J. M. Borwein
and
Q. J. Zhu

Techniques of Variational Analysis

An Introduction

October 8, 2004

Springer
Berlin Heidelberg New York
Hong Kong London
Milan Paris Tokyo

To Tova, Naomi, Rachel and Judith.

To Charles and Lilly.

And in fond and respectful memory of Simon Fitzpatrick (1953-2004).

Preface

Variational arguments are classical techniques whose use can be traced back to the early development of the calculus of variations and further. Rooted in the physical principle of least action they have wide applications in diverse fields. The discovery of modern variational principles and nonsmooth analysis further expand the range of applications of these techniques. The motivation to write this book came from a desire to share our pleasure in applying such variational techniques and promoting these powerful tools. Potential readers of this book will be researchers and graduate students who might benefit from using variational methods.

The only broad prerequisite we anticipate is a working knowledge of undergraduate analysis and of the basic principles of functional analysis (e.g., those encountered in a typical introductory functional analysis course). We hope to attract researchers from diverse areas – who may fruitfully use variational techniques – by providing them with a relatively systematical account of the principles of variational analysis. We also hope to give further insight to graduate students whose research already concentrates on variational analysis. Keeping these two different reader groups in mind we arrange the material into relatively independent blocks. We discuss various forms of variational principles early in Chapter 2. We then discuss applications of variational techniques in different areas in Chapters 3–7. These applications can be read relatively independently. We also try to put general principles and their applications together

The recent monograph "Variational Analysis" by Rockafellar and Wets [230] has already provided an authoritative and systematical account of variational analysis in finite dimensional spaces. We hope to supplement this with a concise account of the essential tools of infinite-dimensional first-order variational analysis; these tools are presently scattered in the literature. We also aim to illustrate applications in many different parts of analysis, optimization and approximation, dynamical systems, mathematical economics and elsewhere. Much of the material we present grows out of talks and short lecture series we have given in the past several years. Thus, chapters in this book can easily be arranged to form material for a graduate level topics course. A fair collection of suitable exercises is provided for this purpose. For many reasons, we avoid pursuing maximum generality in the main corpus. We do, however, aim at selecting proofs of results that best represent the general technique.

In addition, in order to make this book a useful reference for researchers who use variational techniques, or think they might, we have included many more extended guided exercises (with corresponding references) that either give useful generalizations of the main text or illustrate significant relationships with other results. Harder problems are marked by a $*$. The forthcoming book "Variational Analysis in Infinite Dimensions" by Boris Mordukhovich [191], to our great pleasure, is a comprehensive complement to the present work.

We are indebted to many of our colleagues and students who read various versions of our manuscript and provided us with valuable suggestions. Particularly, we thank Heinz Bauschke, Kirsty Eisenhart, Ovidiu Furdui, Warren Hare, Marc Lassonde, Yuri Ledyaev, Boris Mordukhovich, Jean Paul Penot, Jay Treiman, Jack Warga, and Herre Wiersma. We also thank Jiongmin Yong for organizing a short lecture series in 2002 at Fudan university which provided an excellent environment for the second author to test preliminary materials for this book.

We hope our readers get as much pleasure from reading this material as we have had during its writing.

Halifax, Nova Scotia, Canada and Kalamazoo, Michigan, USA *Jonathan Borwein*

July 6, 2004 *Qiji Zhu*

Contents

1

Introduction and Notation

1.1 Introduction

In this book, *variational techniques* refer to proofs by way of establishing that an appropriate auxiliary function attains a minimum. This can be viewed as a mathematical form of the principle of least action in physics. Since so many important results in mathematics, in particular, in analysis have their origins in the physical sciences, it is entirely natural that they can be related in one way or another to variational techniques. The purpose of this book is to provide an introduction to this powerful method, and its applications, to researchers who are interested in using this method. The use of variational arguments in mathematical proofs has a long history. This can be traced back to Johann Bernoulli's problem of the Brachistochrone and its solutions leading to the development of the calculus of variations. Since then the method has found numerous applications in various branches of mathematics. A simple illustration of the variational argument is the following example.

Example 1.1.1 (Surjectivity of Derivatives) Suppose that $f\colon \mathbb{R} \to \mathbb{R}$ is differentiable everywhere and suppose that

$$\lim_{|x|\to\infty} f(x)/|x| = +\infty.$$

Then $\{f'(x) \mid x \in \mathbb{R}\} = \mathbb{R}$.

Proof. Let r be an arbitrary real number. Define $g(x) := f(x) - rx$. We easily check that g is coercive, i.e., $g(x) \to +\infty$ as $|x| \to \infty$ and therfore attains a (global) minimum at, say, $\bar{x}$. Then $0 = g'(\bar{x}) = f'(\bar{x}) - r$. ●

Two conditions are essential in this variational argument. The first is *compactness* (to ensure the existence of the minimum) and the second is *differentiability* of the auxiliary function (so that the differential characterization of the results is possible). Two important discoveries in the 1970's led to significant useful relaxation on both conditions. First, the discovery of general variational principles led to the relaxation of the compactness assumptions. Such principles typically assert that any lower semicontinuous (lsc) function, bounded from below, may be perturbed slightly to ensure the existence of the minimum. Second, the development of the nonsmooth analysis made possible the use of nonsmooth auxiliary functions.

The emphasis in this book is on the new developments and applications of variational techniques in the past several decades. Besides the use of variational principles and concepts that generalize that of a derivative for smooth functions, one often needs to combine a variational principle with other suitable tools. For example, a decoupling method that mimics in nonconvex settings the role of Fenchel duality or the Hahn–Banach theorem is an essential element in deriving many calculus rules for subdifferentials; minimax theorems play a crucial role alongside the variational principle in several important results in nonlinear functional analysis; and the analysis of spectral functions is a combination of the variational principles with the symmetric property of these functions with respect to certain groups. This is reflected in our arrangement of the chapters. An important feature of the new variational techniques is that they can handle nonsmooth functions, sets and multifunctions equally well. In this book we emphasize the role of nonsmooth, most of the time extended valued lower semicontinuous functions and their subdifferential. We illustrate that sets and multifunctions can be handled by using related nonsmooth functions. Other approaches are possible. For example Mordukhovich [191] starts with variational geometry on closed sets and deals with functions and multifunctions by examining their epigraphs and graphs.

Our intention in this book is to provide a concise introduction to the essential tools of infinite-dimensional first-order variational analysis, tools that are presently scattered in the literature. We also aim to illustrate applications in many different parts of analysis, optimization and approximation, dynamic systems and mathematical economics. To make the book more appealing to readers who are not experts in the area of variational analysis we arrange the applications right after general principles wherever possible. Materials here can be used flexibly for a short lecture series or a topics course for graduate students. They can also serve as a reference for researchers who are interested in the theory or applications of the variational analysis methods.

1.2 Notation

We introduce some common notations in this section.

Let (X, d) be a metric space. We denote the closed ball centered at x with radius r by $B_r(x)$. We will often work in a real Banach space. When X is a Banach space we use X^* and $\langle \cdot, \cdot \rangle$ to denote its (topological) dual and the duality pairing, respectively. The closed unit ball of a Banach space X is often denoted by B_X or B when the space is clear from the context.

Let $\mathbb{R}$ be the real numbers. Consider an extended-real-valued function $f\colon X \to \mathbb{R} \cup \{+\infty\}$. The *domain* of f is the set where it is finite and is denoted by $\operatorname{dom} f := \{x \mid f(x) < +\infty\}$. The *range* of f is the set of all the values of f and is denoted by $\operatorname{range} f := \{f(x) \mid x \in \operatorname{dom} f\}$. We call an extended-valued function f *proper* provided that its domain is nonempty. We say $f\colon X \to \mathbb{R} \cup \{+\infty\}$ is *lower semicontinuous* (lsc) at x provided that $\liminf_{y\to x} f(y) \geq f(x)$. We say that f is lsc if it is lsc everywhere in its domain.

A subset S of a metric space (X, d) can often be better studied by using related functions. The extended-valued *indicator function* of S,

$$\iota_S(x) = \iota(S; x) := \begin{cases} 0 & x \in S, \\ +\infty & \text{otherwise}, \end{cases}$$

characterizes S. We also use the *distance function*

$$d_S(x) = d(S; x) := \inf\{d(x, y) \mid y \in S\}.$$

The distance function determines closed sets as shown in Exercises 1.3.1 and 1.3.2. On the other hand, to study a function $f\colon X \to \mathbb{R} \cup \{+\infty\}$ it is often equally helpful to examine its *epigraph* and *graph*, related sets in $X \times \mathbb{R}$, defined by

$$\operatorname{epi} f := \{(x, r) \in X \times \mathbb{R} \mid f(x) \leq r\}$$

and

$$\operatorname{graph} f := \{(x, f(x)) \in X \times \mathbb{R} \mid x \in \operatorname{dom} f\}.$$

We denote the *preimage* of $f\colon X \to \mathbb{R} \cup \{+\infty\}$ of a subset S in $\mathbb{R}$ by

$$f^{-1}(S) := \{x \in X \mid f(x) \in S\}.$$

Two special cases which will be used often are $f^{-1}((-\infty, a])$, the *sublevel set*, and $f^{-1}(a)$, the *level set*, of f at $a \in \mathbb{R}$. For a set S in a Banach space X, we denote by $\operatorname{int} S$, $\overline{S}$, $\operatorname{bd} S$, $\operatorname{conv} S$, $\overline{\operatorname{conv}} S$ its *interior, closure, boundary, convex hull, closed convex hull*, respectively, and we denote by $\operatorname{diam}(S) := \sup\{\|x - y\| \mid x, y \in S\}$ its *diameter* and by $B_r(S) := \{x \in X \mid d(S; x) \leq r\}$ its *r-enlargement*. Closed sets and lsc functions are closely related as illustrated in Exercises 1.3.3, 1.3.4 and 1.3.5.

Another valuable tool in studying lsc functions is the *inf-convolution* of two functions f and g on a Banach space X defined by $(f \Box g)(x) := \inf_{y\in X}[f(y) + g(x - y)]$. Exercise 1.3.7 shows how this operation generates nice functions.

Multifunctions (*set-valued functions*) are equally interesting and useful. Denote by 2^Y the collection of all subsets of Y. A multifunction $F\colon X \to 2^Y$ maps each $x \in X$ to a subset $F(x)$ of Y. It is completely determined by its *graph*,

$$\text{graph } F := \{(x, y) \in X \times Y \mid y \in F(x)\},$$

a subset of the product space $X \times Y$ and, hence, by the indicator function $\iota_{\text{graph}\,F}$. The domain of a multifunction F is defined by $\text{dom}F := \{x \in X \mid F(x) \neq \emptyset\}$. The *inverse of a multifunction* $F\colon X \to 2^Y$ is defined by

$$F^{-1}(y) = \{x \in X \mid y \in F(x)\}.$$

Note that F^{-1} is a multifunction from Y to X. We say a multifunction F is *closed-valued* provided that for every $x \in \text{dom}F$, $F(x)$ is a closed set. We say the multifunction is *closed* if indeed the graph is a closed set in the product space. These two concepts are different (Exercise 1.3.8).

The ability to use extended-valued functions to relate sets, functions and multifunctions is one of the great advantages of the variational technique which is designed to deal fluently with such functions. In this book, for the most part, we shall focus on the theory for extended-valued functions. Corresponding results for sets and multifunctions are most often derivable by reducing them to appropriate function formulations.

1.3 Exercises

Exercise 1.3.1 Show that $x \in \overline{S}$ if and only if $d_S(x) = 0$.

Exercise 1.3.2 Suppose that S_1 and S_2 are two subsets of X. Show that $d_{S_1} = d_{S_2}$ if and only if $\overline{S_1} = \overline{S_2}$.

Exercise 1.3.3 Prove that S is a closed set if and only if ι_S is lsc.

Exercise 1.3.4 Prove that f is lsc if and only if $\text{epi}\, f$ is closed.

Exercise 1.3.5 Prove that f is lsc if and only if its sublevel set at a, $f^{-1}((-\infty, a])$, is closed for all $a \in \mathbb{R}$.

These results can be used to show the supremum of lsc functions is lsc.

Exercise 1.3.6 Let $\{f_a\}_{a \in A}$ be a family of lsc functions. Prove that $f := \sup\{f_a, a \in A\}$ is lsc. Hint: $\text{epi}\, f = \bigcap_{a \in A} \text{epi}\, f_a$.

Exercise 1.3.7 Let f be a lsc function bounded from below. Prove that if g is Lipschitz with rank L, then so is $f \Box g$.

Exercise 1.3.8 Let $F\colon X \to 2^Y$ be a multifunction. Show that if F has a closed graph then F is closed-valued, but the converse is not true.

2

Variational Principles

A lsc function on a noncompact set may well not attain its minimum. Roughly speaking, a variational principle asserts that, for any extended-valued lsc function which is bounded below, one can add a small perturbation to make it attain a minimum. Variational principles allow us to apply the variational technique to extended-valued lsc functions systematically, and therefore significantly extend the power of the variational technique. Usually, in a variational principle the better the geometric (smoothness) property of the underlying space the nicer the perturbation function. There are many possible settings. In this chapter, we focus on two of them: the Ekeland variational principle which holds in any complete metric space and the Borwein–Preiss smooth variational principle which ensures a smooth perturbation suffices in any Banach space with a smooth norm. We will also present a variant of the Borwein–Preiss variational principle derived by Deville, Godefroy and Zizler with an elegant category proof.

These variational principles provide powerful tools in modern variational analysis. Their applications cover numerous areas in both theory and applications of analysis including optimization, Banach space geometry, nonsmooth analysis, economics, control theory and game theory, to name a few. As a first taste we discuss some of their applications; these require minimum prerequisites in Banach space geometry, fixed point theory, an analytic proof of the Gordan theorem of the alternative, a characterization of the level sets associated with majorization and a variational proof of Birkhoff's theorem on the doubly stochastic matrices. Many other applications will be discussed in subsequent chapters.

2.1 Ekeland Variational Principles

2.1.1 The Geometric Picture

Consider a lsc function f bounded below on a Banach space $(X, \|\cdot\|)$. Clearly f may not attain its minimum or, to put it geometrically, f may not have

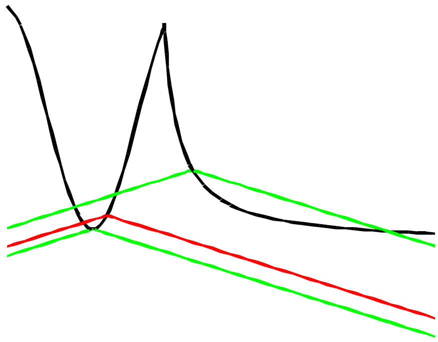

Fig. 2.1. Ekeland variational principle.
Top cone: $f(x_0) - \varepsilon|x - x_0|$
Middle cone: $f(x_1) - \varepsilon|x - x_1|$
Lower cone: $f(y) - \varepsilon|x - y|$

a supporting hyperplane. Ekeland's variational principle provides a kind of approximate substitute for the attainment of a minimum by asserting that, for any $\varepsilon > 0$, f must have a supporting cone of the form $f(y) - \varepsilon\|x - y\|$. One way to see how this happens geometrically is illustrated by Figure 2.1. We start with a point z_0 with $f(z_0) < \inf_X f + \varepsilon$ and consider the cone $f(z_0) - \varepsilon\|x - z_0\|$. If this cone does not support f then one can always find a point $z_1 \in S_0 := \{x \in X \mid f(x) \le f(z) - \varepsilon\|x - z\|)\}$ such that

$$f(z_1) < \inf_{S_0} f + \frac{1}{2}[f(z_0) - \inf_{S_0} f].$$

If $f(z_1) - \varepsilon\|x - z_1\|$ still does not support f then we repeat the above process. Such a procedure either finds the desired supporting cone or generates a sequence of nested closed sets (S_i) whose diameters shrink to 0. In the latter case, $f(y) - \varepsilon\|x - y\|$ is a supporting cone of f, where $\{y\} = \bigcap_{i=1}^{\infty} S_i$. This line of reasoning works similarly in a complete metric space. Moreover, it also provides a useful estimate on the distance between y and the initial ε-minimum z_0.

2.1.2 The Basic Form

We now turn to the analytic form of the geometric picture described above – the Ekeland variational principle and its proof.

Theorem 2.1.1 (Ekeland Variational Principle) *Let (X, d) be a complete metric space and let $f: X \to \mathbb{R} \cup \{+\infty\}$ be a lsc function bounded from below. Suppose that $\varepsilon > 0$ and $z \in X$ satisfy*

$$f(z) < \inf_X f + \varepsilon.$$

Then there exists $y \in X$ such that

(i) $d(z, y) \leq 1$,
(ii) $f(y) + \varepsilon d(z, y) \leq f(z)$, *and*
(iii) $f(x) + \varepsilon d(x, y) \geq f(y)$, *for all $x \in X$.*

Proof. Define a sequence (z_i) by induction starting with $z_0 := z$. Suppose that we have defined z_i. Set

$$S_i := \{x \in X \mid f(x) + \varepsilon d(x, z_i) \leq f(z_i)\}$$

and consider two possible cases: (a) $\inf_{S_i} f = f(z_i)$. Then we define $z_{i+1} := z_i$. (b) $\inf_{S_i} f < f(z_i)$. We choose $z_{i+1} \in S_i$ such that

$$f(z_{i+1}) < \inf_{S_i} f + \frac{1}{2}[f(z_i) - \inf_{S_i} f] = \frac{1}{2}[f(z_i) + \inf_{S_i} f] < f(z_i). \quad (2.1.1)$$

We show that (z_i) is a Cauchy sequence. In fact, if (a) ever happens then z_i is stationary for i large. Otherwise,

$$\varepsilon d(z_i, z_{i+1}) \leq f(z_i) - f(z_{i+1}). \quad (2.1.2)$$

Adding (2.1.2) up from i to $j - 1 > i$ we have

$$\varepsilon d(z_i, z_j) \leq f(z_i) - f(z_j). \quad (2.1.3)$$

Observe that the sequence $(f(z_i))$ is decreasing and bounded from below by $\inf_X f$, and therefore convergent. We conclude from (2.1.3) that (z_i) is Cauchy. Let $y := \lim_{i\to\infty} z_i$. We show that y satisfies the conclusions of the theorem. Setting $i = 0$ in (2.1.3) we have

$$\varepsilon d(z, z_j) + f(z_j) \leq f(z). \quad (2.1.4)$$

Taking limits as $j \to \infty$ yields (ii). Since $f(z) - f(y) \leq f(z) - \inf_X f < \varepsilon$, (i) follows from (ii). It remains to show that y satisfies (iii). Fixing i in (2.1.3) and taking limits as $j \to \infty$ yields $y \in S_i$. That is to say

$$y \in \bigcap_{i=1}^{\infty} S_i.$$

On the other hand, if $x \in \bigcap_{i=1}^{\infty} S_i$ then, for all $i = 1, 2, \ldots$,

$$\varepsilon d(x, z_{i+1}) \leq f(z_{i+1}) - f(x) \leq f(z_{i+1}) - \inf_{S_i} f. \quad (2.1.5)$$

It follows from (2.1.1) that $f(z_{i+1}) - \inf_{S_i} f \leq f(z_i) - f(z_{i+1})$, and therefore $\lim_i[f(z_{i+1}) - \inf_{S_i} f] = 0$. Taking limits in (2.1.5) as $i \to \infty$ we have $\varepsilon d(x, y) = 0$. It follows that

$$\bigcap_{i=1}^{\infty} S_i = \{y\}. \tag{2.1.6}$$

Notice that the sequence of sets (S_i) is nested, i.e., for any i, $S_{i+1} \subset S_i$. In fact, for any $x \in S_{i+1}$, $f(x) + \varepsilon d(x, z_{i+1}) \le f(z_{i+1})$ and $z_{i+1} \in S_i$ yields

$$\begin{aligned} f(x) + \varepsilon d(x, z_i) &\le f(x) + \varepsilon d(x, z_{i+1}) + \varepsilon d(z_i, z_{i+1}) \\ &\le f(z_{i+1}) + \varepsilon d(z_i, z_{i+1}) \le f(z_i), \end{aligned} \tag{2.1.7}$$

which implies that $x \in S_i$. Now, for any $x \neq y$, it follows from (2.1.6) that when i sufficiently large $x \notin S_i$. Thus, $f(x) + \varepsilon d(x, z_i) \ge f(z_i)$. Taking limits as $i \to \infty$ we arrive at (iii). ●

2.1.3 Other Forms

Since $\varepsilon > 0$ is arbitrary the supporting cone in the Ekeland's variational principle can be made as "flat" as one wishes. It turns out that in many applications such a flat supporting cone is enough to replace the possibly non-existent support plane. Another useful geometric observation is that one can trade between a flatter supporting cone and a smaller distance between the supporting point y and the initial ε-minimum z. The following form of this tradeoff can easily be derived from Theorem 2.1.1 by an analytic argument.

Theorem 2.1.2 *Let (X, d) be a complete metric space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a lsc function bounded from below. Suppose that $\varepsilon > 0$ and $z \in X$ satisfy*

$$f(z) < \inf_X f + \varepsilon.$$

Then, for any $\lambda > 0$ there exists y such that

(i) $d(z, y) \le \lambda$,
(ii) $f(y) + (\varepsilon/\lambda) d(z, y) \le f(z)$, *and*
(iii) $f(x) + (\varepsilon/\lambda) d(x, y) > f(y)$, *for all* $x \in X \setminus \{y\}$.

Proof. Exercise 2.1.1. ●

The constant λ in Theorem 2.1.2 makes it very flexible. A frequent choice is to take $\lambda = \sqrt{\varepsilon}$ and so to balance the perturbations in (ii) and (iii).

Theorem 2.1.3 *Let (X, d) be a complete metric space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a lsc function bounded from below. Suppose that $\varepsilon > 0$ and $z \in X$ satisfy*

$$f(z) < \inf_X f + \varepsilon.$$

Then, there exists y such that

(i) $d(z, y) \le \sqrt{\varepsilon}$,

(ii) $f(y) + \sqrt{\varepsilon} d(z, y) \le f(z)$, *and*
(iii) $f(x) + \sqrt{\varepsilon} d(x, y) > f(y)$, *for all* $x \in X \setminus \{y\}$.

Proof. Set $\lambda = \sqrt{\varepsilon}$ in Theorem 2.1.2. ●

When the approximate minimization point z in Theorem 2.1.2 is not explicitly known or is not important the following weak form of the Ekeland variational principle is useful.

Theorem 2.1.4 *Let (X, d) be a complete metric space and let $f: X \to \mathbb{R} \cup \{+\infty\}$ be a lsc function bounded from below. Then, for any $\varepsilon > 0$, there exists y such that*

$$f(x) + \sqrt{\varepsilon} d(x, y) > f(y).$$

Proof. Exercise 2.1.6. ●

2.1.4 Commentary and Exercises

Ekeland's variational principle, appeared in [102], is inspired by the Bishop–Phelps Theorem [22, 23] (see the next section). The original proof of the Ekeland variational principle in [102] is similar to that of the Bishop–Phelps Theorem using Zorn's lemma. J. Lasry pointed out transfinite induction is not needed and the proof given here is taken from the survey paper [103] and was credited to M. Crandall. As an immediate application we can derive a version of the results in Example 1.1.1 in infinite dimensional spaces (Exercises 2.1.2).

The lsc condition on f in the Ekeland variational principle can be relaxed somewhat. We leave the details in Exercises 2.1.4 and 2.1.5.

Exercise 2.1.1 Prove Theorem 2.1.2. Hint: Apply Theorem 2.1.1 with the metric $d(\cdot, \cdot)/\lambda$.

Exercise 2.1.2 Let X be a Banach space and let $f: X \to \mathbb{R}$ be a Fréchet differentiable function (see Section 3.1.1). Suppose that f is bounded from below on any bounded set and satisfies

$$\lim_{\|x\| \to \infty} \frac{f(x)}{\|x\|} = +\infty.$$

Then the range of f', $\{f'(x) \mid x \in X\}$, is dense in X^*.

Exercise 2.1.3 As a comparison, show that in Exercise 2.1.2, if X is a finite dimensional Banach space, then f' is onto. (Note also the assumption that f bounded from below on bounded sets is not necessary in finite dimensional spaces).

Exercise 2.1.4 We say a function f is partially lower semicontinuous (plsc) at x provided that, for any $x_i \to x$ with $f(x_i)$ monotone decreasing, one has $f(x) \leq \lim f(x_i)$. Prove that in Theorems 2.1.1 and 2.1.2, the assumption that f is lsc can be replaced by the weaker condition that f is plsc.

