CMS Books in Mathematics

Jonathan M. Borwein
Adrian S. Lewis

Convex Analysis and Nonlinear Optimization

Theory and Examples

$$g^*(y) = \log \sum_i \exp \langle a^i, y \rangle$$

Canadian Mathematical Society
Société mathématique du Canada

Jonathan M. Borwein Adrian S. Lewis

Convex Analysis and Nonlinear Optimization

Theory and Examples

Springer

Jonathan M. Borwein
Centre for Experimental
and Constructive Mathematics
Department of Mathematics
and Statistics
Simon Fraser University
Burnaby, British Columbia V5A 1S6
Canada
jborwein@cs.sfu.ca

Adrian S. Lewis
Department of Combinatorics
and Optimization
University of Waterloo
Waterloo, Ontario N2L 3G1
Canada
aslewis@orion.uwaterloo.ca

Mathematics Subject Classification (2000): 90C25, 49J53, 52A41, 46N10, 47H10

Library of Congress Cataloging-in-Publication Data
Borwein, Jonathan M.
Convex analysis and nonlinear optimization / Jonathan M. Borwein, Adrian S. Lewis.
p. cm. — (CMS books in mathematics)
Includes bibliographical references and index.
ISBN 0-387-98940-4 (hardcover : alk. paper)
1. Convex functions. 2. Mathematical optimization. 3. Nonlinear theories. I. Lewis, Adrian S. II. Title. III. Series.
QA331.5 .B65 2000
515′.8—dc21 00-020826

Printed on acid-free paper.

Production managed by Michael Koy; manufacturing supervised by Jerome Basma.
Photocomposed copy prepared from the authors' LaTeX2ε files.
Printed and bound by Edwards Brothers, Inc., Ann Arbor, MI.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-98940-4 Springer-Verlag New York Berlin Heidelberg SPIN 10745937

Canadian Mathematical Society
Société mathématique du Canada

Editors-in-Chief
Rédacteurs-en-chef
Jonathan M. Borwein
Peter Borwein

Springer
New York
Berlin
Heidelberg
Barcelona
Hong Kong
London
Milan
Paris
Singapore
Tokyo

CMS Books in Mathematics
Ouvrages de mathématiques de la SMC

1 HERMAN/KUČERA/ŠIMŠA Equations and Inequalities
2 ARNOLD Abelian Groups and Representations of Finite Partially Ordered Sets
3 BORWEIN/LEWIS Convex Analysis and Nonlinear Optimization
4 LEVIN/LUBINSKY Orthogonal Polynomials for Exponential Weights
5 KANE Reflection Groups and Invariant Theory
6 PHILLIPS Two Millennia of Mathematics: From Archimedes to Gauss

To our families

Preface

Optimization is a rich and thriving mathematical discipline. Properties of minimizers and maximizers of functions rely intimately on a wealth of techniques from mathematical analysis, including tools from calculus and its generalizations, topological notions, and more geometric ideas. The theory underlying current computational optimization techniques grows ever more sophisticated—duality-based algorithms, interior point methods, and control-theoretic applications are typical examples. The powerful and elegant language of convex analysis unifies much of this theory. Hence our aim of writing a concise, accessible account of convex analysis and its applications and extensions, for a broad audience.

For students of optimization and analysis, there is great benefit to blurring the distinction between the two disciplines. Many important analytic problems have illuminating optimization formulations and hence can be approached through our main variational tools: subgradients and optimality conditions, the many guises of duality, metric regularity and so forth. More generally, the idea of convexity is central to the transition from classical analysis to various branches of modern analysis: from linear to nonlinear analysis, from smooth to nonsmooth, and from the study of functions to multifunctions. Thus, although we use certain optimization models repeatedly to illustrate the main results (models such as linear and semidefinite programming duality and cone polarity), we constantly emphasize the power of abstract models and notation.

Good reference works on finite-dimensional convex analysis already exist. Rockafellar's classic *Convex Analysis* [149] has been indispensable and ubiquitous since the 1970s, and a more general sequel with Wets, *Variational Analysis* [150], appeared recently. Hiriart–Urruty and Lemaréchal's *Convex Analysis and Minimization Algorithms* [86] is a comprehensive but gentler introduction. Our goal is not to supplant these works, but on the contrary to promote them, and thereby to motivate future researchers. This book aims to make converts.

