)definite, we implicitly assume that the matrix is symmetric*, although we usually add the term "symmetric" for clarity.

Proposition 1.1.11:

(a) A square matrix is symmetric and positive definite if and only if it is invertible and its inverse is symmetric and positive definite.

(b) The sum of two symmetric positive semidefinite matrices is positive semidefinite. If one of the two matrices is positive definite, the sum is positive definite.

(c) If A is a symmetric positive semidefinite $n \times n$ matrix and T is an $m \times n$ matrix, then the matrix TAT' is positive semidefinite. If A is positive definite and T is invertible, then TAT' is positive definite.

(d) If A is a symmetric positive definite $n \times n$ matrix, there exist positive scalars $\underline{\gamma}$ and $\overline{\gamma}$ such that

$$\underline{\gamma}\|x\|^2 \leq x'Ax \leq \overline{\gamma}\|x\|^2, \qquad \forall\ x \in \Re^n.$$

(e) If A is a symmetric positive definite $n \times n$ matrix, there exists a unique symmetric positive definite matrix that yields A when multiplied with itself. This matrix is called the *square root of* A. It is denoted by $A^{1/2}$, and its inverse is denoted by $A^{-1/2}$.

1.1.4 Derivatives

Let $f : \Re^n \mapsto \Re$ be some function, fix some $x \in \Re^n$, and consider the expression

$$\lim_{\alpha \to 0} \frac{f(x + \alpha e_i) - f(x)}{\alpha},$$

where e_i is the ith unit vector (all components are 0 except for the ith component which is 1). If the above limit exists, it is called the ith *partial derivative* of f at the vector x and it is denoted by $(\partial f/\partial x_i)(x)$ or $\partial f(x)/\partial x_i$ (x_i in this section will denote the ith component of the vector x). Assuming all of these partial derivatives exist, the *gradient* of f at x is defined as the column vector

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}.$$

For any $y \in \Re^n$, we define the one-sided *directional derivative* of f in the direction y, to be

$$f'(x; y) = \lim_{\alpha \downarrow 0} \frac{f(x + \alpha y) - f(x)}{\alpha},$$

provided that the limit exists.

If the directional derivative of f at a vector x exists in all directions y and $f'(x; y)$ is a linear function of y, we say that f is *differentiable* at x. This type of differentiability is also called *Gateaux differentiability*. It is seen that f is differentiable at x if and only if the gradient $\nabla f(x)$ exists and satisfies

$$\nabla f(x)'y = f'(x; y), \qquad \forall\ y \in \Re^n.$$

The function f is called *differentiable over a subset U of $\Re^n$* if it is differentiable at every $x \in U$. The function f is called *differentiable* (without qualification) if it is differentiable at all $x \in \Re^n$.

If f is differentiable over an open set U and $\nabla f(\cdot)$ is continuous at all $x \in U$, f is said to be *continuously differentiable over U*. It can then be shown that

$$\lim_{y \to 0} \frac{f(x + y) - f(x) - \nabla f(x)'y}{\|y\|} = 0, \qquad \forall\ x \in U, \tag{1.1}$$

where $\|\cdot\|$ is an arbitrary vector norm. If f is continuously differentiable over $\Re^n$, then f is also called a *smooth* function. If f is not smooth, it is referred to as being *nonsmooth*.

The preceding equation can also be used as an alternative definition of differentiability. In particular, f is called *Frechet differentiable* at x if there exists a vector g satisfying Eq. (1.1) with $\nabla f(x)$ replaced by g. If such a vector g exists, it can be seen that all the partial derivatives $(\partial f / \partial x_i)(x)$ exist and that $g = \nabla f(x)$. Frechet differentiability implies (Gateaux) differentiability but not conversely (see for example Ortega and Rheinboldt [OrR70] for a detailed discussion). In this book, when dealing with a differentiable function f, we will always assume that f is continuously differentiable over some open set [$\nabla f(\cdot)$ is a continuous function over that set], in which case f is both Gateaux and Frechet differentiable, and the distinctions made above are of no consequence.

The definitions of differentiability of f at a vector x only involve the values of f in a neighborhood of x. Thus, these definitions can be used for functions f that are not defined on all of $\Re^n$, but are defined instead in a neighborhood of the vector at which the derivative is computed. In particular, for functions $f : X \mapsto \Re$, where X is a strict subset of $\Re^n$, we use the above definition of differentiability of f at a vector x, *provided x is an interior point of the domain X*. Similarly, we use the above definition of continuous differentiability of f over a subset U, *provided U is an open subset of the domain X*. Thus any mention of continuous differentiability of a function over a subset implicitly assumes that this subset is open.

Differentiation of Vector-Valued Functions

A vector-valued function $f : \Re^n \mapsto \Re^m$ is called differentiable (or smooth) if each component f_i of f is differentiable (or smooth, respectively). The *gradient matrix* of f, denoted $\nabla f(x)$, is the $n \times m$ matrix whose ith column is the gradient $\nabla f_i(x)$ of f_i. Thus,

$$\nabla f(x) = \Big[\nabla f_1(x) \cdots \nabla f_m(x)\Big].$$

The transpose of ∇f is called the *Jacobian* of f and is a matrix whose ijth entry is equal to the partial derivative $\partial f_i / \partial x_j$.

Now suppose that each one of the partial derivatives of a function $f : \Re^n \mapsto \Re$ is a smooth function of x. We use the notation $(\partial^2 f / \partial x_i \partial x_j)(x)$ to indicate the ith partial derivative of $\partial f / \partial x_j$ at a vector $x \in \Re^n$. The *Hessian* of f is the matrix whose ijth entry is equal to $(\partial^2 f / \partial x_i \partial x_j)(x)$, and is denoted by $\nabla^2 f(x)$. We have $(\partial^2 f / \partial x_i \partial x_j)(x) = (\partial^2 f / \partial x_j \partial x_i)(x)$ for every x, which implies that $\nabla^2 f(x)$ is symmetric.

If $f : \Re^{m+n} \mapsto \Re$ is a function of (x, y), where $x \in \Re^m$ and $y \in \Re^n$, and $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$ denote the components of x and y, respectively, we write

$$\nabla_x f(x,y) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x,y)}{\partial x_m} \end{bmatrix}, \qquad \nabla_y f(x,y) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial y_1} \\ \vdots \\ \frac{\partial f(x,y)}{\partial y_n} \end{bmatrix}.$$

We denote by $\nabla^2_{xx} f(x,y)$, $\nabla^2_{xy} f(x,y)$, and $\nabla^2_{yy} f(x,y)$ the matrices with components

$$\big[\nabla^2_{xx} f(x,y)\big]_{ij} = \frac{\partial^2 f(x,y)}{\partial x_i \partial x_j}, \qquad \big[\nabla^2_{xy} f(x,y)\big]_{ij} = \frac{\partial^2 f(x,y)}{\partial x_i \partial y_j},$$

$$\big[\nabla^2_{yy} f(x,y)\big]_{ij} = \frac{\partial^2 f(x,y)}{\partial y_i \partial y_j}.$$

If $f : \Re^{m+n} \mapsto \Re^r$, and $f_1, f_2, \ldots, f_r$ are the component functions of f, we write

$$\nabla_x f(x,y) = \big[\nabla_x f_1(x,y) \cdots \nabla_x f_r(x,y)\big],$$

$$\nabla_y f(x,y) = \big[\nabla_y f_1(x,y) \cdots \nabla_y f_r(x,y)\big].$$

Let $f : \Re^k \mapsto \Re^m$ and $g : \Re^m \mapsto \Re^n$ be smooth functions, and let h be their composition, i.e.,

$$h(x) = g\big(f(x)\big).$$

Then, the *chain rule* for differentiation states that

$$\nabla h(x) = \nabla f(x)\nabla g\big(f(x)\big), \qquad \forall\ x \in \Re^k.$$

Some examples of useful relations that follow from the chain rule are:

$$\nabla\big(f(Ax)\big) = A'\nabla f(Ax), \qquad \nabla^2\big(f(Ax)\big) = A'\nabla^2 f(Ax)A,$$

where A is a matrix,

$$\nabla_x\Big(f\big(h(x), y\big)\Big) = \nabla h(x)\nabla_h f\big(h(x), y\big),$$

$$\nabla_x\Big(f\big(h(x), g(x)\big)\Big) = \nabla h(x)\nabla_h f\big(h(x), g(x)\big) + \nabla g(x)\nabla_g f\big(h(x), g(x)\big).$$

Differentiation Theorems

We now state some theorems relating to differentiable functions that will be useful for our purposes.

Proposition 1.1.12: (Mean Value Theorem) Let $f : \Re^n \mapsto \Re$ be continuously differentiable over an open sphere S, and let x be a vector in S. Then for all y such that $x + y \in S$, there exists an $\alpha \in [0, 1]$ such that

$$f(x + y) = f(x) + \nabla f(x + \alpha y)'y.$$

Proposition 1.1.13: (Second Order Expansions) Let $f : \Re^n \mapsto \Re$ be twice continuously differentiable over an open sphere S, and let x be a vector in S. Then for all y such that $x + y \in S$:

(a) There exists an $\alpha \in [0, 1]$ such that

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

(b) We have

$$f(x + y) = f(x) + y'\nabla f(x) + \tfrac{1}{2}y'\nabla^2 f(x)y + o\big(\|y\|^2\big).$$

Proposition 1.1.14: (Implicit Function Theorem) Consider a function $f : \Re^{n+m} \mapsto \Re^m$ of $x \in \Re^n$ and $y \in \Re^m$ such that:

(1) $f(\overline{x}, \overline{y}) = 0$.

(2) f is continuous, and has a continuous and nonsingular gradient matrix $\nabla_y f(x, y)$ in an open set containing $(\overline{x}, \overline{y})$.

Then there exist open sets $S_{\overline{x}} \subset \Re^n$ and $S_{\overline{y}} \subset \Re^m$ containing $\overline{x}$ and $\overline{y}$, respectively, and a continuous function $\phi : S_{\overline{x}} \mapsto S_{\overline{y}}$ such that $\overline{y} = \phi(\overline{x})$ and $f\big(x, \phi(x)\big) = 0$ for all $x \in S_{\overline{x}}$. The function ϕ is unique in the sense that if $x \in S_{\overline{x}}$, $y \in S_{\overline{y}}$, and $f(x, y) = 0$, then $y = \phi(x)$. Furthermore, if for some integer $p > 0$, f is p times continuously differentiable the same is true for ϕ, and we have

$$\nabla \phi(x) = -\nabla_x f\big(x, \phi(x)\big) \Big(\nabla_y f\big(x, \phi(x)\big)\Big)^{-1}, \qquad \forall\, x \in S_{\overline{x}}.$$

As a final word of caution to the reader, let us mention that one can easily get confused with gradient notation and its use in various formulas, such as for example the order of multiplication of various gradients in the chain rule and the Implicit Function Theorem. Perhaps the safest guideline to minimize errors is to remember our conventions:

(a) A vector is viewed as a column vector (an $n \times 1$ matrix).

(b) The gradient ∇f of a scalar function $f : \Re^n \mapsto \Re$ is also viewed as a column vector.

(c) The gradient matrix ∇f of a vector function $f : \Re^n \mapsto \Re^m$ with components $f_1, \ldots, f_m$ is the $n \times m$ matrix whose columns are the vectors $\nabla f_1, \ldots, \nabla f_m$.

With these rules in mind, one can use "dimension matching" as an effective guide to writing correct formulas quickly.

1.2 CONVEX SETS AND FUNCTIONS

In this and the subsequent sections of this chapter, we introduce some of the basic notions relating to convex sets and functions. This material permeates all subsequent developments in this book, and will be used in the next chapter for the discussion of important issues in optimization.

We first define convex sets (see also Fig. 1.2.1).

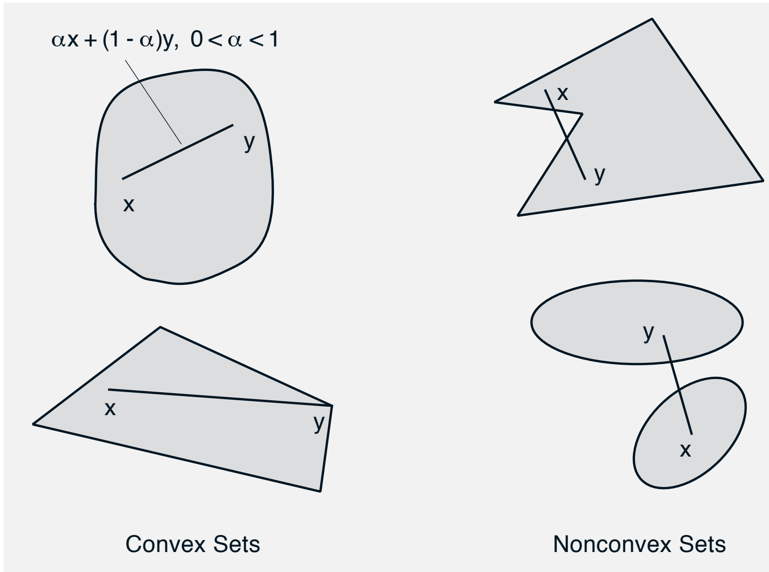


Figure 1.2.1. Illustration of the definition of a convex set. For convexity, linear interpolation between any two points of the set must yield points that lie within the set.

Definition 1.2.1: 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 lists some operations that preserve convexity of a set.

Proposition 1.2.1:

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

(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. For example, 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 Exercise 1.1. The proof of part (e) is similar to the proof of part (b).

To prove part (d), we take 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 take 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 X, 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 X. Hence $\alpha x + (1 - \alpha)y$ belongs to the interior of C, showing that the interior of C is convex. **Q.E.D.**

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. 1.2.2). Several of the results of the preceding proposition have analogs for cones (see Exercise 1.2).

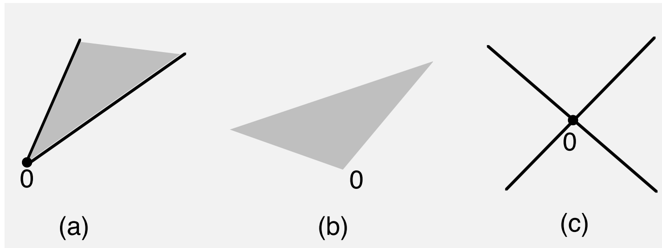


Figure 1.2.2. Illustration of convex and nonconvex cones. Cones (a) and (b) are convex, while cone (c), which consists of two lines passing through the origin, is not convex. Cone (b) does not contain the origin.

Convex Functions

The notion of a convex function is defined below and is illustrated in Fig. 1.2.3.

Definition 1.2.2: Let C be a convex subset of $\Re^n$. A function $f : C \mapsto \Re$ is called *convex* if

$$f\big(\alpha x + (1-\alpha)y\big) \leq \alpha f(x) + (1-\alpha)f(y), \qquad \forall\ x, y \in C,\ \forall\ \alpha \in [0,1]. \tag{1.2}$$

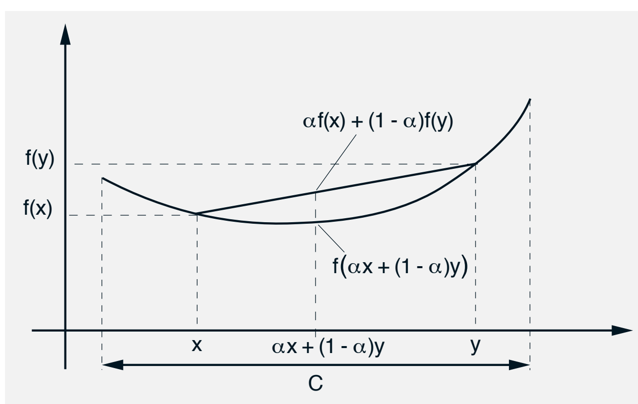


Figure 1.2.3. Illustration of the definition of a function $f : C \mapsto \Re$ that is convex. The linear interpolation $\alpha f(x) + (1-\alpha)f(y)$ overestimates the function value $f\big(\alpha x + (1-\alpha)y\big)$ for all $\alpha \in [0,1]$.

We introduce some more definitions that involve variations of the basic definition of convexity. A convex function $f : C \mapsto \Re$ is called *strictly convex* if the inequality (1.2) 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 $-f$ is convex.

Note that, according to our definition, convexity of the domain C is a prerequisite for calling a function $f : C \mapsto \Re$ "convex." Sometimes we will deal with functions $f : X \mapsto \Re$ that are defined over a (possibly nonconvex) domain X but are convex when restricted to a convex subset of their domain. The following definition formalizes this case.

Definition 1.2.3: Let C and X be subsets of $\Re^n$ such that C is nonempty and convex, and $C \subset X$. A function $f : X \mapsto \Re$ is called *convex over* C if Eq. (1.2) holds, i.e., when the domain of f is restricted to C, f becomes convex.

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, and we have

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

by the convexity of f. Similarly, we can show 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). However, in some situations, prominently arising in the context of optimization and duality, we will encounter operations on real-valued functions that produce extended real-valued functions. 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. The same is true of conjugate convex functions to be discussed in Chapter 7.

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 $f : (0, \infty) \mapsto \Re$ 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 of $-\infty$ and ∞ at some points. Such functions can be characterized using the notions of epigraph and effective domain, which we now introduce.

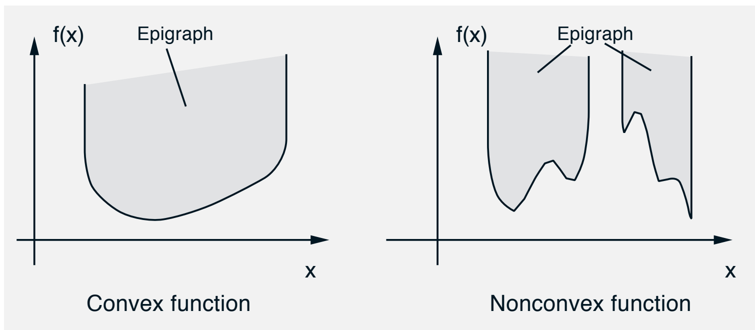


Figure 1.2.4. Illustration of the epigraphs of extended real-valued convex and nonconvex functions.

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

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

(see Fig. 1.2.4). We define the *effective domain* of f to be the set

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

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 the set of x for which $f(x) < \infty$, its epigraph remains unaffected. Similarly, if we enlarge the domain of f by defining $f(x) = \infty$ for $x \notin X$, the epigraph remains 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 may arise on occasion in proofs or other analyses, so we do not wish to exclude *a priori* such functions). The epigraph provides an effective way of dealing with this difficulty.

Definition 1.2.4: Let C be a convex subset of $\Re^n$. An extended real-valued function $f : C \mapsto [-\infty, \infty]$ is called *convex* if $\text{epi}(f)$ 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 $\text{dom}(f)$ and its level sets $\big\{x \in C \mid f(x) \leq \gamma\big\}$ 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, we have

$$f\big(\alpha x + (1-\alpha)y\big) \leq \alpha f(x) + (1-\alpha) f(y), \qquad \forall\, x, y \in C,\ \forall\, \alpha \in [0,1], \quad (1.3)$$

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

A convex function $f : C \mapsto (-\infty, \infty]$ is called *strictly convex* if the inequality (1.3) is strict for all $x, y \in \text{dom}(f)$ with $x \neq y$, 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 1.2.4.

The following definition deals with the case where an extended real-valued function becomes convex when restricted to a subset of its domain.

Definition 1.2.5: Let C and X be subsets of $\Re^n$ such that C is nonempty and convex, and $C \subset X$. An extended real-valued function $f : X \mapsto [-\infty, \infty]$ is called *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.