Exercise 2.1.5 Construct a class of plsc functions that are not lsc.

Exercise 2.1.6 Prove Theorem 2.1.4.

2.2 Geometric Forms Of the Variational Principle

In this section we discuss the Bishop–Phelps Theorem, the flower-petal theorem and the drop theorem. They capture the essence of the Ekeland variational principle from a geometric perspective.

2.2.1 The Bishop–Phelps Theorem

Among the three, the Bishop–Phelps Theorem [22, 23] is the closest to the Ekeland variational principle in its geometric explanation.

Let X be a Banach space. For any $x^* \in X^*\backslash\{0\}$ and any $\varepsilon > 0$ we say that

$$K(x^*, \varepsilon) := \{x \in X \mid \varepsilon\|x^*\|\|x\| \le \langle x^*, x\rangle\}$$

is a *Bishop–Phelps cone* associated with x^* and ε. We illustrate this in Figure 2.2 with the classic "ice cream cone" in three dimensions, in which $\varepsilon = 1/\sqrt{2}$ and $x^* = (-1, 1, 1)$.

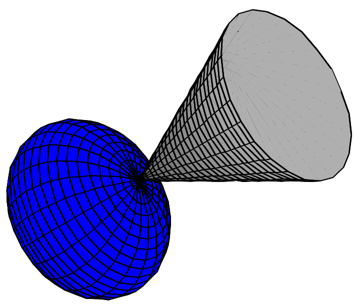

Fig. 2.2. A Bishop–Phelps cone.

Theorem 2.2.1 (Bishop–Phelps Theorem)*Let X be a Banach space and let S be a closed subset of X. Suppose that $x^* \in X^*$ is bounded on S. Then, for every $\varepsilon > 0$, S has a $K(x^*, \varepsilon)$ support point y, i.e.,*

$$\{y\} = S \cap [K(x^*, \varepsilon) + y].$$

Proof. Apply the Ekeland variational principle of Theorem 2.1.1 to the lsc function $f := -x^*/\|x^*\| + \iota_S$. We leave the details as an exercise. ●

The geometric picture of the Bishop–Phelps Theorem and that of the Ekeland variational principle are almost the same: the Bishop–Phelps cone $K(x^*, \varepsilon) + y$ in Theorem 2.2.1 plays a role similar to that of $f(y) - \varepsilon d(x, y)$ in Theorem 2.1.1. One can easily derive a Banach space version of the Ekeland variational principle by applying the Bishop–Phelps Theorem to the epigraph of a lsc function bounded from below (Exercise 2.2.2).

If we have additional information, e.g., known points inside and/or outside the given set, then the supporting cone can be replaced by more delicately constructed bounded sets. The flower-petal theorem and the drop theorem discussed in the sequel are of this nature.

2.2.2 The Flower-Petal Theorem

Let X be a Banach space and let $a, b \in X$. We say that

$$P_\gamma(a, b) := \{x \in X \mid \gamma\|a - x\| + \|x - b\| \le \|b - a\|\}$$

is a *flower petal* associated with $\gamma \in (0, +\infty)$ and $a, b \in X$. A flower petal is always convex, and interesting flower petals are formed when $\gamma \in (0, 1)$ (see Exercises 2.2.3 and 2.2.4).

Figure 2.3 draws the petals $P_\gamma((0,0),(1,0))$ for $\gamma = 1/3$, and $\gamma = 1/2$.

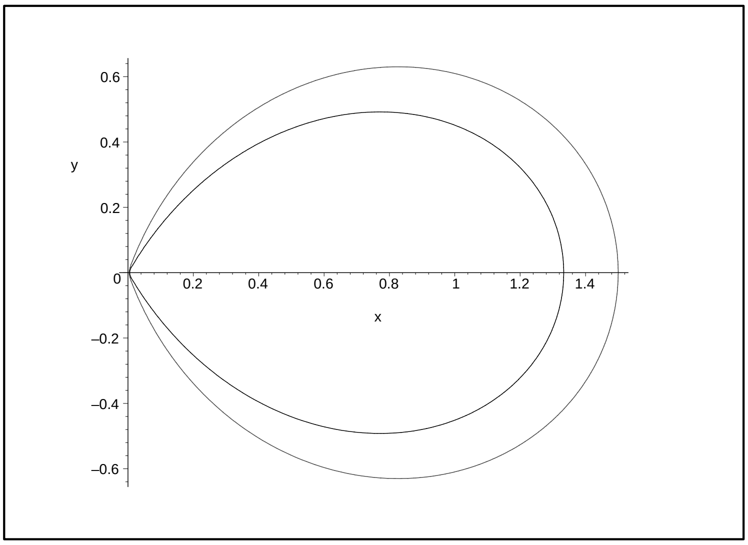

Fig. 2.3. Two flower petals.

Theorem 2.2.2 (Flower Petal Theorem)*Let X be a Banach space and let S be a closed subset of X. Suppose that $a \in S$ and $b \in X \backslash S$ with $r \in (0, d(S; b))$ and $t = \|b - a\|$. Then, for any $\gamma > 0$, there exists $y \in S \cap P_\gamma(a, b)$ satisfying $\|y - a\| \leq (t - r)/\gamma$ such that $P_\gamma(y, b) \cap S = \{y\}$.*

Proof. Define $f(x) := \|x - b\| + \iota_S(x)$. Then

$$f(a) < \inf_X f + (t - r).$$

Applying the Ekeland variational principle of Theorem 2.1.2 to the function $f(x)$ with and $\varepsilon = t - r$ and $\lambda = (t - r)/\gamma$, we have that there exists $y \in S$ such that $\|y - a\| < (t - r)/\gamma$ satisfying

$$\|y - b\| + \gamma\|a - y\| \leq \|a - b\|$$

and

$$\|x - b\| + \gamma\|x - y\| > \|y - b\|, \text{ for all } x \in S\backslash\{y\}.$$

The first inequality says $y \in P_\gamma(a, b)$ while the second implies that $P_\gamma(y, b) \cap S = \{y\}$. ●

2.2.3 The Drop Theorem

Let X be a Banach space, let C be a convex subset of X and let $a \in X$. We say that

$$[a, C] := \operatorname{conv}(\{a\} \cup C) = \{a + t(c - a) \mid c \in C\}$$

is the *drop* associated with a and C.

The following lemma provides useful information on the relationship between drops and flower petals. This is illustrated in Figure 2.4 and the easy proof is left as an exercise.

Lemma 2.2.3 (Drop and Flower Petal) *Let X be a Banach space, let $a, b \in X$ and let $\gamma \in (0, 1)$. Then*

$$B_{\|a-b\|(1-\gamma)/(1+\gamma)}(b) \subset P_\gamma(a, b),$$

so that

$$[a, B_{\|a-b\|(1-\gamma)/(1+\gamma)}(b)] \subset P_\gamma(a, b).$$

Proof. Exercise 2.2.5. ●

Now we can deduce the drop theorem from the flower petal theorem.

Theorem 2.2.4 (The Drop Theorem)*Let X be a Banach space and let S be a closed subset of X. Suppose that $b \in X\backslash S$ and $r \in (0, d(S; b))$. Then, for any $\varepsilon > 0$, there exists $y \in \operatorname{bd}(S)$ satisfying $\|y - b\| \leq d(S; b) + \varepsilon$ such that $[y, B_r(b)] \cap S = \{y\}$.*

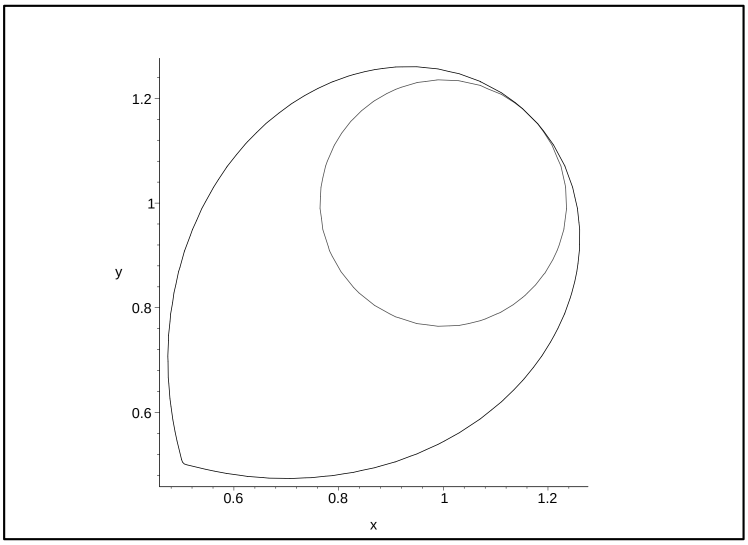

Fig. 2.4. A petal capturing a ball.

Proof. Choose $a \in S$ satisfying $\|a-b\| < d(S;b)+\varepsilon$ and choose

$$\gamma = \frac{\|a-b\|-r}{\|a-b\|+r} \in (0,1).$$

It follows from Theorem 2.2.2 that there exists $y \in S \cap P_\gamma(a,b)$ such that $P_\gamma(y,b) \cap S = \{y\}$. Clearly, $y \in \mathrm{bd}(S)$. Moreover, $y \in P_\gamma(a,b)$ implies that $\|y-b\| < \|a-b\| < d(S;y)+\varepsilon$. Finally, it follows from Lemma 2.2.3 and $r = \frac{1-\gamma}{1+\gamma}\|a-b\|$ that $[y, B_r(b)] \cap S = \{y\}$. ●

2.2.4 The Equivalence with Completeness

Actually, all the results discussed in this section and the Ekeland variational principle are equivalent provided that one states them in sufficiently general form (see e.g. [130]). In the setting of a general metric space, the Ekeland variational principle is more flexible in various applications. More importantly it shows that completeness, rather than the linear structure of the underlying space, is the essential feature. In fact, the Ekeland variational principle characterizes the completeness of a metric space.

Theorem 2.2.5 (Ekeland Variational Principle and Completeness) *Let (X,d) be a metric space. Then X is complete if and only if for every lsc function $f\colon X \to \mathbb{R} \cup \{+\infty\}$ bounded from below and for every $\varepsilon > 0$ there exists a point $y \in X$ satisfying*

$$f(y) \le \inf_X f + \varepsilon,$$

and

$$f(x) + \varepsilon d(x, y) \geq f(y), \quad \textit{for all } x \in X.$$

Proof. The "if" part follows from Theorem 2.1.4. We prove the "only if" part. Let (x_i) be a Cauchy sequence. Then, the function $f(x) := \lim_{i\to\infty} d(x_i, x)$ is well-defined and nonnegative. Since the distance function is Lipschitz with respect to x we see that f is continuous. Moreover, since (x_i) is a Cauchy sequence we have $f(x_i) \to 0$ as $i \to \infty$ so that $\inf_X f = 0$. For $\varepsilon \in (0, 1)$ choose y such that $f(y) \leq \varepsilon$ and

$$f(y) \leq f(x) + \varepsilon d(x, y), \quad \text{for all } x \in X \tag{2.2.1}$$

Letting $x = x_i$ in (2.2.1) and taking limits as $i \to \infty$ we obtain $f(y) \leq \varepsilon f(y)$ so that $f(y) = 0$. That is to say $\lim_{i\to\infty} x_i = y$. ●

2.2.5 Commentary and Exercises

The Bishop–Phelps theorem is the earliest of this type [22, 23]. In fact, this important result in Banach space geometry is the main inspiration for Ekeland's variational principle (see [103]). The drop theorem was discovered by Danes [91]. The relationship among the Ekeland variational principle, the drop theorem and the flower-petal theorem were discussed in Penot [211] and Rolewicz [232]. The book [136] by Hyers, Isac and Rassias is a nice reference containing many other variations and applications of the Ekeland variational principle.

Exercise 2.2.1 Provide details for the proof of Theorem 2.2.1.

Exercise 2.2.2 Deduce the Ekeland variational principle in a Banach space by applying the Bishop–Phelps Theorem to the epigraph of a lsc function.

Exercise 2.2.3 Show that, for $\gamma > 1$, $P_\gamma(a, b) = \{a\}$ and $P_1(a, b) = \{\lambda a + (1 - \lambda)b \mid \lambda \in [0, 1]\}$.

Exercise 2.2.4 Prove that $P_\gamma(a, b)$ is convex.

Exercise 2.2.5 Prove Lemma 2.2.3.

2.3 Applications to Fixed Point Theorems

Let X be a set and let f be a map from X to itself. We say x is a *fixed point* of f if $f(x) = x$. Fixed points of a mapping often represent equilibrium states of some underlying system, and they are consequently of great importance. Therefore, conditions ensuring the existence and uniqueness of fixed point(s) are the subject of extensive study in analysis. We now use Ekeland's variational principle to deduce several fixed point theorems.

2.3.1 The Banach Fixed Point Theorem

Let (X, d) be a complete metric space and let ϕ be a map from X to itself. We say that ϕ is a *contraction* provided that there exists $k \in (0, 1)$ such that

$$d(\phi(x), \phi(y)) \le kd(x, y), \quad \text{for all } x, y \in X.$$

Theorem 2.3.1 (Banach Fixed Point Theorem) *Let (X, d) be a complete metric space. Suppose that $\phi\colon X \to X$ is a* contraction*. Then ϕ has a unique fixed point.*

Proof. Define $f(x) := d(x, \phi(x))$. Applying Theorem 2.1.1 to f with $\varepsilon \in (0, 1-k)$, we have $y \in X$ such that

$$f(x) + \varepsilon d(x, y) \ge f(y), \quad \text{for all } x \in X.$$

In particular, setting $x = \phi(y)$ we have

$$d(y, \phi(y)) \le d(\phi(y), \phi^2(y)) + \varepsilon d(y, \phi(y)) \le (k+\varepsilon) d(y, \phi(y)).$$

Thus, y must be a fixed point. The uniqueness follows directly from the fact that ϕ is a contraction and is left as an exercise. ●

2.3.2 Clarke's Refinement

Clarke observed that the argument in the proof of the Banach fixed point theorem works under weaker conditions. Let (X, d) be a complete metric space. For $x, y \in X$ we define the *segment* between x and y by

$$[x, y] := \{z \in X \mid d(x, z) + d(z, y) = d(x, y)\}. \tag{2.3.1}$$

Definition 2.3.2 (Directional Contraction)*Let (X, d) be a complete metric space and let ϕ be a map from X to itself. We say that ϕ is a* directional contraction *provided that*

(i) *ϕ is continuous, and*
(ii) *there exists $k \in (0,1)$ such that, for any $x \in X$ with $\phi(x) \neq x$ there exists $z \in [x, \phi(x)] \backslash \{x\}$ such that*

$$d(\phi(x), \phi(z)) \leq kd(x, z).$$

Theorem 2.3.3 *Let (X, d) be a complete metric space. Suppose that $\phi \colon X \to X$ is a* directional contraction. *Then ϕ admits a fixed point.*

Proof. Define

$$f(x) := d(x, \phi(x)).$$

Then f is continuous and bounded from below (by 0). Applying the Ekeland variational principle of Theorem 2.1.1 to f with $\varepsilon \in (0, 1-k)$ we conclude that there exists $y \in X$ such that

$$f(y) \leq f(x) + \varepsilon d(x, y), \quad \text{for all } x \in X. \tag{2.3.2}$$

If $\phi(y) = y$, we are done. Otherwise, since ϕ is a directional contraction there exists a point $z \neq y$ with $z \in [y, \phi(y)]$, i.e.,

$$d(y, z) + d(z, \phi(y)) = d(y, \phi(y)) = f(y) \tag{2.3.3}$$

satisfying

$$d(\phi(z), \phi(y)) \leq kd(z, y). \tag{2.3.4}$$

Letting $x = z$ in (2.3.2) and using (2.3.3) we have

$$d(y, z) + d(z, y) \leq d(z, \phi(z)) + \varepsilon d(z, y)$$

or

$$d(y, z) \leq d(z, \phi(z)) - d(z, \phi(y)) + \varepsilon d(z, y) \tag{2.3.5}$$

By the triangle inequality and (2.3.4) we have

$$d(z, \phi(z)) - d(z, \phi(y)) \leq d(\phi(y), \phi(z)) \leq kd(y, z). \tag{2.3.6}$$

Combining (2.3.5) and (2.3.6) we have

$$d(y, z) \leq (k + \varepsilon) d(y, z),$$

a contradiction. ●

Clearly any contraction is a directional contraction. Therefore, Theorem 2.3.3 generalizes the Banach fixed point theorem. The following is an example where Theorem 2.3.3 applies when the Banach contraction theorem does not.

Example 2.3.4 Consider $X = \mathbb{R}^2$ with a metric induced by the norm $\|x\| = \|(x_1, x_2)\| = |x_1| + |x_2|$. A segment between two points (a_1, a_2) and (b_1, b_2) consists of the closed rectangle having the two points as diagonally opposite corners. Define

$$\phi(x_1, x_2) = \Big(\frac{3x_1}{2} - \frac{x_2}{3}, x_1 + \frac{x_2}{3}\Big).$$

Then ϕ is a directional contraction. Indeed, if $y = \phi(x) \neq x$. Then $y_2 \neq x_2$ (for otherwise we will also have $y_1 = x_1$). Now the set $[x, y]$ contains points of the form (x_1, t) with t arbitrarily close to x_2 but not equal to x_2. For such points we have

$$d(\phi(x_1, t), \phi(x_1, x_2)) = \frac{2}{3} d((x_1, t), (x_1, x_2)),$$

so that ϕ is a directional contraction. We can directly check that the fixed points of ϕ are all points of the form $(x, 3x/2)$. Since ϕ has more than one fixed point clearly the Banach fixed point theorem does not apply to this mapping.

2.3.3 The Caristi–Kirk Fixed Point Theorem

A similar argument can be used to prove the Caristi–Kirk fixed point theorem for multifunctions. For a multifunction $F\colon X \to 2^X$, we say that x is a fixed point for F provided that $x \in F(x)$.

Theorem 2.3.5 (Caristi–Kirk Fixed Point Theorem) *Let (X, d) be a complete metric space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a proper lsc function bounded below. Suppose $F\colon X \to 2^X$ is a multifunction with a closed graph satisfying*

$$f(y) \leq f(x) - d(x, y), \quad \textit{for all } (x, y) \in \operatorname{graph} F. \tag{2.3.7}$$

Then F has a fixed point.

Proof. Define a metric ρ on $X \times X$ by $\rho((x_1, y_1), (x_2, y_2)) := d(x_1, x_2) + d(y_1, y_2)$ for any $(x_1, y_1), (x_2, y_2) \in X \times X$. Then $(X \times X, \rho)$ is a complete metric space. Let $\varepsilon \in (0, 1/2)$ and define $g\colon X \times X \to \mathbb{R} \cup \{+\infty\}$ by $g(x, y) := f(x) - (1 - \varepsilon) d(x, y) + \iota_{\operatorname{graph} F}(x, y)$. Then g is a lsc function bounded below (exercise). Applying the Ekeland variational principle of Theorem 2.1.1 to g we see that there exists $(x^*, y^*) \in \operatorname{graph} F$ such that

$$g(x^*, y^*) \leq g(x, y) + \varepsilon\rho((x, y), (x^*, y^*)), \quad \text{for all } (x, y) \in X \times X.$$

So for all $(x, y) \in \operatorname{graph} F$,

$$\begin{aligned} &f(x^*) - (1 - \varepsilon) d(x^*, y^*) \\ &\qquad \leq f(x) - (1 - \varepsilon) d(x, y) + \varepsilon(d(x, x^*) + d(y, y^*)). \end{aligned} \tag{2.3.8}$$

Suppose $z^* \in F(y^*)$. Letting $(x, y) = (y^*, z^*)$ in (2.3.8) we have

$$f(x^*) - (1-\varepsilon)d(x^*, y^*) \le f(y^*) - (1-\varepsilon)d(y^*, z^*) + \varepsilon(d(y^*, x^*) + d(z^*, y^*)).$$

It follows that

$$0 \le f(x^*) - f(y^*) - d(x^*, y^*) \le -(1-2\varepsilon)d(y^*, z^*),$$

so we must have $y^* = z^*$. That is to say y^* is a fixed point of F. •

We observe that it follows from the above proof that $F(y^*) = \{y^*\}$.

2.3.4 Commentary and Exercises

The variational proof of the Banach fixed point theorem appeared in [103]. While the variational argument provides an elegant confirmation of the existence of the fixed point it does not, however, provide an algorithm for finding such a fixed point as Banach's original proof does. For comparison, a proof using an interactive algorithm is outlined in the guided exercises below. Clarke's refinement is taken from [87]. Theorem 2.3.5 is due to Caristi and Kirk [155] and applications of this theorem can be found in [101]. A very nice general reference book for the metric fixed point theory is [123].

Exercise 2.3.1 Let X be a Banach space and let $x, y \in X$. Show that the segment between x and y defined in (2.3.1) has the following representation:

$$[x, y] = \{\lambda x + (1-\lambda)y \mid \lambda \in [0, 1]\}.$$

Exercise 2.3.2 Prove the uniqueness of the fixed point in Theorem 2.3.1.

Exercise 2.3.3 Let $f\colon \mathbb{R}^N \to \mathbb{R}^N$ be a C^1 mapping. Show that f is a contraction if and only if $\sup\{\|f'(x)\| : x \in \mathbb{R}^N\} < 1$.

Exercise 2.3.4 Prove that Kepler's equation

$$x = a + b\sin(x),\ b \in (0, 1)$$

has a unique solution.

Exercise 2.3.5 (Iteration Method) Let (X, d) be a complete metric space and let $\phi\colon X \to X$ be a contraction. Define for an arbitrarily fixed $x_0 \in X$, $x_1 = \phi(x_0), \ldots, x_i = \phi(x_{i-1})$. Show that (x_i) is a Cauchy sequence and $x = \lim_{i\to\infty} x_i$ is a fixed point for ϕ.

Exercise 2.3.6 (Error Estimate) Let (X, d) be a complete metric space and let $\phi\colon X \to X$ be a contraction with contraction constant $k \in (0, 1)$. Establish the following error estimate for the iteration method in Exercise 2.3.5.

$$\|x_i - x\| \le \frac{k^i}{1-k}\|x_1 - x_0\|.$$

Exercise 2.3.7 Deduce the Banach fixed point theorem from the Caristi–Kirk fixed point theorem. Hint: Define $f(x) = d(x, \phi(x))/(1-k)$.

2.4 Variational Principles in Finite Dimensional Spaces

One drawback of the Ekeland variational principle is that the perturbation involved therein is intrinsically nonsmooth. This is largely overcome in the smooth variational principle due to Borwein and Preiss. We discuss a Euclidean space version in this section to illustrate the nature of this result. The general version will be discussed in the next section.

2.4.1 Smooth Variational Principles in Euclidean Spaces

Theorem 2.4.1 (Smooth Variational Principle in a Euclidean Space) *Let $f\colon \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ be a lsc function bounded from below, let $\lambda > 0$ and let $p \geq 1$. Suppose that $\varepsilon > 0$ and $z \in X$ satisfy*

$$f(z) \leq \inf_X f + \varepsilon.$$

Then, there exists $y \in X$ such that

(i) $\|z - y\| \leq \lambda$,
(ii) $f(y) + \frac{\varepsilon}{\lambda^p}\|y - z\|^p \leq f(z)$, *and*
(iii) $f(x) + \frac{\varepsilon}{\lambda^p}\|x - z\|^p \geq f(y) + \frac{\varepsilon}{\lambda^p}\|y - z\|^p$, *for all $x \in X$.*

Proof. Observing that the function $x \to f(x) + \frac{\varepsilon}{\lambda^p}\|x - z\|^p$ approaches $+\infty$ as $\|x\| \to \infty$, it must attain its minimum at some $y \in X$. It is an easy matter to check that y satisfies the conclusion of the theorem. ●

This very explicit formulation which is illustrated in Figure 2.5 – for $f(x) = 1/x, z = 1, \varepsilon = 1, \lambda = 1/2$, with $p = 3/2$ and $p = 2$ – can be mimicked in Hilbert space and many other classical reflexive Banach spaces [46]. It is interesting to compare this result with the Ekeland variational principle geometrically. The Ekeland variational principle says that one can support a lsc function f near its approximate minimum point by a cone with small slope while the Borwein–Preiss variational principle asserts that under stronger conditions this cone can be replaced by a parabolic function with a small derivative at the supporting point. We must caution the readers that although this picture is helpful in understanding the naturalness of the Borwein–Preiss variational principle it is not entirely accurate in the general case, as the support function is usually the sum of an infinite sequence of parabolic functions.

This result can also be stated in the form of an approximate Fermat principle in the Euclidean space $\mathbb{R}^N$.