We try to be succinct rather than systematic, avoiding becoming bogged down in technical details. Our style is relatively informal; for example, the text of each section creates the context for many of the result statements. We value the variety of independent, self-contained approaches over a single, unified, sequential development. We hope to showcase a few memorable principles rather than to develop the theory to its limits. We discuss no algorithms. We point out a few important references as we go, but we make no attempt at comprehensive historical surveys.

Optimization in infinite dimensions lies beyond our immediate scope. This is for reasons of space and accessibility rather than history or application: convex analysis developed historically from the calculus of variations, and has important applications in optimal control, mathematical economics, and other areas of infinite-dimensional optimization. However, rather like Halmos's *Finite Dimensional Vector Spaces* [81], ease of extension beyond finite dimensions substantially motivates our choice of approach. Where possible, we have chosen a proof technique permitting those readers familiar with functional analysis to discover for themselves how a result extends. We would, in part, like this book to be an entrée for mathematicians to a valuable and intrinsic part of modern analysis. The final chapter illustrates some of the challenges arising in infinite dimensions.

This book can (and does) serve as a teaching text, at roughly the level of first year graduate students. In principle we assume no knowledge of real analysis, although in practice we expect a certain mathematical maturity. While the main body of the text is self-contained, each section concludes with an often extensive set of optional exercises. These exercises fall into three categories, marked with zero, one, or two asterisks, respectively, as follows: examples that illustrate the ideas in the text or easy expansions of sketched proofs; important pieces of additional theory or more testing examples; longer, harder examples or peripheral theory.

We are grateful to the Natural Sciences and Engineering Research Council of Canada for their support during this project. Many people have helped improve the presentation of this material. We would like to thank all of them, but in particular Patrick Combettes, Guillaume Haberer, Claude Lemaréchal, Olivier Ley, Yves Lucet, Hristo Sendov, Mike Todd, Xianfu Wang, and especially Heinz Bauschke.

JONATHAN M. BORWEIN
ADRIAN S. LEWIS

Gargnano, Italy
September 1999

Contents

Chapter 1

Background

1.1 Euclidean Spaces

We begin by reviewing some of the fundamental algebraic, geometric and analytic ideas we use throughout the book. **Our setting, for most of the book, is an arbitrary Euclidean space E**, by which we mean a finite-dimensional vector space over the reals $\mathbf{R}$, equipped with an inner product $\langle \cdot, \cdot \rangle$. We would lose no generality if we considered only the space $\mathbf{R}^n$ of real (column) n-vectors (with its standard inner product), but a more abstract, coordinate-free notation is often more flexible and elegant.

We define the *norm* of any point x in $\mathbf{E}$ by $\|x\| = \sqrt{\langle x, x \rangle}$, and the *unit ball* is the set

$$B = \{x \in \mathbf{E} \mid \|x\| \leq 1\}.$$

Any two points x and y in $\mathbf{E}$ satisfy the *Cauchy–Schwarz inequality*

$$|\langle x, y \rangle| \leq \|x\| \|y\|.$$

We define the sum of two sets C and D in $\mathbf{E}$ by

$$C + D = \{x + y \mid x \in C,\ y \in D\}.$$

The definition of $C - D$ is analogous, and for a subset Λ of $\mathbf{R}$ we define

$$\Lambda C = \{\lambda x \mid \lambda \in \Lambda,\ x \in C\}.$$

Given another Euclidean space $\mathbf{Y}$, we can consider the Cartesian product Euclidean space $\mathbf{E} \times \mathbf{Y}$, with inner product defined by $\langle (e, x), (f, y) \rangle = \langle e, f \rangle + \langle x, y \rangle$.

We denote the nonnegative reals by $\mathbf{R}_+$. If C is nonempty and satisfies $\mathbf{R}_+ C = C$ we call it a *cone*. (Notice we require that cones contain the origin.) Examples are the positive orthant

$$\mathbf{R}^n_+ = \{x \in \mathbf{R}^n \mid \text{each } x_i \geq 0\},$$

and the cone of vectors with nonincreasing components

$$\mathbf{R}^n_{\geq} = \{x \in \mathbf{R}^n \mid x_1 \geq x_2 \geq \cdots \geq x_n\}.$$

The smallest cone containing a given set $D \subset \mathbf{E}$ is clearly $\mathbf{R}_+ D$.