Note that 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, the entire subject 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].

Generally, functions that are real-valued over the entire space $\Re^n$ are more convenient (and even essential) in numerical algorithms and also in optimization analyses where a calculus-oriented approach based on differentiability is adopted. This is typically the case in nonconvex optimization, where nonlinear equality and nonconvex inequality constraints are involved (see Chapter 5). On the other hand, extended real-valued functions offer notational advantages in convex optimization, and in fact may be more natural because some basic constructions around duality involve extended real-valued functions (see Chapters 6 and 7). Since we plan to deal with nonconvex as well as convex problems, and with duality theory as well as numerical methods, we will adopt a flexible approach, and use both real-valued and extended real-valued functions.

Lower Semicontinuity and Closedness of Convex Functions

An extended real-valued function $f : X \mapsto [-\infty, \infty]$ is called *lower semicontinuous* at a vector $x \in X$ if $f(x) \le \liminf_{k\to\infty} f(x_k)$ for every sequence $\{x_k\} \subset X$ with $x_k \to x$. This is consistent with the corresponding definition for real-valued functions [cf. Definition 1.1.4(c)]. If f is lower semicontinuous at every x in a subset U of X, we say that f is *lower semicontinuous over* U. The following proposition relates lower semicontinuity of f with closedness of its epigraph and its level sets.

Proposition 1.2.2: For a function $f : \Re^n \mapsto [-\infty, \infty]$, the following are equivalent:

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

(ii) f is lower semicontinuous over $\Re^n$.

(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 $\big\{x \mid f(x) \le \gamma\big\}$ 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 $k \in \mathcal{K}$. Since the set $\big\{x \mid f(x) \leq \gamma\big\}$ is closed, $\overline{x}$ must belong to this set, 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 $\big\{(x_k, w_k)\big\} \subset \text{epi}(f)$. 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, $(\overline{x}, \overline{w}) \in \text{epi}(f)$ and $\text{epi}(f)$ is closed.

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

If the epigraph of a function $f : X \mapsto [-\infty, \infty]$ is a closed set, we say that f is a *closed* function. To understand the relation between closedness and lower semicontinuity, let us extend the domain of f to $\Re^n$ and consider the function $\tilde{f} : \Re^n \mapsto [-\infty, \infty]$ given by

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

Then, we see that f and $\tilde{f}$ have the same epigraph, and according to the preceding proposition, *f is closed if and only if $\tilde{f}$ is lower semicontinuous over $\Re^n$.*

Note, however, that if f is lower semicontinuous over $\text{dom}(f)$, it is not necessarily closed; take for example f to be constant for x in some nonclosed set and ∞ otherwise. Furthermore, if f is closed, $\text{dom}(f)$ need not be closed; for example, the function

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

is closed but $\text{dom}(f)$ is the open half-line of positive numbers. On the other hand, if $\text{dom}(f)$ is closed and f is lower semicontinuous over $\text{dom}(f)$, then f is closed because $\text{epi}(f)$ is closed, as can be seen by reviewing the proof that (ii) implies (iii) in Prop. 1.2.2. We state this as a proposition.

Proposition 1.2.3: Let $f : X \mapsto [-\infty, \infty]$ be a function. If $\text{dom}(f)$ is closed and f is lower semicontinuous over $\text{dom}(f)$, then f 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

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

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.

Recognizing Convex Functions

We can verify the convexity of a given function in a number of ways. Several commonly encountered functions are convex. For example, affine functions and norms are convex; this is straightforward to verify using the definition of convexity. In particular, for any $x, y \in \Re^n$ and any $\alpha \in [0, 1]$, by using the triangle inequality, we have

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

so the norm function $\|\cdot\|$ is convex. The exercises provide further examples of useful convex functions.

Starting with some known convex functions, we can generate other convex functions by using some common algebraic operations that preserve convexity of a function. The following proposition provides some of the necessary machinery.

Proposition 1.2.4:

(a) Let $f_i : \Re^n \mapsto (-\infty, \infty]$, $i = 1, \ldots, m$, be given functions, let $\lambda_1, \ldots, \lambda_m$ be positive scalars, and consider the function $g : \Re^n \mapsto (-\infty, \infty]$ given by

$$g(x) = \lambda_1 f_1(x) + \cdots + \lambda_m f_m(x).$$

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

(b) Let $f : \Re^m \mapsto (-\infty, \infty]$ be a given function, let A be an $m \times n$ matrix, and consider the function $g : \Re^n \mapsto (-\infty, \infty]$ given by

$$g(x) = f(Ax).$$

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

(c) Let $f_i : \Re^n \mapsto (-\infty, \infty]$ be given functions for $i \in I$, where I is an arbitrary index set, and consider the function $g : \Re^n \mapsto (-\infty, \infty]$ given by

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

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

Proof: (a) Let $f_1, \ldots, f_m$ be convex. We use the definition of convexity to write for any $x, y \in \Re^n$ and $\alpha \in [0, 1]$,

$$\begin{aligned} g\big(\alpha x + (1-\alpha)y\big) &= \sum_{i=1}^m \lambda_i f_i\big(\alpha x + (1-\alpha)y\big) \\ &\le \sum_{i=1}^m \lambda_i \big(\alpha f_i(x) + (1-\alpha) f_i(y)\big) \\ &= \alpha \sum_{i=1}^m \lambda_i f_i(x) + (1-\alpha) \sum_{i=1}^m \lambda_i f_i(y) \\ &= \alpha g(x) + (1-\alpha) g(y). \end{aligned}$$

Hence g is convex.

Let the functions $f_1, \ldots, f_m$ be closed. Then they are lower semicontinuous at every $x \in \Re^n$ (cf. Prop. 1.2.2), so for every sequence $\{x_k\}$ converging to x, we have $f_i(x) \le \liminf_{k \to \infty} f_i(x_k)$ for all i. Hence

$$g(x) \le \sum_{i=1}^m \lambda_i \liminf_{k \to \infty} f_i(x_k) \le \liminf_{k \to \infty} \sum_{i=1}^m \lambda_i f_i(x_k) = \liminf_{k \to \infty} g(x_k),$$

where we have used the assumption $\lambda_i > 0$ and Prop. 1.1.4(d) (the sum of the lower limits of sequences is less than or equal to the lower limit of the sum sequence). Therefore, g is lower semicontinuous at all $x \in \Re^n$, so by Prop. 1.2.2, it is closed.

(b) This is straightforward, along the lines of the proof of part (a).

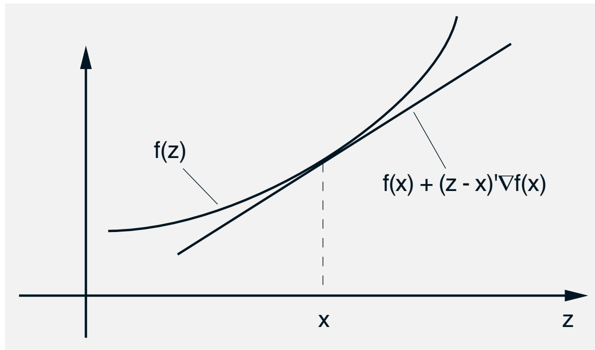


Figure 1.2.5. Characterization of convexity in terms of first derivatives. The condition $f(z) \geq f(x) + (z-x)'\nabla f(x)$ states that a linear approximation, based on the gradient, underestimates a convex function.

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

$$\mathrm{epi}(g) = \cap_{i\in I}\mathrm{epi}(f_i).$$

If the f_i are convex, the epigraphs $\mathrm{epi}(f_i)$ are convex, so $\mathrm{epi}(g)$ is convex, and g is convex. If the f_i are closed, then, by definition, the epigraphs $\mathrm{epi}(f_i)$ are closed, so $\mathrm{epi}(g)$ is closed, and g is closed. **Q.E.D.**

For once or twice differentiable functions, there are some additional useful criteria for verifying convexity, as we now proceed to discuss.

Characterizations of Differentiable Convex Functions

For differentiable functions, a useful alternative characterization of convexity is given in the following proposition and is illustrated in Fig. 1.2.5.

Proposition 1.2.5: Let C be a convex subset of $\Re^n$ and let $f : \Re^n \mapsto \Re$ be differentiable over $\Re^n$.

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

$$f(z) \geq f(x) + (z-x)'\nabla f(x), \qquad \forall\, x, z \in C. \tag{1.4}$$

(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. 1.2.6. We prove (a) and (b) simultaneously. Assume that the inequality (1.4) holds. Choose any $x, y \in C$ and $\alpha \in [0,1]$, and let $z = \alpha x + (1-\alpha)y$. Using the inequality (1.4) twice, we obtain

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

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

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

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

which proves that f is convex. If the inequality (1.4) 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.

Conversely, assume that f is convex, let x and z be any vectors in C with $x \neq z$, and for $\alpha \in (0,1)$, consider the function

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

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

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

with strict inequality if g is strictly monotonically increasing, thereby showing that the desired inequality (1.4) 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

$$\overline{\alpha} = \frac{\alpha_1}{\alpha_2}, \qquad \overline{z} = x + \alpha_2(z-x). \tag{1.5}$$

We have

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

or

$$\frac{f\big(x + \overline{\alpha}(\overline{z}-x)\big) - f(x)}{\overline{\alpha}} \leq f(\overline{z}) - f(x), \tag{1.6}$$

and the above inequalities are strict if f is strictly convex. Substituting the definitions (1.5) in Eq. (1.6), we obtain after a straightforward calculation

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

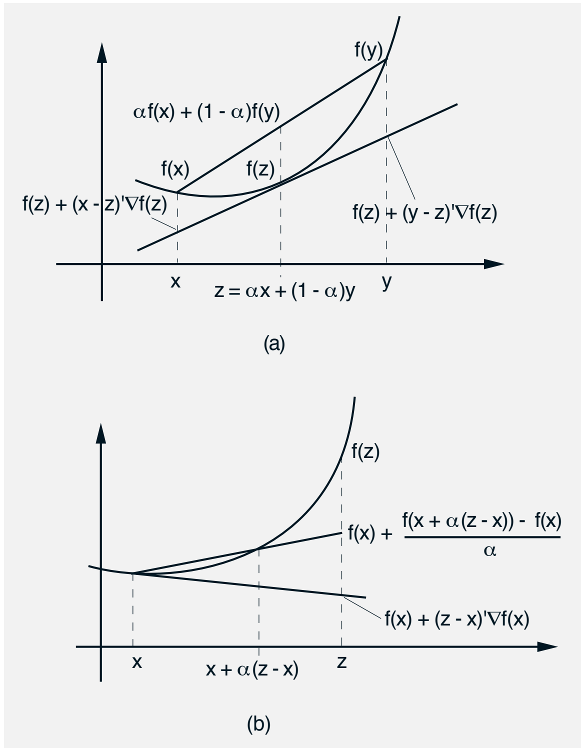


Figure 1.2.6. Geometric illustration of the ideas underlying the proof of Prop. 1.2.5. In figure (a), we linearly approximate f at $z = \alpha x + (1-\alpha)y$. The inequality (1.4) implies that

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

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

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 $f(x) + (z-x)'\nabla f(x)$. It follows that $f(z) \geq f(x) + (z-x)'\nabla f(x)$.

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. 1.2.5(a): if $f : \Re^n \mapsto \Re$ is a 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.

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

Proposition 1.2.6: Let C be a convex subset of $\Re^n$ and let $f : \Re^n \mapsto \Re$ be twice continuously differentiable over $\Re^n$.

(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) By Prop. 1.1.13(b), for all $x, y \in C$ we have

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

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

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

From Prop. 1.2.5(a), we conclude that f is convex.

(b) Similar to the proof of part (a), we have $f(y) > f(x) + (y - x)'\nabla f(x)$ for all $x, y \in C$ with $x \neq y$, and the result follows from Prop. 1.2.5(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 Prop. 1.1.13(b), we obtain $f(x + z) < f(x) + z'\nabla f(x)$, which, in view of Prop. 1.2.5(a), contradicts the convexity of f over C. **Q.E.D.**

As an example, consider the quadratic function

$$f(x) = x'Qx + a'x,$$

where Q is a symmetric $n \times n$ matrix and b is a vector in $\Re^n$. Since $\nabla^2 f(x) = 2Q$, it follows by using Prop. 1.2.6, that f is convex if and only if Q is positive semidefinite, and it is strictly convex if and only if Q is positive definite.

If f is convex over a strict subset $C \subset \Re^n$, it is not necessarily true that $\nabla^2 f(x)$ is positive semidefinite at any point of C [take for example $n = 2$, $C = \{(x_1, 0) \mid x_1 \in \Re\}$, and $f(x) = x_1^2 - x_2^2$]. The relation of convexity and twice differentiability is further considered in Exercises 1.8 and 1.9. In particular, it can be shown that the conclusion of Prop. 1.2.6(c) also holds if C has nonempty interior instead of being open.

1.3 CONVEX AND AFFINE HULLS

Let X be a nonempty subset of $\Re^n$. 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 if X is convex, then a convex combination belongs to X (see the construction of Fig. 1.3.1).

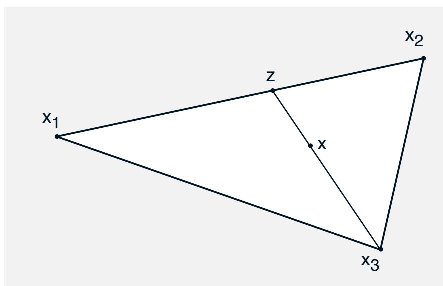


Figure 1.3.1. Illustration of the construction of a convex combination of m vectors by forming a sequence of $m - 1$ convex combinations of pairs of vectors. For example, we have

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

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. The construction shows among other things that a convex combination of a collection of vectors from a convex set belongs to the convex set.

For any function $f : \Re^n \mapsto \Re$ that is convex over X, we have

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

This follows by using repeatedly the definition of convexity together with the construction of Fig. 1.3.1. The preceding relation is a special case of a relation known as *Jensen's inequality*, and can be used to prove a number of interesting relations in applied mathematics and probability theory.

The *convex hull* of a set X, denoted $\text{conv}(X)$, is the intersection of all convex sets containing X, and is a convex set by Prop. 1.2.1(a). It is straightforward to verify that the set of all convex combinations of elements of X is convex, and is equal to $\text{conv}(X)$ (Exercise 1.14). In particular, if X consists of a finite number of vectors $x_1, \ldots, x_m$, its convex hull is

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

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. If X is a subset of $\Re^n$, the *affine hull* of X, denoted $\text{aff}(X)$, is the intersection of all affine sets containing X. Note that $\text{aff}(X)$ is itself an affine set and that it contains $\text{conv}(X)$. It can be seen that

$$\text{aff}(X) = \text{aff}\big(\text{conv}(X)\big) = \text{aff}\big(\text{cl}(X)\big),$$

(see Exercise 1.18). Furthermore, in the case where $0 \in X$, $\text{aff}(X)$ is the subspace generated by X. For a convex set C, the *dimension* of C is defined to be the dimension of $\text{aff}(C)$.

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 $\text{cone}(X)$, is the set of all nonnegative combinations of elements of X. It is easily seen that $\text{cone}(X)$ is a convex cone containing the origin, although it need not be closed even if X is compact, as shown in Fig. 1.3.2 [it can be proved that $\text{cone}(X)$ is closed in special cases, such as when X consists of a finite number of elements – this is one of the central results of polyhedral convexity, which will be shown in Section 3.2].

The following is a fundamental characterization of convex hulls (see Fig. 1.3.3).

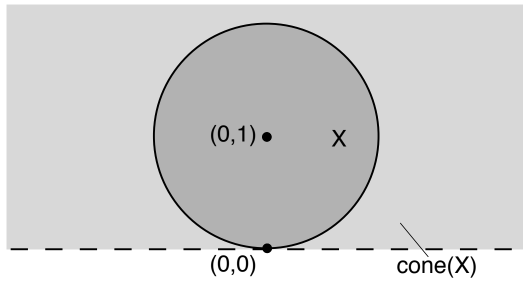


Figure 1.3.2. An example in $\Re^2$ where X is convex and compact, but cone(X) is not closed. Here

$$X = \left\{(x_1, x_2) \mid x_1^2 + (x_2 - 1)^2 \leq 1\right\},$$

$$\text{cone}(X) = \left\{(x_1, x_2) \mid x_2 > 0\right\} \cup \left\{(0, 0)\right\}.$$

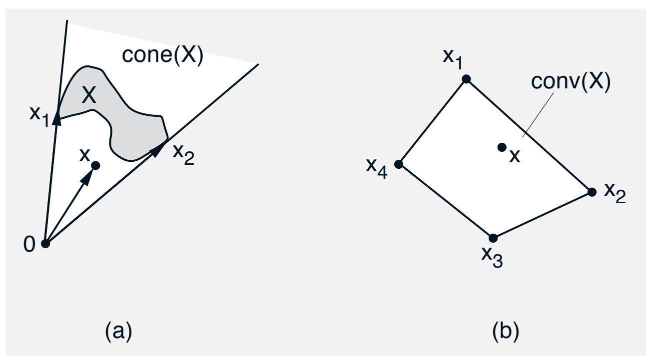


Figure 1.3.3. Illustration of Caratheodory's Theorem. In (a), X is a nonconvex set in $\Re^2$, and a point $x \in \text{cone}(X)$ is represented as a positive combination of the two linearly independent vectors $x_1, x_2 \in X$. In (b), X consists of four points x_1, x_2, x_3, x_4 in $\Re^2$, and the 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 that the vectors $x_2 - x_1, x_3 - x_1$ are linearly independent. Note also that x can alternatively be represented as a convex combination of the vectors x_1, x_2, x_4, so the representation is not unique.

Proposition 1.3.1: (Caratheodory's Theorem) Let X be a nonempty subset of $\Re^n$.

(a) Every $x \neq 0$ in cone(X) can be represented as a positive combination of vectors $x_1, \ldots, x_m$ from X that are linearly independent.

(b) Every $x \notin X$ that belongs to $\text{conv}(X)$ can be represented as a convex combination of vectors $x_1, \ldots, x_m$ from X such that $x_2 - x_1, \ldots, x_m - x_1$ are linearly independent.

Proof: (a) Let x be a nonzero vector in $\text{cone}(X)$, 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 $i = 1, \ldots, m$. If the vectors x_i were linearly dependent, there would 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 $\alpha_i - \gamma\lambda_i \geq 0$ 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.

(b) The proof will be obtained by applying part (a) to the subset of $\Re^{n+1}$ given by

$$Y = \big\{(x, 1) \mid x \in X\big\}.$$

If $x \in \text{conv}(X)$, then for some positive integer I and some positive scalars γ_i, $i = 1, \ldots, I$, with $1 = \sum_{i=1}^I \gamma_i$, we have $x = \sum_{i=1}^I \gamma_i x_i$, 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 positive scalars $\alpha_1, \ldots, \alpha_m$ and some linearly independent vectors $(x_1, 1), \ldots, (x_m, 1)$, with $x_1, \ldots, x_m \in \Re^m$ and $m \geq 2$, i.e.,

$$x = \sum_{i=1}^m \alpha_i x_i, \qquad 1 = \sum_{i=1}^m \alpha_i.$$

Assume, to arrive at a contradiction, that $x_2 - x_1, \ldots, x_m - x_1$ are linearly dependent, so that there exist $\lambda_2, \ldots, \lambda_m$, not all 0, with

$$\sum_{i=2}^m \lambda_i (x_i - x_1) = 0.$$

Equivalently, defining $\lambda_1 = -(\lambda_2 + \cdots + \lambda_m)$, we have

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

which contradicts the linear independence of $(x_1, 1), \ldots, (x_m, 1)$. **Q.E.D.**

Note that in view of the linear independence assertions in Caratheodory's Theorem, a vector in $\text{cone}(X)$ [or $\text{conv}(X)$] may be represented by no more than n (or $n + 1$, respectively) vectors of X. Note also that the proof of the theorem suggests an algorithm to obtain a representation of a vector $x \in \text{cone}(X)$ in terms of linearly independent vectors. The typical step in this algorithm is the proof's construction, which starts with a representation involving linearly dependent vectors, and yields another representation involving fewer vectors.