Lemma 2.4.2 (Approximate Fermat Principle for Smooth Functions) *Let $f\colon \mathbb{R}^N \to \mathbb{R}$ be a smooth function bounded from below. Then there exists a sequence $x_i \in \mathbb{R}^N$ such that $f(x_i) \to \inf_{\mathbb{R}^N} f$ and $f'(x_i) \to 0$.*

Proof. Exercise 2.4.3. ●

We delay the discussion of the general form of the Borwein–Preiss variational principle until the next section and digress to some applications.

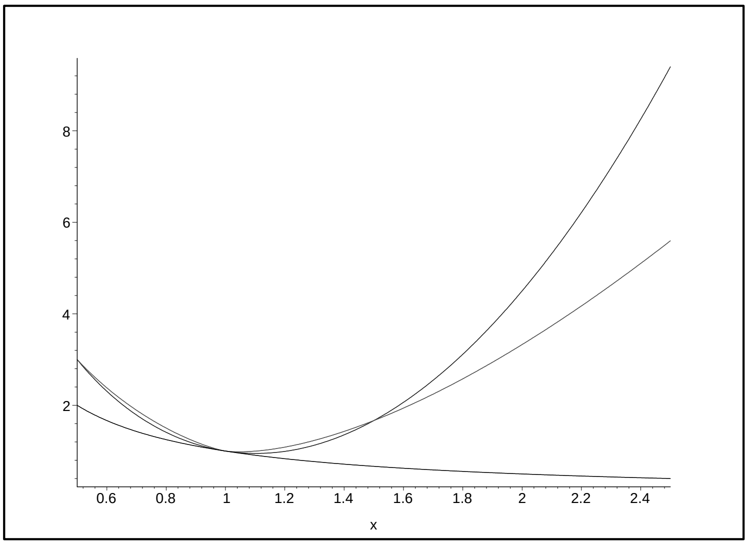

Fig. 2.5. Smooth attained perturbations of $1/x$

2.4.2 Gordan Alternatives

We start with an analytical proof of the Gordan alternative.

Theorem 2.4.3 (Gordan Alternative) *Let $a_1, \dots, a_M \in \mathbb{R}^N$. Then, exactly one of the following systems has a solution:*

$$\sum_{m=1}^{M} \lambda_m a_m = 0, \sum_{m=1}^{M} \lambda_m = 1, 0 \le \lambda_m,\ m = 1, \dots, M, \tag{2.4.1}$$

$$\langle a_m, x \rangle < 0 \textit{ for } m = 1, \dots, M,\ x \in \mathbb{R}^N. \tag{2.4.2}$$

Proof. We need only prove the following statements are equivalent:

(i) The function

$$f(x) := \ln \Big(\sum_{m=1}^{M} \exp \ \langle a_m, x \rangle \Big)$$

is bounded below.

(ii) System (2.4.1) is solvable.

(iii) System (2.4.2) is unsolvable.

The implications (ii)$\Rightarrow$ (iii) $\Rightarrow$ (i) are easy and left as exercises. It remains to show (i) $\Rightarrow$ (ii). Applying the approximate Fermat principle of Lemma 2.4.2 we deduce that there is a sequence (x_i) in $\mathbb{R}^N$ satisfying

$$\|f'(x_i)\| = \left\| \sum_{m=1}^{M} \lambda_m^i a_m \right\| \to 0, \tag{2.4.3}$$

where the scalars

$$\lambda_m^i = \frac{\exp\ \langle a_m, x_i \rangle}{\sum_{l=0}^{M} \exp\ \langle a_l, x_i \rangle} > 0,\ m = 1, \ldots, M$$

satisfy $\sum_{m=1}^{M} \lambda_m^i = 1$. Without loss of generality we may assume that $\lambda_m^i \to \lambda_m$, $m = 1, \ldots, M$. Taking limits in (2.4.3) we see that λ_m, $m = 1, \ldots, M$ is a set of solutions of (2.4.1). ●

2.4.3 Majorization

For a vector $x = (x_1, \ldots, x_N) \in \mathbb{R}^N$, we use $x^\downarrow$ to denote the vector derived from x by rearranging its components in nonincreasing order. For $x, y \in \mathbb{R}^N$, we say that x is *majorized* by y, denoted by $x \prec y$, provided that $\sum_{n=1}^{N} x_n = \sum_{n=1}^{N} y_n$ and $\sum_{n=1}^{k} x_n^\downarrow \le \sum_{n=1}^{k} y_n^\downarrow$ for $k = 1, \ldots, N$.

Example 2.4.4 Let $x \in \mathbb{R}^N$ be a vector with nonnegative components satisfying $\sum_{n=1}^{N} x_n = 1$. Then

$$(1/N, 1/N, \ldots, 1/N) \prec x \prec (1, 0, \ldots, 0).$$

The concept of majorization arises naturally in physics and economics. For example, if we use $x \in \mathbb{R}_+^N$ (the nonnegative orthant of $\mathbb{R}^N$) to represent the distribution of wealth within an economic system, then $x \prec y$ means the distribution represented by x is more even than that of y. Example 2.4.4 then describes the two extremal cases of wealth distribution.

Given a vector $y \in \mathbb{R}^N$ the level set of y with respect to the majorization defined by $l(y) := \{x \in \mathbb{R}^N \mid x \prec y\}$ is often of interest. It turns out that this level set is the convex hull of all the possible vectors derived from permuting the components of y. We will give a variational proof of this fact using a method similar to that of the variational proof of the Gordon alternatives. To do so we will need the following characterization of majorization.

Lemma 2.4.5 *Let $x, y \in \mathbb{R}^N$. Then $x \prec y$ if and only if, for any $z \in \mathbb{R}^N$, $\langle z^\downarrow, x^\downarrow \rangle \le \langle z^\downarrow, y^\downarrow \rangle$.*

Proof. Using Abel's formula we can write

$$\begin{aligned}\langle z^\downarrow, y^\downarrow \rangle - \langle z^\downarrow, x^\downarrow \rangle &= \langle z^\downarrow, y^\downarrow - x^\downarrow \rangle \\ &= \sum_{k=1}^{N-1} \Big((z_k^\downarrow - z_{k+1}^\downarrow) \times \sum_{n=1}^{k} (y_n^\downarrow - x_n^\downarrow) \Big) + z_N^\downarrow \sum_{n=1}^{N} (y_n^\downarrow - x_n^\downarrow).\end{aligned}$$

Now to see the necessity we observe that $x \prec y$ implies $\sum_{n=1}^{k}(y_n^{\downarrow} - x_n^{\downarrow}) \geq 0$ for $k = 1, \ldots, N-1$ and $\sum_{n=1}^{N}(y_n^{\downarrow} - x_n^{\downarrow}) = 0$. Thus, the last term in the right hand side of the previous equality is 0. Moreover, in the remaining sum each term is the product of two nonnegative factors, and therefore it is nonnegative. We now prove sufficiency. Suppose that, for any $z \in \mathbb{R}^N$,

$$0 \leq \langle z^{\downarrow}, y^{\downarrow}\rangle - \langle z^{\downarrow}, x^{\downarrow}\rangle = \sum_{k=1}^{N-1}\Big((z_k^{\downarrow} - z_{k+1}^{\downarrow}) \times \sum_{n=1}^{k}(y_n^{\downarrow} - x_n^{\downarrow})\Big) + z_N^{\downarrow}\sum_{n=1}^{N}(y_n^{\downarrow} - x_n^{\downarrow}).$$

Setting $z = \sum_{n=1}^{k} e_n$ for $k = 1, \ldots, N-1$ (where $\{e_n : n = 1, \ldots, N\}$ is the standard basis of $\mathbb{R}^N$) we have $\sum_{n=1}^{k} y_n^{\downarrow} \geq \sum_{n=1}^{k} x_n^{\downarrow}$, and setting $z = \pm\sum_{n=1}^{N} e_n$ we have $\sum_{n=1}^{N} y_n = \sum_{n=1}^{N} x_n$. ●

Let us denote by $P(N)$ the set of $N \times N$ permutation matrices (those matrices derived by permuting the rows or the columns of the identity matrix). Then we can state the characterization of the level set of a vector with respect to majorization as follows.

Theorem 2.4.6 (Representation of Level Sets of the Majorization) *Let $y \in \mathbb{R}^N$. Then*

$$l(y) = \operatorname{conv}\{Py : P \in P(N)\}.$$

Proof. It is not hard to check that $l(y)$ is convex and, for any $P \in P(N)$, $Py \in l(y)$. Thus, $\operatorname{conv}\{Py : P \in P(N)\} \subset l(y)$ (Exercise 2.4.8).

We now prove the reversed inclusion. For any $x \prec y$, by Lemma 2.4.5 there exists $P = P(z) \in P(N)$ satisfies

$$\langle z, Py\rangle = \langle z^{\downarrow}, y^{\downarrow}\rangle \geq \langle z^{\downarrow}, x^{\downarrow}\rangle \geq \langle z, x\rangle. \tag{2.4.4}$$

Observe that $P(N)$ is a finite set (with $N!$ elements to be precise). Thus, the function

$$f(z) := \ln\Big(\sum_{P\in P(N)} \exp\langle z, Py - x\rangle\Big).$$

is defined for all $z \in \mathbb{R}^N$, is differentiable, and is bounded from below by 0. By the approximate Fermat principle of Lemma 2.4.2 we can select a sequence (z_i) in $\mathbb{R}^N$ such that

$$0 = \lim_{i\to\infty} f'(z_i) = \sum_{P\in P(N)} \lambda_P^i (Py - x). \tag{2.4.5}$$

where

$$\lambda_P^i = \frac{\exp\langle z_i, Py - x\rangle}{\sum_{P\in P(N)} \exp\langle z_i, Py - x\rangle}.$$

Clearly, $\lambda_P^i > 0$ and $\sum_{P\in P(N)} \lambda_P^i = 1$. Thus, taking a subsequence if necessary we may assume that, for each $P \in P(N)$, $\lim_{i\to\infty} \lambda_P^i = \lambda_P \geq 0$ and $\sum_{P\in P(N)} \lambda_P = 1$. Now taking limits as $i \to \infty$ in (2.4.5) we have

$$\sum_{P\in P(N)} \lambda_P(Py - x) = 0.$$

Thus, $x = \sum_{P\in P(N)} \lambda_P Py$, as was to be shown. •

2.4.4 Doubly Stochastic Matrices

We use $E(N)$ to denote the Euclidean space of all real N by N square matrices with inner product

$$\langle A, B\rangle = \mathrm{tr}(B^\top A) = \sum_{n,m=1}^{N} a_{nm}b_{nm}, \quad A, B \in E(N).$$

A matrix $A = (a_{nm}) \in E(N)$ is *doubly stochastic* provided that the entries of A are all nonnegative, $\sum_{n=1}^N a_{nm} = 1$ for $m = 1, \dots, N$ and $\sum_{m=1}^N a_{nm} = 1$ for $n = 1, \dots, N$. Clearly every $P \in P(N)$ is doubly stochastic and they provide the simplest examples of doubly stochastic matrices. Birkhoff's theorem asserts that any doubly stochastic matrix can be represented as a convex combination of permutation matrices. We now apply the method in the previous section to give a variational proof of Birkhoff's thoerem.

For $A = (a_{nm}) \in E(N)$, we denote $r_n(A) = \{m \mid a_{nm} \neq 0\}$, the set of indices of columns containing nonzero elements of the nth row of A and we use $\#(S)$ to signal the number of elements in set S. Then a doubly stochastic matrix has the following interesting property.

Lemma 2.4.7 *Let $A \in E(N)$ be a doubly stochastic matrix. Then, for any $1 \le n_1 < n_2 < \cdots < n_K \le N$,*

$$\#\Big(\bigcup_{k=1}^{K} r_{n_k}(A)\Big) \ge K. \tag{2.4.6}$$

Proof. We prove by contradiction. Suppose (2.4.6) is violated for some K. Permuting the rows of A if necessary we may assume that

$$\#\Big(\bigcup_{k=1}^{K} r_k(A)\Big) < K. \tag{2.4.7}$$

Rearranging the order of the columns of A if needed we may assume

$$A = \begin{pmatrix} O & B \\ C & D \end{pmatrix},$$

where O is a K by L submatrix of A with all entries equal to 0. By (2.4.7) we have $L > N - K$. On the other hand, since A is doubly stochastic, every column of C and every row of B add up to 1. That leads to $L + K \le N$, a contradiction. ●

Condition (2.4.6) actually ensures a matrix has a diagonal with all elements nonzero which is made precise in the next lemma.

Lemma 2.4.8 *Let $A \in E(N)$. Suppose that A satisfies condition (2.4.6). Then for some $P \in P(N)$, the entries in A corresponding to the 1's in P are all nonzero. In particular, any doubly stochastic matrix has the above property.*

Proof. We use induction on N. The lemma holds trivially when $N = 1$. Now suppose that the lemma holds for any integer less than N. We prove it is true for N. First suppose that, for any $1 \le n_1 < n_2 < \cdots < n_K \le N$, $K < N$

$$\#\Big(\bigcup_{k=1}^{K} r_{n_k}(A)\Big) \ge K + 1. \tag{2.4.8}$$

Then pick a nonzero element of A, say a_{NN} and consider the submatrix A' of A derived by eliminating the Nth row and Nth column of A. Then A' satisfies condition (2.4.6), and therefore there exists $P' \in P(N-1)$ such that the entries in A' corresponding to the 1's in P' are all nonzero. It remains to define $P \in P(N)$ as

$$P = \begin{pmatrix} P' & 0 \\ 0 & 1 \end{pmatrix}.$$

Now consider the case when (2.4.8) fails so that there exist $1 \le n_1 < n_2 < \cdots < n_K \le N$, $K < N$ satisfying

$$\#\Big(\bigcup_{k=1}^{K} r_{n_k}(A)\Big) = K. \tag{2.4.9}$$

By rearranging the rows and columns of A we may assume that $n_k = k, k = 1, \ldots, K$ and $\bigcup_{k=1}^{K} r_k(A) = \{1, \ldots, K\}$. Then

$$A = \begin{pmatrix} B & O \\ C & D \end{pmatrix},$$

where $B \in E(K)$, $D \in E(N-K)$ and O is a K by $N-K$ submatrix with all entries equal to 0. Observe that for any $1 \le n_1 < \cdots < n_L \le K$,

$$\bigcup_{l=1}^{L} r_{n_l}(B) = \bigcup_{l=1}^{L} r_{n_l}(A).$$

Thus,

$$\#\Big(\bigcup_{l=1}^{L} r_{n_l}(B)\Big) \ge L,$$

and therefore B satisfies condition (2.4.6). On the other hand for any $K+1 \leq n_1 < \cdots < n_L \leq N$,

$$\Big[\bigcup_{k=1}^{K} r_k(A)\Big] \cup \Big[\bigcup_{l=1}^{L} r_{n_l}(A)\Big] = \{1, \ldots, K\} \cup \Big[\bigcup_{l=1}^{L} r_{n_l}(D)\Big].$$

Thus, D also satisfies condition (2.4.6). By the induction hypothesis we have $P_1 \in P(K)$ and $P_2 \in P(N-K)$ such that the elements in B and D corresponding to the 1's in P_1 and P_2, respectively, are all nonzero. It follows that

$$P = \begin{pmatrix} P_1 & O \\ O & P_2 \end{pmatrix} \in P(N),$$

and the elements in A corresponding to the 1's in P are all nonzero. •

We now establish the following analogue of (2.4.4).

Lemma 2.4.9 *Let $A \in E(N)$ be a doubly stochastic matrix. Then for any $B \in E(N)$ there exists $P \in P(N)$ such that*

$$\langle B, A - P\rangle \geq 0.$$

Proof. We use an induction argument on the number of nonzero elements of A. Since every row and column of A sums to 1, A has at least N nonzero elements. If A has exactly N nonzero elements then they must all be 1, so that A itself is a permutation matrix and the lemma holds trivially. Suppose now that A has more than N nonzero elements. By Lemma 2.4.8 there exists $P \in P(N)$ such that the entries in A corresponding to the 1's in P are all nonzero. Let $t \in (0, 1)$ be the minimum of these N positive elements. Then we can verify that $A_1 = (A - tP)/(1 - t)$ is a doubly stochastic matrix and has at least one fewer nonzero elements than A. Thus, by the induction hypothesis there exists $Q \in P(N)$ such that

$$\langle B, A_1 - Q\rangle \geq 0.$$

Multiplying the above inequality by $1 - t$ we have $\langle B, A - tP - (1 - t)Q\rangle \geq 0$, and therefore at least one of $\langle B, A - P\rangle$ or $\langle B, A - Q\rangle$ is nonnegative. •

Now we are ready to present a variational proof for the Birkhoff theorem.

Theorem 2.4.10 (Birkhoff)*Let $\mathcal{A}(N)$ be the set of all $N \times N$ doubly stochastic matrices. Then*

$$\mathcal{A}(N) = \operatorname{conv}\{P \mid P \in P(N)\}.$$

Proof. It is an easy matter to verify that $\mathcal{A}(N)$ is convex and $P(N) \subset \mathcal{A}(N)$. Thus, $\operatorname{conv} P(N) \subset \mathcal{A}(N)$.

To prove the reversed inclusion, define a function f on $E(N)$ by

$$f(B) := \ln\Big(\sum_{P\in P(N)} \exp\langle B, A-P\rangle\Big).$$

Then f is defined for all $B \in E(N)$, is differentiable and is bounded from below by 0. By the approximate Fermat principle of Theorem 2.4.2 we can select a sequence (B_i) in $E(N)$ such that

$$0 = \lim_{i\to\infty} f'(B_i) = \lim_{i\to\infty} \sum_{P\in P(N)} \lambda^i_P(A-P). \tag{2.4.10}$$

where

$$\lambda^i_P = \frac{\exp\langle B_i, A-P\rangle}{\sum_{P\in P(N)} \exp\langle B_i, A-P\rangle}.$$

Clearly, $\lambda^i_P > 0$ and $\sum_{P\in P(N)} \lambda^i_P = 1$. Thus, taking a subsequence if necessary we may assume that for each $P \in P(N)$, $\lim_{i\to\infty} \lambda^i_P = \lambda_P \geq 0$ and $\sum_{P\in P(N)} \lambda_P = 1$. Now taking limits as $i \to \infty$ in (2.4.10) we have

$$\sum_{P\in P(N)} \lambda_P(A-P) = 0.$$

It follows that $A = \sum_{P\in P(N)} \lambda_P P$, as was to be shown. •

Majorization and doubly stochastic matrices are closely related. Their relationship is described in the next theorem.

Theorem 2.4.11 (Doubly Stochastic Matrices and Majorization) *A nonnegative matrix A is doubly stochastic if and only if $Ax \prec x$ for any vector $x \in \mathbb{R}^N$.*

Proof. We use $e_n, n = 1,\dots,N$, to denote the standard basis of $\mathbb{R}^N$.

Let $Ax \prec x$ for all $x \in \mathbb{R}^N$. Choosing x to be $e_n, n = 1,\dots,N$ we can deduce that the sum of elements of each column of A is 1. Next let $x = \sum_{n=1}^N e_n$; we can conclude that the sum of elements of each row of A is 1. Thus, A is doubly stochastic.

Conversely, let A be doubly stochastic and let $y = Ax$. To prove $y \prec x$ we may assume, without loss of generality, that the coordinates of both x and y are in nonincreasing order. Now note that for any k, $1 \leq k \leq N$, we have

$$\sum_{m=1}^k y_m = \sum_{m=1}^k \sum_{n=1}^N a_{mn} x_n.$$

If we put $t_n = \sum_{m=1}^k a_{mn}$, then $t_n \in [0,1]$ and $\sum_{n=1}^N t_n = k$. We have

$$
\begin{aligned}
\sum_{m=1}^{k} y_m - \sum_{m=1}^{k} x_m &= \sum_{n=1}^{N} t_n x_n - \sum_{m=1}^{k} x_m \\
&= \sum_{n=1}^{N} t_n x_n - \sum_{m=1}^{k} x_m + (k - \sum_{n=1}^{N} t_n) x_k \\
&= \sum_{n=1}^{k} (t_n - 1)(x_n - x_k) + \sum_{n=k+1}^{N} t_n (x_n - x_k) \\
&\leq 0.
\end{aligned}
$$

Further, when $k = N$ we must have equality here simply because A is doubly stochastic. Thus, $y \prec x$. ●

Combining Theorems 2.4.6, 2.4.11 and 2.4.10 we have

Corollary 2.4.12 *Let $y \in \mathbb{R}^N$. Then $l(y) = \{Ay \mid A \in \mathcal{A}(N)\}$.*

2.4.5 Commentary and Exercises

Theorem 2.4.1 is a finite dimensional form of the Borwein–Preiss variational principle [46]. The approximate Fermat principle of Lemma 2.4.2 was suggested by [132]. The variational proof of Gordan's alternative is taken from [62] which can also be used in other related problems (Exercises 2.4.4 and 2.4.5).

Geometrically, Gordan's alternative [125] is clearly a consequence of the separation theorem: it says either 0 is contained in the convex hull of $a_0, \ldots, a_M$ or it can be strictly separated from this convex hull. Thus, the proof of Theorem 2.4.3 shows that with an appropriate auxiliary function variational method can be used in the place of a separation theorem – a fundamental result in analysis.

Majorization and doubly stochastic matrices are import concepts in matrix theory with many applications in physics and economics. Ando [3], Bhatia [20] and Horn and Johnson [133, 134] are excellent sources for the background and preliminaries for these concepts and related topics. Birkhoff's theorem appeared in [21]. Lemma 2.4.8 is a matrix form of Hall's matching condition [129]. Lemma 2.4.7 was established in König [158]. The variational proofs for the representation of the level sets with respect to the majorization and Birkhoff's theorem given here follow [267].

Exercise 2.4.1 Supply the details for the proof of Theorem 2.4.1.

Exercise 2.4.2 Prove the implications (ii) $\Rightarrow$ (iii) $\Rightarrow$ (i) in the proof of the Gordan Alternative of Theorem 2.4.3.

Exercise 2.4.3 Prove Lemma 2.4.2.

$*$**Exercise 2.4.4** (Ville's Theorem) Let $a_1, \ldots, a_M \in \mathbb{R}^N$ and define $f\colon \mathbb{R}^N \to \mathbb{R}$ by

$$f(x) := \ln\Big(\sum_{m=1}^{M} \exp \langle a_m, x\rangle\Big).$$

Consider the optimization problem

$$\inf\{f(x) \mid x \geq 0\} \tag{2.4.11}$$

and its relationship with the two systems

$$\sum_{m=1}^{M} \lambda_m a_m = 0, \sum_{m=1}^{M} \lambda_m = 1, 0 \leq \lambda_m,\ m = 1, \ldots, M, \tag{2.4.12}$$

$$\langle a_m, x\rangle < 0 \text{ for } m = 1, \ldots, M,\ x \in \mathbb{R}^N_+. \tag{2.4.13}$$

Imitate the proof of Gordan's alternatives to prove the following are equivalent:

(i) Problem (2.4.11) is bounded below.
(ii) System (2.4.12) is solvable.
(iii) System (2.4.13) is unsolvable.

Generalize by considering the problem $\inf\{f(x) \mid x_m \geq 0, m \in K\}$, where K is a subset of $\{1, \ldots, M\}$.

$*$**Exercise 2.4.5** (Stiemke's Theorem) Let $a_1, \ldots, a_M \in \mathbb{R}^N$ and define $f\colon \mathbb{R}^N \to \mathbb{R}$ by

$$f(x) := \ln\Big(\sum_{m=1}^{M} \exp \langle a_m, x\rangle\Big).$$

Consider the optimization problem

$$\inf\{f(x) \mid x \in \mathbb{R}^N\} \tag{2.4.14}$$

and its relationship with the two systems

$$\sum_{m=1}^{M} \lambda_m a_m = 0, 0 < \lambda_m,\ m = 1, \ldots, M, \tag{2.4.15}$$

and

$$\langle a_m, x\rangle \leq 0 \text{ for } m = 1, \ldots, M, \text{ not all } 0,\ x \in \mathbb{R}^N. \tag{2.4.16}$$

Prove the following are equivalent:

(i) Problem (2.4.14) has an optimal solution.

(ii) System (2.4.15) is solvable.
(iii) System (2.4.16) is unsolvable.

