DUALITY INEQUALITIES IN NONSMOOTH OPTIMIZATION

by

Herre Wiersma
B.Sc., Acadia University 1999

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
in the Department
of
Mathematics

SIMON FRASER UNIVERSITY
November 2002

APPROVAL

Name: Herre Wiersma

Degree: Master of Science

Title of thesis: Duality Inequalities in Nonsmooth Optimization

Examining Committee: Dr. Alistair Lachlan
Chair

Dr. Jonathan Borwein, Senior Supervisor

Dr. Rustum Choksi

Dr. Petr Lisonek

Dr. Adrian Lewis

Date Approved: December 3, 2002

PARTIAL COPYRIGHT LICENSE

Title of Thesis/Project/Extended Essay

Duality Inequalities in Nonsmooth Optimization

Author: ______________________
(signature)

Herre Wiersma
(name)

Dec. 4, 2002
(date)

Abstract

Duality inequalities are pervasive in modern optimization; Fenchel duality and the Mean Value theorem are two prominent examples. This thesis surveys some recent duality results pertaining to nonsmooth functions, and examines some interesting corollaries thereof. One of these results is a somewhat surprising nonsmooth generalization of both the classical Mean Value theorem and the standard Fenchel duality theorem. Another gives rise to a variety of nonsmooth analogs to Rolle's theorem. Fixed point theory is central to the development of these results, and it is interesting to ask whether variational proofs might exist for some duality results. The answer to this question is mixed: some results admit variational proofs, whereas for others such a proof is unlikely. In particular, we show by counterexample that a certain Rolle-type duality theorem does not hold, even in $\mathbb{R}^2$.

Acknowledgments

I would like to thank my supervisor, Dr. Jonathan Borwein, for introducing me to the field of nonsmooth analysis and optimization, and for the interesting experiences and opportunities I have had at the CECM. I would also like to thank my friends at SFU for helping to make my stay here a pleasant one, both academically and otherwise. Finally, I would like to thank my family and my girlfriend, Laurie, for helping me through a difficult year.

I am indebted as well to the Natural Sciences and Engineering Research Council for its generous funding.

Contents

List of Figures

Chapter 1

Introduction and Preliminaries

A *duality inequality*, broadly speaking, is an inequality that expresses a relationship between elements in a linear space X and elements of the topological dual space X^* of all continuous real linear functions on X. For instance, we may consider the derivative $f'(x_0)$ of a function $f : \mathbb{R} \to \mathbb{R}$ at x_0 to be a linear functional that, when translated, approximates f near x_0. Thus one central duality equality in analysis is the classical mean value theorem (see, e.g. [Wad95]):

Theorem 1.1 (Classical Mean Value Theorem) *Let f be continuous on a closed bounded nondegenerate interval $[a, b]$ and differentiable on (a, b). Then*

$$f(b) - f(a) = \langle f'(x_0), b - a\rangle$$

for some $x_0 \in (a, b)$.

In the above, $\langle x^*, x\rangle$ for $x \in X$ and $x^* \in X^*$ denotes $x^*(x)$. The mean value theorem, of course, extends in a straightforward way to multiple dimensions:

Theorem 1.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable on a neighborhood of $[a, b] = \{\lambda a + (1 - \lambda)b \mid \lambda \in [0, 1]\}$. Then*

$$f(b) - f(a) = \langle \nabla f(x_0), b - a\rangle$$

for some $x_0 \in (a, b) = \lambda a + (1 - \lambda)b \mid \lambda \in (0, 1)\}$.

Another example of a duality inequality comes from convex analysis (see [Roc97], [BL00], [HUL93a], [HUL93b]) in the form of the classical Fenchel duality theorem. Given functions

f and h on X, we consider the problem

$$p = \inf_{x \in X} \{f(x) + h(x)\}.$$

The *dual problem* to finding p is given by

$$d = \sup_{x^* \in X^*} \{-f^*(x^*) - h^*(-x^*)\}$$

where f^* and h^* are certain convex functions on X^*. Weak Fenchel duality asserts that $p \geq d$, and strong Fenchel duality says $p = d$ provided f and h satisfy certain constraints.

In this thesis, we examine several recent duality inequalities from the field of nonlinear and nonsmooth analysis. In Chapter 2 we examine a powerful nonsmooth duality result developed in [CL94], [LR96] and [BF01] that extends both Fenchel duality and the mean value theorem. In Chapter 3, we consider a related result by Borwein and Fitzpatrick that leads to a variety of nonsmooth Rolle-type inequalities. As in [BF01], we then use Ekeland's variational principle to improve certain results. We conclude with Chapter 4, in which a counterexample (see [BKW02]) is constructed to a conjectured Rolle-type duality inequality.