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

> **Proposition 1.3.2:** 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 will 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

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

is bounded, it has a limit point $\left\{(\alpha_1, \ldots, \alpha_{n+1}, x_1, \ldots, x_{n+1})\right\}$, 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 $\text{conv}(X)$, is a limit point of the sequence $\left\{\sum_{i=1}^{n+1} \alpha_i^k x_i^k\right\}$, showing that $\text{conv}(X)$ is compact. **Q.E.D.**

Note that it is not generally true that the convex hull of a closed set is closed. As an example, for the closed subset of $\Re^2$

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

the convex hull is

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

which is not closed.

1.4 RELATIVE INTERIOR, CLOSURE, AND CONTINUITY

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 is also a nonempty convex set (Prop. 1.2.1). While the interior of C may be empty, it turns out 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 $S \cap \mathrm{aff}(C) \subset C$, i.e., x is an interior point of C relative to $\mathrm{aff}(C)$. The set of all relative interior points of C is called the *relative interior of* C, and is denoted by $\mathrm{ri}(C)$. The set C is said to be *relatively open* if $\mathrm{ri}(C) = C$. A vector in the closure of C which is not a relative interior point of C is said to be a *relative boundary point* of C. The set of all relative boundary points of C is called the *relative boundary* of C.

For example, if C is a line segment connecting two distinct points in the plane, then $\mathrm{ri}(C)$ consists of all points of C except for the two end points. The relative boundary of C consists of the two end points.

The following proposition gives some basic facts about relative interior points.

Proposition 1.4.1: Let C be a nonempty convex set.

(a) (*Line Segment Principle*) If $x \in \mathrm{ri}(C)$ and $\overline{x} \in \mathrm{cl}(C)$, then all points on the line segment connecting x and $\overline{x}$, except possibly $\overline{x}$, belong to $\mathrm{ri}(C)$.

(b) (*Nonemptiness of Relative Interior*) $\mathrm{ri}(C)$ is a nonempty convex set, and has the same affine hull as C. In fact, 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 $\mathrm{aff}(C)$.

(c) $x \in \mathrm{ri}(C)$ 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 $\gamma > 1$ such that $x + (\gamma - 1)(x - \overline{x}) \in C$].

Proof: (a) For the case where $\overline{x} \in C$, the proof is given in Fig. 1.4.1. Consider the case where $\overline{x} \notin C$. To show that for any $\alpha \in (0, 1]$ we have $x_\alpha = \alpha x + (1 - \alpha)\overline{x} \in \mathrm{ri}(C)$, 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. 1.4.1, we see that $\{z \mid \|z - x_{k,\alpha}\| < \alpha\epsilon\} \cap \mathrm{aff}(C) \subset C$ for all k, where ϵ is such that the open sphere $S = \{z \mid \|z - x\| < \epsilon\}$ satisfies $S \cap \mathrm{aff}(C) \subset C$. 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 $\{z \mid \|z - x_\alpha\| < \alpha\epsilon/2\} \cap \mathrm{aff}(C) \subset C$, which shows that $x_\alpha \in \mathrm{ri}(C)$.

(b) Convexity of $\mathrm{ri}(C)$ follows from the Line Segment Principle of part (a). By using a translation argument if necessary, we assume without loss of generality that $0 \in C$. Then, the affine hull of C is a subspace whose dimension will be denoted by m. If $m = 0$, then C and $\mathrm{aff}(C)$ consist of a single point, which is a unique relative interior point. If $m > 0$, we

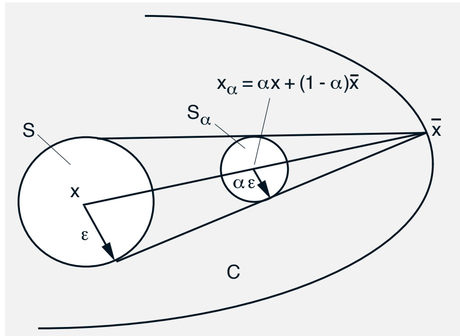


Figure 1.4.1. Proof of the Line Segment Principle for the case where $\overline{x} \in C$. Since $x \in \mathrm{ri}(C)$, there exists an open sphere $S = \{z \mid \|z - x\| < \epsilon\}$ such that $S \cap \mathrm{aff}(C) \subset C$. For all $\alpha \in (0, 1]$, let $x_\alpha = \alpha x + (1 - \alpha)\overline{x}$ and let $S_\alpha = \{z \mid \|z - x_\alpha\| < \alpha\epsilon\}$. It can be seen that each point of $S_\alpha \cap \mathrm{aff}(C)$ is a convex combination of $\overline{x}$ and some point of $S \cap \mathrm{aff}(C)$. Therefore, by the convexity of C, $S_\alpha \cap \mathrm{aff}(C) \subset C$, implying that $x_\alpha \in \mathrm{ri}(C)$.

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. 1.4.2). 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 by Prop. 1.1.11(d), we have for some positive scalar γ, which is independent of x and $\overline{x}$,

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

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

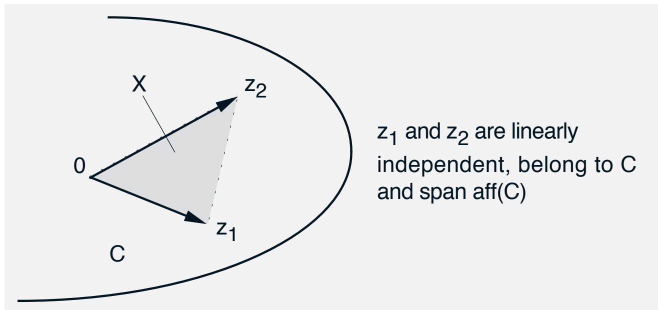


Figure 1.4.2. Construction of the relatively open set X in the proof of nonemptiness of the relative interior of a convex set C that contains the origin, assuming that $m > 0$. We choose m linearly independent vectors $z_1, \ldots, z_m \in C$, where m is the dimension of $\text{aff}(C)$, and we let

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

X contains the intersection of $\text{aff}(C)$ and an open ball centered at $\overline{x}$, so X is open relative to $\text{aff}(C)$. It follows that all points of X are relative interior points of C, so that $\text{ri}(C)$ is nonempty. Also, since by construction, $\text{aff}(X) = \text{aff}(C)$ and $X \subset \text{ri}(C)$, we see that $\text{ri}(C)$ and C have the same affine hull.

To show the last assertion of part (b), consider vectors

$$x_0 = \alpha \sum_{i=1}^{m} z_i, \qquad x_i = x_0 + \alpha z_i, \quad i = 1, \ldots, m,$$

where α is a positive scalar such that $\alpha(m+1) < 1$. The vectors $x_0, \ldots, x_m$ belong to X, and since $X \subset \text{ri}(C)$, they also belong to $\text{ri}(C)$. Furthermore, because $x_i - x_0 = \alpha z_i$ for all i and the vectors $z_1, \ldots, z_m$ span $\text{aff}(C)$, the vectors $x_1 - x_0, \ldots, x_m - x_0$ also span $\text{aff}(C)$.

(c) 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. We will show that $x \in \text{ri}(C)$. By part (b), there exists a vector $\overline{x} \in \text{ri}(C)$. We may assume that $\overline{x} \neq x$, since otherwise we are done. By the given condition, since $\overline{x}$ is in C, there is a $\gamma > 1$ such that $y = x + (\gamma - 1)(x - \overline{x}) \in C$. Then we have $x = (1 - \alpha)\overline{x} + \alpha y$, where $\alpha = 1/\gamma \in (0, 1)$, so by the Line Segment Principle, we obtain $x \in \text{ri}(C)$. **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 1.4.2: 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 Prop. 1.4.1(c), there exists a $\gamma > 1$ such that the vector

$$\hat{x} = x^* + (\gamma - 1)(x^* - x)$$

belongs to X, implying that

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

By the concavity of the function f, we have

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

and since $f(\hat{x}) \geq f(x^*)$ and $f(x) \geq f(x^*)$, we obtain

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

Hence $f(x) = f(x^*)$. **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 constraint set, since such a function cannot be constant over an open sphere.

Operations with Relative Interiors and Closures

To deal with set operations such as intersection, vector sum, linear transformation in the analysis of convex optimization problems, we need tools for calculating the corresponding relative interiors and closures. These tools are provided in the next three propositions.

Proposition 1.4.3: Let C be a nonempty convex set.

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

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

(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. 1.4.1(b)], and assume that $\overline{x} \neq x$ (otherwise we are done). By the Line Segment Principle [Prop. 1.4.1(a)], 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 $\mathrm{aff}(C) = \mathrm{aff}\big(\mathrm{cl}(C)\big)$ (see Exercise 1.18). 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. 1.4.1(b), there exists an $x \in \mathrm{ri}(C)$. We may assume that $x \neq z$ (otherwise we are done). We choose $\gamma > 1$, with γ sufficiently close to 1 so that the vector $y = z + (\gamma - 1)(z - x)$ belongs to $\mathrm{cl}(C)$ [cf. Prop. 1.4.1(c)]. Then we have $z = (1-\alpha)x + \alpha y$ where $\alpha = 1/\gamma \in (0,1)$, so by the Line Segment Principle [Prop. 1.4.1(a)], 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.**

Proposition 1.4.4: 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 $A \cdot \mathrm{cl}(C) = \mathrm{cl}(A \cdot C)$.

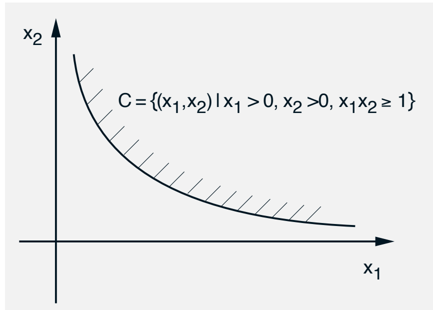


Figure 1.4.3. An example of a closed convex set C whose image $A \cdot C$ under a linear transformation A is not closed. Here

$$C = \Big\{(x_1, x_2) \mid x_1 > 0,\, x_2 > 0,\, x_1 x_2 \geq 1\Big\}$$

and A acts as projection on the horizontal axis, i.e., $A \cdot (x_1, x_2) = (x_1, 0)$. Then $A \cdot C$ is the (nonclosed) halfline $\big\{(x_1, x_2) \mid x_1 > 0,\, x_2 = 0\big\}$.

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. 1.4.3(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. 1.4.3(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 Prop. 1.4.1(c), there exists a $\gamma > 1$ such that the vector $\overline{y} = \overline{z} + (\gamma - 1)(\overline{z} - \overline{x})$ belongs to C. Thus we have $A\overline{y} \in A \cdot C$ and $A\overline{y} = z + (\gamma - 1)(z - x)$, so by Prop. 1.4.1(c), 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 above proposition]. An example is given in Fig. 1.4.3.

Proposition 1.4.5: Let C_1 and C_2 be nonempty convex sets.

(a) 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

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

(b) 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: (a) Take any $x \in \mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$ and any $y \in C_1 \cap C_2$. By Prop. 1.4.1(c), 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 the same proposition, 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

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

Assume now that $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$ is nonempty. 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. 1.4.1(a)], the vector $\alpha x + (1-\alpha)y$ belongs to $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$ for all $\alpha \in (0, 1]$. 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. 1.4.3(c), they have the same relative interior, implying that

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

We showed earlier the reverse inclusion, so we obtain $\mathrm{ri}(C_1 \cap C_2) = \mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$.

(b) Consider the linear transformation $A : \Re^{2n} \mapsto \Re^n$ given by $A(x_1, x_2) = x_1 + x_2$ for all $x_1, x_2 \in \Re^n$. The relative interior of the Cartesian product $C_1 \times C_2$ (viewed as a subset of $\Re^{2n}$) is $\mathrm{ri}(C_1) \times \mathrm{ri}(C_2)$ (see Exercise 1.37). Since $A(C_1 \times C_2) = C_1 + C_2$, from Prop. 1.4.4(a), we obtain $\mathrm{ri}(C_1 + C_2) = \mathrm{ri}(C_1) + \mathrm{ri}(C_2)$.

Similarly, the closure of the Cartesian product $C_1 \times C_2$ is $\mathrm{cl}(C_1) \times \mathrm{cl}(C_2)$ (see Exercise 1.37). From Prop. 1.4.4(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, $\mathrm{cl}(C_1) + \mathrm{cl}(C_2) \subset \mathrm{cl}(C_1 + C_2)$.

Finally, we show the reverse inclusion, assuming that C_1 is bounded. Indeed, if $x \in \mathrm{cl}(C_1 + C_2)$, 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 $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2) \neq \emptyset$ is essential in part (a) of the preceding proposition. As an example, consider the following subsets of $\Re$:

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

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 $\Re$:

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

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

The requirement that at least one of the sets C_1 and C_2 be bounded is essential in part (b) of the preceding proposition. This is illustrated by the example of Fig. 1.4.4.

We note that the results regarding the closure of the image of a closed set under a linear transformation [cf. Prop. 1.4.4(b)] and the related result regarding the closure of the vector sum of two closed sets [cf. Prop. 1.4.5(b)] will be refined in the next section by using the machinery of recession cones, which will be developed in that section.

Continuity of Convex Functions

We close this section with a basic result on the continuity properties of convex functions.

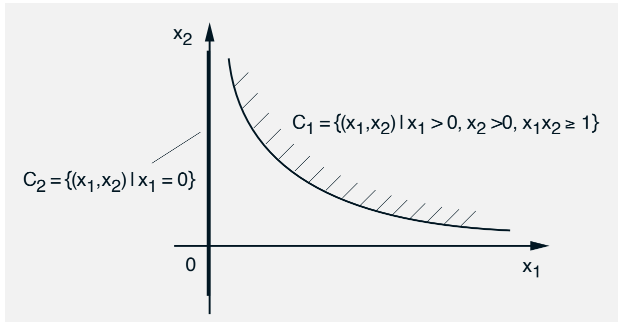


Figure 1.4.4. An example where the sum of two closed convex sets C_1 and C_2 is not closed. Here

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

and $C_1 + C_2$ is the open halfspace $\{(x_1, x_2) \mid x_1 > 0\}$.

Proposition 1.4.6: 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 $\text{dom}(f)$. 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 $f(x_k) \to f(0)$.

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

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

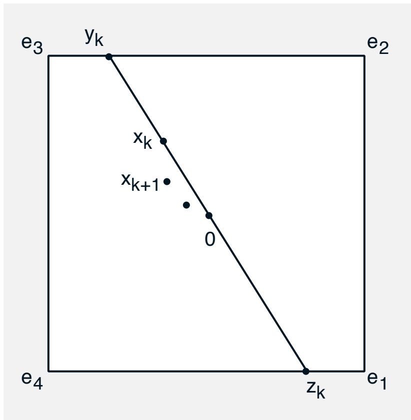


Figure 1.4.5. Construction for proving continuity of a convex function (cf. Prop. 1.4.6).

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

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

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

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

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

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

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

A straightforward consequence of the continuity of a real-valued function f that is convex over $\Re^n$ is that its epigraph as well as the level sets $\big\{x \mid f(x) \le \gamma\big\}$ for all scalars γ are closed and convex (cf. Prop. 1.2.2). Thus, a real-valued convex function is closed.

1.5 RECESSION CONES

Some of the preceding results [Props. 1.3.2, 1.4.4(b)] have illustrated how closedness and compactness of convex sets are affected by various operations such as linear transformations. In this section we take a closer look at this issue. In the process, we develop some important convexity topics that are broadly useful in optimization.

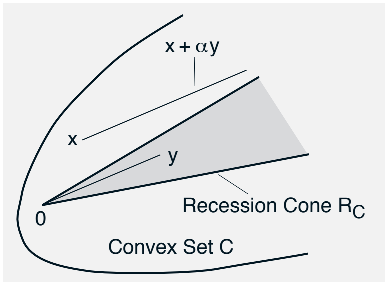


Figure 1.5.1. Illustration of the recession cone R_C of a convex set C. A direction of recession y has the property that $x+\alpha y \in C$ for all $x \in C$ and $\alpha \geq 0$.

We will first introduce the principal notions of this section: directions of recession and lineality space. We will then show how these concepts play an important role in various conditions that guarantee:

(a) The nonemptiness of the intersection of a sequence of closed convex sets.

(b) The closedness of the image of a closed convex set under a linear transformation.

It turns out that these issues lie at the heart of important questions relating to the existence of solutions of convex optimization problems, to minimax theory, and to duality theory, as we will see in subsequent chapters.

Given a nonempty convex set C, we say that a vector y is a *direction of recession* of C if $x + \alpha y \in C$ for all $x \in C$ and $\alpha \geq 0$. Thus, y is a direction of recession of C if starting at any x in C and going indefinitely along y, 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. 1.5.1). The following proposition gives some properties of recession cones.

Proposition 1.5.1: (Recession Cone Theorem) Let C be a nonempty closed convex set.

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

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

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

(d) The recession cones of C and $\text{ri}(C)$ are equal.

(e) If D is another closed convex set such that $C \cap D \neq \emptyset$, we have

$$R_{C\cap D} = R_C \cap R_D.$$

More generally, for any collection of closed convex sets C_i, $i \in I$, where I is an arbitrary index set and $\cap_{i\in I} C_i$ is nonempty, we have

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

Proof: (a) If y_1, y_2 belong to R_C and λ_1, λ_2 are positive scalars such that $\lambda_1 + \lambda_2 = 1$, we have for any $x \in C$ and $\alpha \geq 0$

$$x + \alpha(\lambda_1 y_1 + \lambda_2 y_2) = \lambda_1(x + \alpha y_1) + \lambda_2(x + \alpha y_2) \in C,$$

where the last inclusion holds because C is convex, and $x+\alpha y_1$ and $x+\alpha y_2$ belong to C by the definition of R_C. Hence $\lambda_1 y_1 + \lambda_2 y_2 \in R_C$, implying that R_C is convex.

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

(b) If $y \in R_C$, every vector $x \in C$ has the required property by the definition of R_C. Conversely, let y be such that there exists a vector $x \in C$ with $x+\alpha y \in C$ for all $\alpha \geq 0$. With no loss of generality, we assume that $y \neq 0$. We fix $\overline{x} \in C$ and $\alpha > 0$, and we show that $\overline{x} + \alpha y \in C$. It is sufficient to show that $\overline{x} + y \in C$, 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 y with y/α.

Let

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

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

$$y_k = \frac{z_k - \overline{x}}{\|z_k - \overline{x}\|}\|y\|, \qquad k = 1, 2, \ldots$$

so that $\overline{x} + y_k$ lies on the line that starts at $\overline{x}$ and passes through z_k (see the construction of Fig. 1.5.2).