Hint: To prove (iii) implies (i), show that if problem (2.4.14) has no optimal solution then neither does the problem

$$\inf\Big\{\sum_{m=1}^{M}\exp y_m \mid y\in K\Big\}, \tag{2.4.17}$$

where K is the subspace $\{(\langle a^1,x\rangle,\ldots,\langle a^M,x\rangle)\mid x\in\mathbb{R}^N\}\subset\mathbb{R}^M$. Hence, by considering a minimizing sequence for (2.4.17), deduce system (2.4.16) is solvable.

$*$**Exercise 2.4.6** Prove the following

Lemma 2.4.13 (Farkas Lemma) *Let $a_1,\ldots,a_M$ and let $b\neq 0$ in $\mathbb{R}^N$. Then exactly one of the following systems has a solution:*

$$\sum_{m=1}^{M}\lambda_m a_m = b,\ 0\le\lambda_m,\ m=1,\ldots,M, \tag{2.4.18}$$

$$\langle a_m,x\rangle\le 0 \text{ for } m=1,\ldots,M,\ \langle b,x\rangle>0,\ x\in\mathbb{R}^N \tag{2.4.19}$$

Hint: Use the Gordan alternatives and induction.

Exercise 2.4.7 Verify Example 2.4.4.

Exercise 2.4.8 Let $y\in\mathbb{R}^N$. Verify that $l(y)$ is a convex set and, for any $P\in P(N)$, $Py\in l(y)$.

Exercise 2.4.9 Give an alternative proof of Birkhoff's theorem by going through the following steps.

(i) Prove $P(N)=\{(a_{mn})\in\mathcal{A}(N)\mid a_{mn}=0 \text{ or } 1 \text{ for all } m,n\}$.
(ii) Prove $P(N)\subset\mathrm{ext}(\mathcal{A}(N))$,where $\mathrm{ext}(S)$ signifies *extreme points* of set S.
(iii) Suppose $(a_{mn})\in\mathcal{A}(N)\backslash P(N)$. Prove there exist sequences of distinct indices $m_1,m_2,\ldots,m_k$ and $n_1,n_2,\ldots,n_k$ such that

$$0<a_{m_rn_r},a_{m_{r+1}n_r}<1(r=1,\ldots,k)$$

(where $m_{k+1}=m_1$). For these sequences, show the matrix (a'_{mn}) defined by

$$a'_{mn}-a_{mn}=\begin{cases}\varepsilon & \text{if } (m,n)=(m_r,n_r) \text{ for some } r,\\ -\varepsilon & \text{if } (m,n)=(m_{r+1},n_r) \text{ for some } r,\\ 0 & \text{otherwise,}\end{cases}$$

is doubly stochastic for all small real ε. Deduce $(a_{mn})\notin\mathrm{ext}(\mathcal{A}(N))$.
(iv) Deduce $\mathrm{ext}(\mathcal{A}(N))=P(N)$. Hence prove Birkhoff's theorem.
(v) Use Carathéodory's theorem [73] to bound the number of permutation matrices needed to represent a doubly stochastic matrix in Birkhoff's theorem.

2.5 Borwein–Preiss Variational Principles

Now we turn to a general form of the Borwein–Preiss smooth variational principle and a variation thereof derived by Deville, Godefroy and Zizler with a category proof.

2.5.1 The Borwein–Preiss Principle

Definition 2.5.1 *Let (X, d) be a metric space. We say that a continuous function $\rho\colon X \times X \to [0, \infty]$ is a* gauge-type function *on a complete metric space (X, d) provided that*

(i) $\rho(x, x) = 0$, *for all $x \in X$,*
(ii) *for any $\varepsilon > 0$ there exists $\delta > 0$ such that for all $y, z \in X$ we have $\rho(y, z) \le \delta$ implies that $d(y, z) < \varepsilon$.*

Theorem 2.5.2 (Borwein–Preiss Variational Principle) *Let (X, d) be a complete metric space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a lsc function bounded from below. Suppose that ρ is a gauge-type function and $(\delta_i)_{i=0}^{\infty}$ is a sequence of positive numbers, and suppose that $\varepsilon > 0$ and $z \in X$ satisfy*

$$f(z) \le \inf_X f + \varepsilon.$$

Then there exist y and a sequence $\{x_i\} \subset X$ such that

(i) $\rho(z, y) \le \varepsilon/\delta_0$, $\rho(x_i, y) \le \varepsilon/(2^i \delta_0)$,
(ii) $f(y) + \sum_{i=0}^{\infty} \delta_i \rho(y, x_i) \le f(z)$, *and*
(iii) $f(x) + \sum_{i=0}^{\infty} \delta_i \rho(x, x_i) > f(y) + \sum_{i=0}^{\infty} \delta_i \rho(y, x_i)$, *for all $x \in X \backslash \{y\}$.*

Proof. Define sequences (x_i) and (S_i) inductively starting with $x_0 := z$ and

$$S_0 := \{x \in X \mid f(x) + \delta_0 \rho(x, x_0) \le f(x_0)\}. \tag{2.5.1}$$

Since $x_0 \in S_0$, S_0 is nonempty. Moreover it is closed because both f and $\rho(\cdot, x_0)$ are lsc functions. We also have that, for all $x \in S_0$,

$$\delta_0 \rho(x, x_0) \le f(x_0) - f(x) \le f(z) - \inf_X f \le \varepsilon. \tag{2.5.2}$$

Take $x_1 \in S_0$ such that

$$f(x_1) + \delta_0 \rho(x_1, x_0) \le \inf_{x \in S_0} [f(x) + \delta_0 \rho(x, x_0)] + \frac{\delta_1 \varepsilon}{2\delta_0}. \tag{2.5.3}$$

and define similarly

$$S_1 := \left\{ x \in S_0 \;\middle|\; f(x) + \sum_{k=0}^{1} \delta_k \rho(x, x_k) \le f(x_1) + \delta_0 \rho(x_1, x_0) \right\}. \tag{2.5.4}$$

In general, suppose that we have defined x_j, S_j for $j = 0, 1, \ldots, i-1$ satisfying

$$f(x_j) + \sum_{k=0}^{j-1} \delta_k \rho(x_j, x_k) \le \inf_{x \in S_{j-1}} \Big[f(x) + \sum_{k=0}^{j-1} \delta_k \rho(x, x_k) \Big] + \frac{\varepsilon \delta_j}{2^j \delta_0} \quad (2.5.5)$$

and

$$S_j := \Big\{ x \in S_{j-1} \;\Big|\; f(x) + \sum_{k=0}^{j} \delta_k \rho(x, x_k) \le f(x_j) + \sum_{k=0}^{j-1} \delta_k \rho(x_j, x_k) \Big\}. \quad (2.5.6)$$

We choose $x_i \in S_{i-1}$ such that

$$f(x_i) + \sum_{k=0}^{i-1} \delta_k \rho(x_i, x_k) \le \inf_{x \in S_{i-1}} \Big[f(x) + \sum_{k=0}^{i-1} \delta_k \rho(x, x_k) \Big] + \frac{\varepsilon \delta_i}{2^i \delta_0} \quad (2.5.7)$$

and we define

$$S_i := \Big\{ x \in S_{i-1} \;\Big|\; f(x) + \sum_{k=0}^{i} \delta_k \rho(x, x_k) \le f(x_i) + \sum_{k=0}^{i-1} \delta_k \rho(x_i, x_k) \Big\}. \quad (2.5.8)$$

We can see that for every $i = 1, 2, \ldots,$ S_i is a closed and nonempty set. It follows from (2.5.7) and (2.5.8) that, for all $x \in S_i$,

$$\begin{aligned} \delta_i \rho(x, x_i) &\le \Big[f(x_i) + \sum_{k=0}^{i-1} \delta_k \rho(x_i, x_k) \Big] - \Big[f(x) + \sum_{k=0}^{i-1} \delta_k \rho(x, x_k) \Big] \\ &\le \Big[f(x_i) + \sum_{k=0}^{i-1} \delta_k \rho(x_i, x_k) \Big] - \inf_{x \in S_{i-1}} \Big[f(x) + \sum_{k=0}^{i-1} \delta_k \rho(x, x_k) \Big] \\ &\le \frac{\varepsilon \delta_i}{2^i \delta_0}, \end{aligned}$$

which implies that

$$\rho(x, x_i) \le \frac{\varepsilon}{2^i \delta_0}, \quad \text{for all } x \in S_i. \quad (2.5.9)$$

Since ρ is a gauge-type function, inequality (2.5.9) implies that $d(x, x_i) \to 0$ uniformly, and therefore $\operatorname{diam}(S_i) \to 0$. Since X is complete, by Cantor's intersection theorem there exists a unique $y \in \bigcap_{i=0}^{\infty} S_i$, which satisfies (i) by (2.5.2) and (2.5.9). Obviously, we have $x_i \to y$. For any $x \ne y$, we have that $x \notin \bigcap_{i=0}^{\infty} S_i$, and therefore for some j,

$$\begin{aligned} f(x) + \sum_{k=0}^{\infty} \delta_k \rho(x, x_k) &\ge f(x) + \sum_{k=0}^{j} \delta_k \rho(x, x_k) \\ &> f(x_j) + \sum_{k=0}^{j-1} \delta_k \rho(x_j, x_k). \end{aligned} \quad (2.5.10)$$

On the other hand, it follows from (2.5.1), (2.5.8) and $y \in \bigcap_{i=0}^{\infty} S_i$ that, for any $q \geq j$,

$$\begin{aligned} f(x_0) &\geq f(x_j) + \sum_{k=0}^{j-1} \delta_k \rho(x_j, x_k) \\ &\geq f(x_q) + \sum_{k=0}^{q-1} \delta_k \rho(x_q, x_k) \\ &\geq f(y) + \sum_{k=0}^{q} \delta_k \rho(y, x_k). \end{aligned} \tag{2.5.11}$$

Taking limits in (2.5.11) as $q \to \infty$ we have

$$\begin{aligned} f(z) = f(x_0) &\geq f(x_j) + \sum_{k=0}^{j-1} \delta_k \rho(x_j, x_k) \\ &\geq f(y) + \sum_{k=0}^{\infty} \delta_k \rho(y, x_k), \end{aligned} \tag{2.5.12}$$

which verifies (ii). Combining (2.5.10) and (2.5.12) yields (iii). ●

We shall frequently use the following normed space form of the Borwein–Preiss variational principle, especially in spaces with a Fréchet smooth renorm, in which case we may deduce first-order (sub)differential information from the conclusion.

Theorem 2.5.3 *Let X be a Banach space with norm $\|\cdot\|$ and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a lsc function bounded from below, let $\lambda > 0$ and let $p \geq 1$. Suppose that $\varepsilon > 0$ and $z \in X$ satisfy*

$$f(z) < \inf_X f + \varepsilon.$$

Then there exist y and a sequence (x_i) in X with $x_1 = z$ and a function $\varphi_p\colon X \to \mathbb{R}$ of the form

$$\varphi_p(x) := \sum_{i=1}^{\infty} \mu_i \|x - x_i\|^p,$$

where $\mu_i > 0$ for all $i = 1, 2, \ldots$ and $\sum_{i=1}^{\infty} \mu_i = 1$ such that

(i) $\|x_i - y\| \leq \lambda, n = 1, 2, \ldots,$
(ii) $f(y) + (\varepsilon/\lambda^p)\varphi_p(y) \leq f(z)$, *and*
(iii) $f(x) + (\varepsilon/\lambda^p)\varphi_p(x) > f(y) + (\varepsilon/\lambda^p)\varphi_p(y)$, *for all $x \in X \setminus \{y\}$.*

Proof. Exercise 2.5.1. ●

Note that when $\|\cdot\|$ is Fréchet smooth so is φ_p for $p > 1$.

2.5.2 The Deville–Godefroy–Zizler Principle

An important counterpart of the Borwein–Preiss variational principle subsequently found by Deville, Godefroy and Zizler [94] is given below. It is interesting to see how the Baire category theorem is used in the proof. Recall that the Baire category theorem states that in a complete metric space every countable intersection of dense open sets is dense: a set containing such a dense G_δ set is called *generic* or *residual* and the complement of such a set is *meager*. We say a function $f\colon X \to \mathbb{R} \cup \{+\infty\}$ attains a *strong minimum* at $x \in X$ if $f(x) = \inf_X f$ and $\|x_i - x\| \to 0$ whenever $x_i \in X$ and $f(x_i) \to f(x)$. If f is bounded on X, we define $\|f\|_\infty := \sup\{|f(x)| \mid x \in X\}$. We say that $\phi\colon X \to \mathbb{R}$ is a *bump function* if ϕ is bounded and has bounded nonempty *support* $\operatorname{supp}(\phi) := \{x \in X \mid \phi(x) \neq 0\}$.

Theorem 2.5.4 (The Deville–Godefroy–Zizler Variational Principle) *Let X be a Banach space and Y a Banach space of continuous bounded functions g on X such that*

(i) $\|g\|_\infty \leq \|g\|_Y$ *for all* $g \in Y$.
(ii) *For each $g \in Y$ and $z \in X$, the function $x \to g_z(x) = g(x+z)$ is in Y and $\|g_z\|_Y = \|g\|_Y$.*
(iii) *For each $g \in Y$ and $a \in \mathbb{R}$, the function $x \to g(ax)$ is in Y.*
(iv) *There exists a bump function in Y.*

If $f\colon X \to \mathbb{R} \cup \{+\infty\}$ is a proper lsc function and bounded below, then the set G of all $g \in Y$ such that $f + g$ attains a strong minimum on X is residual (in fact a dense G_δ set).

Proof. Given $g \in Y$, define $S(g; a) := \{x \in X \mid g(x) \leq \inf_X g + a\}$ and $U_i := \{g \in Y \mid \operatorname{diam} S(f + g; a) < 1/i, \text{ for some } a > 0\}$. We show that each of the sets U_i is dense and open in Y and that their intersection is the desired set G.

To see that U_i is open, suppose that $g \in U_i$ with a corresponding $a > 0$. Then, for any $h \in Y$ such that $\|g - h\|_Y < a/3$, we have $\|g - h\|_\infty < a/3$. Now, for any $x \in S(f + h; a/3)$,

$$(f + h)(x) \leq \inf_X (f + h) + \frac{a}{3}.$$

It is an easy matter to estimate

$$\begin{aligned}(f + g)(x) &\leq (f + h)(x) + \|g - h\|_\infty \leq \inf_X (f + h) + \frac{a}{3} + \|g - h\|_\infty \\ &\leq \inf_X (f + g) + \frac{a}{3} + 2\|g - h\|_\infty \leq \inf_X (f + g) + a.\end{aligned}$$

This shows that $S(f + h; a/3) \subset S(f + g; a)$. Thus, $h \in U_i$.

To see that each U_i is dense in Y, suppose that $g \in Y$ and $\varepsilon > 0$; it suffices to produce $h \in Y$ such that $\|h\|_Y < \varepsilon$ and for some $a > 0$ $\operatorname{diam} S(f + g +$

$h; a) < 1/i$. By hypothesis (iv), Y contains a bump function ϕ. Without loss of generality we may assume that $\|\phi\|_Y < \varepsilon$. By hypothesis (ii) we can assume that $\phi(0) \neq 0$, and therefore that $\phi(0) > 0$. Moreover, by hypothesis (iii) we can assume that $\mathrm{supp}(\phi) \subset B(0, 1/2i)$. Let $a = \phi(0)/2$ and choose $\bar{x} \in X$ such that

$$(f+g)(\bar{x}) < \inf_X (f+g) + \phi(0)/2.$$

Define h by $h(x) := -\phi(x-\bar{x})$; by hypothesis (ii), $h \in Y$ and $\|h\|_Y = \|\phi\|_Y < \varepsilon$ and $h(\bar{x}) = -\phi(0)$. To show that $\mathrm{diam}\, S(f+g+h; a) < 1/i$, it suffices to show that this set is contained in the ball $B(\bar{x}, 1/2i)$; that is, if $\|x - \bar{x}\| > 1/2i$, then $x \notin S(f+g+h; a)$, the latter being equivalent to

$$(f+g+h)(x) > \inf_X (f+g+h) + a.$$

Now, $\mathrm{supp}(h) \subset B(\bar{x}, 1/2i)$, so $h(x) = 0$ if $\|x - \bar{x}\| > 1/2i$ hence

$$\begin{aligned}(f+g+h)(x) &= (f+g)(x) \geq \inf_X (f+g) > (f+g)(\bar{x}) - a \\ &= (f+g+h)(\bar{x}) + \phi(0) - \phi(0)/2 \geq \inf_X (f+g+h) + a.\end{aligned}$$

as was to be shown.

Finally we show $\bigcap_{i=1}^{\infty} U_i = G$. The easy part of $G \subset \bigcap_{i=1}^{\infty} U_i$ is left as an exercise. Let $g \in \bigcap_{i=1}^{\infty} U_i$. We will show that $g \in G$; that is, $f + g$ attains a strong minimum on X. First, for all i there exists $a_i > 0$ such that $\mathrm{diam}\, S(f+g; a_i) < 1/i$ and hence there exists a unique point $\bar{x} \in \bigcap_{i=1}^{\infty} S(f+g; a_i)$. Suppose that $x_k \in X$ and that $(f+g)(x_k) \to \inf_X (f+g)$. Given $i > 0$ there exists i_0 such that $(f+g)(x_k) \leq \inf_X (f+g) + a_i$ for all $i \geq i_0$, therefore $x_k \in S(f+g; a_i)$ for all $i \geq i_0$ and hence $\|x_k - \bar{x}\| \leq \mathrm{diam}\, S(f+g; a_i) < 1/i$ if $k \geq i_0$. Thus, $x_k \to \bar{x}$, and therefore $g \in G$. ●

2.5.3 Commentary and Exercises

The Borwein–Preiss smooth variational principle appeared in [46]. The proof here is adapted from Li and Shi [176]. Their original proof leads to a clean generalization of both the Ekeland and Borwein–Preiss variational principle (see Exercises 2.5.2 and 2.5.3). The Deville–Godefroy–Zizler variational principle and its category proof is from [94]. Another very useful variational principle due to Stegall, is given in Section 6.3.

Exercise 2.5.1 Deduce Theorem 2.5.3 from Theorem 2.5.2. Hint: Set $\rho(x,y) = \|x-y\|^p$.

Exercise 2.5.2 Check that, with $\delta_0 := 1$, $\delta_i := 0, i = 1, 2, \dots$ and $\rho := \varepsilon d$, the procedure in the proof of Theorem 2.5.2 reduces to a proof of the Ekeland variational principle.

If one works harder, the two variational principles can be unified.

$*$**Exercise 2.5.3** Adapt the proof of Theorem 2.5.2 for a nonnegative sequence $(\delta_i)_{i=0}^{\infty}$, $\delta_0 > 0$ to derive the following generalization for both the Ekeland and the Borwein–Preiss variational principles.

Theorem 2.5.5 *Let (X, d) be a complete metric space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a lsc function bounded from below. Suppose that ρ is a gauge-type function and $(\delta_i)_{i=0}^{\infty}$ is a sequence of nonnegative numbers with $\delta_0 > 0$. Then, for every $\varepsilon > 0$ and $z \in X$ satisfying*

$$f(z) \le \inf_X f + \varepsilon,$$

there exists a sequence $\{x_i\} \subset X$ converging to some $y \in X$ such that

(i) $\rho(z, y) \le \varepsilon/\delta_0$,
(ii) $f(y) + \sum_{i=0}^{\infty} \delta_i \rho(y, x_i) \le f(z)$, *and*
(iii) $f(x) + \sum_{i=0}^{\infty} \delta_i \rho(x, x_i) > f(y) + \sum_{i=0}^{\infty} \delta_i \rho(y, x_i)$, *for all* $x \in X \setminus \{y\}$.

Moreover, if $\delta_k > 0$ and $\delta_l = 0$ for all $l > k \ge 0$, then (iii) *may be replaced by*

(iii$'$) *for all $x \in X \setminus \{y\}$, there exists $j \ge k$ such that*

$$f(x) + \sum_{i=0}^{k-1} \delta_i \rho(x, x_i) + \delta_k \rho(x, x_j) > f(y) + \sum_{i=0}^{k-1} \delta_i \rho(y, x_i) + \delta_k \rho(y, x_j).$$

The Ekeland variational principle, the Borwein–Preiss variational principle and the Deville–Godefroy–Zizler variational principle are related in the following exercises.

Exercise 2.5.4 Deduce the following version of Ekeland's variational principle from Theorem 2.5.4.

Theorem 2.5.6 *Let X be a Banach space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a proper lsc function and bounded below. Then for all $\varepsilon > 0$ there exists $\bar{x} \in X$ such that*

$$f(\bar{x}) \le \inf_X f + 2\varepsilon$$

and the perturbed function $x \to f(x) + \varepsilon\|x - \bar{x}\|$ attains a strong minimum at $\bar{x}$.

Hint: Let Y be the space of all bounded Lipschitz continuous functions g on X with norm

$$\|g\|_Y := \|g\|_\infty + \sup\Big\{\frac{|g(x) - g(y)|}{\|x - y\|} \;\Big|\; x, y \in X, x \ne y\Big\}.$$

Exercise 2.5.5 Deduce the following version of the smooth variational principle from Theorem 2.5.4.

Theorem 2.5.7 *Let X be a Banach space with a Lipschitz Fréchet smooth bump function and let $f: X \to \mathbb{R} \cup \{+\infty\}$ be a proper lsc function and bounded below. Then there exists a constant $a > 0$ (depending only on X) such that for all $\varepsilon \in (0, 1)$ and for any $y \in X$ satisfying $f(y) < \inf_X f + a\varepsilon^2$, there exist a Lipschitz Fréchet differentiable function g and $x \in X$ such that*

(i) *$f + g$ has a strong minimum at x,*
(ii) *$\|g\|_\infty < \varepsilon$ and $\|g'\|_\infty < \varepsilon$,*
(iii) *$\|x - y\| < \varepsilon$.*

$*$**Exercise 2.5.6** (Range of Bump Functions)Let $b: \mathbb{R}^N \to \mathbb{R}$ be a C^1 bump function.

(i) Show that $0 \in \operatorname{int} \operatorname{range}(b')$ by applying the smooth variational principle.
(ii) Find an example where $\operatorname{range}(b')$ is not simply connected.

Reference: [31].

3

Variational Techniques in Subdifferential Theory

For problems of smooth variation we can usually apply arguments based on Fermat's principle – that a differentiable function has a vanishing derivative at its minima (maxima). However, nonsmooth functions and mappings arise intrinsically in many applications. The following are several such examples of intrinsic nonsmoothness.

Example 3.0.1 (Max Function) Let $f_n\colon X \to \mathbb{R} \cup \{+\infty\}, n = 1, \ldots, N$ be lsc functions. Then so is

$$f = \max(f_1, \ldots, f_N).$$

However, this maximum is often nonsmooth even if all $f_n, n = 1, \ldots, N$ are smooth functions. For example,

$$|x| = \max(x, -x).$$

is nonsmooth at $x = 0$.

Example 3.0.2 (Optimal Value Functions) Consider the simple constrained minimization problem of minimizing $f(x)$ subject to $g(x) = a$, $x \in \mathbb{R}$. Here $a \in \mathbb{R}$ is a parameter allowing for perturbation of the constraint. In practice it is often important to know how the model responds to the perturbation a. For this we need to consider, for example, the *optimal value*

$$v(a) := \inf\{f(x) : g(x) = a\}$$

as a function of a. Consider a concrete example, illustrated in Figure 3.1, of the two smooth functions $f(x) := 1 - \cos x$ and $g(x) := \sin(6x) - 3x$, and $a \in [-\pi/2, \pi/2]$ which corresponds to $x \in [-\pi/6, \pi/6]$. It is easy to show that the optimal value function v is not smooth, in fact, not even continuous.

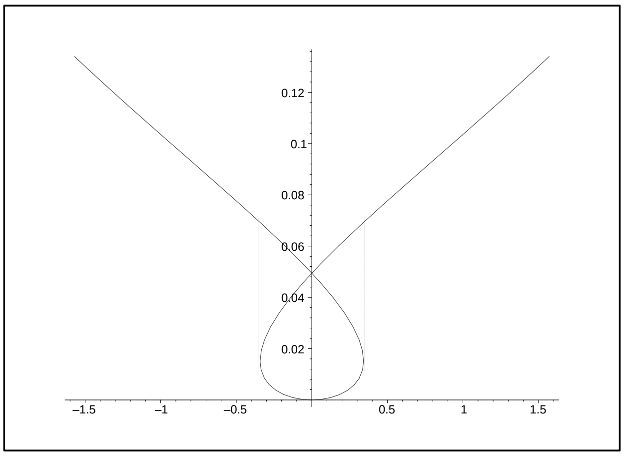

Fig. 3.1. Smooth becomes nonsmooth: g (vertical) plotted against f.