The fundamental geometric idea of this book is *convexity*. A set C in $\mathbf{E}$ is *convex* if the line segment joining any two points x and y in C is contained in C: algebraically, $\lambda x + (1-\lambda)y \in C$ whenever $0 \leq \lambda \leq 1$. An easy exercise shows that intersections of convex sets are convex.

Given any set $D \subset \mathbf{E}$, the *linear span* of D, denoted $\operatorname{span}(D)$, is the smallest linear subspace containing D. It consists exactly of all linear combinations of elements of D. Analogously, the *convex hull* of D, denoted $\operatorname{conv}(D)$, is the smallest convex set containing D. It consists exactly of all *convex combinations* of elements of D, that is to say points of the form $\sum_{i=1}^m \lambda_i x^i$, where $\lambda_i \in \mathbf{R}_+$ and $x^i \in D$ for each i, and $\sum \lambda_i = 1$ (see Exercise 2).

The language of elementary point-set topology is fundamental in optimization. A point x lies in the *interior* of the set $D \subset \mathbf{E}$ (denoted $\operatorname{int} D$) if there is a real $\delta > 0$ satisfying $x + \delta B \subset D$. In this case we say D is a *neighbourhood* of x. For example, the interior of $\mathbf{R}^n_+$ is

$$\mathbf{R}^n_{++} = \{x \in \mathbf{R}^n \mid \text{each } x_i > 0\}.$$

We say the point x in $\mathbf{E}$ is the *limit* of the sequence of points $x^1, x^2, \ldots$ in $\mathbf{E}$, written $x^j \to x$ as $j \to \infty$ (or $\lim_{j\to\infty} x^j = x$), if $\|x^j - x\| \to 0$. The *closure* of D is the set of limits of sequences of points in D, written $\operatorname{cl} D$, and the *boundary* of D is $\operatorname{cl} D \setminus \operatorname{int} D$, written $\operatorname{bd} D$. The set D is *open* if $D = \operatorname{int} D$, and is *closed* if $D = \operatorname{cl} D$. Linear subspaces of $\mathbf{E}$ are important examples of closed sets. Easy exercises show that D is open exactly when its complement D^c is closed, and that arbitrary unions and finite intersections of open sets are open. The interior of D is just the largest open set contained in D, while $\operatorname{cl} D$ is the smallest closed set containing D. Finally, a subset G of D is *open in D* if there is an open set $U \subset \mathbf{E}$ with $G = D \cap U$.

Much of the beauty of convexity comes from *duality* ideas, interweaving geometry and topology. The following result, which we prove a little later, is both typical and fundamental.

Theorem 1.1.1 (Basic separation) *Suppose that the set $C \subset \mathbf{E}$ is closed and convex, and that the point y does not lie in C. Then there exist real b and a nonzero element a of $\mathbf{E}$ satisfying $\langle a, y\rangle > b \geq \langle a, x\rangle$ for all points x in C.*

Sets in $\mathbf{E}$ of the form $\{x \mid \langle a, x\rangle = b\}$ and $\{x \mid \langle a, x\rangle \leq b\}$ (for a nonzero element a of $\mathbf{E}$ and real b) are called *hyperplanes* and *closed halfspaces*, respectively. In this language the above result states that the point y is *separated* from the set C by a hyperplane. In other words, C is contained in a certain closed halfspace whereas y is not. Thus there is a "dual" representation of C as the intersection of all closed halfspaces containing it.

The set D is *bounded* if there is a real k satisfying $kB \supset D$, and it is *compact* if it is closed and bounded. The following result is a central tool in real analysis.

Theorem 1.1.2 (Bolzano–Weierstrass) *Bounded sequences in* $\mathbf{E}$ *have convergent subsequences.*

Just as for sets, geometric and topological ideas also intermingle for the functions we study. Given a set D in $\mathbf{E}$, we call a function $f : D \to \mathbf{R}$ *continuous* (on D) if $f(x^i) \to f(x)$ for any sequence $x^i \to x$ in D. In this case it easy to check, for example, that for any real α the *level set* $\{x \in D \mid f(x) \leq \alpha\}$ is closed providing D is closed.