We have

$$\frac{y_k}{\|y\|} = \frac{\|z_k - x\|}{\|z_k - \overline{x}\|} \cdot \frac{z_k - x}{\|z_k - x\|} + \frac{x - \overline{x}}{\|z_k - \overline{x}\|} = \frac{\|z_k - x\|}{\|z_k - \overline{x}\|} \cdot \frac{y}{\|y\|} + \frac{x - \overline{x}}{\|z_k - \overline{x}\|}.$$

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

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

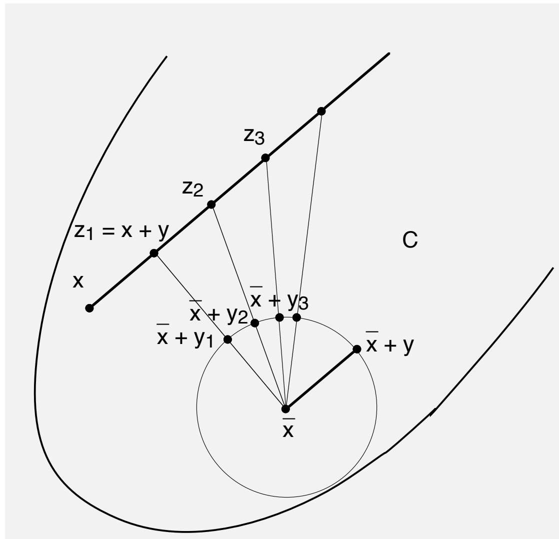


Figure 1.5.2. Construction for the proof of Prop. 1.5.1(b).

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

(c) Assuming that C is unbounded, we will show that R_C contains a nonzero direction (the reverse implication is clear). Choose any $\overline{x} \in C$ and any unbounded sequence $\{z_k\} \subset C$. Consider the sequence $\{y_k\}$, where

$$y_k = \frac{z_k - \overline{x}}{\|z_k - \overline{x}\|},$$

and let y be a limit point of $\{y_k\}$ (compare with the construction of Fig. 1.5.2). For any fixed $\alpha \geq 0$, the vector $\overline{x} + \alpha y_k$ lies between $\overline{x}$ and z_k in the line segment connecting $\overline{x}$ and z_k for all k such that $\|z_k - \overline{x}\| \geq \alpha$. Hence by the convexity of C, we have $\overline{x} + \alpha y_k \in C$ for all sufficiently large k. Since $\overline{x} + \alpha y$ is a limit point of $\{\overline{x} + \alpha y_k\}$, and C is closed, we have $\overline{x} + \alpha y \in C$. Hence, using also part (b), it follows that the nonzero vector y is a direction of recession.

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

(e) By the definition of direction of recession, $y \in R_{C\cap D}$ implies that $x + \alpha y \in C \cap D$ for all $x \in C \cap D$ and all $\alpha \geq 0$. By part (b), this in turn implies that $y \in R_C$ and $y \in R_D$, so that $R_{C\cap D} \subset R_C \cap R_D$. Conversely, by the definition of direction of recession, if $y \in R_C \cap R_D$ and $x \in C \cap D$, we have $x + \alpha y \in C \cap D$ for all $\alpha > 0$, so $y \in R_{C\cap D}$. Thus, $R_C \cap R_D \subset R_{C\cap D}$. The preceding argument can also be simply adapted to show that $R_{\cap_{i\in I} C_i} = \cap_{i\in I} R_{C_i}$. **Q.E.D.**

It is essential to assume that the set C is closed in the above 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. For an example where parts (b)-(e) fail, 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 of direction of recession, it can be verified that C has no nonzero directions of recession, so parts (b) and (c) of the proposition fail. It can also be verified that $(0, 1)$ is a direction of recession of $\mathrm{ri}(C)$, so part (d) 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 part (e) fails as well.

Note that part (e) 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 (e) to write $R_C = R_{C\cap D} = R_C \cap R_D$, from which we obtain $R_C \subset R_D$. It is essential that the sets C and D be closed in order for this property to hold.

Note also that part (c) of the above proposition yields a characterization of compact and convex sets, namely that a closed convex set C is bounded if and only if $R_C = \{0\}$. The following is a useful generalization.

Proposition 1.5.2: Let C be a nonempty closed convex subset of $\Re^n$, let W be a nonempty convex compact subset of $\Re^m$, and let A be an $m \times n$ matrix. Consider the set

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

and assume that it is nonempty. Then, V is closed and convex, and its recession cone is $R_C \cap N(A)$, where $N(A)$ is the nullspace of A. Furthermore, V is compact if and only if

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

Proof: We note that $V = C \cap \overline{V}$, where $\overline{V}$ is the set

$$\overline{V} = \{x \in \Re^n \mid Ax \in W\},$$

which is closed and convex since it is the inverse image of the closed convex set W under the continuous linear transformation A. Hence, V is closed and convex.

The recession cone of $\overline{V}$ is $N(A)$ [clearly $N(A) \subset R_{\overline{V}}$; if $y \notin N(A)$ but $y \in R_{\overline{V}}$, then for all $x \in \overline{V}$, we must have

$$Ax + \alpha Ay \in W, \qquad \forall\ \alpha > 0,$$

which contradicts the boundedness of W since $Ay \neq 0$]. Hence, since $V = C \cap \overline{V}$, V is nonempty, and the sets C and $\overline{V}$ are closed and convex, by Prop. 1.5.1(e), the recession cone of V is $R_C \cap N(A)$. Since V is closed and convex, by Prop. 1.5.1(c), it follows that V is compact if and only if $R_C \cap N(A) = \{0\}$. **Q.E.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 y whose opposite, $-y$, are also directions of recession:

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

Thus, if $y \in L_C$, then for every $x \in C$, the line $\{x + \alpha y \mid \alpha \in \Re\}$ is contained in C.

The lineality space inherits several of the properties of the recession cone that we have shown in the preceding two propositions. We collect these properties in the following proposition.

Proposition 1.5.3: Let C be a nonempty closed convex subset of $\Re^n$.

(a) The lineality space of C is a subspace of $\Re^n$.

(b) The lineality spaces of C and $\text{ri}(C)$ are equal.

(c) If D is another closed convex set such that $C \cap D \neq \emptyset$, we have

$$L_{C \cap D} = L_C \cap L_D.$$

More generally, for any collection of closed convex sets C_i, $i \in I$, where I is an arbitrary index set and $\cap_{i \in I} C_i$ is nonempty, we have

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

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

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

is nonempty, it is closed and convex, and its lineality space is $L_C \cap N(A)$, where $N(A)$ is the nullspace of A.

Proof: (a) Let y_1 and y_2 belong to L_C, and let α_1 and α_2 be nonzero scalars. We will show that $\alpha_1 y_1 + \alpha_2 y_2$ belongs to L_C. Indeed, we have

$$\begin{aligned} \alpha_1 y_1 + \alpha_2 y_2 &= |\alpha_1|\big(\mathrm{sgn}(\alpha_1) y_1\big) + |\alpha_2|\big(\mathrm{sgn}(\alpha_2) y_2\big) \\ &= \big(|\alpha_1| + |\alpha_2|\big)\big(\alpha \overline{y}_1 + (1-\alpha)\overline{y}_2\big), \end{aligned} \tag{1.9}$$

where

$$\alpha = \frac{|\alpha_1|}{|\alpha_1| + |\alpha_2|}, \qquad \overline{y}_1 = \mathrm{sgn}(\alpha_1) y_1, \quad \overline{y}_2 = \mathrm{sgn}(\alpha_2) y_2,$$

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 $\overline{y}_1$ and $\overline{y}_2$ belong to L_C, any positive multiple of a convex combination of $\overline{y}_1$ and $\overline{y}_2$ belongs to L_C. It follows from Eq. (1.9) that $\alpha_1 y_1 + \alpha_2 y_2 \in L_C$.

(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. 1.5.1(d).

(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. 1.5.1(e).

(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. 1.5.2. **Q.E.D.**

Let us also 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. 1.5.3).

Proposition 1.5.4: (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 the sum of the subspace S and its orthogonal complement $S^\perp$. Let $x \in C$, so that $x = y + z$ for some $y \in S$ and $z \in S^\perp$. Because $-y \in S$ and $S \subset L_C$, the vector $-y$ is a direction of recession of C, so the vector $x - y$, which is equal to z, belongs to C. Thus, $z \in C \cap S^\perp$, and we have $x = y + z$ with $y \in S$ and $z \in C \cap S^\perp$. This shows that $C \subset S + (C \cap S^\perp)$.

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

1.5.1 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 the intersection of a nested sequence of nonempty compact sets is nonempty and compact [cf. Prop. 1.1.6(h)]. Another property is that the image of a compact set under a linear transformation is compact [cf. Prop. 1.1.9(d)]. These properties fail for general closed sets, but it turns out that they hold under some assumptions involving convexity and directions of recession.

In what follows in this section, we will generalize the properties of compact sets just mentioned to closed convex sets. In subsequent chapters,

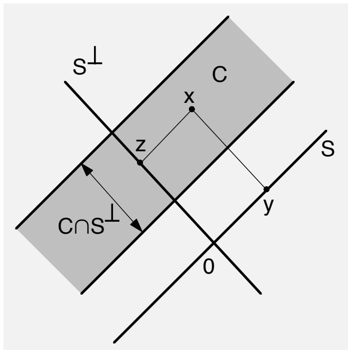


Figure 1.5.3. Illustration of the decomposition of a convex set C as

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

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

we will translate these properties into important results relating to the existence of solutions of convex optimization problems, and to fundamental issues in minimax theory and duality theory. For a glimpse into this connection, note that the set of minimizing points of a function is equal to the intersection of its nonempty level sets, so the question of existence of a solution to an optimization problem reduces to a question of nonemptiness of a set intersection.

We 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.

The following result makes no assumptions about the structure of the sets C_k, other than closedness and convexity.

Proposition 1.5.5: Let $\{C_k\}$ be a sequence of nonempty closed convex subsets of $\Re^n$ such that $C_{k+1} \subset C_k$ for all k. Let R_k and L_k be the recession cone and the lineality space of C_k, respectively, and let

$$R = \cap_{k=0}^{\infty} R_k, \qquad L = \cap_{k=0}^{\infty} L_k.$$

Assume that

$$R = L.$$

Then the intersection $\cap_{k=0}^{\infty} C_k$ is nonempty and has the form

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

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

Proof: Since the sets C_k are nested, the lineality spaces L_k are also nested [cf. Prop. 1.5.1(e)]. Since each L_k is a subspace, it follows that for all k sufficiently large, we have $L_k = L$. Thus, we may assume without loss of generality that

$$L_k = L, \qquad \forall\ k.$$

We next show by contradiction that for all sufficiently large k, we have $R_k \cap L^\perp = \{0\}$. Indeed, suppose that this is not so. Then since the R_k are nested [cf. Prop. 1.5.1(e)], for each k there exists some $y_k \in R_k \cap L^\perp$ such that $\|y_k\| = 1$. Hence the set $\{y \mid \|y\| = 1\} \cap R_k \cap L^\perp$ is nonempty, and since it is also compact, the intersection $\{y \mid \|y\| = 1\} \cap (\cap_{k=0}^\infty R_k) \cap L^\perp$ is nonempty. This intersection is equal to $\{y \mid \|y\| = 1\} \cap L \cap L^\perp$, since, by hypothesis, we have $\cap_{k=0}^\infty R_k = R = L$. But this is a contradiction since $L \cap L^\perp = \{0\}$. We may thus assume without loss of generality that

$$R_k \cap L^\perp = \{0\}, \qquad \forall\ k.$$

By the Recession Cone Theorem [Prop. 1.5.1(e)], for each k, the recession cone of $C_k \cap L^\perp$ is given by

$$R_{C_k \cap L^\perp} = R_k \cap R_{L^\perp},$$

and since $R_{L^\perp} = L^\perp$ and $R_k \cap L^\perp = \{0\}$, it follows that

$$R_{C_k \cap L^\perp} = \{0\}, \qquad \forall\ k.$$

Hence, by the Recession Cone Theorem [Prop. 1.5.1(c)], the sets $C_k \cap L^\perp$ are compact, as well as nested, so that their intersection

$$\tilde{C} = \cap_{k=0}^\infty (C_k \cap L^\perp) \tag{1.10}$$

is nonempty and compact, which implies in particular that the intersection $\cap_{k=0}^\infty C_k$ is nonempty. Furthermore, since L is the lineality space of all the sets C_k, it is also the lineality space of $\cap_{k=0}^\infty C_k$ [cf. Prop. 1.5.3(c)]. Therefore, by using the decomposition property of Prop. 1.5.4, we have

$$\cap_{k=0}^\infty C_k = L + \big(\cap_{k=0}^\infty C_k\big) \cap L^\perp,$$

implying, by Eq. (1.10), that $\cap_{k=0}^\infty C_k = L + \tilde{C}$, as required. **Q.E.D.**

Note that in the special case where $\cap_{k=0}^\infty R_k = \{0\}$, the preceding proposition shows that the intersection $\cap_{k=0}^\infty C_k$ is nonempty and compact. In fact, the proof of the proposition shows that the set C_k is compact for all sufficiently large k.

In the following two propositions, we consider the intersection of sets that are defined, at least in part, in terms of linear and/or quadratic inequalities.

Proposition 1.5.6: Let $\{C_k\}$ be a sequence of closed convex subsets of $\Re^n$, and let X be a subset of $\Re^n$ specified by linear inequality constraints, i.e.,

$$X = \{x \mid a_j'x \leq b_j,\ j = 1, \ldots, r\}, \tag{1.11}$$

where a_j are vectors in $\Re^n$ and b_j are scalars. Assume that:

(1) $C_{k+1} \subset C_k$ for all k.

(2) The intersection $X \cap C_k$ is nonempty for all k.

(3) We have

$$R_X \cap R \subset L,$$

where R_X is the recession cone of X, and

$$R = \cap_{k=0}^{\infty} R_k, \qquad L = \cap_{k=0}^{\infty} L_k,$$

with R_k and L_k denoting the recession cone and the lineality space of C_k, respectively.

Then the intersection $X \cap \big(\cap_{k=0}^{\infty} C_k\big)$ is nonempty.

Proof: We use induction on the dimension of the set X. Suppose that the dimension of X is 0. Then, X consists of a single point. By assumption (2), this point belongs to $X \cap C_k$ for all k, and hence belongs to the intersection $X \cap \big(\cap_{k=0}^{\infty} C_k\big)$.

Assume that, for some $l < n$, the intersection $\overline{X} \cap \big(\cap_{k=0}^{\infty} C_k\big)$ is nonempty for every set $\overline{X}$ of dimension less than or equal to l that is specified by linear inequality constraints, and is such that $\overline{X} \cap C_k$ is nonempty for all k and $R_{\overline{X}} \cap R \subset L$. Let X be of the form (1.11), be such that $X \cap C_k$ is nonempty for all k, satisfy $R_X \cap R \subset L$, and have dimension $l+1$. We will show that the intersection $X \cap \big(\cap_{k=0}^{\infty} C_k\big)$ is nonempty.

If $L_X \cap L = R_X \cap R$, then by Prop. 1.5.5 applied to the sets $X \cap C_k$, we have that $X \cap \big(\cap_{k=0}^{\infty} C_k\big)$ is nonempty, and we are done. We may thus assume that $L_X \cap L \neq R_X \cap R$.

Since we always have $L_X \cap L \subset R_X \cap R$, from the assumption $R_X \cap R \subset L$ it follows that there exists a nonzero direction $\overline{y} \in R_X \cap R$ such that $\overline{y} \notin L_X$, i.e.,

$$\overline{y} \in R_X, \qquad -\overline{y} \notin R_X, \qquad \overline{y} \in L.$$

Using Prop. 1.5.1(e), it is seen that the recession cone of X is

$$R_X = \{y \mid a_j'y \leq 0,\ j = 1, \ldots, r\},$$

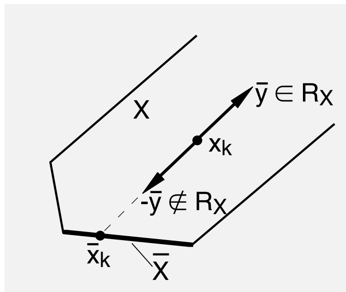


Figure 1.5.4. Construction used in the induction proof of Prop. 1.5.6.

so the fact $\overline{y} \in R_X$ implies that

$$a_j'\overline{y} \le 0, \qquad \forall\ j = 1, \ldots, r,$$

while the fact $-\overline{y} \notin R_X$ implies that the index set

$$J = \{j \mid a_j'\overline{y} < 0\}$$

is nonempty.

By assumption (2), we may select a sequence $\{x_k\}$ such that

$$x_k \in X \cap C_k, \quad \forall\ k.$$

We then have

$$a_j'x_k \le b_j, \qquad \forall\ j = 1, \ldots, r, \quad \forall\ k.$$

We may assume that

$$a_j'x_k < b_j, \qquad \forall\ j \in J, \quad \forall\ k;$$

otherwise we can replace x_k with $x_k + \overline{y}$, which belongs to $X \cap C_k$ (since $\overline{y} \in R_X$ and $\overline{y} \in L$).

Suppose that for each k, we start at x_k and move along $-\overline{y}$ as far as possible without leaving the set X, up to the point where we encounter the vector

$$\overline{x}_k = x_k - \beta_k \overline{y},$$

where β_k is the positive scalar given by

$$\beta_k = \min_{j \in J} \frac{a_j'x_k - b_j}{a_j'\overline{y}}$$

(see Fig. 1.5.4). Since $a_j'\overline{y} = 0$ for all $j \notin J$, we have $a_j'\overline{x}_k = a_j'x_k$ for all $j \notin J$, so the number of linear inequalities of X that are satisfied by $\overline{x}_k$ as equalities is strictly larger than the number of those satisfied by x_k. Thus, there exists $j_0 \in J$ such that $a_{j_0}'\overline{x}_k = b_{j_0}$ for all k in an infinite index set $\mathcal{K} \subset \{0, 1, \ldots\}$. By reordering the linear inequalities if necessary, we can assume that $j_0 = 1$, i.e.,

$$a_1'\overline{x}_k = b_1, \qquad a_1'x_k < b_1, \qquad \forall\ k \in \mathcal{K}.$$

To apply the induction hypothesis, consider the set

$$\overline{X} = \{x \mid a_1'x = b_1,\ a_j'x \le b_j,\ j = 2, \ldots, r\},$$

and note that $\{\overline{x}_k\}_{\mathcal{K}} \subset \overline{X}$. Since $\overline{x}_k = x_k - \beta_k \overline{y}$ with $x_k \in C_k$ and $\overline{y} \in L$, we have $\overline{x}_k \in C_k$ for all k, implying that $\overline{x}_k \in \overline{X} \cap C_k$ for all $k \in \mathcal{K}$. Thus, $\overline{X} \cap C_k \neq \emptyset$ for all k. Because the sets C_k are nested, so are the sets $\overline{X} \cap C_k$. Furthermore, the recession cone of $\overline{X}$ is

$$R_{\overline{X}} = \{y \mid a_1'y = 0,\ a_j'y \le 0,\ j = 2, \ldots, r\},$$

which is contained in R_X, so that

$$R_{\overline{X}} \cap R \subset R_X \cap R \subset L.$$

Finally, to show that the dimension of $\overline{X}$ is smaller than the dimension of X, note that the set $\{x \mid a_1'x = b_1\}$ contains $\overline{X}$, so that a_1 is orthogonal to the subspace $S_{\overline{X}}$ that is parallel to $\mathrm{aff}(\overline{X})$. Since $a_1'\overline{y} < 0$, it follows that $\overline{y} \notin S_{\overline{X}}$. On the other hand, $\overline{y}$ belongs to S_X, the subspace that is parallel to $\mathrm{aff}(X)$, since for all k, we have $x_k \in X$ and $x_k - \beta_k \overline{y} \in X$.