Example 3.0.3 (Penalization Functions) Constrained optimization problems occur naturally in many applications. A simplified form of such a problem is

$$\mathcal{P} \qquad \begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && x \in S, \end{aligned}$$

where S is a closed subset of X often referred to as the feasible set. One often wishes to convert such a problem to a simpler one without constraint. The use of nonsmooth functions makes this conversion easier. For example, if f is Lipschitz with a Lipschitz constant L then, for any $\mu > L$, problem $\mathcal{P}$ is equivalent to

$$\text{minimize } f + \mu d_S.$$

This is often referred to as *exact penalization.* If f is lsc then $\mathcal{P}$ is equivalent to

$$\text{minimize } f + \iota_S.$$

Example 3.0.4 (Spectral Functions) The maximum eigenvalue of a matrix often plays an important role in problems related to a matrix. When the matrix contains one or more parameters, the maximum eigenvalue then becomes a function of those parameters. This maximum eigenvalue function is often intrinsically nonsmooth. For example, consider the 2 by 2 matrix with a parameter x,

$$\begin{bmatrix} 1 & x \\ x & 1 \end{bmatrix}.$$

Then the maximum eigenvalue is $1 + |x|$, a nonsmooth function.

This intrinsic nonsmoothness motivated the development of nonsmooth analysis. Concepts generalizing that of the derivative for smooth functions have been introduced which enable us to apply the variational technique to nonsmooth functions. There are many competing concepts of subdifferentials; we mainly focus on the Fréchet subdifferential which is a natural fit for the variational technique.

3.1 The Fréchet Subdifferential and Normal Cones

3.1.1 The Fréchet Subdifferential

To generate the Fréchet subdifferential at a nondifferentiable point of a lsc function, we use the collection of all the (Fréchet) derivatives of smooth "osculating" functions (functions lying below and touching at the point in question), if they exist, to replace the missing derivative. More often than not, this simple contrivance is sufficient. Moreover, in the language of analysis, we are led to study a local minimum of the difference of two functions which fits very well with techniques of variational analysis. The geometric concept of the Fréchet normal cone to a closed set is then introduced through the subdifferential of the indicator function of the set – an extended-valued lsc function.

Let X be a Banach space. We say a function f on X is *Fréchet differentiable* at x and $f'(x) \in X^*$ is the *Fréchet derivative* of f at x provided that

$$\lim_{\|h\|\to 0} \frac{|f(x+h) - f(x) - \langle f'(x), h\rangle|}{\|h\|} = 0.$$

We say f is C^1 at x if $f'\colon X \to X^*$ is norm continuous at x. We say a Banach space is *Fréchet smooth* provided that it has an equivalent norm that is differentiable, indeed C^1, for all $x \neq 0$.

Definition 3.1.1 (Fréchet Subdifferential) *Let X be a real Banach space. Let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a proper lsc function. We say f is* Fréchet-subdifferentiable *and x^* is a* Fréchet-subderivative *of f at x if $x \in \operatorname{dom} f$ and*

$$\liminf_{\|h\|\to 0} \frac{f(x+h) - f(x) - \langle x^*, h\rangle}{\|h\|} \geq 0. \tag{3.1.1}$$

We denote the set of all Fréchet-subderivatives of f at x by $\partial_F f(x)$ and call this object the Fréchet subdifferential *of f at x. For convenience we define $\partial_F f(x) = \emptyset$ if $x \notin \operatorname{dom} f$.*

Definition 3.1.2 (Viscosity Fréchet Subdifferential) *Let X be a real Banach space. Let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a proper lsc function. We say f is* viscosity Fréchet-subdifferentiable *and x^* is a* viscosity Fréchet-subderivative *of f at x if $x \in \operatorname{dom} f$ and there exists a C^1 function g such that $g'(x) = x^*$ and*

$f-g$ attains a local minimum at x. We denote the set of all viscosity Fréchet-subderivatives of f at x by $\partial_{VF} f(x)$ and call this object the viscosity Fréchet subdifferential *of f at x. For convenience we define $\partial_{VF} f(x) = \emptyset$ if $x \notin \mathrm{dom} f$.*

Since shifting g by a constant does not influence its derivative we can require that $f-g$ attains a local minimum of 0 at x in the above definition.

The following relationship between the Fréchet subdifferential and the viscosity Fréchet subdifferential is easy and useful.

Proposition 3.1.3 *Let X be a Banach space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a lsc function. Then $\partial_{VF} f(x) \subset \partial_F f(x)$.*

Proof. Exercise 3.1.1. ●

In fact, with some additional effort one can show that in a Fréchet-smooth Banach space $\partial_{VF} f(x) = \partial_F f(x)$ [95]. Since we work mostly in Fréchet smooth Banach spaces in this book, we will use ∂_F for both Fréchet and viscosity Fréchet subdifferentials unless pointed out otherwise.

If f is Fréchet differentiable at x then it is not hard to show that $\partial_F f(x) = \{f'(x)\}$. The converse is not true (Exercises 3.1.3). In general, $\partial_F f(x)$ may be empty even if $x \in \mathrm{dom} f$. An easy example is $\partial_F(-\|\cdot\|)(0) = \emptyset$. However, a variational argument leads to the following important result about the existence of the Fréchet subdifferential.

Theorem 3.1.4 *Let X be a Fréchet smooth Banach space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a lsc function. Then $\{x \in X \mid \partial_F f(x) \neq \emptyset\}$ is dense in $\mathrm{dom} f$.*

Proof. Let $\bar{x} \in \mathrm{dom} f$ and let ε be an arbitrary positive number. We show f is Fréchet subdifferentiable at some point $y \in B_\varepsilon(\bar{x})$. Since f is lsc at $\bar{x}$ there exists $\delta > 0$ such that $f(x) > f(\bar{x}) - 1$ for all $x \in B_\delta(\bar{x})$. Define $\tilde{f} := f + \iota_{B_\delta(\bar{x})}$. Then, $\tilde{f}$ is lsc and

$$\tilde{f}(\bar{x}) = f(\bar{x}) < \inf_{B_\delta(\bar{x})} f + 1 = \inf_X \tilde{f} + 1.$$

Applying the Borwein–Preiss Variational Principle of Theorem 2.5.3, using the asserted Fréchet smooth renorm with $\lambda < \min(\delta, \varepsilon)$, we conclude that there exists $y \in B_\lambda(\bar{x}) \subset \mathrm{int}(B_\delta(\bar{x}) \cap B_\varepsilon(\bar{x}))$ and $\varphi_2(x) := \sum_{i=1}^\infty \mu_i \|x - x_i\|^2$ where (x_i) is a sequence converging to y and (μ_i) is a sequence of positive numbers satisfying $\sum_{i=1}^\infty \mu_i = 1$ such that $\tilde{f} + \lambda^{-2}\varphi_2$ attains a minimum at y. Since y is an interior point of $B_\delta(\bar{x})$, $f + \lambda^{-2}\varphi_2$ attains a local minimum at y. After checking that φ_2 is Fréchet differentiable, we see that f is Fréchet subdifferentiable at $y \in B_\varepsilon(\bar{x})$. ●

We put meat on the bones of the last result by recalling that Hilbert space and $L_p(1 < p < \infty)$ are Fréchet smooth in their original norms while every reflexive space has a Fréchet smooth renorm [46, 95].

Note that the subdifferential is usually a set. The following are subdifferentials of several nonsmooth functions at typical nonsmooth points that can easily be verified.

Example 3.1.5

$$\partial_F |\cdot|(0) = [-1, 1],$$
$$\partial_F \sqrt{|\cdot|}(0) = (-\infty, \infty),$$
$$\partial_F \max(\cdot, 0)(0) = [0, 1],$$

and

$$\partial_F \iota_{[0,1]}(0) = (-\infty, 0].$$

3.1.2 The Fréchet Normal Cone

The central geometric concept of the normal cone to a closed set can now be defined through the indicator function of the set.

Definition 3.1.6 (Fréchet Normal Cone) *Let S be a closed subset of X. We define the Fréchet normal cone of S at x to be $N_F(S;x) := \partial_F \iota_S(x)$.*

Some easy facts directly follow from the definition. It is easy to verify that $N_F(S;x)$ is a cone that always contains $\{0\}$ and when $x \in \operatorname{int} S$, $N_F(S;x) = \{0\}$ (Exercises 3.1.6, 3.1.8 and 3.1.9). Moreover, consider the constrained minimization problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in S \subset X. \end{aligned} \tag{3.1.2}$$

We have an easy and useful necessary optimality condition in terms of the normal cone of S.

Proposition 3.1.7 *Let X be a Fréchet smooth Banach space, let f be a C^1 function on X and let S be a closed subset of X. Suppose that $\bar{x}$ is a solution of the constained minimization problem* (3.1.2). *Then*

$$0 \in f'(\bar{x}) + N_F(S;\bar{x}).$$

Proof. Exercise 3.1.13. ●

Recall that for a C^1 function f, $v = f'(x)$ if and only if $(v, -1)$ is a normal vector for the graph of f at $(x, f(x))$. Our next theorem is a Fréchet subdifferential version of this fact which characterizes the Fréchet subdifferential of a function in terms of the normal cone to its epigraph.

Theorem 3.1.8 *Let X be a Fréchet smooth Banach space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a lsc function. Then $x^* \in \partial_F f(x)$ if and only if*

$$(x^*, -1) \in N_F(\operatorname{epi} f; (x, f(x))).$$

Proof. (a) The "only if" part. Let $x^* \in \partial_F f(x)$. Then there exists a C^1 function g such that $g'(x) = x^*$ and $f - g$ attains a minimum at x. Define $h(y, r) := g(y) - r$. We have $hx, f(x)) = (x^*, -1)$ and

$$\iota_{\mathrm{epi} f}(y, r) - h(y, r) \geq \iota_{\mathrm{epi} f}(x, f(x)) - h(x, f(x)). \tag{3.1.3}$$

Thus, $(x^*, -1) \in N_F(\mathrm{epi} f; (x, f(x)))$.

(b) The "if" part. Let $(x^*, -1) \in N_F(\mathrm{epi} f; (x, f(x)))$. Then there exists a C^1 function h such that $h'(x, f(x)) = (x^*, -1)$ and $h(y, r) \leq h(x, f(x)) = 0$ for any $(y, r) \in \mathrm{epi} f$. By the implicit function theorem (see e.g. [264]) there exists a C^1 function $g\colon X \to \mathbb{R}$ such that in a neighborhood of x, $h(y, g(y)) = 0$, $g(x) = f(x)$ and $g'(x) = x^*$. Since h is C^1 and the second component of $h'(x, f(x))$ is negative there exists $a > 0$ such that $h(y, r) < h(y, r')$, for any $y \in B_a(x)$ and $f(x) - a < r' < r < f(x) + a$. Take $b \in (0, a)$ such that for any $y \in B_b(x)$, $g(y) \in (f(x) - a, f(x) + a)$ and $f(y) > f(x) - a$. Then, for any $y \in B_b(x)$, we have $f(y) - g(y) \geq 0 = f(x) - g(x)$. In fact, the inequality is obvious when $f(y) \geq f(x) + a$. If $f(y) < f(x) + a$ then it follows from $h(y, f(y)) \leq 0 = h(y, g(y))$. ●

The normal cone to the epigraph of a function has the following special properties.

Lemma 3.1.9 *Let f be a lsc function. Then*

(i) *for any $(x, r) \in \mathrm{epi} f$, $N_F(\mathrm{epi} f; (x, r)) \subset N_F(\mathrm{epi} f; (x, f(x)))$,*
(ii) *if $(x^*, -\lambda) \in N_F(\mathrm{epi} f; (x, f(x)))$ and $\lambda \neq 0$ then $\lambda > 0$ and $x^* \in \lambda \partial_F f(x)$.*

Proof. Exercise 3.1.10. ●

Thus, Theorem 3.1.8 also characterizes $(x^*, \lambda) \in N_F(\mathrm{epi} f; (x, f(x)))$ when $\lambda \neq 0$ in terms of the subdifferentials of f. The characterization of $(x^*, 0) \in N_F(\mathrm{epi} f; (x, f(x)))$ in terms of the subdifferentials of f is more delicate and will be discussed later after we have developed the subdifferential calculus.

3.1.3 The Subdifferential Form of the Variational Principle

We conclude this section with a subdifferential version of the Borwein–Preiss Variational Principle. This is the form most frequently used in applications involving subdifferentials. The easy proof is left as an exercise.

Theorem 3.1.10 *Let X be a Banach space with a Fréchet smooth norm $\|\cdot\|$ and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a lsc function bounded from below, $\lambda > 0$ and $p > 1$. Then, for every $\varepsilon > 0$ and $z \in X$ satisfying*

$$f(z) < \inf_X f + \varepsilon,$$

there exists a point $y \in X$ such that $\|z-y\| \le \lambda$ and a C^1 function φ with $|\varphi(y)| < \varepsilon/\lambda$ and $\|\varphi'(y)\| < p\varepsilon/\lambda$ such that $f+\varphi$ attains a minimum at y. Consequently,

$$\partial_F f(y) \cap \frac{p\varepsilon}{\lambda} B_{X^*} \neq \emptyset.$$

Proof. Exercise 3.1.12. ●

3.1.4 Commentary and Exercises

Although the use of generalized (one-sided) derivatives dates back explicitly to Dini and before, especially in the context of integration theory, the systematic study of such concepts for variational analysis, especially off the real line, is quite recent. Consistent theory was developed first for certain classes of functions, e.g., the convex subdifferential for convex functions (see [229]) and the quasi-differential for quasi-differentiable functions (see [217]). Clarke's pioneering work [85] on the generalized gradient opened the door to methodical study of general nonsmooth problems. Many competing concepts of generalized derivatives were introduced in the ensuing past several decades. Several frequently used concepts are Halkin's screen [128], the limiting subdifferential developed by Mordukhovich [189, 192, 194], Ioffe's approximate and G-subdifferential [137, 140, 141], Michel and Penot's subdifferential [187], Treiman's linear subdifferential [243, 245], Warga's derivative container [259, 256] and Sussmann's semidifferential [239, 240].

The last decade has witnessed a unification and reconciliation of much of this work in two directions. One is along the ideas pioneered by Warga to study abstract subdifferentials that satisfy a set of axioms so as to provide basic properties of many different subdifferentials alluded to above with a unified framework. The other, which is more relevant to this book, is to turn our attention to the simpler smooth subdifferentials based on the fact that many of the above subdifferentials can be represented by such smooth subdifferentials in spaces with a reasonable geometric property [60, 88, 141, 179]. In this book we primarily consider the Fréchet subdifferential in Fréchet smooth Banach spaces. It was introduced by Bazaraa, Goode and Nashed in finite dimensions [18] and developed in detail in infinite dimensions by Borwein and Strojwas [49], Kruger [159, 160], Kruger and Mordukhovich [161] and others. This allows us to illustrate variational techniques without too many technical assumptions. Most of the results apply to more general bornological smooth subdifferentials or s-Hölder subdifferentials [46, 215] with minor changes. Systematic accounts of nonsmooth analysis and its applications can be found in [8, 87, 88, 83, 179, 145, 194, 202, 230, 259, 256].

Unlike derivatives, subdifferentials do not determine functions up to a constant, even on well connected sets. Thus, we do not have an "integration" theory corresponding to the subdifferentials (see guided Exercises 3.1.19, 3.1.20, 3.1.21 and 3.1.22 for details).

Exercise 3.1.1 Prove Proposition 3.1.3.

Exercise 3.1.2 Verify the Fréchet subdifferentials in Example 3.1.5.

Exercise 3.1.3 Show that

(i) If f is Fréchet differentiable at x then $\partial_F f(x) = \{f'(x)\}$.
(ii) A function can have a unique Fréchet subdifferential without being differentiable.
(iii) There exists a Lipschitz function having the properties described in (ii).

Hint: Consider $f(x) := |x|(\sin(\log(|x|)) + 1), x \neq 0$ and $f(0) := 0$.

Exercise 3.1.4 (Fréchet Superdifferential) Let $f\colon X \to \mathbb{R} \cup \{-\infty\}$ be an upper semicontinuous function (i.e., $-f$ is lsc). We define the Fréchet superdifferential of f at x to be $\partial^F f(x) = -\partial_F(-f)(x)$. Prove that f is Fréchet differentiable at x if and only if $\partial^F f(x) = \partial_F(f)(x) = \{f'(x)\}$. Indeed it suffices that $\partial^F f(x) \cap \partial_F f(x) \neq \emptyset$.

Exercise 3.1.5 Show that for any $\lambda > 0$, $\partial_F(\lambda f)(x) = \lambda \partial_F f(x)$. Care must be taken with zero, when $\partial_F f(x)$ is empty.

Exercise 3.1.6 Verify that for any closed set S and $x \in S$, $N_F(S; x)$ is a cone, i.e., for any $x^* \in N_F(S; x)$ and any $r \geq 0$, $rx^* \in N_F(S; x)$.

Exercise 3.1.7 Construct a set $S \subset \mathbb{R}^2$ such that $N_F(S; (0, 0))$ is neither open nor closed.

Exercise 3.1.8 Show that if $s \in \mathrm{int}S$, then $N_F(S; s) = \{0\}$.

Exercise 3.1.9 Let $\{e_i\}$ be the standard orthonormal basis of ℓ_2 and let $S := \overline{\mathrm{conv}}\{\pm e_i/i\}_{i=1}^{\infty}$. Show that $0 \notin \mathrm{int}S$ yet $N_F(S, 0) = \{0\}$.

Exercise 3.1.10 Prove Lemma 3.1.9.

Exercise 3.1.11 Show that in Definition 3.1.2 we can require that $f - g$ attains a local minimum of 0 at x.

Exercise 3.1.12 Suppose that f is a lsc function and that g is a C^1 function. Show that $\partial_F(f + g)(x) = \partial_F f(x) + g'(x)$.

Exercise 3.1.13 Prove Proposition 3.1.13.

Exercise 3.1.14 Prove that if f is a Lipschitz function with rank L then, for any x, $x^* \in \partial_F f(x)$ implies that $\|x^*\| \leq L$.

$*$**Exercise 3.1.15** Let X be a Fréchet smooth Banach space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a lsc function. Prove that f is Lipschitz with rank L if and only if, for any x, $x^* \in \partial_F f(x)$ implies that $\|x^*\| \leq L$.

$*$**Exercise 3.1.16** Let X be a Fréchet smooth Banach space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a lsc function. Prove that $\partial_{VF} f(x) = \partial_F f(x)$. Reference: [95].

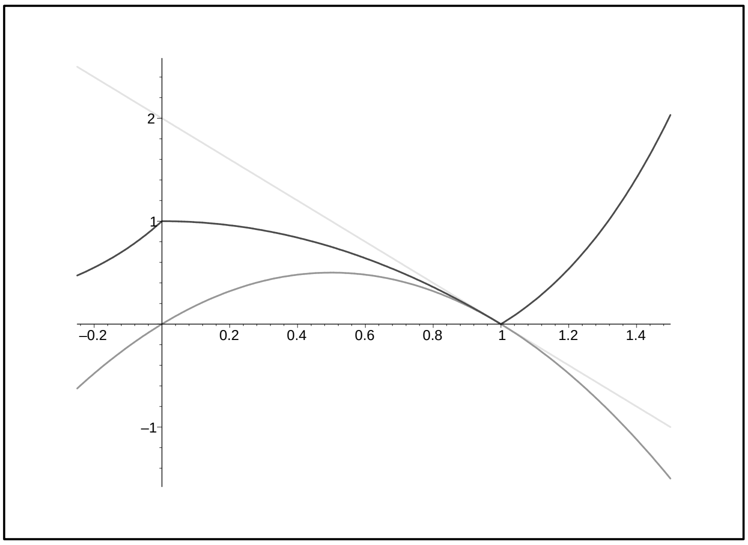

Fig. 3.2. Every Fréchet subdifferential is a “viscosity” subdifferential.

$*$**Exercise 3.1.17** Let X be a Banach space with a Fréchet smooth equivalent norm and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a lsc function. Prove that $x^* \in \partial_F f(x)$ if and only if there exists a *concave* C^1 function g such that $g'(x) = x^*$ and $f - g$ attains a local minimum at x, as drawn in Figure 3.2. Reference: [69, Remark 1.4].

Exercise 3.1.18 Prove Theorem 3.1.10.

Exercise 3.1.19 Construct two lsc functions on $\mathbb{R}$ with the identical Fréchet subdifferential yet their difference is not a constant. Hint: Consider $f = 1 - \chi_{[0,1]}$ and $2f$ where χ_S is the characteristic function of set S defined by $\chi_S(x) = 1$ for $x \in S$ and $\chi_S(x) = 0$ for $x \notin S$.

Exercise 3.1.20 Construct two continuous functions on $\mathbb{R}$ with the identical Fréchet subdifferential yet their difference is not a constant. Hint: Consider the Cantor function f and $2f$ (see [69] and also Exercise 3.5.5).

Exercise 3.1.21 Prove that if two Lipschitz functions on $\mathbb{R}$ have the identical Fréchet subdifferential then they differ only by a constant.

$*$**Exercise 3.1.22** The conclusion in Exercise 3.1.21 fails if the Fréchet subdifferential is replaced by the proximal subdifferential. Recall the proximal subdifferential is defined as follows.

Definition 3.1.11 (Proximal Subdifferential) *Let X be a real Hilbert space. Let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a proper lsc function. We say f is a* proximal subdifferentiable *and x^* is a* proximal subderivative *of f at x if $x \in \operatorname{dom} f$*

and there exists a constant $c \geq 0$ *such that* $f(y) - \langle x^*, y \rangle - c\|y - x\|^2$ *attains a local minimum at* x*. We denote the set of all proximal-subderivatives of* f *at* x *by* $\partial_P f(x)$ *and call this object the* proximal subdifferential *of* f *at* x*. For convenience we define* $\partial_P f(x) = \emptyset$ *if* $x \notin \mathrm{dom} f$*.*

Precisely prove the following theorem.

Theorem 3.1.12 *There exists uncountably many different Lipschitz function* $f\colon \mathbb{R} \to \mathbb{R}$ *with* $f(0) = 0$ *such that* $\partial_P f(x) = (-1, 1)$ *when* x *is a dyadic rational, and* $\partial_P f(x) = \emptyset$ *when* x *is not a dyadic rational.*

One can start with the construction in the following proposition for a function on $[0, 1]$ and then extend it periodically to $\mathbb{R}$.

Proposition 3.1.13 *Let* (a_i) *be a sequence satisfying* $0 < a_1 < a_2 < \cdots < 1$, $a_i \to 1$ *and* $2^i(1 - a_i) \to \infty$*. Then there exists a Lipschitz function* $f\colon [0, 1] \to \mathbb{R}$ *with Lipschitz constant* 1 *satisfying* $f(0) = f(1) = 0$ *and* $f(1/2) = a_1/2$ *such that* $\partial_P f(x) = (-1, 1)$ *when* $x \in (0, 1)$ *is a dyadic rational, and* $\partial_P f(x) = \emptyset$ *when* $x \in (0, 1)$ *is not a dyadic rational.*

Hint: Define $f = \lim_i f_i$ where f_i are affine on the intervals $[n/2^i, (n + 1)/2^i]$ for $n = 0, 1, \ldots, 2^i - 1$. Denote the slope of f_i on this interval by $s_{n,i}$ and define $f_i(0) = 0$ and

$$\begin{aligned} &s_{2n,i} := a_i, && s_{2n+1,i} := 2s_{n,i-1} - a_i, && \text{if } s_{n,i-1} \geq 0, \\ &s_{2n,i} := 2s_{n,i-1} + a_i, && s_{2n+1,i} := -a_i, && \text{if } s_{n,i-1} \leq 0. \end{aligned}$$

Then show that

(i) For all $i = 1, 2, \ldots,$ f_i is defined and Lipschitz on $[0, 1]$ and $f_i(2n/2^i) = f_{i-1}(n/2^{i-1})$ for $n = 0, \ldots, 2^{i-1}$ and $i = 2, 3, \ldots$.
(ii) $s_{n,i} \in [-a_i, a_i]$ for all $n = 0, \ldots, 2^{i-1}$ and $i = 1, 2, \ldots$.
(iii) The sequence (f_i) uniformly converges to a Lipschitz function f with a Lipschitz constant 1.
(iv) $\partial_P f(x) = (-1, 1)$ when $x \in (0, 1)$ is a dyadic rational.
(v) $\partial_P f(x) = \emptyset$ when $x \in (0, 1)$ is not a dyadic rational.
(vi) Verify that $f(0) = f(1) = 0$ and $f(1/2) = a_1/2$.
(vii) Extend f periodically to $\mathbb{R}$ and check $\partial_P f(x) = (-1, 1)$ when x is an integer.