Given another Euclidean space $\mathbf{Y}$, we call a map $A : \mathbf{E} \to \mathbf{Y}$ *linear* if any points x and z in $\mathbf{E}$ and any reals λ and μ satisfy $A(\lambda x + \mu z) = \lambda Ax + \mu Az$. In fact any linear function from $\mathbf{E}$ to $\mathbf{R}$ has the form $\langle a, \cdot \rangle$ for some element a of $\mathbf{E}$. Linear maps and *affine* functions (linear functions plus constants) are continuous. Thus, for example, closed halfspaces are indeed closed. A *polyhedron* is a finite intersection of closed halfspaces, and is therefore both closed and convex. The *adjoint* of the map A above is the linear map $A^* : \mathbf{Y} \to \mathbf{E}$ defined by the property

$$\langle A^*y, x \rangle = \langle y, Ax \rangle \text{ for all points } x \text{ in } \mathbf{E} \text{ and } y \text{ in } \mathbf{Y}$$

(whence $A^{**} = A$). The *null space* of A is $N(A) = \{x \in \mathbf{E} \mid Ax = 0\}$. The *inverse image* of a set $H \subset \mathbf{Y}$ is the set $A^{-1}H = \{x \in \mathbf{E} \mid Ax \in H\}$ (so for example $N(A) = A^{-1}\{0\}$). Given a subspace G of $\mathbf{E}$, the *orthogonal complement* of G is the subspace

$$G^\perp = \{y \in \mathbf{E} \mid \langle x, y \rangle = 0 \text{ for all } x \in G\},$$

so called because we can write $\mathbf{E}$ as a direct sum $G \oplus G^\perp$. (In other words, any element of $\mathbf{E}$ can be written uniquely as the sum of an element of G and an element of $G^\perp$.) Any subspace G satisfies $G^{\perp\perp} = G$. The range of any linear map A coincides with $N(A^*)^\perp$.

Optimization studies properties of minimizers and maximizers of functions. Given a set $\Lambda \subset \mathbf{R}$, the *infimum* of Λ (written $\inf \Lambda$) is the greatest lower bound on Λ, and the *supremum* (written $\sup \Lambda$) is the least upper bound. To ensure these are always defined, it is natural to append $-\infty$ and $+\infty$ to the real numbers, and allow their use in the usual notation for open and closed intervals. Hence, $\inf \emptyset = +\infty$ and $\sup \emptyset = -\infty$, and for example

$(-\infty, +\infty]$ denotes the interval $\mathbf{R} \cup \{+\infty\}$. We try to avoid the appearance of $+\infty - \infty$, but when necessary we use the convention $+\infty - \infty = +\infty$, so that any two sets C and D in $\mathbf{R}$ satisfy $\inf C + \inf D = \inf(C + D)$. We also adopt the conventions $0 \cdot (\pm\infty) = (\pm\infty) \cdot 0 = 0$. A *(global) minimizer* of a function $f : D \to \mathbf{R}$ is a point $\bar{x}$ in D at which f attains its infimum

$$\inf_D f = \inf f(D) = \inf\{f(x) \mid x \in D\}.$$

In this case we refer to $\bar{x}$ as an *optimal solution* of the *optimization problem* $\inf_D f$.

For a positive real δ and a function $g : (0, \delta) \to \mathbf{R}$, we define

$$\liminf_{t \downarrow 0} g(t) = \lim_{t \downarrow 0} \inf_{(0,t)} g$$

and

$$\limsup_{t \downarrow 0} g(t) = \lim_{t \downarrow 0} \sup_{(0,t)} g.$$

The limit $\lim_{t \downarrow 0} g(t)$ exists if and only if the above expressions are equal.

The question of *attainment*, or in other words the *existence* of an optimal solution for an optimization problem is typically topological. The following result is a prototype. The proof is a standard application of the Bolzano–Weierstrass theorem above.

Proposition 1.1.3 (Weierstrass) *Suppose that the set $D \subset \mathbf{E}$ is nonempty and closed, and that all the level sets of the continuous function $f : D \to \mathbf{R}$ are bounded. Then f has a global minimizer.*

Just as for sets, convexity of functions will be crucial for us. Given a convex set $C \subset \mathbf{E}$, we say that the function $f : C \to \mathbf{R}$ is *convex* if

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

for all points x and y in C and $0 \leq \lambda \leq 1$. The function f is *strictly convex* if the inequality holds strictly whenever x and y are distinct in C and $0 < \lambda < 1$. It is easy to see that a strictly convex function can have at most one minimizer.