Based on the preceding, we can use the induction hypothesis to assert that the intersection $\overline{X} \cap \big(\cap_{k=0}^{\infty} C_k\big)$ is nonempty. Since $\overline{X} \subset X$, it follows that $X \cap \big(\cap_{k=0}^{\infty} C_k\big)$ is nonempty. **Q.E.D.**

Figure 1.5.5 illustrates the need for the assumptions of the preceding proposition.

Proposition 1.5.7: Let $\{C_k\}$ be a sequence of subsets of $\Re^n$ given by

$$C_k = \big\{x \mid x'Qx + a'x + b \le w_k\big\},$$

where Q is a symmetric positive semidefinite $n \times n$ matrix, a is a vector in $\Re^n$, b is a scalar, and $\{w_k\}$ is a nonincreasing scalar sequence that converges to 0. Let also X be a subset of $\Re^n$ of the form

$$X = \big\{x \mid x'Q_jx + a_j'x + b_j \le 0,\ j = 1, \ldots, r\big\}, \tag{1.12}$$

where Q_j are symmetric positive semidefinite $n \times n$ matrices, a_j are vectors in $\Re^n$, and b_j are scalars. Assume further that $X \cap C_k$ is nonempty for all k. Then, the intersection $X \cap \big(\cap_{k=0}^{\infty} C_k\big)$ is nonempty.

Proof: We note that X and all the sets C_k are closed, and that by the positive semidefiniteness of Q and Q_j, the set X and all the sets C_k are

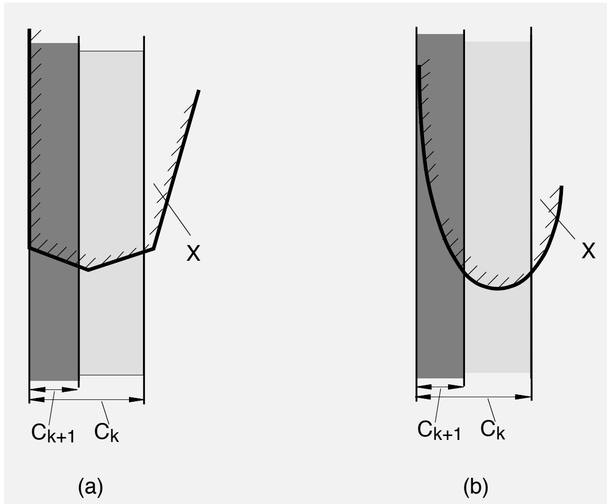


Figure 1.5.5. Illustration of the issues regarding the nonemptiness of the intersection $X \cap \big(\cap_{k=0}^{\infty} C_k\big)$ in Prop. 1.5.6, under the assumption $R_X \cap R \subset L$. Here the intersection $\cap_{k=0}^{\infty} C_k$ is equal to the left vertical line. In the figure on the left, X is specified by linear constraints and the intersection $X \cap \big(\cap_{k=0}^{\infty} C_k\big)$ is nonempty. In the figure on the right, X is specified by a nonlinear constraint, and the intersection $X \cap \big(\cap_{k=0}^{\infty} C_k\big)$ is empty.

convex [cf. Prop. 1.2.6(a)]. Furthermore, all the sets C_k have the same recession cone R and the same lineality space L, which are given by

$$R = \{y \mid Qy = 0,\ a'y \le 0\}, \qquad L = \{y \mid Qy = 0,\ a'y = 0\}.$$

[To see this, note that $y \in R$ if and only if $x + \alpha y \in R$ for all $x \in C_k$ and $\alpha > 0$, or equivalently, $(x+\alpha y)'Q(x+\alpha y) + a'(x+\alpha y) + b \le w_k$, i.e.,

$$x'Qx + a'x + \alpha(2x'Qy + a'y) + \alpha^2 y'Qy + b \le w_k, \qquad \forall\ \alpha \ge 0,\ \forall\ x \in C_k.$$

Since Q is positive semidefinite, this relation implies that $y'Qy = 0$ and that y is in the nullspace of Q, so that we must also have $a'y \le 0$. Conversely, if $y'Qy = 0$ and $a'y \le 0$, the above relation holds and $y \in R$.] Similarly, the recession cone of X is

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

We will prove the result by induction on the number r of quadratic functions that define X. For $r = 0$, we have $X \cap C_k = C_k$ for all k, and by our assumption that $X \cap C_k \ne \emptyset$, we have $C_k \ne \emptyset$ for all k. Furthermore, since $\{w_k\}$ is nonincreasing, the sets C_k are nested. If $R = L$, then by Prop.

1.5.5, it follows that $\cap_{k=0}^\infty C_k$ is nonempty and we are done, so assume that $R \neq L$. Since we always have $L \subset R$, there exists a nonzero vector $\overline{y}$ such that $\overline{y} \in R$ while $-\overline{y} \notin R$, i.e.,

$$Q\overline{y} = 0, \qquad a'\overline{y} < 0.$$

Consider a point of the form $x + \alpha\overline{y}$ for some $x \in \Re^n$ and $\alpha > 0$. We have

$$(x + \alpha\overline{y})'Q(x + \alpha\overline{y}) + a'(x + \alpha\overline{y}) = x'Qx + a'x + \alpha\, a'\overline{y}.$$

Since $a'\overline{y} < 0$, we can choose $\alpha > 0$ so that

$$x'Qx + a'x + \alpha\, a'\overline{y} + b < 0,$$

and because $\{w_k\}$ is nonincreasing, it follows that $x + \alpha\overline{y} \in \cap_{k=0}^\infty C_k$.

Assume that $\overline{X} \cap \left(\cap_{k=0}^\infty C_k\right)$ is nonempty whenever $\overline{X}$ is given by Eq. (1.12) in terms of at most r convex quadratic functions and is such that $\overline{X} \cap C_k$ is nonempty for all k. Suppose that the set X is given by Eq. (1.12) in terms of $r+1$ convex quadratic functions. If $R_X \cap R = L_X \cap L$, by using Prop. 1.5.5, we see that $X \cap \left(\cap_{k=0}^\infty C_k\right) \neq \emptyset$ and we are done, so assume that there exists a direction $\overline{y}$ such that $\overline{y} \in R_X \cap R$ but $-\overline{y} \notin R_X \cap R$. For any $x \in X$, and any $\overline{y} \in R_X \cap R$ and $\alpha > 0$, we have

$$x + \alpha\overline{y} \in X,$$

$$(x + \alpha\overline{y})'Q(x + \alpha\overline{y}) + a'(x + \alpha\,\overline{y}) = x'Qx + a'x + \alpha a'\overline{y}.$$

If $-\overline{y} \notin R$, i.e., $a'\overline{y} < 0$, it follows that for some sufficiently large α,

$$x'Qx + a'x + \alpha\, a'\overline{y} + b \leq 0,$$

implying that $x + \alpha\overline{y} \in \cap_{k=0}^\infty C_k$. Since $x + \alpha\overline{y} \in X$ for all α, it follows that $x + \alpha\overline{y} \in X \cap \left(\cap_{k=0}^\infty C_k\right)$. Thus, if $\overline{y} \in R_X \cap R$ but $-\overline{y} \notin R$, then $X \cap \left(\cap_{k=0}^\infty C_k\right)$ is nonempty.

Assume now that $\overline{y} \in R_X \cap R$ and $-\overline{y} \notin R_X$. Then $Q_j\overline{y} = 0$ and $a_j'\overline{y} \leq 0$ for all j, while $a_j'\overline{y} < 0$ for at least one j. For convenience, let us reorder the inequalities of X so that

$$Q_j\overline{y} = 0, \qquad a_j'\overline{y} = 0, \qquad \forall\ j = 1, \ldots, \overline{r}, \tag{1.13}$$

$$Q_j\overline{y} = 0, \qquad a_j'\overline{y} < 0, \qquad \forall\ j = \overline{r}+1, \ldots, r+1, \tag{1.14}$$

where $\overline{r}$ is an integer with $0 \leq \overline{r} < r+1$.

Consider now the set

$$\overline{X} = \left\{x \mid x'Q_jx + a_j'x + b_j \leq 0,\ j = 1, \ldots, \overline{r}\right\},$$

where $\overline{X} = \Re^n$ if $\overline{r} = 0$. Since $X \subset \overline{X}$ and $X \cap C_k \neq \emptyset$ for all k, the set $\overline{X} \cap C_k$ is nonempty for all k. Thus, by the induction hypothesis, it follows that the intersection

$$\overline{X} \cap \big(\cap_{k=0}^{\infty} C_k\big)$$

is nonempty. Let $\overline{x}$ be a point in this set. Since $\overline{x} \in \cap_{k=0}^{\infty} C_k$ and $\overline{y} \in R$, it follows that for all $\alpha \geq 0$,

$$\overline{x} + \alpha \overline{y} \in \cap_{k=0}^{\infty} C_k.$$

Furthermore, since $\overline{x} \in \overline{X}$, by Eq. (1.13), we have that for all $\alpha \geq 0$,

$$(\overline{x} + \alpha\overline{y})' Q_j (\overline{x} + \alpha\overline{y}) + a_j'(\overline{x} + \alpha\overline{y}) + b_j \leq 0, \qquad \forall\ j = 1, \ldots, \overline{r}.$$

Finally, in view of Eq. (1.14), we can choose a sufficiently large $\overline{\alpha} > 0$ so that

$$(\overline{x} + \overline{\alpha}\,\overline{y})' Q_j (\overline{x} + \overline{\alpha}\,\overline{y}) + a_j'(\overline{x} + \overline{\alpha}\,\overline{y}) + b_j \leq 0, \qquad \forall\ j = \overline{r}+1, \ldots, r+1.$$

The preceding three relations imply that $\overline{x} + \overline{\alpha}\,\overline{y} \in X \cap \big(\cap_{k=0}^{\infty} C_k\big)$, showing that $X \cap \big(\cap_{k=0}^{\infty} C_k\big)$ is nonempty. **Q.E.D.**

Note that it is essential to require that $\{w_k\}$ is convergent in Prop. 1.5.7. As an example, consider the subsets of $\Re^2$ given by

$$X = \big\{(x_1, x_2) \mid x_1^2 \leq x_2\big\}, \qquad C_k = \big\{(x_1, x_2) \mid x_1 \leq -k\big\}, \quad k = 0, 1, \ldots$$

Then all the assumptions of Prop. 1.5.7 are satisfied, except that the right-hand side, $-k$, of the quadratic inequality that defines C_k does not converge to a scalar. It can be seen that the intersection $X \cap \big(\cap_{k=0}^{\infty} C_k\big)$ is empty, since $\cap_{k=0}^{\infty} C_k$ is empty.

1.5.2 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, $A\,C$, of a closed convex set C under a linear transformation A. This is the subject of the following proposition.

Proposition 1.5.8: Let C be a nonempty closed convex subset of $\Re^n$, and let A be an $m \times n$ matrix with nullspace denoted by $N(A)$.

(a) If $R_C \cap N(A) \subset L_C$, then the set $A\,C$ is closed.

(b) Let X be a nonempty subset of $\Re^n$ specified by linear inequality constraints, i.e.,

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

where a_j are vectors in $\Re^n$ and b_j are scalars. If

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

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

(c) Let C be specified by convex quadratic inequalities, i.e.,

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

where Q_j are symmetric positive semidefinite $n \times n$ matrices, a_j are vectors in $\Re^n$, and b_j are scalars. Then the set $A\,C$ is closed.

Proof: (a) Let $\{y_k\}$ be a sequence of points in $A\,C$ converging to some $\overline{y} \in \Re^n$. We will prove that $A\,C$ is closed by showing that $\overline{y} \in A\,C$.

We introduce the sets

$$W_k = \big\{z \mid \|z - \overline{y}\| \le \|y_k - \overline{y}\|\big\},$$

and

$$C_k = \{x \in C \mid Ax \in W_k\}$$

(see Fig. 1.5.6). To show that $\overline{y} \in A\,C$, it is sufficient to prove that the intersection $\cap_{k=0}^{\infty} C_k$ is nonempty, since every $\overline{x} \in \cap_{k=0}^{\infty} C_k$ satisfies $\overline{x} \in C$ and $A\overline{x} = \overline{y}$ (because $y_k \to \overline{y}$). To show that $\cap_{k=0}^{\infty} C_k$ is nonempty, we will use Prop. 1.5.5.

Each set C_k is nonempty (since $y_k \in A\,C$ and $y_k \in W_k$), and it is convex and closed by Prop. 1.5.2. By taking an appropriate subsequence if necessary, we may assume that the sets C_k are nested. It can be seen that all C_k have the same recession cone, denoted by R, and the same lineality space, denoted by L, which by Props. 1.5.2 and 1.5.3(d), are given by

$$R = R_C \cap N(A), \qquad L = L_C \cap N(A). \tag{1.15}$$

Since $R_C \cap N(A) \subset L_C$, we have $R_C \cap N(A) \subset L_C \cap N(A)$, and in view of the relation $L_C \cap N(A) \subset R_C \cap N(A)$, which always holds, it follows that

$$R_C \cap N(A) = L_C \cap N(A).$$

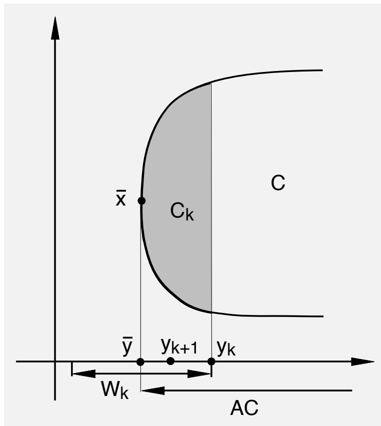


Figure 1.5.6. Construction used in the proof of Prop. 1.5.8(a). Here A is the projection on the horizontal axis of points in the plane.

This relation and Eq. (1.15) imply that $R = L$, so that by Prop. 1.5.5, the intersection $\cap_{k=0}^{\infty} C_k$ is nonempty.

(b) We use a similar argument to the one used for part (a), except that we assume that $\{y_k\} \subset A(X \cap C)$, and we use Prop. 1.5.6 to show that $X \cap \big(\cap_{k=0}^{\infty} C_k\big)$ is nonempty.

Let W_k and C_k be defined as in part (a). By our choice of $\{y_k\}$, the sets C_k are nested, so that assumption (1) of Prop. 1.5.6 is satisfied. Since $y_k \in A(X \cap C)$ and $y_k \in W_k$, it follows that $X \cap C_k$ is nonempty for all k. Thus assumption (2) of Prop. 1.5.6 is also satisfied.

Since $R_X \cap R_C \cap N(A) \subset L_C$, we also have

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

In view of this relation and Eq. (1.15), it follows that $R_X \cap R \subset L$, thus implying that assumption (3) of Prop. 1.5.6 is satisfied. Therefore, by applying Prop. 1.5.6 to the sets $X \cap C_k$, we see that the intersection $X \cap \big(\cap_{k=0}^{\infty} C_k\big)$ is nonempty. Every point in this intersection is such that $x \in X$ and $x \in C$ with $Ax = \overline{y}$, showing that $\overline{y} \in A(X \cap C)$.

(c) Similar to part (a), we let $\{y_k\}$ be a sequence in $A\,C$ converging to some $\overline{y} \in \Re^n$. We will show that $\overline{y} \in A\,C$. We let

$$C_k = \big\{x \mid \|Ax - \overline{y}\|^2 \le \|y_k - \overline{y}\|^2\big\},$$

or equivalently

$$C_k = \big\{x \mid x'A'Ax - 2(A'\overline{y})'x + \|\overline{y}\|^2 \le \|y_k - \overline{y}\|^2\big\}.$$

Thus, C_k has the form given in Prop. 1.5.7, with

$$Q = A'A, \qquad a = -2A'\overline{y}, \qquad b = \|\overline{y}\|^2, \qquad w_k = \|y_k - \overline{y}\|^2,$$

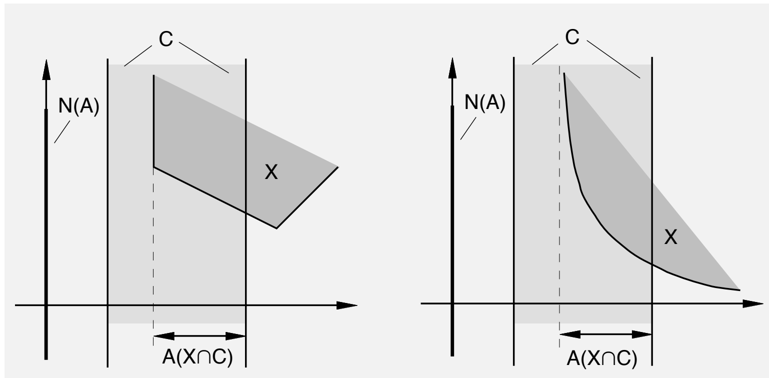


Figure 1.5.7. Illustration of the need to assume that the set X is specified by linear inequalities in Prop. 1.5.8(b). In both examples shown, the matrix A is the projection onto the horizontal axis, and its nullspace is the vertical axis. The condition $R_X \cap R_C \cap N(A) \subset L_C$ is satisfied. However, in the example on the right, X is not specified by linear inequalities, and the set $A(X \cap C)$ is not closed.

and $w_k \to 0$. By applying Prop. 1.5.7, with $X = C$, we see that the intersection $X \cap \big(\cap_{k=0}^{\infty} C_k\big)$ is nonempty. For any x in this intersection, we have $x \in C$ and $Ax = \overline{y}$ (since $y_k \to \overline{y}$), showing that $\overline{y} \in A\,C$, and implying that $A\,C$ is closed. **Q.E.D.**

Figure 1.5.7 illustrates the need for the assumptions of part (b) of the preceding proposition. Part (a) of the proposition implies that 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, then $A\,C$ is closed. This fact can be applied 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 1.5.9: Let $C_1, \ldots, C_m$ be nonempty closed convex subsets of $\Re^n$ such that the equality $y_1 + \cdots + y_m = 0$ for some vectors $y_i \in R_{C_i}$ implies that $y_i = 0$ for all $i = 1, \ldots, m$. Then the vector sum $C_1 + \cdots + C_m$ is a closed set.

Proof: Let C be the Cartesian product $C_1 \times \cdots \times C_m$ viewed as a subset of $\Re^{mn}$, and let A be the linear transformation that maps a vector $(x_1, \ldots, x_m) \in \Re^{mn}$ into $x_1 + \cdots + x_m$. Then it can be verified that C is closed and convex [see Exercise 1.37(a)]. We have

$$R_C = R_{C_1} \times \cdots \times R_{C_m}$$

[see Exercise 1.37(c)] and

$$N(A) = \big\{(y_1, \ldots, y_m) \mid y_1 + \cdots + y_m = 0,\, y_i \in \Re^n\big\},$$

so under the given condition, we obtain $R_C \cap N(A) = \{0\}$. Since $A\,C = C_1 + \cdots + C_m$, the result follows from Prop. 1.5.8(a). **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 result can be generalized by using Prop. 1.5.8(a). In particular, by using the argument of Prop. 1.5.9, we can show that if C_1 and C_2 are closed convex sets, the set $C_1 - C_2$ is closed if

$$R_{C_1} \cap R_{C_2} = L_{C_1} \cap L_{C_2}$$

(see Exercise 1.43).

Some other conditions asserting the closedness of vector sums can be derived from Prop. 1.5.8. For example, by applying Prop. 1.5.8(b), we can show that if X is specified by linear inequality constraints, 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. Furthermore, if the sets $C_1, \ldots, C_m$ are specified by convex quadratic inequalities as in Prop. 1.5.8(c), then, similar to Prop. 1.5.9, we can show that the vector sum $C_1 + \cdots + C_m$ is closed (see Exercise 1.44).