Reference: see [38] for details and check [19, 51, 84] for related earlier examples.

4

Variational Techniques in Convex Analysis

Convex analysis is now a rich branch of modern analysis. The purpose of this chapter is merely to point out the applications of variational techniques in convex analysis. In most of the cases direct proofs in the convex case lead to sharper results.

4.1 Convex Functions and Sets

4.1.1 Definitions and Basic Properties

Let X be a Banach space. We say that a subset C of X is *convex* if, for any $x, y \in C$ and any $\lambda \in [0,1]$, $\lambda x + (1-\lambda)y \in C$. We say an extended-valued function $f\colon X \to \mathbb{R} \cup \{+\infty\}$ is *convex* if its domain is convex and for any $x, y \in \operatorname{dom} f$ and any $\lambda \in [0,1]$, one has

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

We call a function $f\colon X \to [-\infty, +\infty)$ *concave* if $-f$ is convex. In some sense convex functions are the simplest functions next to linear functions. Convex functions and convex sets are intrinsically related. For example, if C is a convex set then ι_C and d_C are convex functions. On the other hand if f is a convex function then $\operatorname{epi} f$ and $f^{-1}((-\infty, a])$, $a \in \mathbb{R}$ are convex sets (Exercises 4.1.1, 4.1.2 and 4.1.3). Two other important functions related to a convex set C are the *gauge function* defined by

$$\gamma_C(x) := \inf\{r > 0 \mid x \in rC\},$$

and the *support function* defined on the dual space X^* by

$$\sigma_C(x^*) = \sigma(C; x^*) := \sup\{\langle x, x^* \rangle \mid x \in C\}.$$

Several useful properties of the gauge function and the support function are discussed in Exercises 4.1.6 and 4.1.10.

4.1.2 Local Lipschitz Property of Convex Functions

Lower semicontinuous convex functions are actually locally Lipschitz in the interior of their domains. This is, in fact, a combination of two facts: (a) a convex function f locally bounded above is locally Lipschitz in $\operatorname{int}\operatorname{dom} f$ and (b) a lsc convex function f is locally bounded above in $\operatorname{int}\operatorname{dom} f$. Fact (a) is quite useful itself and we describe it in two propositions.

Proposition 4.1.1 *Let X be a Banach space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a convex function. Suppose that f is locally bounded above at $\bar{x} \in D :=$ $\operatorname{int}(\operatorname{dom} f)$. Then f is locally bounded at $\bar{x}$.*

Proof. Suppose f is bounded above by M, say, in $B_r(x) \subset \operatorname{int}(\operatorname{dom} f)$ for some $r > 0$, then it is bounded below in $B_r(x)$. Indeed, if $y \in B_r(x)$ then so is $2x - y$ and

$$f(x) \le \frac{1}{2}[f(y) + f(2x - y)] \le \frac{1}{2}[f(y) + M]$$

so $f(y) \ge 2f(x) - M$ for all $y \in B_r(x)$. ●

Proposition 4.1.2 *Let X be a Banach space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a convex function. Suppose that f is locally bounded at $\bar{x} \in D := \operatorname{int}(\operatorname{dom} f)$. Then f is locally Lipschitz at $\bar{x}$.*

Proof. Suppose that $|f|$ is bounded by M over $B_{2r}(\bar{x}) \subset D$. Consider distinct points $x, y \in B_r(\bar{x})$. Let $a = \|y - x\|$ and let $z = y + (r/a)(y - x)$. Then $z \in B_{2r}(\bar{x})$. Since

$$y = \frac{a}{a+r} z + \frac{r}{a+r} x$$

is a convex combination lying in $B_{2r}(\bar{x})$, we have

$$f(y) \le \frac{a}{a+r} f(z) + \frac{r}{a+r} f(x).$$

Thus,

$$f(y) - f(x) \le \frac{a}{a+r}(f(z) - f(x)) \le \frac{2Ma}{r} = \frac{2M}{r}\|y - x\|.$$

Interchange x and y gives

$$|f(y) - f(x)| \le \frac{2M}{r}\|y - x\|.$$

●

Theorem 4.1.3 (Lipschitz Property of Convex Functions) *Let X be a Banach space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a lsc convex function. Then f is locally Lipschitz on* $\operatorname{int}(\operatorname{dom} f)$.

Proof. By Propositions 4.1.1 and 4.1.2 we need only show f is locally bounded above. For each natural number i, define $D_i := \{x \in X : f(x) \leq i\}$. The sets D_i are closed and $D \subset \bigcup_{i=1}^{\infty} D_i$. Since D is an open set, by Baire's category theorem, we must have for some i, $\operatorname{int} D_i$ is nonempty. Suppose that $B_s(x) \subset \operatorname{int} D_i$. Then f is bounded above by i over $B_s(x)$. Also since D is open, if $y \in D$ and $y \neq x$, then there exist $\mu > 1$ such that $z := x+\mu(y-x) \in D$. Let $\lambda = 1/\mu \in (0,1)$. The set $U = \{\lambda z + (1-\lambda)b : b \in B_s(x)\}$ is a neighborhood of y in D. For any point $u = \lambda z + (1-\lambda)b \in U$ (where $b \in B_s(x)$) we have

$$f(u) \leq \lambda f(z) + (1-\lambda)i,$$

so f is bounded above in U and therefore locally Lipschitz at y. ●

4.1.3 Convex Series Closed Sets

The condition in Theorem 4.1.3 can be weakened. To understand this deeper result we need the following concepts.

Definition 4.1.4 (Convex Series Closed and Compact) *Let X be a Banach space and let C be a subset of X. We say that C is* convex series closed (cs-closed) *if $\bar{x} = \sum_{i=1}^{\infty} \lambda_i x_i$ with $\lambda_i \geq 0$, $\sum_{i=1}^{\infty} \lambda_i = 1$ and $x_i \in C$ implies $\bar{x} \in C$. We say that C is* convex series compact (cs-compact) *if for any sequence $x_i \in C, i = 1, 2, \dots,$ and any sequence $\lambda_i \geq 0, i = 1, 2, \dots,$ with $\sum_{i=1}^{\infty} \lambda_i = 1$ we have $\sum_{i=1}^{\infty} \lambda_i x_i$ converges to a point of C.*

Some simple yet useful facts related to the cs-closed and cs-compact sets are given below.

Lemma 4.1.5 *Closed convex sets, open convex sets and G_δ convex sets in a Banach space are cs-closed.*

Proof. We prove the lemma for open convex sets and the proofs for the other two cases are left as exercises. Let C be a convex open set in a Banach space and let $\bar{x} = \sum_{i=1}^{\infty} \lambda_i x_i$ with $\lambda_i \geq 0$, $\sum_{i=1}^{\infty} \lambda_i = 1$ and $x_i \in C$. We show that $\bar{x} \in C$. Suppose on the contrary that $\bar{x} \notin C$. Then according to the Hahn–Banach separation theorem there exists a nonzero linear functional $x^* \in X^*$ such that $\langle x^*, c\rangle > \langle x^*, \bar{x}\rangle$ for all $c \in C$. In particular, $0 > \langle x^*, \bar{x} - x_i\rangle$ for $i = 1, 2, \dots,$ and therefore for any $\lambda_i > 0$, $0 > \langle x^*, \lambda_i(\bar{x} - x_i)\rangle$. This leads to $0 > \langle x^*, \bar{x} - \sum_{i=1}^{\infty} \lambda_i x_i\rangle = 0$, a contradiction. ●

Lemma 4.1.6 *Let X and Y be two Banach spaces and let $A\colon X \to Y$ be a continuous linear mapping. Suppose that C is a cs-compact subset of X. Then $A(C)$ is cs-closed.*

Proof. Exercise 4.1.13. ●

An important fact about cs-closed sets is that they share their interior points with their closure.

Theorem 4.1.7 (Open Mapping Theorem: cs-Closed Sets) *Let S be a cs-closed subset of a Banach space X. Then*

$$\text{int } S = \text{int } \overline{S}.$$

Proof. We consider the nontrivial case when $\text{int } \overline{S} \neq \emptyset$. Let $x \in \text{int } \overline{S}$. Shifting S and multiplying it by a constant if necessary we may assume (see Exercise 4.1.16) that

$$0 = x \in B_X \subset \overline{S} \subset S + \frac{1}{2} B_X. \tag{4.1.1}$$

For $i = 1, 2, \ldots$ multiplying (4.1.1) by $1/2^i$ we have

$$\frac{1}{2^i} B_X \subset \frac{1}{2^i} S + \frac{1}{2^{i+1}} B_X. \tag{4.1.2}$$

It follows from (4.1.2) that for any $i = 1, 2, \ldots,$

$$\frac{1}{2} B_X \subset \frac{1}{2} S + \frac{1}{4} S + \cdots + \frac{1}{2^i} S + \frac{1}{2^{i+1}} B_X. \tag{4.1.3}$$

That is to say, for any $u \in B_X/2$ there exist $s_1, \ldots, s_i \in S$ such that

$$u \in \frac{1}{2} s_1 + \frac{1}{4} s_2 + \cdots + \frac{1}{2^i} s_i + \frac{1}{2^{i+1}} B_X. \tag{4.1.4}$$

Taking limits as $i \to \infty$ in (4.1.4) we have $u = \sum_{i=1}^{\infty} s_i/2^i \in S$ because S is cs-closed. Thus, $0 \in 2^{-1} B_X \subset S$, and therefore $\text{int}\, \overline{S} \subset \text{int}\, S$. Hence $\text{int } S = \text{int } \overline{S}$. ●

We now turn to the promised sharper results on the local Lipschitz property for a convex function. Let S be a subset of a Banach space X. We say s is in the *core* of S, denote $s \in \text{core}(S)$, provided that $\bigcup_{\lambda > 0} \lambda(S - s) = X$. Clearly, $\text{int}(S) \subset \text{core}(S)$ and the inclusion could be proper (Exercises 4.1.17 and 4.1.18). Our next result says that if S is the domain of a lsc convex function then the interior and the core of S coincide. The importance of this result is due to the fact that it is much easier to verify that a point belongs to the core than to the interior.

Theorem 4.1.8 *Let X be a Banach space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a lsc convex function. Then*

$$\text{core}(\text{dom}\, f) = \text{int}(\text{dom}\, f).$$

Proof. We need only show that

$$\mathrm{core}(\mathrm{dom}\, f) \subset \mathrm{int}(\mathrm{dom}\, f).$$

Suppose that $\bar{x} \in \mathrm{core}(\mathrm{dom}\ f)$. For each natural number i, define $D_i := \{x \in X : f(x) \le i\}$. The sets D_i are closed and

$$X = \bigcup_{j=1}^{\infty} j(\mathrm{dom}\ f - \bar{x}) = \bigcup_{j,i=1}^{\infty} j(D_i - \bar{x}). \tag{4.1.5}$$

By Baire's category theorem, $\mathrm{int}(D_i - \bar{x})$ (and therefore $\mathrm{int}\ D_i$) is nonempty for some i. Suppose that $B_r(x) \subset \mathrm{int}\ D_i$. Then f is bounded above by i over $B_r(x)$. Moreover, by (4.1.5) there exist integers $j, k > 0$ such that $\bar{x} - x \in j(D_k - \bar{x})$. Letting $\mu = (1+1/j)$, we have $z := x+\mu(\bar{x}-x) \in D_k$. Note that D_k and D_i are contained in the convex set $D_{\max(i,k)}$. Let $\lambda = 1/\mu \in (0,1)$. The set $U = \{\lambda z + (1-\lambda)b : b \in B_r(x)\}$ is a neighborhood of $\bar{x}$ in $D_{\max(i,k)} \subset \mathrm{dom}\, f$. ●

4.1.4 Commentary and Exercises

Although there is a long history of using the convexity of both functions and sets in analysis, the systematical study of convex functions and sets starts in the 1950's associated with the names of Fenchel, Moreau, and Rockafellar. A classical reference for convex analysis is Rockafellar [229]. For a nice short introduction that provides details missed in this chapter we recommend Phelps [215]. More discussion on convex series closed and compact sets can be found in Jameson [147].

Exercise 4.1.1 Let C be a convex subset of a Banach space. Show that d_C and ι_C are convex functions.

Exercise 4.1.2 Let f be a convex function on a Banach space. Show that for any $a \in \mathbb{R}$, $f^{-1}((-\infty, a])$ is a convex set.

Exercise 4.1.3 Let X be a Banach space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be an extended-valued function. Show that f is convex if and only if $\mathrm{epi}\, f$ is a convex subset of $X \times \mathbb{R}$.

Exercise 4.1.4 Show that the intersection of a family of arbitary convex sets is convex. Conclude that $f(x) := \sup\{f_\alpha(x) : \alpha \in A\}$ is convex (and lsc) when $\{f_\alpha\}_{\alpha\in A}$ is a collection of convex (and lsc) functions.

Exercise 4.1.5 Calculate the gauge function for $C := \mathrm{epi}\, 1/x \cap \mathbb{R}^2_+$ and conclude that a gauge function is not necessarily lsc.

Exercise 4.1.6 Let C be a convex subset of a Banach space X and let γ_C be the gauge function of C.

Exercise 4.1.19 Show that in the proof of Theorem 4.1.8 the set U can be expressed explicitly as $U = B_{r(1-\lambda)}(\bar{x})$.

5

Variational Techniques and Multifunctions

Multifunctions arise naturally in many situations. Some frequently encountered examples are: the level sets and sublevel sets of a function, various subdifferentials of nonsmooth functions, the solution sets of an optimization problem depending on some parameters and the vector field of a control system. Here we give a concise discussion on how to apply the technique of variational analysis to problems involving multifunctions. We also discuss subdifferentials as multifunctions.

5.1 Multifunctions

5.1.1 Multifunctions and Related Functions

Let X and Y be two sets. A multifunction from X to Y is a mapping $F\colon X \to 2^Y$, where 2^Y represents the collection of all subsets of Y. We define the domain, range and graph of F by $\operatorname{dom} F := \{x \in X \mid F(x) \neq \emptyset\}$, $\operatorname{range} F := \{y \in Y \mid y \in F(x) \text{ for some } x \in X\}$ and $\operatorname{graph} F := \{(x,y) \in X \times Y \mid y \in F(x)\}$, respectively. The inverse of a multifunction $F\colon X \to 2^Y$ is a multifunction $F^{-1}\colon Y \to 2^X$ defined by $F^{-1}(y) := \{x \in X \mid y \in F(x)\}$. Clearly the domain of F is the range of F^{-1} and the range of F is the domain of F^{-1}. A multifunction is completely characterized by its graph. Moreover, we have the following symmetric relationship between F, F^{-1} and the graph of F: $F(x) = \{y \in Y \mid (x,y) \in \operatorname{graph} F\}$ and $F^{-1}(y) = \{x \in X \mid (x,y) \in \operatorname{graph} F\}$. The following are some examples of multifunctions.

Example 5.1.1 Let X be a Fréchet smooth Banach space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a lsc function. Then $\partial_F f$ is a multifunction from X to X^*.

Example 5.1.2 Let X and Y be metric spaces and let $f\colon X \times Y \to \mathbb{R} \cup \{+\infty\}$ be a lsc function. Then the solution set to the parametric minimization problem of minimizing $x \to f(x,y)$,

$$\operatorname{argmin}(y) := \big\{x \in X \mid f(x,y) = \inf\{f(x',y) \mid x' \in X\}\big\},$$

is a multifunction from Y to X.

Example 5.1.3 Let X be a metric space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a lsc function. Then the sublevel set

$$f^{-1}((-\infty, r]) = \{x \in X \mid f(x) \le r\}$$

and the level set

$$f^{-1}(r) = \{x \in X \mid f(x) = r\}$$

are multifunctions from $\mathbb{R} \to X$.

Example 5.1.4 Let X be a metric space and let $f\colon X \to \mathbb{R}$ be a lsc function. Then the epigraphical profile mapping

$$E_f(x) = \{r \in \mathbb{R} \mid f(x) \le r\}$$

is a multifunction from $X \to \mathbb{R}$. We can see that $\operatorname{graph} E_f = \operatorname{epi} f$ (Exercise 5.1.3).

One can often study a multifunction $F\colon X \to 2^Y$ through related functions. Clearly, $\iota_{\operatorname{graph} F}$ completely characterizes F. When both X and Y are topological spaces, $\iota_{\operatorname{graph} F}$ is a lsc function on $X \times Y$ if and only if $\operatorname{graph} F$ is a closed subset of $X \times Y$. This is an important condition when we analyze a multifunction with variational techniques. Thus, we define a multifunction to be closed if its graph is closed. We say that multifunction F is closed (open, compact, convex) valued if, for every $x \in \operatorname{dom} F$, the set $F(x)$ is closed (open, compact, convex). Note that a closed multifunction is always closed valued yet the converse is not true (Exercise 5.1.2). When Y has additional structure other functions can be used to study a multifunction $F\colon X \to 2^Y$. For example when Y is a metric space we can use $(x,y) \to d(F(x);y)$ and when Y is a Banach space we can use $(x,x^*) \to \sigma(F(x);x^*)$. These functions are in general nonsmooth. We will emphasize the use of variational tools in studying multifunctions by their related nonsmooth functions.

5.1.2 An Example: The Convex Subdifferential

Subdifferentials are multifunctions from X to X^*. In Section 3.4 we have seen the interplay of properties of a function and its (Fréchet) subdifferential. Here we further discuss the subdifferential of a convex function to illustrate various nice properties of the subdifferential as a multifunction inherited from the convexity of the underlying function.

We say a multifunction $F\colon X \to 2^{X^*}$ is monotone provided that for any $x, y \in X$, $x^* \in F(x)$ and $y^* \in F(y)$,

$$\langle y^* - x^*, y - x \rangle \geq 0.$$

The convex subdifferential of a convex lsc function is a typical example of a monotone multifunction.

Theorem 5.1.5 *Let X be a Banach space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a lsc convex function. Then ∂f is a monotone multifunction.*

Proof. Let $x^* \in \partial f(x)$ and $y^* \in \partial f(y)$. It follows from the definition of the convex subdifferential that

$$f(y) - f(x) \geq \langle x^*, y - x \rangle \tag{5.1.1}$$

and

$$f(x) - f(y) \geq \langle y^*, x - y \rangle \tag{5.1.2}$$

Adding (5.1.1) and (5.1.2) we have

$$\langle y^* - x^*, y - x \rangle \geq 0.$$

●

In fact the monotonicity of the subdifferential characterizes the convexity of the underlying function.

Theorem 5.1.6 (Convexity) *Let X be a Fréchet smooth Banach space and let $f\colon X \to \mathbb{R}\cup\{+\infty\}$ be a lsc function. Suppose that $\partial_F f$ is monotone. Then f is convex.*

Proof. If $\partial_F f$ is monotone then for each $x^* \in X^*$ the operator $x \to \partial_F f(x) + x^* = \partial_F(f + x^*)(x)$ is monotone, hence quasi-monotone. By Theorem 3.4.12, for each $x^* \in X^*$, the function $f + x^*$ is quasi-convex. This implies the convexity of f (Exercise 5.1.4). ●

Recall that a monotone multifunction $F\colon X \to 2^{X^*}$ is said to be maximal monotone if graph F is not properly contained in the graph of any monotone multifunction. It is not hard to check that a maximal monotone multifunction is convex valued and closed (Exercise 5.1.5). We can further prove the maximal monotonicity of a monotone Fréchet subdifferential of a lsc function (which must be convex by Theorem 5.1.6).

Theorem 5.1.7 (Maximal Monotonicity) *Let X be a Fréchet smooth Banach space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a proper lsc function. Suppose that* $\operatorname{dom} f \neq \emptyset$ *and $\partial_F f$ is monotone. Then $\partial_F f$ is maximal monotone.*

Proof. Let $b \in X$ and $b^* \in X^*$ be such that $b^* \notin \partial_F f(b)$. We need to show that there exists $x \in X$ and $x^* \in \partial_F f(x)$ such that $\langle x^* - b^*, x - b \rangle < 0$. Observing that $0 \notin \partial_F(f - b^*)(b)$, and therefore b is not a minimum of

(i) Show that γ_C is convex and when $0 \in C$ it is subadditive.
(ii) Show that if $x \in \operatorname{core} C$ then $\operatorname{dom} \gamma_{C-x} = X$.
(iii) Suppose $0 \in \operatorname{core} C$. Prove that $\operatorname{cl} C \subset \{x \in X \mid \gamma_C(x) \le 1\}$.

Exercise 4.1.7 Let X be a Banach space and let C be a cs-closed subset of X. Prove that $\operatorname{int} C = \operatorname{core} C$.

Exercise 4.1.8 Let X be a Banach space and let C be a convex subset of X. Suppose that C is cs-closed and $0 \in \operatorname{core} C$.

(i) Show that $\operatorname{int} C = \{x \in X \mid \gamma_C(x) < 1\}$.
(ii) Deduce that γ_C is definied on X and is continuous.

∗**Exercise 4.1.9** Construct an example showing that the conclusion in Exercise 4.1.8 fails when C is not cs-closed. Hint: Use the existence of a Hamel basis in a vector space to show that in every infinite dimensional Banach space there is a finite linear functional, ϕ which is (everywhere) discontinuous. Deduce that $C := \phi^{-1}[-1, 1]$ is a symmetric convex set with a nonempty core that contains 0 but an empty interior. Yet $\gamma_C(0) = 0 < 1$.

Exercise 4.1.10 Let C_1 and C_2 be closed convex subsets of a Banach space X. Then $C_1 \subset C_2$ if and only if, for any $x^* \in X^*$, $\sigma(C_1; x^*) \le \sigma(C_2; x^*)$. Thus, a closed convex set is characterized by its support function.

Exercise 4.1.11 Prove that if f is a convex lsc function then $\partial f(x) = \partial_F f(x)$.

Exercise 4.1.12 Prove Lemma 4.1.5 for the cases of closed convex sets and convex G_δ sets.

Exercise 4.1.13 Prove Lemma 4.1.6.

Exercise 4.1.14 Let X be a Banach space and let C be a subset of X. Show that C is cs-compact if and only if C is cs-closed and bounded. In particular, both the open and closed unit balls in a Banach space are cs-compact.

Exercise 4.1.15 Let X be a Banach space and let A and B be subsets of X. Suppose that A is cs-compact and B is cs-closed. Then $A+B$ and $\operatorname{conv}(A \cup B)$ are cs-closed.

Exercise 4.1.16 Suppose that S is cs-closed and $\bar{x} \in S$. Show that for any $\delta > 0$ $(S - \bar{x})/\delta$ is also cs-closed.

Exercise 4.1.17 Let S be a subset of a Banach space. Show that $\operatorname{int}(S) \subset \operatorname{core}(S)$.

Exercise 4.1.18 (Core Versus Interior) Consider the set in $\mathbb{R}^2$

$$S = \{(x, y) \mid y = 0 \text{ or } |y| \ge x^2\}.$$

Prove $0 \in \operatorname{core}(S) \setminus \operatorname{int}(S)$.

$f - b^*$, there exists $a \in X$ such that $(f - b^*)(a) < (f - b^*)(b)$. Then it follows from Theorem 3.4.6 that there exists a sequence (x_i) converging to $c \in [a, b)$ and $x_i^* \in \partial_F f(x_i)$ such that $y_i^* := x_i^* - b^* \in \partial_F (f - b^*)(x_i)$ satisfying $\liminf_{i\to\infty}\langle y_i^*, c - x_i\rangle \geq 0$ and $\liminf_{i\to\infty}\langle y_i^*, b - a\rangle > 0$. It follows that

$$\begin{aligned}\liminf_{i\to\infty}\langle x_i^* - b^*, b - x_i\rangle &\geq \liminf_{i\to\infty}\langle y_i^*, b - c\rangle + \liminf_{i\to\infty}\langle y_i^*, c - x_i\rangle \\ &\geq \frac{\|b-c\|}{\|b-a\|}\liminf_{i\to\infty}\langle y_i^*, b - a\rangle + \liminf_{i\to\infty}\langle y_i^*, c - x_i\rangle > 0\end{aligned}$$

It remains to set $x := x_i$ and $x^* := x_i^*$ for i sufficiently large. •

We have seen in Proposition 4.1.2 and Theorem 4.1.8 that a lsc convex function is locally Lipschitz in the core of its domain. Consequently the subdifferential of a lsc convex function is locally bounded in the core of its domain. We will show this holds true in general for a maximal monotone multifunction. The proof actually reduces this more general situation to the continuity of a convex function in the core of its domain.