Requiring the function f to have bounded level sets is a "growth condition". Another example is the stronger condition

$$\liminf_{\|x\| \to \infty} \frac{f(x)}{\|x\|} > 0, \tag{1.1.4}$$

where we define

$$\liminf_{\|x\| \to \infty} \frac{f(x)}{\|x\|} = \lim_{r \to +\infty} \inf \left\{ \frac{f(x)}{\|x\|} \;\middle|\; 0 \neq x \in C \cap rB \right\}.$$

Surprisingly, for *convex* functions these two growth conditions are equivalent.

Proposition 1.1.5 *For a convex set $C \subset \mathbf{E}$, a convex function $f : C \to \mathbf{R}$ has bounded level sets if and only if it satisfies the growth condition (1.1.4).*

The proof is outlined in Exercise 10.

Exercises and Commentary

Good general references are [156] for elementary real analysis and [1] for linear algebra. Separation theorems for convex sets originate with Minkowski [129]. The theory of the relative interior (Exercises 11, 12, and 13) is developed extensively in [149] (which is also a good reference for the recession cone, Exercise 6).

1. Prove the intersection of an arbitrary collection of convex sets is convex. Deduce that the convex hull of a set $D \subset \mathbf{E}$ is well-defined as the intersection of all convex sets containing D.

2. (a) Prove that if the set $C \subset \mathbf{E}$ is convex and if

$$x^1, x^2, \ldots, x^m \in C, \; 0 \le \lambda_1, \lambda_2, \ldots, \lambda_m \in \mathbf{R},$$

and $\sum \lambda_i = 1$ then $\sum \lambda_i x^i \in C$. Prove, furthermore, that if $f : C \to \mathbf{R}$ is a convex function then $f(\sum \lambda_i x^i) \le \sum \lambda_i f(x^i)$.

   (b) We see later (Theorem 3.1.11) that the function $-\log$ is convex on the strictly positive reals. Deduce, for any strictly positive reals $x^1, x^2, \ldots, x^m$, and any nonnegative reals $\lambda_1, \lambda_2, \ldots, \lambda_m$ with sum 1, the *arithmetic-geometric mean inequality*

$$\sum_i \lambda_i x^i \ge \prod_i (x^i)^{\lambda_i}.$$

   (c) Prove that for any set $D \subset \mathbf{E}$, $\operatorname{conv} D$ is the set of all convex combinations of elements of D.

3. Prove that a convex set $D \subset \mathbf{E}$ has convex closure, and deduce that $\operatorname{cl}(\operatorname{conv} D)$ is the smallest closed convex set containing D.

4. **(Radstrom cancellation)** Suppose sets $A, B, C \subset \mathbf{E}$ satisfy

$$A + C \subset B + C.$$

   (a) If A and B are convex, B is closed, and C is bounded, prove

$$A \subset B.$$

   (Hint: Observe $2A + C = A + (A + C) \subset 2B + C$.)

   (b) Show this result can fail if B is not convex.

5. * **(Strong separation)** Suppose that the set $C \subset \mathbf{E}$ is closed and convex, and that the set $D \subset \mathbf{E}$ is compact and convex.

   (a) Prove the set $D - C$ is closed and convex.

   (b) Deduce that if in addition D and C are disjoint then there exists a nonzero element a in $\mathbf{E}$ with $\inf_{x \in D}\langle a, x\rangle > \sup_{y \in C}\langle a, y\rangle$. Interpret geometrically.

   (c) Show part (b) fails for the closed convex sets in $\mathbf{R}^2$,

$$\begin{aligned} D &= \{x \mid x_1 > 0,\ x_1 x_2 \geq 1\}, \\ C &= \{x \mid x_2 = 0\}. \end{aligned}$$

6. ** **(Recession cones)** Consider a nonempty closed convex set $C \subset \mathbf{E}$. We define the *recession cone* of C by

$$0^+(C) = \{d \in \mathbf{E} \mid C + \mathbf{R}_+ d \subset C\}.$$

   (a) Prove $0^+(C)$ is a closed convex cone.

   (b) Prove $d \in 0^+(C)$ if and only if $x + \mathbf{R}_+ d \subset C$ for some point x in C. Show this equivalence can fail if C is not closed.

   (c) Consider a family of closed convex sets C_γ $(\gamma \in \Gamma)$ with nonempty intersection. Prove $0^+(\cap C_\gamma) = \cap 0^+(C_\gamma)$.