As an illustration of the need for the assumptions of Prop. 1.5.9, consider the example of Fig. 1.4.4, where C_1 and C_2 are closed sets, but $C_1 + C_2$ is not closed. In this example, the set C_1 has a nonzero direction of recession, which is the opposite of a direction of recession of C_2.

1.6 NOTES, SOURCES, AND EXERCISES

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 [Neu28], [Neu37] 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 introduced several of the topics that are fundamental in our exposition, such as the theory of subdifferentiability and the theory of conjugate convex functions.

There are several books that relate to both convex analysis and optimization. The book by Rockafellar [Roc70], widely viewed as the classic convex analysis text, contains a detailed development of convexity and convex optimization (it does not cross over into nonconvex optimization). 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. The normal cone, introduced by Mordukhovich [Mor76] and discussed in Chapter 4, and the work of Clarke on nonsmooth analysis [Cla83] play a central role in this subject.

Among other books with detailed accounts of convexity and optimization, Stoer and Witzgall [StW70] discuss similar topics as Rockafellar [Roc70] but less comprehensively. Ekeland and Temam [EkT76] 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. 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. Schrijver [Sch86] provides an extensive account of polyhedral convexity with applications to integer programming and combinatorial optimization, and gives many historical references. Ben-Tal and Nemirovski [BeN01] focus on conic and semidefinite programming. Auslender and Teboulle [AuT03] emphasize the question of existence of solutions for convex as well as nonconvex optimization problems, and associated issues in duality theory and variational inequalities. Finally, let us note a few books that focus primarily on the geometry and other properties of convex sets, but have limited connection with duality, game theory, and optimization: Bonnesen and Fenchel [BoF34], Eggleston [Egg58], Grunbaum [Gru67], Klee [Kle63], Valentine [Val64], Webster [Web94], and Barvinok [Bar02].

The development of this chapter mostly follows well-established lines. The only exception are the conditions guaranteeing the nonemptiness of a closed set intersection and the closedness of the image of a closed convex set under a linear transformation (Props. 1.5.5-1.5.8). The associated line of analysis, together with its use in minimax theory and duality theory in subsequent chapters (Sections 2.3, 2.6, and 6.5), have not received much attention, and are largely new (Nedić and Bertsekas [NeB02]; see also Bertsekas and Tseng [BeT06]). Our Props. 1.5.5 and 1.5.6 are new in the level of generality given here. Our Prop. 1.5.7 is due to Luo (see Luo and Zhang [LuZ99]). A generalization of this result to nonquadratic functions is given in Exercise 2.7 of Chapter 2 (see the material on bidirectionally flat functions). Our Prop. 1.5.6 may be derived from a special form of Helly's Theorem (Th. 21.5 in Rockafellar [Roc70], which deals with the intersection of a possibly uncountable family of sets; see also Rockafellar [Roc65]). Our induction proof of Prop. 1.5.6 is more elementary, and relies on our assumption that the family of sets is countable, which is sufficient for the analyses of this book. We note that the history and range of applications of Helly's Theorem are discussed, among others, by Danzer, Grunbaum, and Klee [DGK63], and Valentine [Val63], [Val64]. The use of recession cones in proving refined versions of Helly's Theorem and closedness of images of sets under a linear transformation was first studied by Fenchel [Fen51].

E X E R C I S E S

1.1

Let C be a nonempty subset of $\Re^n$, and let λ_1 and λ_2 be positive scalars. Show that if C is convex, then $(\lambda_1 + \lambda_2)C = \lambda_1 C + \lambda_2 C$ [cf. Prop. 1.2.1(c)]. Show by example that this need not be true when C is not convex.

1.2 (Properties of Cones)

Show that:

(a) The intersection $\cap_{i\in I} C_i$ of a collection $\{C_i \mid i \in I\}$ of cones is a cone.

(b) The Cartesian product $C_1 \times C_2$ of two cones C_1 and C_2 is a cone.

(c) The vector sum $C_1 + C_2$ of two cones C_1 and C_2 is a cone.

(d) The closure of a cone is a cone.

(e) The image and the inverse image of a cone under a linear transformation is a cone.

1.3 (Lower Semicontinuity under Composition)

(a) Let $f : \Re^n \mapsto \Re^m$ be a continuous function and $g : \Re^m \mapsto \Re$ be a lower semicontinuous function. Show that the function h defined by $h(x) = g\big(f(x)\big)$ is lower semicontinuous.

(b) Let $f : \Re^n \mapsto \Re$ be a lower semicontinuous function, and $g : \Re \mapsto \Re$ be a lower semicontinuous and monotonically nondecreasing function. Show that the function h defined by $h(x) = g\big(f(x)\big)$ is lower semicontinuous. Give an example showing that the monotonic nondecrease assumption is essential.

1.4 (Convexity under Composition)

Let C be a nonempty convex subset of $\Re^n$.

(a) Let $f : C \mapsto \Re$ be a convex function, and $g : \Re \mapsto \Re$ be a function that is convex and monotonically nondecreasing over a convex set that contains the set of values that f can take, $\big\{f(x) \mid x \in C\big\}$. Show that the function h defined by $h(x) = g\big(f(x)\big)$ is convex over C. In addition, if g is monotonically increasing and f is strictly convex, then h is strictly convex.

(b) Let $f = (f_1, \ldots, f_m)$, where each $f_i : C \mapsto \Re$ is a convex function, and let $g : \Re^m \mapsto \Re$ be a function that is convex and monotonically nondecreasing over a convex set that contains the set $\big\{f(x) \mid x \in C\big\}$, in the sense that for all $u, \overline{u}$ in this set such that $u \leq \overline{u}$, we have $g(u) \leq g(\overline{u})$. Show that the function h defined by $h(x) = g\big(f(x)\big)$ is convex over C.

1.5 (Examples of Convex Functions)

Show that the following functions from $\Re^n$ to $(-\infty, \infty]$ are convex:

(a)

$$f_1(x_1, \ldots, x_n) = \begin{cases} -(x_1 x_2 \cdots x_n)^{\frac{1}{n}} & \text{if } x_1 > 0, \ldots, x_n > 0, \\ \infty & \text{otherwise.} \end{cases}$$

(b) $f_2(x) = \ln\big(e^{x_1} + \cdots + e^{x_n}\big)$.

(c) $f_3(x) = \|x\|^p$ with $p \geq 1$.

(d) $f_4(x) = \frac{1}{f(x)}$, where f is concave and $f(x)$ is a positive number for all x.

(e) $f_5(x) = \alpha f(x) + \beta$, where $f : \Re^n \mapsto \Re$ is a convex function, and α and β are scalars such that $\alpha \geq 0$.

(f) $f_6(x) = e^{\beta x' A x}$, where A is a positive semidefinite symmetric $n \times n$ matrix and β is a positive scalar.

(g) $f_7(x) = f(Ax + b)$, where $f : \Re^m \mapsto \Re$ is a convex function, A is an $m \times n$ matrix, and b is a vector in $\Re^m$.

1.6 (Ascent/Descent Behavior of a Convex Function)

Let $f : \Re \mapsto \Re$ be a convex function.

(a) *(Monotropic Property)* Use the definition of convexity to show that f is "turning upwards" in the sense that if x_1, x_2, x_3 are three scalars such that $x_1 < x_2 < x_3$, then

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq \frac{f(x_3) - f(x_2)}{x_3 - x_2}.$$

(b) Use part (a) to show that there are four possibilities as x increases to ∞: (1) $f(x)$ decreases monotonically to $-\infty$, (2) $f(x)$ decreases monotonically to a finite value, (3) $f(x)$ reaches some value and stays at that value, (4) $f(x)$ increases monotonically to ∞ when $x \geq \overline{x}$ for some $\overline{x} \in \Re$.

1.7 (Characterization of Differentiable Convex Functions)

Let $f : \Re^n \mapsto \Re$ be a differentiable function. Show that f is convex over a nonempty convex set C if and only if

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

Note: The condition above says that the function f, restricted to the line segment connecting x and y, has monotonically nondecreasing gradient.

1.8 (Characterization of Twice Continuously Differentiable Convex Functions)

Let C be a nonempty convex subset of $\Re^n$ and let $f : \Re^n \mapsto \Re$ be twice continuously differentiable over $\Re^n$. Let S be the subspace that is parallel to the affine hull of C. Show that f is convex over C if and only if $y'\nabla^2 f(x)y \geq 0$ for all $x \in C$ and $y \in S$. [In particular, when C has nonempty interior, f is convex over C if and only if $\nabla^2 f(x)$ is positive semidefinite for all $x \in C$.]

1.9 (Strong Convexity)

Let $f : \Re^n \mapsto \Re$ be a differentiable function. We say that f *is strongly convex with coefficient* α if

$$\big(\nabla f(x) - \nabla f(y)\big)'(x - y) \geq \alpha \|x - y\|^2, \qquad \forall\ x, y \in \Re^n, \tag{1.16}$$

where α is some positive scalar.

(a) Show that if f is strongly convex with coefficient α, then f is strictly convex.

(b) Assume that f is twice continuously differentiable. Show that strong convexity of f with coefficient α is equivalent to the positive semidefiniteness of $\nabla^2 f(x) - \alpha I$ for every $x \in \Re^n$, where I is the identity matrix.

1.10 (Posynomials)

A *posynomial* is a function of positive scalar variables $y_1, \ldots, y_n$ of the form

$$g(y_1, \ldots, y_n) = \sum_{i=1}^{m} \beta_i y_1^{a_{i1}} \cdots y_n^{a_{in}},$$

where a_{ij} and β_i are scalars, such that $\beta_i > 0$ for all i. Show the following:

(a) A posynomial need not be convex.

(b) By a logarithmic change of variables, where we set

$$f(x) = \ln\big(g(y_1, \ldots, y_n)\big), \qquad b_i = \ln \beta_i, \ \forall\ i, \qquad x_j = \ln y_j, \ \forall\ j,$$

we obtain a convex function

$$f(x) = \ln \exp(Ax + b), \qquad \forall\ x \in \Re^n,$$

where $\exp(z) = e^{z_1} + \cdots + e^{z_m}$ for all $z \in \Re^m$, A is an $m \times n$ matrix with entries a_{ij}, and $b \in \Re^m$ is a vector with components b_i.

(c) Every function $g : \Re^n \mapsto \Re$ of the form

$$g(y) = g_1(y)^{\gamma_1} \cdots g_r(y)^{\gamma_r},$$

where g_k is a posynomial and $\gamma_k > 0$ for all k, can be transformed by a logarithmic change of variables into a convex function f given by

$$f(x) = \sum_{k=1}^{r} \gamma_k \ln \exp(A_k x + b_k),$$

with the matrix A_k and the vector b_k being associated with the posynomial g_k for each k.

1.11 (Arithmetic-Geometric Mean Inequality)

Show that if $\alpha_1, \ldots, \alpha_n$ are positive scalars with $\sum_{i=1}^{n} \alpha_i = 1$, then for every set of positive scalars $x_1, \ldots, x_n$, we have

$$x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n} \leq \alpha_1 x_1 + a_2 x_2 + \cdots + \alpha_n x_n,$$

with equality if and only if $x_1 = x_2 = \cdots = x_n$. *Hint*: Show that $-\ln x$ is a strictly convex function on $(0, \infty)$.

1.12 (Young and Holder Inequalities)

Use the result of Exercise 1.11 to verify Young's inequality

$$xy \le \frac{x^p}{p} + \frac{y^q}{q}, \qquad \forall\, x \ge 0,\ \forall\, y \ge 0,$$

where $p > 0$, $q > 0$, and

$$1/p + 1/q = 1.$$

Then, use Young's inequality to verify Holder's inequality

$$\sum_{i=1}^{n} |x_i y_i| \le \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \left(\sum_{i=1}^{n} |y_i|^q\right)^{1/q}.$$

1.13

Let C be a nonempty convex set in $\Re^{n+1}$, and let $f : \Re^n \mapsto [-\infty, \infty]$ be the function defined by

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

Show that f is convex.

1.14

Show that the convex hull of a nonempty set coincides with the set of all convex combinations of its elements.

1.15

Let C be a nonempty convex subset of $\Re^n$. Show that

$$\text{cone}(C) = \cup_{x \in C} \{\gamma x \mid \gamma \ge 0\}.$$

1.16 (Convex Cones)

Show that:

(a) For any collection of vectors $\{a_i \mid i \in I\}$, the set $C = \{x \mid a_i' x \le 0,\ i \in I\}$ is a closed convex cone.

(b) A cone C is convex if and only if $C + C \subset C$.

(c) For any two convex cones C_1 and C_2 containing the origin, we have

$$C_1 + C_2 = \text{conv}(C_1 \cup C_2),$$

$$C_1 \cap C_2 = \bigcup_{\alpha \in [0,1]} \big(\alpha C_1 \cap (1-\alpha) C_2\big).$$

1.17

Let $\{C_i \mid i \in I\}$ be an arbitrary collection of convex sets in $\Re^n$, and let C be the convex hull of the union of the collection. Show that

$$C = \bigcup_{\overline{I} \subset I,\ \overline{I}:\ \text{finite set}} \left\{ \sum_{i \in \overline{I}} \alpha_i C_i \;\middle|\; \sum_{i \in \overline{I}} \alpha_i = 1,\ \alpha_i \geq 0,\ \forall\ i \in \overline{I} \right\},$$

i.e., the convex hull of the union of the C_i is equal to the set of all convex combinations of vectors from the C_i.

1.18 (Convex Hulls, Affine Hulls, and Generated Cones)

Let X be a nonempty set. Show that:

(a) X, $\text{conv}(X)$, and $\text{cl}(X)$ have the same affine hull.

(b) $\text{cone}(X) = \text{cone}\big(\text{conv}(X)\big)$.

(c) $\text{aff}\big(\text{conv}(X)\big) \subset \text{aff}\big(\text{cone}(X)\big)$. Give an example where the inclusion is strict, i.e., $\text{aff}\big(\text{conv}(X)\big)$ is a strict subset of $\text{aff}\big(\text{cone}(X)\big)$.

(d) If the origin belongs to $\text{conv}(X)$, then $\text{aff}\big(\text{conv}(X)\big) = \text{aff}\big(\text{cone}(X)\big)$.

1.19

Let $\{f_i \mid i \in I\}$ be an arbitrary collection of proper convex functions $f_i : \Re^n \mapsto (-\infty, \infty]$. Define

$$f(x) = \inf \left\{ w \mid (x, w) \in \text{conv}\big(\cup_{i \in I} \text{epi}(f_i)\big) \right\}, \qquad x \in \Re^n.$$

Show that $f(x)$ is given by

$$f(x) = \inf \left\{ \sum_{i \in \overline{I}} \alpha_i f_i(x_i) \;\middle|\; \sum_{i \in \overline{I}} \alpha_i x_i = x,\ x_i \in \Re^n,\ \sum_{i \in \overline{I}} \alpha_i = 1,\ \alpha_i \geq 0,\ \forall\ i \in \overline{I}, \right.$$
$$\left. \overline{I} \subset I,\ \overline{I}:\ \text{finite} \right\}.$$

1.20 (Convexification of Nonconvex Functions)

Let X be a nonempty subset of $\Re^n$ and let $f : X \mapsto \Re$ be a function that is bounded below over X. Define the function $F : \text{conv}(X) \mapsto \Re$ by

$$F(x) = \inf \left\{ w \mid (x, w) \in \text{conv}\big(\text{epi}(f)\big) \right\}.$$

Show that:

(a) F is convex over $\mathrm{conv}(X)$ and it is given by

$$F(x) = \inf \left\{ \sum_i \alpha_i f(x_i) \;\middle|\; \sum_i \alpha_i x_i = x,\ x_i \in X,\ \sum_i \alpha_i = 1,\ \alpha_i \geq 0,\ \forall\, i \right\},$$

where the infimum is taken over all representations of x as a convex combination of elements of X (i.e., with finitely many nonzero coefficients α_i).

(b)

$$\inf_{x \in \mathrm{conv}(X)} F(x) = \inf_{x \in X} f(x).$$

(c) Every $x^* \in X$ that attains the minimum of f over X, i.e., $f(x^*) = \inf_{x \in X} f(x)$, also attains the minimum of F over $\mathrm{conv}(X)$.

1.21 (Minimization of Linear Functions)

Show that minimization of a linear function over a set is equivalent to minimization over its convex hull. In particular, if $X \subset \Re^n$ and $c \in \Re^n$, then

$$\inf_{x \in \mathrm{conv}(X)} c'x = \inf_{x \in X} c'x.$$

Furthermore, the infimum in the left-hand side above is attained if and only if the infimum in the right-hand side is attained.

1.22 (Extension of Caratheodory's Theorem)

Let X_1 and X_2 be nonempty subsets of $\Re^n$, and let $X = \mathrm{conv}(X_1) + \mathrm{cone}(X_2)$. Show that every vector x in X can be represented in the form

$$x = \sum_{i=1}^{k} \alpha_i x_i + \sum_{i=k+1}^{m} \alpha_i y_i,$$

where m is a positive integer with $m \leq n+1$, the vectors $x_1, \ldots, x_k$ belong to X_1, the vectors $y_{k+1}, \ldots, y_m$ belong to X_2, and the scalars $\alpha_1, \ldots, \alpha_m$ are nonnegative with $\alpha_1 + \cdots + \alpha_k = 1$. Furthermore, the vectors $x_2 - x_1, \ldots, x_k - x_1, y_{k+1}, \ldots, y_m$ are linearly independent.

1.23

Let X be a nonempty bounded subset of $\Re^n$. Show that

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

In particular, if X is compact, then $\mathrm{conv}(X)$ is compact (cf. Prop. 1.3.2).

1.24 (Radon's Theorem)

Let $x_1, \ldots, x_m$ be vectors in $\Re^n$, where $m \geq n+2$. Show that there exists a partition of the index set $\{1, \ldots, m\}$ into two disjoint sets I and J such that

$$\text{conv}\big(\{x_i \mid i \in I\}\big) \cap \text{conv}\big(\{x_j \mid j \in J\}\big) \neq \emptyset.$$

Hint: The system of $n+1$ equations in the m unknowns $\lambda_1, \ldots, \lambda_m$,

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

has a nonzero solution λ^*. Let $I = \{i \mid \lambda_i^* \geq 0\}$ and $J = \{j \mid \lambda_j^* < 0\}$.

1.25 (Helly's Theorem [Hel21])

Consider a finite collection of convex subsets of $\Re^n$, and assume that the intersection of every subcollection of $n+1$ (or fewer) sets has nonempty intersection. Show that the entire collection has nonempty intersection. *Hint*: Use induction. Assume that the conclusion holds for every collection of M sets, where $M \geq n+1$, and show that the conclusion holds for every collection of $M+1$ sets. In particular, let $C_1, \ldots, C_{M+1}$ be a collection of $M+1$ convex sets, and consider the collection of $M+1$ sets $B_1, \ldots, B_{M+1}$, where

$$B_j = \cap_{\substack{i=1,\ldots,M+1, \\ i \neq j}} C_i, \qquad j = 1, \ldots, M+1.$$

Note that, by the induction hypothesis, each set B_j is the intersection of a collection of M sets that have the property that every subcollection of $n+1$ (or fewer) sets has nonempty intersection. Hence each set B_j is nonempty. Let x_j be a vector in B_j. Apply Radon's Theorem (Exercise 1.24) to the vectors $x_1, \ldots, x_{M+1}$. Show that any vector in the intersection of the corresponding convex hulls belongs to the intersection of $C_1, \ldots, C_{M+1}$.