Theorem 5.1.8 (Boundedness of Monotone Multifunctions) *Let* $F\colon X \to 2^{X^*}$ *be a monotone multifunction. Suppose that* $x \in \operatorname{core}(\operatorname{dom} F)$. *Then* F *is locally bounded at* x.

Proof. By choosing any $x^* \in F(x)$ and replacing F by the monotone multifunction $y \to F(y + x) - x^*$, we lose no generality in assuming that $x = 0$ and that $0 \in F(0)$. Define, for $x \in X$,

$$f(x) := \sup\{\langle y^*, x - y\rangle : y \in \operatorname{dom} F, \|y\| \leq 1 \text{ and } y^* \in F(y)\}.$$

As the supreme of affine continuous functions, f is convex and lower semicontinuous. We show that $\operatorname{dom} f$ is an absorbing set. First, since $0 \in F(0)$, we must have $f \geq 0$. Second, whenever $y \in \operatorname{dom} F$ and $y^* \in F(y)$, monotonicity implies that $0 \leq \langle y^* - 0, y - 0\rangle$, so $f(0) \leq 0$. Thus, $f(0) = 0$. Suppose $x \in X$. By hypothesis, $\operatorname{dom} F$ is absorbing so there exists $t > 0$ such that $F(tx) \neq \emptyset$. Choose any element $u^* \in F(tx)$. If $y \in \operatorname{dom} F$ and $y^* \in F(y)$, then by monotonicity

$$\langle y^*, tx - y\rangle \leq \langle u^*, tx - y\rangle.$$

Consequently,

$$f(tx) \leq \sup\{\langle u^*, tx - y\rangle : y \in \operatorname{dom} F, \|y\| \leq 1\} < \langle u^*, tx\rangle + \|u^*\| < +\infty.$$

By virtue of Proposition 4.1.2 and Theorem 4.1.8, f is continuous at 0 and hence there exists $\eta > 0$ such that $f(x) < 1$ for all $x \in 2\eta B_X$. Equivalently, if $x \in 2\eta B_X$, then $\langle y^*, x\rangle \leq \langle y^*, y\rangle + 1$ whenever $y \in \operatorname{dom} F$, $\|y\| \leq 1$ and $y^* \in F(y)$. Thus, if $y \in \eta B_X \cap \operatorname{dom} F$ and $y^* \in F(y)$, then

$$2\eta\|y^*\| = \sup\{\langle y^*, x\rangle : x \in 2\eta B_X\} \leq \|y^*\| \times \|y\| + 1 \leq \eta\|y^*\| + 1,$$

so $\|y^*\| \le 1/\eta$. ●

Note that Theorem 5.1.8 does not require that the domain of F be convex.

5.1.3 Limits of Sequences of Sets

Having defined multifunctions we turn to their limits and continuity. We will take a sequential approach, and therefore need to study the limits of sequences of sets.

Definition 5.1.9 *Let Y be a Hausdorff topological space and let (F_i) be a sequence of subsets of Y. The sequential* lower *and* upper *limits of F_i are defined by*

$$\liminf_{i\to\infty} F_i = \{\lim_{i\to\infty} y_i \mid y_i \in F_i \text{ for all } i = 1, 2, \ldots\}$$

and

$$\limsup_{i\to\infty} F_i = \{\lim_{k\to\infty} y_{i_k} \mid y_{i_k} \in F_{i_k} \text{ for some } i_k \to \infty\}.$$

Clearly $\liminf_{i\to\infty} F_i \subset \limsup_{i\to\infty} F_i$. When they are equal we define the common set to be the Painlevé–Kuratowski limit of the sequence (F_i) and denote it by $\lim_{i\to\infty} F_i$. In a metric space both the sequential lower and upper limits are closed. However, this is not true in general (Exercise 5.1.8).

When Y is a metric space the lower and upper limits can be represented alternatively as

$$\liminf_{i\to\infty} F_i = \bigcap_{k=1}^{\infty} \bigcup_{j=1}^{\infty} \bigcap_{i=j}^{\infty} B_{\frac{1}{k}}(F_i) \tag{5.1.3}$$

and

$$\limsup_{i\to\infty} F_i = \bigcap_{k=1}^{\infty} \bigcap_{j=1}^{\infty} \bigcup_{i=j}^{\infty} B_{\frac{1}{k}}(F_i). \tag{5.1.4}$$

We leave the proofs of these alternative representations as Exercise 5.1.6.

These lower and upper limits can also be described by using the distance between a set and a point.

Lemma 5.1.10 *Let Y be a metric space and let (F_i) be a sequence of subsets in Y. Then*

$$\liminf_{i\to\infty} F_i = \{y \in Y \mid \limsup_{i\to\infty} d(F_i; y) = 0\}$$

and

$$\limsup_{i\to\infty} F_i = \{y \in Y \mid \liminf_{i\to\infty} d(F_i; y) = 0\}.$$

Proof. Exercise 5.1.7. ●

Lemma 5.1.10 is a special case of the following more general characterization of the upper and lower limits of a sequence of sets.

Lemma 5.1.11 *Let Y be a metric space, let F be a closed subset of X and let (F_i) be a sequence of subsets in Y. Then*

$$F \subset \liminf_{i\to\infty} F_i \tag{5.1.5}$$

if and only if for any $y \in Y$,

$$\limsup_{i\to\infty} d(F_i; y) \le d(F; y); \tag{5.1.6}$$

and

$$\limsup_{i\to\infty} F_i \subset F \tag{5.1.7}$$

if and only if for any $y \in Y$,

$$\liminf_{i\to\infty} d(F_i; y) \ge d(F; y). \tag{5.1.8}$$

Consequently,

$$\lim_{i\to\infty} F_i = F$$

if and only if for any $y \in Y$,

$$\lim_{i\to\infty} d(F_i; y) = d(F; y).$$

Proof. We prove the equivalence of (5.1.5) and (5.1.6). It follows from Lemma 5.1.10 that (5.1.6) implies (5.1.5). Now suppose (5.1.5) holds and let $y \in Y$ be an arbitrary element. For any $\varepsilon > 0$ choose $x \in F$ such that $d(F; y) + \varepsilon \ge d(x, y)$ and let (x_i) be a sequence converges to x with $x_i \in F_i$. Then

$$d(F_i; y) \le d(x_i, y).$$

Taking $\limsup$ as $i \to \infty$ we have

$$\limsup_{i\to\infty} d(F_i; y) \le d(x, y) \le d(F; y) + \varepsilon.$$

Since ε is arbitrary we obtain (5.1.6).

The proof of the equivalence of (5.1.7) and (5.1.8) is similar and left as Exercise 5.1.9. ●

Applying the Painlevé–Kuratowski limit to the epigraph of a sequence of functions leads to the concept of epi-convergence. This is particularly useful in analyzing approximations of functions when minimizing the function is a primary concern.

Definition 5.1.12 (Epi-convergence) *Let X be a metric space and let $f_i\colon X \to \mathbb{R}$ be a sequence of lsc functions. The lower epi-limit* $\operatorname{e-lim\,inf}_{i\to\infty} f_i$ *is the function with*

$$\operatorname{epi}(\operatorname*{e-lim\,inf}_{i\to\infty} f_i) = \limsup_{i\to\infty} \operatorname{epi} f_i,$$

and the upper epi-limit $\operatorname{e-lim\,sup}_{i\to\infty} f_i$ *is the function with*

$$\operatorname{epi}(\operatorname*{e-lim\,sup}_{i\to\infty} f_i) = \liminf_{i\to\infty} \operatorname{epi} f_i.$$

When these two functions coincide we say that f_i epi-converges to its epi-limit

$$\operatorname*{e-lim}_{i\to\infty} f_i = \operatorname*{e-lim\,inf}_{i\to\infty} f_i = \operatorname*{e-lim\,sup}_{i\to\infty} f_i.$$

Note that both lower and upper epi-limits are lsc functions, and so is the epi-limit when it exists (Exercise 5.1.11). Epi-limits have the following easy yet useful characterization, whose proof is left as an exercise.

Lemma 5.1.13 *Let X be a metric space and let $f_i\colon X \to \mathbb{R}$ be a sequence of lsc functions. Then $f = \operatorname{e-lim}_{i\to\infty} f_i$ if and only if at each point $x \in X$ one has*

$$\liminf_{i\to\infty} f_i(x_i) \geq f(x) \text{ for every sequence } x_i \to x \tag{5.1.9}$$

and

$$\limsup_{i\to\infty} f_i(x_i) \leq f(x) \text{ for some sequence } x_i \to x \tag{5.1.10}$$

Proof. Exercise 5.1.12. ●

We end this subsection with a result that illuminates the usefulness of epi-convergence in minimization problems.

Theorem 5.1.14 *Let X be a metric space and let $f_i\colon X \to \mathbb{R}$ be a sequence of lsc functions. Suppose that $f = \operatorname{e-lim}_{i\to\infty} f_i$ and that $\operatorname{dom} f, \operatorname{dom} f_i \subset E$, $i = 1, 2, \ldots$ for some compact subset E of X. Then*

$$\lim_{i\to\infty} \inf f_i = \inf f, \tag{5.1.11}$$

and

$$\limsup_{i\to\infty} \operatorname{argmin} f_i \subset \operatorname{argmin} f. \tag{5.1.12}$$

Proof. Let $\bar{x} \in \operatorname{argmin} f$. Then

$$(\bar{x}, f(\bar{x})) \in \operatorname{epi} f \subset \liminf_{i\to\infty} \operatorname{epi} f_i,$$

so that there exists $(x_i, r_i) \in \operatorname{epi} f_i$ satisfying $\lim_{i\to\infty}(x_i, r_i) = (\bar{x}, f(\bar{x}))$. It follows that

$$\inf f = f(\bar{x}) = \lim_{i\to\infty} r_i \geq \limsup_{i\to\infty} f_i(x_i) \geq \limsup_{i\to\infty} \inf f_i. \tag{5.1.13}$$

On the other hand, let $x_i \in \operatorname{argmin} f_i \subset E$. Since E is compact there exists a subsequence (i_k) of the natural numbers such that for some $x \in E$, $x = \lim_{k\to\infty} x_{i_k}$ and

$$\lim_{k\to\infty} f_{i_k}(x_{i_k}) = \liminf_{i\to\infty} \inf f_i.$$

Thus,

$$(x, \liminf_{i\to\infty} \inf f_i) \in \limsup_{i\to\infty} \operatorname{epi} f_i \subset \operatorname{epi} f,$$

so that

$$\liminf_{i\to\infty} \inf f_i \geq f(x) \geq \inf f. \tag{5.1.14}$$

Combining inequalities (5.1.13) and (5.1.14) we have

$$\lim_{i\to\infty} \inf f_i = \inf f.$$

Finally, let $\bar{x} \in \limsup \operatorname{argmin} f_i$ so that there exists a subsequence (i_k) of the natural numbers and $x_{i_k} \in \operatorname{argmin} f_{i_k}$ such that $\bar{x} = \lim_{k\to\infty} x_{i_k}$. Since

$$\limsup_{i\to\infty} \operatorname{epi} f_i \subset \operatorname{epi} f$$

we have $(\bar{x}, \limsup_{k\to\infty} f_{i_k}(x_{i_k})) \in \operatorname{epi} f$ so that

$$\limsup_{k\to\infty} f_{i_k}(x_{i_k}) \geq f(\bar{x}).$$

Now consider any $x \in \operatorname{dom} f$. Then

$$(x, f(x)) \in \operatorname{epi} f \subset \liminf_{i\to\infty} \operatorname{epi} f_i$$

so that there exists a sequence $(y_i, r_i) \in \operatorname{epi} f_i$ converging to $(x, f(x))$. It follows that

$$\begin{aligned} f(x) = \lim_{i\to\infty} r_i &\geq \limsup_{i\to\infty} f_i(y_i) \\ &\geq \limsup_{k\to\infty} f_{i_k}(x_{i_k})) \geq f(\bar{x}). \end{aligned}$$

Since $x \in \operatorname{dom} f$ is arbitrary, $\bar{x} \in \operatorname{argmin} f$. ●

By carefully examining the proof we can see that the condition that $\operatorname{dom} f$ and $\operatorname{dom} f_i$ are contained in a compact subset E of X is not needed in establishing inclusion (5.1.12). However, without this condition, (5.1.11) is false (Exercise 5.1.13).

5.1.4 Continuity of Multifunctions

The basic definition is given below.

Definition 5.1.15 (Continuity of Multifunction) *Let X and Y be two Hausdorff topological spaces and let $F\colon X \to 2^Y$ be a multifunction. We say that F is upper (lower) semicontinuous at $\bar{x} \in X$ provided that for any open set U in Y with $F(\bar{x}) \subset U$, $(F(\bar{x}) \cap U \neq \emptyset)$,*

$$\{x \in X \mid F(x) \subset U\} \quad (\{x \in X \mid F(x) \cap U \neq \emptyset\})$$

is an open set in X. We say that F is continuous at $\bar{x}$ if it is both upper and lower semicontinuous at $\bar{x}$. We say that F is upper (lower) continuous on $S \subset X$ if it is upper (lower) continuous at every $x \in S$. We omit S when it coincides with the domain of F.

We will also need a sequential approach to limits and continuity of multifunctions. This is mainly for applications in the subdifferential theory because the corresponding topological approach often yields objects that are too big.

Definition 5.1.16 (Sequential Lower and Upper Limits) *Let X and Y be two Hausdorff topological spaces and let $F\colon X \to 2^Y$ be a multifunction. We define the sequential lower and upper limit of F at $\bar{x} \in X$ by*

$$\mathop{\text{s-lim\,inf}}_{x \to \bar{x}} F(x) := \bigcap \{\liminf_{i \to \infty} F(x_i) \mid x_i \to \bar{x}\}$$

and

$$\mathop{\text{s-lim\,sup}}_{x \to \bar{x}} F(x) := \bigcup \{\limsup_{i \to \infty} F(x_i) \mid x_i \to \bar{x}\}.$$

When

$$\mathop{\text{s-lim\,inf}}_{x \to \bar{x}} F(x) = \mathop{\text{s-lim\,sup}}_{x \to \bar{x}} F(x)$$

we call the common set the sequential limit of F at $\bar{x}$ and denote it by $\text{s-lim}_{x \to \bar{x}} F(x)$.

Definition 5.1.17 (Semicontinuity and Continuity) *Let X and Y be two Hausdorff topological spaces and let $F\colon X \to 2^Y$ be a multifunction. We say that F is sequentially lower (upper) semicontinuous at $\bar{x} \in X$ provided that*

$$F(\bar{x}) \subset \mathop{\text{s-lim\,inf}}_{x \to \bar{x}} F(x) \;\; (\mathop{\text{s-lim\,sup}}_{x \to \bar{x}} F(x) \subset F(\bar{x})).$$

When F is both upper and lower semicontinuous at $\bar{x}$ we say it is continuous at $\bar{x}$. In the notation introduced above,

$$F(\bar{x}) = \mathop{\text{s-lim}}_{x \to \bar{x}} F(x).$$

Clearly, when Y is a metric space the sequential and the topological (semi) continuity coincide.

The following example illustrates how the semicontinuity and continuity of multifunctions relate to that of functions.

Example 5.1.18 (Profile Mappings) Let X be a Banach space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a function. Then the epigraphic profile of f, E_f is upper (lower) semicontinuous at $\bar{x}$ if and only if f is lower (upper) semicontinuous at $\bar{x}$. Consequently, E_f is continuous at $\bar{x}$ if and only if f is continuous.

Example 5.1.19 (Sublevel Set Mappings) Let X be a Banach space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a lsc function. Then the sublevel set mapping $S(a) = f^{-1}((-\infty, a])$ is upper semicontinuous.

When X and Y are metric spaces we have the following characterizations of the sequential lower and upper limit.

Theorem 5.1.20 (Continuity and Distance Functions) *Let X and Y be two metric spaces and let $F\colon X \to 2^Y$ be a multifunction. Then F is sequentially lower (upper) semicontinuous at $\bar{x} \in X$ if and only if for every $y \in Y$, the distance function $x \to d(F(x); y)$ is upper (lower) semicontinuous. Consequently, F is continuous at $\bar{x}$ if and only if for every $y \in Y$, the distance function $x \to d(F(x); y)$ is continuous.*

Proof. This follows from Lemma 5.1.11. Details are left as Exercise 5.1.15. ●

5.1.5 Uscos and Cuscos

The acronym *usco* (*cusco*) stands for a (convex) upper semicontinuous nonempty valued compact multifunction. Such multifunctions are interesting because they describe common features of the maximal monotone operators, of the convex subdifferential and of the Clarke generalized gradient.

Definition 5.1.21 *Let X be a Banach space and let Y be a Hausdorff topological vector space. We say $F\colon X \to 2^Y$ is an* usco *(*cusco*) provided that F is a nonempty (convex) compact valued upper semicontinuous multifunction. An usco (cusco) is* minimal *if it does not properly contain any other usco (cusco).*

A particularly useful case is when $Y = X^*$ with its weak-star topology. In this case we use the terminology weak*-usco (-cusco).

Closed multifunctions and uscos have an intimate relationship.

Proposition 5.1.22 *Let X and Y be two Hausdorff topological spaces and let $F\colon X \to 2^Y$ be a multifunction. Suppose that F is an usco. Then it is closed. If in addition,* range F *is compact, then F is an usco if and only if F is closed.*

Proof. It is easy to check that if $F\colon X \to 2^Y$ is an usco, then its graph is closed (Exercise 5.1.16). Now suppose F is closed and range F is compact. Then clearly F is compact valued. We show it is upper semicontinuous. Suppose on the contrary that F is not upper semicontinuous at $\bar{x} \in X$. Then there exists an open set $U \subset Y$ containing $F(\bar{x})$ and a net $x_\alpha \to \bar{x}$ and $y_\alpha \in F(x_\alpha)\backslash U$ for each α. Since range F is compact, we can take subnet (x_β, y_β) of (x_α, y_α) such that $x_\beta \to \bar{x}$ and $y_\beta \to \bar{y} \notin U$. On the other hand it follows from F is closed that $\bar{y} \in F(\bar{x}) \subset U$, a contradiction. ●

An important feature of an usco (cusco) is that it always contains a minimal one.

Proposition 5.1.23 (Existence of Minimal usco) *Let X and Y be two Hausdorff topological spaces and let $F\colon X \to 2^Y$ be an usco (cusco). Then there exists a minimal usco (cusco) contained in F.*

Proof. By virtue of of Zorn's lemma we need only show that any decreasing chain (F_α) of usco (cusco) maps contained in F in terms of set inclusion has a minimal element. For $x \in X$ define $F_0(x) = \bigcap F_\alpha(x)$. Since $F_\alpha(x)$ are compact, $F_0(x)$ is nonempty, (convex) and compact. It remains to show that F_0 is upper semicontinuous. Suppose that $x \in X$, U is open in Y and $F_0(x) \subset U$. Then $F_\alpha(x) \subset U$ for some α. Indeed, if each $F_\alpha(x)\backslash U$ were nonempty then the intersection of these compact nested sets would be a nonempty subset of $F_0(x)\backslash U$, a contradiction. By upper semicontinuity of F_α, there exists an open set V containing x such that $F_0(V) \subset F_\alpha(V) \subset U$. ●

When $Y = \mathbb{R}$ the proposition below provides a procedure of constructing a minimal usco contained in a given usco.

Proposition 5.1.24 *Let X be a Hausdorff topological space and $F\colon X \to 2^{\mathbb{R}}$ an usco. For each $x \in X$, put $f(x) := \min\{r \mid r \in F(x)\}$. Let $G\colon X \to 2^{\mathbb{R}}$ be the closure of f (i.e., the set-valued mapping whose graph is the closure of the graph of f). Now put $g(x) := \max\{r \mid r \in G(x)\}$ for each $x \in X$. Finally let $H\colon X \to 2^{\mathbb{R}}$ be the closure of g. Then H is a minimal usco contained in F.*

Proof. Since the graph of F is closed, G is contained in F, and G is an usco as G is closed and F is an usco. For the same reason H is an usco contained in G.

To show that H is minimal, consider open sets $U \subset X$ and $W \subset \mathbb{R}$, such that there is some $w \in H(U) \cap W$. It is sufficient to find a nonempty open subset of U, whose image under H is entirely contained in W.

Fix some $\varepsilon < d(\mathbb{R} \setminus W; w)$. Since $w \in H(U)$, there is some $x \in U$ such that $g(x) \in (w-\varepsilon; w+\varepsilon)$. This means that $G(x) \subset (-\infty; w+\varepsilon)$ and by upper semi-continuity of G there is an open $V \subset U$, $V \ni x$, such that $G(V) \subset (-\infty\,;\, w+\varepsilon)$.

As $g(x) \in (w-\varepsilon, w+\varepsilon)$, there is some $x' \in V$ with $f(x') \in (w-\varepsilon, w+\varepsilon)$. This means that $F(x') \subset (w-\varepsilon, +\infty)$ and by upper semi-continuity of F there is an open $V' \subset V$, $V' \ni x'$, such that $F(V') \subset (w-\varepsilon, +\infty)$.

Now $H(V') \subset F(V') \cap G(V) \subset (w-\varepsilon, w+\varepsilon) \subset W$. Thus H is a minimal usco. •

Maximal monotone operators, in particular, subdifferentials of convex functions provide interesting examples of w*-cuscos. We leave the verification of the following example as a guided exercise (Exercise 5.1.17).

Example 5.1.25 Let X be a Banach space, let $F\colon X \to 2^{X^*}$ be a maximal monotone multifunction and let S be an open subset of $\operatorname{dom} F$. Then the restriction of F to S is a w*-cusco.

To further explore the relationship of maximal monotone multifunctions and cuscos we need to extend the notion of maximal monotone multifunctions to arbitrary set.

Definition 5.1.26 (Maximal Monotone on a Set) *Let X be a Banach space, let $F\colon X \to 2^{X^*}$ be a monotone multifunction and let S be a subset of X. We say that F is maximal monotone in S provided the monotone set*

$$\operatorname{graph} F \cap (S \times X^*) := \{(x, x^*) \in S \times X^* \mid x \in S \text{ and } x^* \in F(x)\}$$

is maximal under the set inclusion in the family of all monotone sets contained in $S \times X^*$.

It turns out that a monotone cusco on an open set is maximal.

Lemma 5.1.27 *Let X be a Banach space, let $F\colon X \to 2^{X^*}$ be a monotone multifunction and let S be an open subset of X. Suppose that $S \subset \operatorname{dom} F$ and F is a w*-cusco on S. Then F is maximal monotone in S.*

Proof. We need only show that if $(y, y^*) \in S \times X^*$ satisfies

$$\langle y^* - x^*, y - x\rangle \geq 0 \text{ for all } x \in S, x^* \in F(x), \tag{5.1.15}$$

then $y^* \in F(y)$. If not, by the separation theorem there exists $z \in X\backslash\{0\}$ such that $F(y) \subset \{z^* \in X^* \mid \langle z^*, z\rangle < \langle y^*, z\rangle\} = W$. Since W is weak* open and F is w*-upper semicontinuous on S, there exists an $h > 0$ with $B_h(y) \subset S$ such that $F(B_h(y)) \subset W$. Now, for $t \in (0, h/\|z\|)$, we have $y + tz \in B_h(y)$, and therefore $F(y+tz) \subset W$. Applying (5.1.15) to any $u^* \in F(y+tz)$ we get

$$0 \leq \langle y^* - u^*, y - (y+tz)\rangle = -t\langle y^* - u^*, z\rangle,$$

which implies $\langle u^*, z\rangle \geq \langle y^*, z\rangle$, that is $u^* \notin W$, a contradiction. •

As a corollary we have

Corollary 5.1.28 *Let X be a Banach space, let $F\colon X \to 2^{X^*}$ be a maximal monotone multifunction and let S be an open subset of X. Suppose that $S \subset \operatorname{dom} F$. Then F is maximal monotone in S.*

Proof. By Example 5.1.25 the maximal monotonicity of F implies that F is a w*-cusco on S, so the result follows from Lemma 5.1.27. ●

Now we can prove the interesting relation that a maximal monotone multifunction on an open set is a minimal cusco.

Theorem 5.1.29 (Maximal Monotonicity and Minimal cusco) *Let X be a Banach space, let S be an open subset of X and let F be a maximal monotone multifunction in S. Then F is a minimal w*-cusco.*

Proof. We know by Example 5.1.25 that F is a w*-cusco. Suppose that $G\colon S \to 2^{X^*}$ is a w*-cusco and $\operatorname{graph} G \subset \operatorname{graph} F$. By Lemma 5.1.27, G is maximal monotone, and therefore $G = F$. ●

Note that a maximal monotone multifunction need not be a minimal usco. The following example clarifies the difference whose easy proof is left as Exercise 5.1.18.