   (d) For a unit vector u in $\mathbf{E}$, prove $u \in 0^+(C)$ if and only if there is a sequence (x^r) in C satisfying $\|x^r\| \to \infty$ and $\|x^r\|^{-1} x^r \to u$. Deduce C is unbounded if and only if $0^+(C)$ is nontrivial.

   (e) If $\mathbf{Y}$ is a Euclidean space, the map $A : \mathbf{E} \to \mathbf{Y}$ is linear, and $N(A) \cap 0^+(C)$ is a linear subspace, prove AC is closed. Show this result can fail without the last assumption.

   (f) Consider another nonempty closed convex set $D \subset \mathbf{E}$ such that $0^+(C) \cap 0^+(D)$ is a linear subspace. Prove $C - D$ is closed.

7. For any set of vectors $a^1, a^2, \ldots, a^m$ in $\mathbf{E}$, prove the function $f(x) = \max_i \langle a^i, x\rangle$ is convex on $\mathbf{E}$.

8. Prove Proposition 1.1.3 (Weierstrass).

9. **(Composing convex functions)** Suppose that the set $C \subset \mathbf{E}$ is convex and that the functions $f_1, f_2, \ldots, f_n : C \to \mathbf{R}$ are convex, and define a function $f : C \to \mathbf{R}^n$ with components f_i. Suppose further that $f(C)$ is convex and that the function $g : f(C) \to \mathbf{R}$ is convex and *isotone*: any points $y \leq z$ in $f(C)$ satisfy $g(y) \leq g(z)$. Prove the composition $g \circ f$ is convex.

10. * **(Convex growth conditions)**

   (a) Find a function with bounded level sets which does not satisfy the growth condition (1.1.4).

   (b) Prove that any function satisfying (1.1.4) has bounded level sets.

   (c) Suppose the convex function $f : C \to \mathbf{R}$ has bounded level sets but that (1.1.4) fails. Deduce the existence of a sequence (x^m) in C with $f(x^m) \leq \|x^m\|/m \to +\infty$. For a fixed point $\bar{x}$ in C, derive a contradiction by considering the sequence

$$\bar{x} + \frac{m}{\|x^m\|}(x^m - \bar{x}).$$

   Hence complete the proof of Proposition 1.1.5.

The relative interior

Some arguments about finite-dimensional convex sets C simplify and lose no generality if we assume C contains 0 and spans $\mathbf{E}$. The following exercises outline this idea.

11. ** **(Accessibility lemma)** Suppose C is a convex set in $\mathbf{E}$.

   (a) Prove $\operatorname{cl} C \subset C + \epsilon B$ for any real $\epsilon > 0$.

   (b) For sets D and F in $\mathbf{E}$ with D open, prove $D + F$ is open.

   (c) For x in $\operatorname{int} C$ and $0 < \lambda \leq 1$, prove $\lambda x + (1 - \lambda)\operatorname{cl} C \subset C$. Deduce $\lambda \operatorname{int} C + (1 - \lambda)\operatorname{cl} C \subset \operatorname{int} C$.

   (d) Deduce $\operatorname{int} C$ is convex.

   (e) Deduce further that if $\operatorname{int} C$ is nonempty then $\operatorname{cl}(\operatorname{int} C) = \operatorname{cl} C$. Is convexity necessary?

12. ** **(Affine sets)** A set L in $\mathbf{E}$ is *affine* if the entire line through any distinct points x and y in L lies in L: algebraically, $\lambda x + (1 - \lambda)y \in L$ for *any* real λ. The *affine hull* of a set D in $\mathbf{E}$, denoted $\operatorname{aff} D$, is the smallest affine set containing D. An *affine combination* of points $x^1, x^2, \ldots, x^m$ is a point of the form $\sum_1^m \lambda_i x^i$, for reals λ_i summing to one.

   (a) Prove the intersection of an arbitrary collection of affine sets is affine.

   (b) Prove that a set is affine if and only if it is a translate of a linear subspace.

   (c) Prove $\operatorname{aff} D$ is the set of all affine combinations of elements of D.

   (d) Prove $\operatorname{cl} D \subset \operatorname{aff} D$ and deduce $\operatorname{aff} D = \operatorname{aff}(\operatorname{cl} D)$.