1.26

Consider the problem of minimizing over $\Re^n$ the function

$$\max\big\{f_1(x), \ldots, f_M(x)\big\},$$

where $f_i : \Re^n \mapsto (-\infty, \infty]$, $i = 1, \ldots, M$, are convex functions, and assume that the optimal value, denoted f^*, is finite. Show that there exists a subset I of $\{1, \ldots, M\}$, containing no more than $n+1$ indices, such that

$$\inf_{x \in \Re^n} \left\{ \max_{i \in I} f_i(x) \right\} = f^*.$$

Hint: Consider the convex sets $X_i = \big\{x \mid f_i(x) < f^*\big\}$, argue by contradiction, and apply Helly's Theorem (Exercise 1.25).

1.27

Let C be a nonempty convex subset of $\Re^n$, and let $f : \Re^n \mapsto (-\infty, \infty]$ be a convex function such that $f(x)$ is finite for all $x \in C$. Show that if for some scalar γ, we have $f(x) \geq \gamma$ for all $x \in C$, then we also have $f(x) \geq \gamma$ for all $x \in \text{cl}(C)$.

1.28

Let C be a nonempty convex set, and let S be the subspace that is parallel to the affine hull of C. Show that

$$\text{ri}(C) = \text{int}(C + S^{\perp}) \cap C.$$

1.29

Let $x_0, \ldots, x_m$ be vectors in $\Re^n$ such that $x_1 - x_0, \ldots, x_m - x_0$ are linearly independent. The convex hull of $x_0, \ldots, x_m$ is called an *m-dimensional simplex*, and $x_0, \ldots, x_m$ are called the *vertices* of the simplex.

(a) Show that the dimension of a convex set is the maximum of the dimensions of all the simplices contained in the set.

(b) Use part (a) to show that a nonempty convex set has a nonempty relative interior.

1.30

Let C_1 and C_2 be two nonempty convex sets such that $C_1 \subset C_2$.

(a) Give an example showing that $\text{ri}(C_1)$ need not be a subset of $\text{ri}(C_2)$.

(b) Assuming that the sets C_1 and C_2 have the same affine hull, show that $\text{ri}(C_1) \subset \text{ri}(C_2)$.

(c) Assuming that the sets $\text{ri}(C_1)$ and $\text{ri}(C_2)$ have nonempty intersection, show that $\text{ri}(C_1) \subset \text{ri}(C_2)$.

(d) Assuming that the sets C_1 and $\text{ri}(C_2)$ have nonempty intersection, show that the set $\text{ri}(C_1) \cap \text{ri}(C_2)$ is nonempty.

1.31

Let C be a nonempty convex set.

(a) Show the following refinement of Prop. 1.4.1(c): $x \in \text{ri}(C)$ if and only if for every $\overline{x} \in \text{aff}(C)$, there exists a $\gamma > 1$ such that $x + (\gamma - 1)(x - \overline{x}) \in C$.

(b) Assuming that the origin lies in $\text{ri}(C)$, show that $\text{cone}(C)$ coincides with $\text{aff}(C)$.

(c) Show the following extension of part (b) to a nonconvex set: If X is a nonempty set such that the origin lies in the relative interior of $\mathrm{conv}(X)$, then $\mathrm{cone}(X)$ coincides with $\mathrm{aff}(X)$.

1.32

Let C be a nonempty set.

(a) If C is convex and compact, and the origin is not in the relative boundary of C, then $\mathrm{cone}(C)$ is closed.

(b) Give examples showing that the assertion of part (a) fails if C is unbounded or the origin is in the relative boundary of C.

(c) If C is compact and the origin is not in the relative boundary of $\mathrm{conv}(C)$, then $\mathrm{cone}(C)$ is closed. *Hint*: Use part (a) and Exercise 1.18(b).

1.33

(a) Let C be a nonempty convex cone. Show that $\mathrm{ri}(C)$ is also a convex cone.

(b) Let $C = \mathrm{cone}\big(\{x_1, \ldots, x_m\}\big)$. Show that

$$\mathrm{ri}(C) = \left\{ \sum_{i=1}^{m} \alpha_i x_i \;\middle|\; \alpha_i > 0,\ i = 1, \ldots, m \right\}.$$

1.34

Let A be an $m \times n$ matrix and let C be a nonempty convex set in $\Re^m$. Assuming that $A^{-1} \cdot \mathrm{ri}(C)$ is nonempty, show that

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

(Compare these relations with those of Prop. 1.4.4.)

1.35 (Closure of a Convex Function)

Consider a proper convex function $f : \Re^n \mapsto (-\infty, \infty]$ and the function whose epigraph is the closure of the epigraph of f. This function is called the *closure of* f and is denoted by $\mathrm{cl}\, f$. Show that:

(a) $\mathrm{cl}\, f$ is the greatest lower semicontinuous function majorized by f, i.e., if $g : \Re^n \mapsto [-\infty, \infty]$ is lower semicontinuous 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) $\mathrm{cl}\, f$ is a closed proper convex function and

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

(c) If $x \in \text{ri}\big(\text{dom}(f)\big)$ and $y \in \text{dom}(\text{cl}\, f)$, we have

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

(d) Assume that $f = f_1 + \cdots + f_m$, where $f_i : \Re^n \mapsto (-\infty, \infty]$, $i = 1, \ldots, m$, are proper convex functions such that $\cap_{i=1}^m \text{ri}\big(\text{dom}(f_i)\big) \neq \emptyset$. Show that

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

1.36

Let C be a convex set and let M be an affine set such that the intersection $\text{cl}(C) \cap M$ is nonempty and bounded. Show that for every affine set $\overline{M}$ that is parallel to M, the intersection $C \cap \overline{M}$ is bounded.

1.37 (Properties of Cartesian Products)

Given nonempty sets $X_i \subset \Re^{n_i}$, $i = 1, \ldots, m$, let $X = X_1 \times \cdots \times X_m$ be their Cartesian product. Show that:

(a) The convex hull (closure, affine hull) of X is equal to the Cartesian product of the convex hulls (closures, affine hulls, respectively) of the X_i.

(b) If all the sets $X_1, \ldots, X_m$ contain the origin, then

$$\text{cone}(X) = \text{cone}(X_1) \times \cdots \times \text{cone}(X_m).$$

Furthermore, the result fails if one of the sets does not contain the origin.

(c) If all the sets $X_1, \ldots, X_m$ are convex, then the relative interior (recession cone) of X is equal to the Cartesian product of the relative interiors (recession cones, respectively) of the X_i.

1.38 (Recession Cones of Nonclosed Sets)

Let C be a nonempty convex set.

(a) Show that

$$R_C \subset R_{\text{cl}(C)}, \qquad \text{cl}(R_C) \subset R_{\text{cl}(C)}.$$

Give an example where the inclusion $\text{cl}(R_C) \subset R_{\text{cl}(C)}$ is strict.

(b) Let $\overline{C}$ be a closed convex set such that $C \subset \overline{C}$. Show that $R_C \subset R_{\overline{C}}$. Give an example showing that the inclusion can fail if $\overline{C}$ is not closed.

1.39 (Recession Cones of Relative Interiors)

Let C be a nonempty convex set.

(a) Show that $R_{\mathrm{ri}(C)} = R_{\mathrm{cl}(C)}$.

(b) Show that a vector y belongs to $R_{\mathrm{ri}(C)}$ if and only if there exists a vector $x \in \mathrm{ri}(C)$ such that $x + \alpha y \in \mathrm{ri}(C)$ for every $\alpha \geq 0$.

(c) Let $\overline{C}$ be a convex set such that $\overline{C} = \mathrm{ri}(\overline{C})$ and $C \subset \overline{C}$. Show that $R_C \subset R_{\overline{C}}$. Give an example showing that the inclusion can fail if $\overline{C} \neq \mathrm{ri}(\overline{C})$.

1.40

This exercise is a refinement of Prop. 1.5.6. Let $\{X_k\}$ and $\{C_k\}$ be sequences of closed convex subsets of $\Re^n$, such that the intersection

$$X = \cap_{k=0}^{\infty} X_k$$

is specified by linear inequality constraints as in Prop. 1.5.6. Assume that:

(1) $X_{k+1} \subset X_k$ and $C_{k+1} \subset C_k$ for all k.

(2) $X_k \cap C_k$ is nonempty for all k.

(3) We have

$$R_X = L_X, \qquad R_X \cap R_C \subset L_C,$$

where

$$R_X = \cap_{k=0}^{\infty} R_{X_k}, \qquad L_X = \cap_{k=0}^{\infty} L_{X_k},$$
$$R_C = \cap_{k=0}^{\infty} R_{C_k}, \qquad L_C = \cap_{k=0}^{\infty} L_{C_k}.$$

Then the intersection $\cap_{k=0}^{\infty}(X_k \cap C_k)$ is nonempty. *Hint*: Consider the sets $\overline{C}_k = X_k \cap C_k$ and the intersection $X \cap (\cap_{k=0}^{\infty} \overline{C}_k)$. Apply Prop. 1.5.6.

1.41

Let C be a nonempty convex subset of $\Re^n$ and let A be an $m \times n$ matrix. Show that if $R_{\mathrm{cl}(C)} \cap N(A) = \{0\}$, then

$$\mathrm{cl}(A \cdot C) = A \cdot \mathrm{cl}(C), \qquad A \cdot R_{\mathrm{cl}(C)} = R_{A \cdot \mathrm{cl}(C)}.$$

Give an example showing that $A \cdot R_{\mathrm{cl}(C)}$ and $R_{A \cdot \mathrm{cl}(C)}$ may differ when $R_{\mathrm{cl}(C)} \cap N(A) \neq \{0\}$.

1.42

Let C be a nonempty convex subset of $\Re^n$. Show the following refinement of Prop. 1.5.8(a) and Exercise 1.41: if A is an $m \times n$ matrix and $R_{\mathrm{cl}(C)} \cap N(A)$ is a subspace of the lineality space of $\mathrm{cl}(C)$, then

$$\mathrm{cl}(A \cdot C) = A \cdot \mathrm{cl}(C), \qquad A \cdot R_{\mathrm{cl}(C)} = R_{A \cdot \mathrm{cl}(C)}.$$

1.43 (Recession Cones of Vector Sums)

This exercise is a refinement of Prop. 1.5.9.

(a) Let $C_1, \ldots, C_m$ be nonempty closed convex subsets of $\Re^n$ such that the equality $y_1 + \cdots + y_m = 0$ with $y_i \in R_{C_i}$ implies that each y_i belongs to the lineality space of C_i. Then, the vector sum $C_1 + \cdots + C_m$ is a closed set and

$$R_{C_1+\cdots+C_m} = R_{C_1} + \cdots + R_{C_m}.$$

(b) Show the following extension of part (a) to nonclosed sets: Let $C_1, \ldots, C_m$ be nonempty convex subsets of $\Re^n$ such that the equality $y_1 + \cdots + y_m = 0$ with $y_i \in R_{\mathrm{cl}(C_i)}$ implies that each y_i belongs to the lineality space of $\mathrm{cl}(C_i)$. Then, we have

$$\mathrm{cl}(C_1 + \cdots + C_m) = \mathrm{cl}(C_1) + \cdots + \mathrm{cl}(C_m),$$

$$R_{\mathrm{cl}(C_1+\cdots+C_m)} = R_{\mathrm{cl}(C_1)} + \cdots + R_{\mathrm{cl}(C_m)}.$$

1.44

Let $C_1, \ldots, C_m$ be nonempty subsets of $\Re^n$ that are specified by convex quadratic inequalities, i.e., for all $i = 1, \ldots, n$,

$$C_i = \left\{ x \mid x'Q_{ij}x + a_{ij}'x + b_{ij} \le 0,\ j = 1, \ldots, r_i \right\},$$

where Q_{ij} are symmetric positive semidefinite $n \times n$ matrices, a_{ij} are vectors in $\Re^n$, and b_{ij} are scalars. Show that the vector sum $C_1 + \cdots + C_m$ is a closed set.

1.45 (Set Intersection and Helly's Theorem)

Show that the conclusions of Props. 1.5.5 and 1.5.6 hold if the assumption that the sets C_k are nonempty and nested is replaced by the weaker assumption that any subcollection of $n+1$ (or fewer) sets from the sequence $\{C_k\}$ has nonempty intersection. *Hint*: Consider the sets $\overline{C}_k$ given by

$$\overline{C}_k = \cap_{i=1}^k C_i, \qquad \forall\ k = 1, 2, \ldots,$$

and use Helly's Theorem (Exercise 1.25) to show that they are nonempty.

References

[AHU61] Arrow, K. J., Hurwicz, L., and Uzawa, H., 1961. “Constraint Qualifications in Maximization Problems,” Naval Research Logistics Quarterly, Vol. 8, pp. 175-191.

[Aba67] Abadie, J., 1967. “On the Kuhn-Tucker Theorem,” in Nonlinear Programming, Abadie, J., (Ed.), North Holland, Amsterdam.

[Ash72] Ash, R. B., 1972. Real Analysis and Probability, Academic Press, N. Y.

[AuF90] Aubin, J.-P., and Frankowska, H., 1990. Set-Valued Analysis, Birkhauser, Boston.

[AuT03] Auslender, A., and Teboulle, M., 2003. Asymptotic Cones and Functions in Optimization and Variational Inequalities, Springer, N. Y.

[Aus96] Auslender, A., 1996. “Non Coercive Optimization Problems,” Math. of Operations Research, Vol. 21, pp. 769-782.

[Aus97] Auslender, A., 1997. “How to Deal with the Unbounded in Optimization: Theory and Algorithms,” Math. Programing, Vol. 79, pp. 3-18.

[Aus00] Auslender, A., 2000. “Existence of Optimal Solutions and Duality Results Under Weak Conditions,” Math. Programing, Vol. 88, pp. 45-59.

[BGL95] Bonnans, J. F., Gilbert, J. C., Lemaréchal, C., Sagastizábal, S. C., 1995. “A Family of Variable Metric Proximal Methods,” Math. Programming, Vol. 68, pp. 15-47.

[BGS72] Bazaraa, M. S., Goode, J. J., and Shetty, C. M., 1972. “Constraint Qualifications Revisited,” Management Science, Vol. 18, pp. 567-573.

[BMN01] Ben-Tal, A., Margalit, T., and Nemirovski, A., 2001. “The Ordered Subsets Mirror Descent Optimization Method and its Use for the Positron Emission Tomography Reconstruction,” in Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications, Eds., Butnariu, D., Censor, Y., and Reich, S., Elsevier Science, Amsterdam, Netherlands.

[BNO02] Bertsekas, D. P., Nedić, A., and Ozdaglar, A. E., 2002. “Min Common/Max Crossing Duality: A Simple Geometric Framework for Convex

Optimization and Minimax Theory," Lab. for Information and Decision Systems Report LIDS-P-2536, M.I.T., Cambridge, MA; to appear in J. Opt. Theory and Appl.

[BOT06] Bertsekas, D. P., Ozdaglar, A. E., and Tseng, P., 2006. "Enhanced Fritz John Conditions for Convex Programming," SIAM J. on Optimization, Vol. 16, 2006, p. 766-797.

[BaW75] Balinski, M., and Wolfe, P., (Eds.), 1975. Nondifferentiable Optimization, Math. Programming Study 3, North-Holland, Amsterdam.

[Bar02] Barvinok, A., 2002. A Course in Convexity, American Math. Society, Providence, RI.

[BeG62] Berge, C., and Ghouila-Houri, A., 1962. Programmes, Jeux et Reseau de Transport, Dunod, Paris.

[BeM71] Bertsekas, D. P, and Mitter, S. K., 1971. "Steepest Descent for Optimization Problems with Nondifferentiable Cost Functionals," Proc. 5th Annual Princeton Confer. Inform. Sci. Systems, Princeton, N. J., pp. 347-351.

[BeM73] Bertsekas, D. P., and Mitter, S. K., 1973. "A Descent Numerical Method for Optimization Problems with Nondifferentiable Cost Functionals," SIAM J. on Control, Vol. 11, pp. 637-652.

[BeN01] Ben-Tal, A., and Nemirovski, A., 2001. Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications, SIAM, Philadelphia.

[BeO00] Bertsekas, D. P, and Ozdaglar, A. E., 2000. "Enhanced Optimality Conditions and Exact Penalty Functions," Proceedings of Allerton Conference, Allerton Park, Ill.

[BeO02] Bertsekas, D. P, and Ozdaglar, A. E., 2002. "Pseudonormality and a Lagrange Multiplier Theory for Constrained Optimization," J. Opt. Theory and Appl., Vol. 114, pp. 287-343.

[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.

[BeT96] Bertsekas, D. P., and Tsitsiklis, J. N., 1996. Neuro-Dynamic Programming, Athena Scientific, Belmont, MA.

[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.

[Ber75] Bertsekas, D. P., 1975. "Necessary and Sufficient Conditions for a Penalty Method to be Exact," Math. Programming, Vol. 9, pp. 87-99.

[Ber82] Bertsekas, D. P., 1982. Constrained Optimization and Lagrange Multiplier Methods, Academic Press, N. Y; republished by Athena Scientific, Belmont, MA, 1997.

[Ber98] Bertsekas, D. P., 1998. Network Optimization: Continuous and Discrete Models, Athena Scientific, Belmont, MA.

[Ber99a] Bertsekas, D. P., 1999. Nonlinear Programming: 2nd Edition, Athena Scientific, Belmont, MA.

[Ber99b] Bertsekas, D. P., 1999. "A Note on Error Bounds for Convex and Nonconvex Problems", Computational Optimization and Applications, Vol. 12, 1999, pp. 41-51.

[Ber08] Bertsekas, D. P., 2008. "Min Common/Max Crossing Duality: A Geometric View of Conjugacy in Convex Optimization," Lab. for Information and Decision Systems Report LIDS-P-2796, MIT; revised Jan. 2009.

[Ber09] Bertsekas, D. P., 2009. Convex Optimization Theory, Athena Scientific, Belmont, MA.

[Ber10] Bertsekas, D. P., 2010. "Incremental Gradient, Subgradient, and Proximal Methods for Convex Optimization: A Survey", Lab. for Information and Decision Systems Report LIDS-P-2848, MIT; this is an extended version of a paper in the edited volume Optimization for Machine Learning, by S. Sra, S. Nowozin, and S. J. Wright, MIT Press, Cambridge, MA, 2012, pp. 85-119.

[Ber11] Bertsekas, D. P., 2011. "Incremental Proximal Methods for Large Scale Convex Optimization," Math. Programming, Vol. 129, pp. 163-195.

[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.

[Bou30] Bouligand, G., 1930. "Sur les Surfaces Depourvues de Points Hyperlimites," Annales de la Societe Polonaise de Mathematique, Vol. 9, pp. 32-41.

[Bou32] Bouligand, G., 1932. Introduction a la Geometrie Infinitesimale Directe, Gauthiers-Villars, Paris.

[Brä93] Brännlund, U., 1993. "On Relaxation Methods for Nonsmooth Convex Optimization," Doctoral Thesis, Royal Institute of Technology, Stockholm, Sweden.

[Bro64] Bronsted, A., 1964. "Conjugate Convex Functions in Topological Vector Spaces," Fys. Medd. Dans. Vid. Selsk., Vol. 34, pp. 1-26.

[BuQ98] Burke, J. V., and Qian, M., 1998. "A Variable Metric Proximal Point Algorithm for Monotone Operators," SIAM J. on Control and Optimization, Vol. 37, pp. 353-375.