Example 5.1.30 Define monotone multifunctions F_0, F_1 and F_2 from $\mathbb{R} \to 2^{\mathbb{R}}$ by

$$F_0(x) = F_1(x) = F_2(x) = \operatorname{sgn} x \text{ if } x \neq 0,$$

while

$$F_0(0) = \{-1\}, F_1(0) = \{-1, 1\} \text{ and } F_2(0) = [-1, 1].$$

Then $\operatorname{graph} F_0 \subset \operatorname{graph} F_1 \subset \operatorname{graph} F_2$, and they are all distinct. The multifunction F_2 is maximal monotone and minimal cusco, F_1 is minimal usco and F_0 does not have a closed graph.

5.1.6 Monotone Operators and the Fitzpatrick Function

Throughout this subsection, $(X, \|\cdot\|)$ is a reflexive Banach space with dual X^* and $T\colon X \to 2^{X^*}$ is maximal monotone. The *Fitzpatrick function* F_T, associated with T, is the proper closed convex function defined on $X \times X^*$ by

$$\begin{aligned} F_T(x, x^*) &:= \sup_{y^* \in Ty} [\langle y^*, x\rangle + \langle x^*, y\rangle - \langle y^*, y\rangle] \\ &= \langle x^*, x\rangle + \sup_{y^* \in Ty} \langle x^* - y^*, y - x\rangle. \end{aligned}$$

Since T is maximal monotone

$$\sup_{y^* \in Ty} \langle x^* - y^*, y - x\rangle \geq 0$$

and the equality holds if and only if $x^* \in Tx$, it follows that

$$F_T(x, x^*) \geq \langle x^*, x\rangle \tag{5.1.16}$$

with equality holding if and only if $x^* \in Tx$. Thus, we capture much of a maximal monotone multifunction via an associated convex function.

Using only the Fitzpatrick function and the decoupling lemma we can prove the following fundamental result remarkably easily.

Theorem 5.1.31 (Rockafellar) *Let X be a reflexive Banach space and let $T\colon X \to 2^{X^*}$ be a maximal monotone operator. Then* range$(T + J) = X^*$. *Here J is the* duality map *defined by $J(x) := \partial\|x\|^2/2$.*

Proof. The Cauchy inequality and (5.1.16) implies that for all x, x^*,

$$F_T(x, x^*) + \frac{\|x\|^2 + \|x^*\|^2}{2} \geq 0. \tag{5.1.17}$$

Applying the decoupling result of Lemma 4.3.1 to (5.1.17) we conclude that there exists a point $(w^*, w) \in X^* \times X$ such that

$$\begin{aligned} 0 \leq\, & F_T(x, x^*) - \langle w^*, x\rangle - \langle x^*, w\rangle \\ & + \frac{\|y\|^2 + \|y^*\|^2}{2} + \langle w^*, y\rangle + \langle y^*, w\rangle \end{aligned} \tag{5.1.18}$$

Choosing $y \in -Jw^*$ and $y^* \in -Jw$ in inequality (5.1.18) we have

$$F_T(x, x^*) - \langle w^*, x\rangle - \langle x^*, w\rangle \geq \frac{\|w\|^2 + \|w^*\|^2}{2}. \tag{5.1.19}$$

For any $x^* \in Tx$, adding $\langle w^*, w\rangle$ to both sides of the above inequality and noticing $F_T(x, x^*) = \langle x^*, x\rangle$ we obtain

$$\langle x^* - w^*, x - w\rangle \geq \frac{\|w\|^2 + \|w^*\|^2}{2} + \langle w^*, w\rangle \geq 0. \tag{5.1.20}$$

Since (5.1.20) holds for all $x^* \in Tx$ and T is maximal we must have $w^* \in Tw$. Now setting $x^* = w^*$ and $x = w$ in (5.1.20) yields

$$\frac{\|w\|^2 + \|w^*\|^2}{2} + \langle w^*, w\rangle = 0,$$

which implies $-w^* \in Jw$. Thus, $0 \in (T + J)w$. Since the argument applies equally well to all translations of T, we have range$(T + J) = X^*$ as required. ●

There is a tight relationship between nonexpansive mappings and monotone operators in Hilbert spaces, as stated in the next lemma.

Lemma 5.1.32 *Let H be a Hilbert space. Suppose that P and T are two multifunctions from subsets of H to 2^H whose graphs are related by the condition $(x, y) \in$ graph P if and only if $(v, w) \in$ graph T where $x = w + v$ and $y = w - v$. Then*

(i) *P is nonexpansive (and single-valued) if and only if T is monotone.*
(ii) $\operatorname{dom} P = \operatorname{range}(T + I)$.

Proof. Exercise 5.1.29. •

This very easily leads to the Kirszbraun–Valentine theorem [156, 248] on the existence of nonexpansive extensions to all of Hilbert space of nonexpansive mappings on subsets of Hilbert space. The proof is left as a guided exercise.

Theorem 5.1.33 (Kirszbraun–Valentine) *Let H be a Hilbert space and let D be a non-empty subset of H. Suppose that $P\colon D \to H$ is a nonexpansive mapping. Then there exists a nonexpansive mapping $\widehat{P}\colon H \to H$ defined on all of H such that $\widehat{P}|_D = P$.*

Proof. Exercise 5.1.30. •

Alternatively [220], one may directly associate a convex Fitzpatrick function F_P with a non-expansive mapping P, and thereby derive the Kirszbraun–Valentine theorem, see Exercise 5.1.31.

5.1.7 Commentary and Exercises

Multifunctions or set-valued functions have wide applications and have been the subject of intensive research in the past several decades. Our purpose in this short section is merely to provide minimal preliminaries and some interesting examples. Aubin and Frankowska's monograph [8] and Klein and Thompson's book [157] are excellent references for readers who are interested in this subject.

The subdifferential for convex functions is the first generalized differential concept that leads to a multifunction. It has many nice properties later generalized to the classes of usco and cusco multifunctions. The usco and cusco also relate to other concepts of generalized derivative such as the Clarke generalized gradient. Our discussion on usco and cusco here largely follows those in [62, 68, 215].

Maximal monotone operators are generalizations of the convex subdifferential—though they first flourished in partial differential equation theory. Rockafellar's result in Theorem 5.1.31 is in [223]. The original proofs were very extended and quite sophisticated—they used tools such as Brouwer's fixed point theorem and Banach space renorming theory. As with the proof of the local boundedness of Theorem 5.1.8, ultimately the result is reduced to much more accessible geometric convex analysis. These proofs well illustrate the techniques of variational analysis: using a properly constructed auxiliary function, the variational principle with decoupling in the form of a sandwich theorem and followed by an appropriate decoding of the information. Simon Fitzpatrick played a crucial role in this process by constructing the auxillary functions. The proof of Theorem 5.1.8 follows [33]. The short proof of Theorem 5.1.31 is a reworking of that of [234] given in [52] using the Fitzpatrick function discovered in [116].

Exercise 5.1.1 Let F be a multifunction from X to Y.

(i) Show that $\operatorname{dom} F = \operatorname{range} F^{-1}$ and $\operatorname{range} F = \operatorname{dom} F^{-1}$.
(ii) Show that $F(x) = \{y \in Y \mid (x, y) \in \operatorname{graph} F\}$ and $F^{-1}(y) = \{x \in X \mid (x, y) \in \operatorname{graph} F\}$.

Exercise 5.1.2 Let X and Y be Hausdorff topological spaces and let $F\colon X \to 2^Y$ be a multifunction.

(i) Show that if F is closed then it is closed valued.
(ii) Construct a closed valued multifunction whose graph is not closed.

Exercise 5.1.3 Let X be a metric space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a lsc function. Show that $\operatorname{graph} E_f = \operatorname{epi} f$.

Exercise 5.1.4 Let X be a Banach space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a function. Suppose that, for any $x^* \in X^*$, $x \to f(x) + \langle x^*, x\rangle$ is quasi-convex. Show that f is a convex function. Hint: Choose x^* such that $f(x) + \langle x^*, x\rangle = f(y) + \langle x^*, y\rangle$.

Exercise 5.1.5 Let X be a Banach space and let $F\colon X \to 2^{X^*}$ be a maximal monotone multifunction. Show that F is convex valued and closed.

Exercise 5.1.6 Prove the representations of the lower and upper limits of sequence of subsets in (5.1.3) and (5.1.4).

Exercise 5.1.7 Prove Lemma 5.1.10.

Exercise 5.1.8 Prove that in a metric space the sequential lower and upper limits of a sequence of subsets are always closed sets. Give an example showing that this is not the case in a general Hausdorff topological space.

Exercise 5.1.9 Prove the equivalence of (5.1.7) and (5.1.8) in Lemma 5.1.11.

Exercise 5.1.10 (Limits of Monotone and Sandwiched Sequences) Let (F_i) be a sequence in a metric space Y.

(i) Suppose that (F_i) is monotone increasing, i.e, $F_i \subset F_{i+1}$ for $i = 1, 2, \ldots$. Then $\lim_{i\to\infty} F_i = \operatorname{cl} \bigcup_{i=1}^{\infty} F_i$.
(ii) Suppose that (F_i) is monotone decreasing, i.e., $F_{i+1} \subset F_i$ for $i = 1, 2, \ldots$. Then $\lim_{i\to\infty} F_i = \bigcap_{i=1}^{\infty} \operatorname{cl} F_i$.

(iii) Suppose that $F_i \subset G_i \subset H_i$ and $\lim_{i\to\infty} F_i = \lim_{i\to\infty} H_i = G$. Then $\lim_{i\to\infty} G_i = G$.

Exercise 5.1.11 (Lower Semicontinuity of Epi-limits) Let X be a metric space and let $f_i\colon X \to \mathbb{R}$ be a sequence of lsc functions. Then both $\text{e-lim}\inf_{i\to\infty} f_i$ and $\text{e-lim}\sup_{i\to\infty} f_i$ are lsc functions. Therefore, $\text{e-lim}_{i\to\infty} f_i$ is a lsc function when exists.

Exercise 5.1.12 (Characterization of Epi-limits) Prove Lemma 5.1.13.

Exercise 5.1.13 Construct an example on $X = \mathbb{R}$ showing that without the condition that $\operatorname{dom} f$ and $\operatorname{dom} f_i$ belong to a compact subset of X, the conclusion (5.1.11) in Theorem 5.1.14 is false.

Exercise 5.1.14 Prove the claim in Example 5.1.18.

Exercise 5.1.15 Prove Theorem 5.1.20.

Exercise 5.1.16 Let $F\colon X \to 2^Y$ be an usco. Show that $\operatorname{graph} F$ is a closed subset of $X \times Y$.

Exercise 5.1.17 Verify Example 5.1.25. Hint: By Exercise 5.1.5 F is convex valued and closed. The upper semicontinuity of F follows from Theorem 5.1.8 and Proposition 5.1.22.

Exercise 5.1.18 Verify the claims in Example 5.1.30.

Exercise 5.1.19 Construct a multifunction F from $\mathbb{R}$ to $\mathbb{R}^2$ whose projections into $\mathbb{R}$ are both minimal usco mappings yet F itself is not. Hint: Let $F(x) = \{(\operatorname{sgn}(x); \operatorname{sgn}(x))\}$ for $z \neq 0$, while

$$F(0) = \{(-1;-1), (-1;1), (1;-1), (1;1)\}.$$

Exercise 5.1.20 Construct a minimal usco contained in a given usco $F\colon Z \to \mathbb{R}^N$.

Exercise 5.1.21 Deduce that every maximal monotone mapping on a reflexive space which is *coercive* (in the sense that $\inf_{x^*\in Tx}\langle x^*, x\rangle/\|x\| \to \infty$ with $\|x\| \to \infty$) is surjective, by considering the sequence $(T + \frac{1}{i}J)$. Hint: It helps to know that maximal monotone operators (and so their inverses) are sequentially *demi-closed*, that is $x_i \to_* x, y_i \to y, y_i \in Tx_i$ implies $y \in Tx$. This is neatly proved via the Fitzpatrick function.

In a non-reflexive space this fails badly. Indeed the existence of surjective, coercive subgradient mappings forces the space to be reflexive, [117].

Exercise 5.1.22 Show in finite dimensions that a single-valued surjective monotone mapping is *weakly coercive*, meaning that $\|Tx\| \to \infty$ when $\|x\| \to \infty$.

Exercise 5.1.23 Compute the Fitzpatrick function of T when T is a linear maximal monotone mapping.

Exercise 5.1.24 Compute the Fitzpatrick function of $T+S$ when T is maximal monotone and S is a skew bounded linear mapping.

$*$**Exercise 5.1.25** Suppose T is maximal monotone and *skew* – that is, both T and $-T$ are monotone on X. Suppose, on translating if need be that $0 \in T(0)$ and $\mathrm{dom}(T)$ is a dense absorbing set.

Show that in any Banach space, a maximal monotone skew mapping whose domain has non-empty interior extends to a bounded skew affine mapping on the whole space. Hint: Show that $T(x) \subset K(x) := \{x^* \mid F_T(x,x^*) \leq 0\}$, so that K is a convex multifunction. Now check that $K(0) = \{0\}$. Deduce that K is single valued, and therefore $T(x) = K(x)$ on $\mathrm{dom}(T)$.

Exercise 5.1.26 Supposing T is maximal monotone and skew, show that $\mathrm{dom}(T)$ is affine.

$*$**Exercise 5.1.27** Show that every C^1 monotone mapping, T, whose domain is open, can be written as $T = f' + S$ where f is a twice differentiable convex function and S is a skew and bounded linear mapping. Hint: (i) the gradient of T is a linear monotone mapping, and so can be written as $P(x) + S(x)$ where P is positive semi-definite and (ii) the skew monotone part is linear by Exercise 5.1.25.

$*$**Exercise 5.1.28** Monotone mappings such that $T+J$ is surjective are called *hypermonotone*.Show that a hypermonotone mapping on a reflexive space is maximal monotone as soon as J and J^{-1} are both injective, but not necessarily more generally. In Hilbert space this result is due to Minty [188]. Deduce that T is hypermonotone as soon as $T + \alpha J$ is surjective for some $\alpha > 0$.

Exercise 5.1.29 Prove Lemma 5.1.32.

Exercise 5.1.30 Prove Theorem 5.1.33 as follows:

(i) Associate P to a monotone function T as in Lemma 5.1.32.
(ii) Extend T to a maximal monotone multifunction $\widehat{T}$.
(iii) Define $\widehat{P}$ from $\widehat{T}$ using Lemma 5.1.32 and use Rockafellar's theorem to assert $\mathrm{dom}(\widehat{P}) = \mathrm{range}(\widehat{T} + I) = H$.
(iv) Check that $\widehat{P}$ is indeed an extension of P.

$*$**Exercise 5.1.31** Use Lemma 5.1.32 to explicitly define a convex Fitzpatrick function associated with a nonexpansive mapping, and determine its properties.

Exercise 5.1.32 Let H be a Hilbert space and let $T\colon H \to 2^H$ be a monotone multifunction. Show that $Q := (I + T^{-1})^{-1}$ is nonexpansive. Moreover, if T is maximal monotone then $\mathrm{dom}\, Q = H$. Hint: $\mathrm{dom}\, Q = \mathrm{range}(I + T^{-1})$.

Exercise 5.1.33 (Resolvents) Let H be a Hilbert space with $T\colon H \to 2^H$ a maximal monotone multifunction. For $\lambda > 0$, show that the *resolvent* $R_\lambda := (I + \lambda T)^{-1}$ is everywhere defined, with range in the domain of T and non-expansive. Deduce that the *Yosida approximate* $T_\lambda(x) := TR_\lambda$ is an everywhere defined, $(1/\lambda)$-Lipschitz and maximal monotone mapping.

Show for x in the domain of T that $T_\lambda(x)$ converges to the minimal norm member of Tx. What happens when Tx is empty?

Non-expansivity is very definitely a Hilbert space property, but the Yosida approximate remains useful more generally (as in the next exercise) [92]. Hint: Supposing $x^* \in Tx$ and $x_i^* \in T_\lambda(x_i)$ we have $\langle x_i^* - x^*, x_i\rangle \le 0$. Thus $\limsup_{i\to\infty} \|x_i\| = \inf ||Tx\|$. Now use demi-closure.

∗**Exercise 5.1.34** Show that the domain and range of a maximal monotone operator on a reflexive space are *semi-convex* – that is, have a convex closure. It is unknown whether this holds in arbitrary Banach space [235]. Hint: Since X is reflexive it suffices to show the domain is semi-closed.

Fix $y \in \mathrm{dom}(T)$, $y^* \in T(y)$, $x \in X$, and use Rockafellar's theorem to solve

$$0 \in J(x - x_i) + \frac{1}{i}\, T(x_i)$$

for integer $i > 0$. Then for some $x_i^* \in T(x_i)$ and $j_i^* \in J(x - x_i)$, we have

$$\begin{aligned}\|x_i - x\|^2 &= -\,\frac{1}{i}\,\langle x_i^*, x_i - x\rangle \le -\frac{1}{i}\,\langle y^*, x_i - y\rangle - \frac{1}{i}\,\langle x_i^*, x - y\rangle\\ &= \frac{1}{i}\,\langle y^*, x_i - y\rangle + \langle j_i^*, x - y\rangle.\end{aligned}$$

Deduce that $\|j_i^*\| = \|x_i - x\|$ stays bounded and so (j_i^*) has a weak cluster point j^*. In particular, as x_i is in $D = \mathrm{dom}(T)$, one has

$$d_D(x) \le \inf_{y\in D} \langle j^*, x - y\rangle \le \|j^*\|\, d_{\mathrm{co(D)}}(x),$$

for all x in X. Hence, $\mathrm{cl}\,\mathrm{dom}(T)$ is convex as required.

∗**Exercise 5.1.35** (Maximality of the Sum) Let T and U be maximal monotone operators on a Hilbert space, H, and let $\lambda > 0$ be given.

(i) Show that $\mathrm{range}(T_\lambda + U + \mu I) = H$, for $\mu > 1/\lambda$.
(ii) Deduce that $T_\lambda + U$ is maximal monotone.
(iii) Show that $T + U$ is maximal monotone when $\mathrm{dom}(U) \cap \mathrm{int}(\mathrm{dom}\, T) \neq \emptyset$.

Hint: (i) For any $y \in H$, the mapping

$$x \mapsto (S + \mu I)^{-1}[y - T_\lambda(x)]$$

is a Banach contraction. (iii) We may suppose $0 \in T(0) \cap U(0)$ and that 0 is interior to the domain of T. Let $\lambda_i \downarrow 0$. Note that $0 \in T_{\lambda_i}(0)$. Show that the solutions $t_i \in T_{\lambda_i}(x_i), u_i \in U(x_i)$ with $y = t_i + u_i + x_i$ yield a Cauchy sequence (x_i) as follows:

$$\langle x_i - x_j, x_i - x_j \rangle \le -\langle t_i - t_j, \lambda_i\, t_i - \lambda_j\, t_j \rangle \le 2\,(\lambda_i + \lambda_j) \sup \|t_k\|^2.$$

Use monotonicity and the fact that the domains intersect to show $\|x_i\| \le \|y\|$. Now use the interiority hypothesis and the consequent local boundedness at 0 of the monotone operator T to show (t_i) remains bounded and also has a weakly convergent subsequence. Conclude that (x_i) converges in norm.

Finish by taking limits and using demi-closedness.

Note that everything has been reduced to Rockafellar's theorem and so to the Hahn–Banach theorem. An extension of this proof will work in arbitrary reflexive space, but step (i) must be replaced by a finite dimensional approximation argument.

Exercise 5.1.36 Show that for a closed convex set C in a Banach space and $\lambda > 0$ one has

$$(\partial \iota_C)_{2\lambda} = \partial \iota_C \,\square\, \lambda \,\|\cdot\|^2 = \lambda\, d_C^2(x).$$

$*$**Exercise 5.1.37** (Monotone Variational Inequalities) Let T be a maximal monotone operator on a Banach space and let C be a closed convex subset of X.

(i) Show that the solution of the monotone variational inequality:

$$\mathrm{VI}\,(T, C) \qquad \begin{cases} \text{there exist } x \in C \text{ and } t^* \in T(x) \\ \text{such that } \langle t^*, c - x \rangle \ge 0 \text{ for all } c \in C \end{cases}$$

is equivalent to the monotone inclusion

$$0 \in (T + \partial \iota_C)(x).$$

(ii) In particular, if T is coercive on C and the sum $T + \partial \iota_C$ is maximal monotone for which Exercise 5.1.35 gives conditions, then the variational inequality has a solution.

(iii) Specialize this to the cases when T is coercive and (a) $C = i\,B_X$, as $i \to \infty$, or (b) C is a closed convex cone – a so-called complementarity problem.

(iv) Consider two monotone operators T and U on X and Y respectively. Show that $M(x, y) := (Tx, Uy)$ is monotone on $X \times Y$ and is maximal if and only if both T and U are. Denote the diagonal convex set by $\Delta := \{(x, y) \in X \times Y \mid x = y\}$. Check that $0 \in \mathrm{range}(T + U)$ if and only if $\mathrm{VI}\,(M, \Delta)$ has solution.

$*$**Exercise 5.1.38** (Transversality I) Let T be maximal monotone operator on a Hilbert space, H, and let C be a non-empty closed convex subset of H.

(i) Show that when T is coercive on C the condition

$$0 \in \operatorname{core}[\operatorname{dom}(T) - C] \tag{5.1.21}$$

implies VI (T, C) has a solution.
(ii) This remains true in a reflexive Banach space.

Hint: By Exercise 5.1.35, VI $(T_{1/i}, C)$ has a solution:

$$x_i \in C,\, t_i \in T(x_i - \frac{1}{i}\, t_i), \qquad \inf_{c \in C} \langle t_i, c - x \rangle \geq 0.$$

Condition (5.1.21) and the Baire category theorem imply that for some $N > 0$ one has $0 \in \operatorname{cl}[T^{-1}(NB_H) - C \cap NB_H]$. This and coercivity of T suffice to show, much as in Exercise 5.1.35, that (x_i) and (t_i) remain bounded as i goes to infinity. Thence, (x_i) is norm convergent and one may to move to the limit.

$*$**Exercise 5.1.39** (Transversality II) Let T and U be maximal monotone operators on a Hilbert space.

(i) Use Exercises 5.1.37 and 5.1.38 to show that

$$0 \in \operatorname{core}[\operatorname{dom}(T) - \operatorname{dom}(U)]$$

implies $T + U$ is maximal monotone.
(ii) This remains true in a reflexive Banach space.

7

Variational Techniques In the Presence of Symmetry

Symmetry is exploited in many physical and geometrical applications. The focus of this chapter is what happens when we apply variational methods to functions with additional symmetry. The mathematical characterization of symmetry is invariance under certain group actions. Typical examples are the spectral functions associated with a linear transformation, such as the maximum eigenvalue for a matrix. They are in general invariant with respect to the similarity transform. Another example is the distance function on a Riemannian manifold, which is invariant with respect to an isometric transform. It turns out that nonsmooth functions on smooth manifolds provide a convenient mathematical framework for such problems.

7.1 Nonsmooth Functions on Smooth Manifolds

7.1.1 Smooth Manifolds and Submanifolds

We start with a brief review of the smooth manifolds and related notation. In what follows k is either a nonnegative integer or ∞. Let Y be an N-dimensional C^k complex manifold with a C^k atlas $\{(U_a, \psi_a)\}_{a\in A}$. For each a, the N components $(x_a^1, \dots, x_a^N)$ of ψ_a form a *local coordinate system* on (U_a, ψ_a). A function $g\colon Y \to \mathbb{R}$ is C^k at $y \in Y$ if $y \in U_a$ and $g \circ \psi_a^{-1}$ is a C^k function in a neighborhood of $\psi_a(y)$. As usual C^0 represents the collection of continuous functions. It is well known that this definition is independent of the coordinate systems. If g is C^k at all $y \in Y$, we say g is C^k on Y. The collection of all C^k functions on Y is denoted by $C^k(Y)$. A map $v\colon C^1(Y) \to \mathbb{R}$ is called a *tangent vector* of Y at y provided that for any $f, g \in C^1(Y)$,

(i) $v(af + bg) = av(f) + bv(g)$ for all $a, b \in \mathbb{R}$ and
(ii) $v(f \cdot g) = v(f)g(y) + f(y)v(g)$.

The collection of all the tangent vectors of Y at y forms an (N-dimensional) vector space, called the *tangent space* of Y at y and denoted by $T_y(Y)$ or

Index