[CCP98] Cook, W., Cunningham, W., Pulleyblank, W., and Schrijver, A., 1998. Combinatorial Optimization, Wiley, N. Y.

[CFM75] Camerini, P. M., Fratta, L., and Maffioli, F., 1975. "On Improving Relaxation Methods by Modified Gradient Techniques," Math. Programming Studies, Vol. 3, pp. 26-34.

[Cam68] Camion, P., 1968. "Modules Unimodulaires," J. Comb. Theory, Vol. 4, pp. 301-362.

[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.

[ChG59] Cheney, E. W., and Goldstein, A. A., 1959. "Newton's Method for Convex Programming and Tchebycheff Approximation," Numer. Math., Vol. I, pp. 253-268.

[Chv83] Chvatal, V., 1983. Linear Programming, W. H. Freeman and Co., N. Y.

[Cla76] Clarke, F. H., 1976. "A New Approach to Lagrange Multipliers," Math. of Operations Research, Vol. 1, pp. 165-174.

[Cla83] Clarke, F. H., 1983. Optimization and Nonsmooth Analysis, J. Wiley, N. Y.

[DGK63] Danzer, L., Grunbaum, B., and Klee, V. L., 1963. "Helly's Theorem and its Relatives," in Convexity, by Klee, V. L., (Ed.), American Math. Society, Providence, R. I., pp. 101-180.

[Dan63] Dantzig, G. B., 1963. Linear Programming and Extensions, Princeton Univ. Press, Princeton, N. J.

[DeV85] Demjanov, V. F., and Vasilév, L. V., 1985. Nondifferentiable Optimization, Optimization Software Inc., N. Y.

[Dem66] Demjanov, V. F., 1966. "The Solution of Several Minimax Problems," Kibernetika, Vol. 2, pp. 58-66.

[Dem68] Demjanov, V. F., 1968. "Algorithms for Some Minimax Problems," J. of Computer and Systems Science, Vol. 2, pp. 342-380.

[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.

[Eke74] Ekeland, I., 1974. "On the Variational Principle," J. of Math. Analysis and Applications, Vol. 47, pp. 324-353.

[ElM75] Elzinga, J., and Moore, T. G., 1975. "A Central Cutting Plane Algorithm for the Convex Programming Problem," Math. Programming, Vol. 8, pp. 134-145.

[Erm83] Ermoliev, Yu. M., 1983. "Stochastic Quasigradient Methods and Their Application to System Optimization," Stochastics, Vol. 9, pp. 1-36.

[FGH57] Fan, K., Glicksberg, I., and Hoffman, A. J., 1957. "Systems of Inequalities Involving Convex Functions," Proc. Amer. Math. Soc., Vol. 8, pp. 617-622.

[Fan52] Fan, K., 1952. "Fixed-Point and Minimax Theorems in Locally Convex Topological Linear Spaces," Proc. Nat. Acad. Sci. USA, Vol. 38, pp. 121-126.

[Fen49] Fenchel, W., 1949. "On Conjugate Convex Functions," Canad. J. of Math., Vol. 1, pp. 73-77.

[Fen51] Fenchel, W., 1951. "Convex Cones, Sets, and Functions," Mimeographed Notes, Princeton Univ.

[FrW56] Frank, M., and Wolfe, P., 1956. "An Algorithm for Quadratic Programming," Naval Research Logistics Quarterly, Vol. 3, pp. 95-110.

[GKT51] Gale, D., Kuhn, H. W., and Tucker, A. W., 1951. "Linear Programming and the Theory of Games," in Activity Analysis of Production and Allocation, Koopmans, T. C., (Ed.), Wiley, N. Y.

[GoK99] Goffin, J. L., and Kiwiel, K. C., 1999. "Convergence of a Simple Subgradient Level Method," Math. Programming, Vol. 85, pp. 207-211.

[GoT56] Goldman, A. J., and Tucker, A. W., 1956. "Theory of Linear Programming," in Linear Inequalities and Related Systems, H. W. Kuhn and A. W. Tucker, eds., Princeton University Press, Princeton, N.J., pp. 53-97.

[GoT71] Gould, F. J., and Tolle, J., 1971. "A Necessary and Sufficient Condition for Constrained Optimization," SIAM J. Applied Math., Vol. 20, pp. 164-172.

[GoT72] Gould, F. J., and Tolle, J., 1972. "Geometry of Optimality Conditions and Constraint Qualifications," Math. Programming, Vol. 2, pp. 1-18.

[GoV02] Goffin, J. L., and Vial, J. P., 2002. "Convex Nondifferentiable Optimization: A Survey Focussed on the Analytic Center Cutting Plane Method," Optimization Methods and Software, Vol. 17, pp. 805-867.

[Gor73] Gordan, P., 1873. "Uber die Auflosung Linearer Gleichungen mit Reelen Coefficienten," Mathematische Annalen, Vol. 6, pp. 23-28.

[Gri72] Grinold, R. C., 1972. "Steepest Ascent for Large-Scale Linear Programs," SIAM Review, Vol. 14, pp. 447-464.

[Gru67] Grunbaum, B., 1967. Convex Polytopes, Wiley, N. Y.

[Gui69] Guignard, M., 1969. "Generalized Kuhn-Tucker Conditions for Mathematical Programming Problems in a Banach Space," SIAM J. on Control, Vol. 7, pp. 232-241.

[HaM79] Han, S. P., and Mangasarian, O. L., 1979. "Exact Penalty Functions in Nonlinear Programming," Math. Programming, Vol. 17, pp. 251-269.

[Hel21] Helly, E., 1921. "Uber Systeme Linearer Gleichungen mit Unendlich Vielen Unbekannten," Monatschr. Math. Phys., Vol. 31, pp. 60-91.

[Hes75] Hestenes, M. R., 1975. Optimization Theory: The Finite Dimensional Case, 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.

[Joh48] John, F., 1948. "Extremum Problems with Inequalities as Subsidiary Conditions," in Studies and Essays: Courant Anniversary Volume, Friedrichs, K. O., Neugebauer, O. E., and Stoker, J. J., (Eds.), Wiley-Interscience, N. Y., pp. 187-204.

[KaC98] Kaskavelis, C. A., and Caramanis, M. C., 1998. "Efficient Lagrangian Relaxation Algorithms for Industry Size Job-Shop Scheduling Problems," IIE Transactions on Scheduling and Logistics, Vol. 30, pp. 1085–1097.

[Kak41] Kakutani, S., 1941. "A Generalization of Brouwer's Fixed Point Theorem," Duke Mathematical Journal, Vol. 8, pp. 457-459.

[Kar39] Karush, W., 1939. "Minima of Functions of Several Variables with Inequalities as Side Conditions," M.S. Thesis, Department of Math., University of Chicago.

[Kel60] Kelley, J. E., 1960. "The Cutting-Plane Method for Solving Convex Programs," J. Soc. Indust. Appl. Math., Vol. 8, pp. 703-712.

[Kib79] Kibardin, V. M., 1979. "Decomposition into Functions in the Minimization Problem," Automation and Remote Control, Vol. 40, pp. 1311-1323.

[Kiw97] Kiwiel, K. C., 1997. "Efficiency of the Analytic Center Cutting Plane Method for Convex Minimization," SIAM J. on Optimization, Vol. 7, pp. 336-346.

[Kle63] Klee, V. L., (Ed.), 1963. Convexity, American Math. Society, Providence, R. I.

[Kle68] Klee, V. L., 1968. "Maximal Separation Theorems for Convex Sets," Trans. Amer. Math. Soc., Vol. 134, pp. 133-148.

[KrM40] Krein, M., and Millman, D., 1940. "On Extreme Points of Regular Convex Sets," Studia Math., Vol. 9, pp. 133-138.

[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.

[Kuh76] Kuhn, H. W., 1976. "Nonlinear Programming: A Historical View," in Nonlinear Programming, Cottle, R. W., and Lemke, C. E., (Eds.), SIAM-AMS Proc., Vol. IX, American Math. Soc., Providence, RI, pp. 1-26.

[LVB98] Lobo, M. S., Vandenberghe, L., Boyd, S., and Lebret, H., 1998. "Applications of Second-Order Cone Programming," Linear Algebra and Applications, Vol. 284, pp. 193-228.

[LaT85] Lancaster, P., and Tismenetsky, M., 1985. The Theory of Matrices, Academic Press, N. Y.

[LeS93] Lemaréchal, C., and Sagastizábal, C., 1993. "An Approach to Variable Metric Bundle Methods," in Systems Modelling and Optimization, Proc. of the 16th IFIP-TC7 Conference, Compiègne, Henry, J., and Yvon, J.-P., (Eds.), Lecture Notes in Control and Information Sciences 197, pp. 144-162.

[Lem74] Lemaréchal, C., 1974. "An Algorithm for Minimizing Convex Functions," in Information Processing '74, Rosenfeld, J. L., (Ed.), pp. 552-556, North-Holland, Amsterdam.

[Lem75] Lemaréchal, C., 1975. "An Extension of Davidon Methods to Nondifferentiable Problems," Math. Programming Study 3, Balinski, M., and Wolfe, P., (Eds.), North-Holland, Amsterdam, pp. 95-109.

[Lit66] Litvakov, B. M., 1966. "On an Iteration Method in the Problem of Approximating a Function from a Finite Number of Observations," Avtom. Telemech., No. 4, pp. 104-113.

[LuZ99] Luo, Z.-Q., and Zhang, S. Z., 1999. "On the Extension of Frank-Wolfe Theorem," Comput. Optim. Appl., Vol. 13, pp. 87-110.

[MSQ98] Mifflin, R., Sun, D., and Qi, L., 1998. "Quasi-Newton Bundle-Type Methods for Nondifferentiable Convex Optimization, SIAM J. on Optimization, Vol. 8, pp. 583-603.

[MaF67] Mangasarian, O. L., and Fromovitz, S., 1967. "The Fritz John Necessary Optimality Conditions in the Presence of Equality and Inequality Constraints," J. Math. Anal. and Appl., Vol. 17, pp. 37-47.

[McS73] McShane, E. J., 1973. "The Lagrange Multiplier Rule," Amer. Mathematical Monthly, Vol. 80, pp. 922-925.

[Man69] Mangasarian, O. L., 1969. Nonlinear Programming, Prentice-Hall, Englewood Cliffs, N. J.; also SIAM, Classics in Applied Mathematics 10, Phila., PA., 1994.

[Mif96] Mifflin, R., 1996. "A Quasi-Second-Order Proximal Bundle Algorithm," Math. Programming, Vol. 73, pp. 51-72.

[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.

[Mor62] Moreau, J.-J., 1962. "Fonctions Convexes Duales et Points Proximaux dans en Espace Hilbertien," Comptes Rendus de l'Academie des Sciences de Paris, 255, pp. 2897-2899.

[Mor64] Moreau, J.-J., 1964. "Theoremes 'inf-sup'," Comptes Rendus de l'Academie des Sciences de Paris, 258, pp. 2720-2722.

[Mor76] Mordukhovich, B. S., 1976. "Maximum Principle in the Problem of Time Optimal Response with Nonsmooth Constraints," J. of Applied Mathematics and Mechanics, Vol. 40, pp. 960-969.

[Mor88] Mordukhovich, B. S., 1988. Approximation Methods in Problems of Optimization and Control, Nauka, Moscow.

[NBB01] Nedić, A., Bertsekas, D. P., and Borkar, V. S., 2001. "Distributed Asynchronous Incremental Subgradient Methods," in Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications, Butnariu, D., Censor, Y., and Reich, S., (Eds.), Elsevier Science, Amsterdam, Netherlands.

[NeB01a] Nedić, A., and Bertsekas, D. P., 2001. "Incremental Subgradient Methods for Nondifferentiable Optimization," SIAM J. on Optim., Vol. 12, pp. 109-138.

[NeB01b] Nedić, A., and Bertsekas, D. P., 2001. "Convergence Rate of

Incremental Subgradient Algorithms," in Stochastic Optimization: Algorithms and Applications, Uryasev, S., and Pardalos, P. M., (Eds.), Kluwer Academic Publishers, Dordrecht, Netherlands, pp. 223-264.

[NeB02] Nedić, A., and Bertsekas, D. P., 2002. "A Unified Framework for Minimax Theory, and for Existence of Solutions and Absence of Duality Gap in Convex Programs," Unpublished Research.

[NeB10] Nedić, A., and Bertsekas, D. P., 2010. "The Effect of Deterministic Noise in Subgradient Methods," Math. Programming, Ser. A, Vol. 125, pp. 75-99.

[NeW88] Nemhauser, G. L., and Wolsey, L. A., 1988. Integer and Combinatorial Optimization, Wiley, N. Y.

[Ned02] Nedić, A., 2002. Subgradient Methods for Convex Minimization, Ph.D. Thesis, Mass. Institute of Technology, Cambridge, MA.

[Ned11] Nedić, A., 2011. "Random Projection Algorithms for Constraint Minimization Problems," Math. Programming, Ser. B, Vol. 129, pp. 225-253.

[Neu28] Neumann, J. von, 1928. "Zur Theorie der Gesellschaftsspiele," Math. Ann., Vol. 100, pp. 295-320.

[Neu37] Neumann, J. von, 1937. "Ueber ein Okonomisches Gleichungssystem und eine Verallgemeinerung des Brouwerschen Fixpunktsatzes," Engebn. Math. Kolloq. Wien, Vol. 8, pp. 73-83.

[Nev75] Neveu, J., 1975. Discrete Parameter Martingales, North-Holland, Amsterdam.

[Nur77] Nurminskii, E. A., 1977. "The Continuity of ϵ-Subgradient Mappings," (Russian) Kibernetika (Kiev), pp. 148-149.

[OrR70] Ortega, J. M., and Rheinboldt, W. C., 1970. Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, N. Y.

[OzB04] Ozdaglar, A. E., and Bertsekas, D. P., 2004. "The Relation Between Pseudonormality and Quasiregularity in Constrained Optimization," Optimization Methods and Software," Vol. 19, pp. 493-506.

[Ozd03] Ozdaglar, A. E., 2003. Pseudonormality and a Lagrange Multiplier Theory for Constrained Optimization, Ph.D. Thesis, Mass. Institute of Technology, Cambridge, MA.

[Pie69] Pietrzykowski, T., 1969. "An Exact Potential Method for Constrained Maxima," SIAM J. Numer. Anal., Vol. 6, pp. 294-304.

[Pol87] Polyak, B. T., 1987. Introduction to Optimization, Optimization Software Inc., N.Y.

[RoW98] Rockafellar, R. T., and Wets, R. J.-B., 1998. Variational Analysis, Springer-Verlag, Berlin.

[Roc64] Rockafellar, R. T., 1964. "Minimax Theorems and Conjugate Saddle Functions," Math. Scand., Vol. 14, pp. 151-173.

[Roc65] Rockafellar, R. T., 1965. "Helly's Theorem and Minima of Convex Functions," Duke Math. J., Vol. 33, pp. 81-90.

[Roc69] Rockafellar, R. T., 1969. "The Elementary Vectors of a Subspace of R^N," in Combinatorial Mathematics and its Applications, by Bose, R. C., and Dowling, T. A., (Eds.), University of North Carolina Press, pp. 104-127.

[Roc70] Rockafellar, R. T., 1970. Convex Analysis, Princeton Univ. Press, Princeton, N. J.

[Roc71] Rockafellar, R. T., 1971. "Ordinary Convex Programs Without a Duality Gap," J. Opt. Theory and Appl., Vol. 7, pp. 143-148.

[Roc84] Rockafellar, R. T., 1984. Network Flows and Monotropic Optimization, Wiley, N. Y.; republished by Athena Scientific, Belmont, MA, 1998.

[Roc93] Rockafellar, R. T., 1993. "Lagrange Multipliers and Optimality," SIAM Review, Vol. 35, pp. 183-238.

[Rud76] Rudin, W., 1976. Principles of Mathematical Analysis, McGraw-Hill, N. Y.

[Rus89] Ruszczynski, A., 1989. "An Augmented Lagrangian Decomposition Method for Block Diagonal Linear Programming Problems," Operations Res. Letters, Vol. 8, pp. 287-294.

[Sch86] Schrijver, A., 1986. Theory of Linear and Integer Programming, Wiley, N. Y.

[Sho85] Shor, N. Z., 1985. Minimization Methods for Nondifferentiable Functions, Springer-Verlag, Berlin.

[Sho98] Shor, N. Z., 1998. Nondifferentiable Optimization and Polynomial Problems, Kluwer, Dordrecht, the Netherlands.

[Sla50] Slater, M., 1950. "Lagrange Multipliers Revisited: A Contribution to Non-Linear Programming," Cowles Commission Discussion Paper, Math. 403.

[SoZ98] Solodov, M. V., and Zavriev, S. K., 1998. "Error Stability Properties of Generalized Gradient-Type Algorithms," J. Opt. Theory and Appl., Vol. 98, pp. 663–680.

[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 Press, N. Y.

[TBA86] Tsitsiklis, J. N., Bertsekas, D. P., and Athans, M., 1986. "Distributed Asynchronous Deterministic and Stochastic Gradient Optimization Algorithms," IEEE Trans. on Aut. Control, Vol. AC-31, pp. 803-812.

[Ter85] Terlaky, T., 1985. "On l_p Programming," European J. of Operational Research, Vol. 22, pp. 70-100.

[Val63] Valentine, F. A., 1963. "The Dual Cone and Helly Type Theorems," in Convexity, by Klee, V. L., (Ed.), American Math. Society, Providence, R. I, pp. 473-493.

[Val64] Valentine, F. A., 1964. Convex Sets, McGraw-Hill, N.Y.

[Vor08] Voronoi, G., 1908. "Nouvelles Applications des Parametres Continus a la Theorie des Formes Quadratiques - Premier Memoire. Sur Quelques Proprietes des Formes Quadratiques Positives Parfaites," Journal fur die Reine und Angewandte Mathematik, Vol. 133, pp. 97-178.

[WaB12] Wang, M., and Bertsekas, D. P., 2012. "Incremental Constraint Projection Methods for Variational Inequalities," Lab. for Information and Decision Systems Report LIDS-P-2898, MIT.

[Web94] Webster, R., 1994. Convexity, Oxford Univ. Press, Oxford, UK.

[Wet90] Wets, R. J., 1990. "Elementary Constructive Proofs of the Theorems of Farkas, Minkowski and Weyl," in Economic Decision Making: Games, Econometrics and Optimization. Contributions in Honour of Jacques Dreze, Gabszewicz, J. J., Richard, J.-F., and Wolsey, L. A., (Eds.), North-Holland, Elsevier-Science, Amsterdam, pp. 427-432.

[Wey35] Weyl, H., 1935. "Elementare Theorie der Konvexen Polyeder," Commentarii Mathematici Helvetici, Vol. 7, pp. 290-306.

[Wol75] Wolfe, P., 1975. "A Method of Conjugate Subgradients for Minimizing Nondifferentiable Functions," Math. Programming Study 3, Balinski, M., and Wolfe, P., (Eds.), North-Holland, Amsterdam, pp. 145-173.

[Wol98] Wolsey, L. A., 1998. Integer Programming, Wiley, N. Y.

[Ye97] Ye, Y., 1997. Interior Point Algorithms: Theory and Analysis, Wiley Interscience, N. Y.

[ZLW99] Zhao, X., Luh, P. B., and Wang, J., 1999. "Surrogate Gradient Algorithm for Lagrangian Relaxation," J. Opt. Theory and Appl., Vol. 100, pp. 699–712.

[Zan67] Zangwill, W. I., 1967. "Nonlinear Programming via Penalty Functions," Management Sci., Vol. 13, pp. 344-358.