The central results, Theorem 2.1 and Theorem 3.3 draw on many analytical and topological results in their proofs. The rest of this chapter reviews the necessary background material for chapters 2 and 3.

1.1 Notation

The natural setting for many of the results we use is Banach space, that is, a complete normed space. If a proposition is inherently a Banach space fact, we prove it or cite it in that setting, even if we only apply it to $\mathbb{R}^n$. We use X for a general Banach space, and X^* for its dual. Furthermore, elements of the primal space are represented with lower case letters, as in "$x \in X$", and elements of the dual are represented as starred lower case letters, as in "$x^* \in X^*$". We often use this notation even when $X = \mathbb{R}^n$, to stress the role of subgradients and gradients, e.g., as members of the dual space. The norm on X is denoted by $\|\cdot\|$ and the closed unit ball by B. The dual norm and unit ball are denoted by $\|\cdot\|_*$ and B^*, repectively. [FHH$^+$01] is a good reference for Banach space theory.

A function $X \to \overline{\mathbb{R}}$ is *convex* if

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

for any $x, y \in X$ and $\lambda \in [0,1]$. We allow functions to take the extended real values $+\infty$ and $-\infty$, and denote $\mathbb{R} \cup \{+\infty\}$ by $\overline{\mathbb{R}}$. By the *epigraph*, $\mathrm{epi}(f)$, of a real-valued function f, we mean those points in $X \times \mathbb{R}$ that are above the graph of f:

$$\mathrm{epi}(f) := \{(x,t) \in X \times R \mid f(x) \leq t\}.$$

Likewise, the *hypograph* of f are the points below the graph:

$$\mathrm{hyp}(g) := \{(x,t) \in X \times R \mid f(x) \geq t\}.$$

f is said to be *closed* if $\mathrm{epi}(f)$ is a closed set. This is equivalent to f being *lower semicontinuous* on X:

$$\liminf_{n\to\infty} f(x_n) \geq f(x)$$

for every sequence $\{x_n\}$ that converges to x, for every $x \in X$. The *domain* of a function f, $\mathrm{dom}\, f$, is the set of points for which f is finite; we say a function is *proper* if $\mathrm{dom}\, f \neq \emptyset$ and f never takes the value $-\infty$. A point $x \in X$ is a *local minimum* for f if there exists a neighborhood U of x such that

$$f(x) \leq f(x') \text{ for all } x' \in U.$$

The interior of a set S we denote by $\mathrm{int}\, S$, the closure by $\mathrm{cl}\, S$, and the boundary by ∂S. The *convex hull* of S, $\mathrm{conv}\, S$, is the smallest convex set containing S, and the closed convex hull, $\overline{\mathrm{conv}} S$, is the closure of $\mathrm{conv}\, S$. For sets S_1 and S_2, the closed interval $[S_1, S_2]$ is defined to be $\overline{\mathrm{conv}}(S_1 \cup S_2)$. A point x is in the *core* of S, $\mathrm{core}\, S$, if for every direction $d \in X$ there is an $\varepsilon > 0$ such that $[x, x+\varepsilon d] \subset S$. Clearly, $\mathrm{int}\, S \subset \mathrm{core}\, S$. Finally, by $S_1 + S_2$ we mean the set of all sums of an element in S_1 and an element in S_2.

1.2 Convex and Nonsmooth Analysis Facts

A real-valued function $f : X \to \mathbb{R}$ is *Gâteaux differentiable* at a point $x \in X$ if there is a *gradient* $\nabla f(x) \in X^*$ such that

$$\lim_{t \searrow 0} \frac{f(x+td) - f(x)}{t} = \langle \nabla f(x), d \rangle$$

for all $d \in X$. If this limit exists uniformly for d in bounded sets, then f is *Fréchet differentiable* at x. We call the limit on the left hand side of this equation the *directional* *derivative* of x in the direction d, and we denote it by $f'(x;d)$. We say that f is *continuously differentiable* or C^1 if the gradient mapping $x \to \nabla f(x)$ is a continuous mapping from X to X^*. These notions of differentiability can also be applied to functions $f : X \to Y$, where Y is an arbitrary Banach space; in this case, the gradient becomes the derivative $f'(x)$ at x, where $f'(x)$ is a linear transformation from X to Y.

The classical derivative imposes a strong restriction on functions at points of differentiability: there must be a unique tangent functional at such points. This condition may fail for otherwise well-behaved functions, e.g. $f(x) = |x|$ for $x = 0$. The natural solution to this problem is to allow a set-valued derivative, or *subdifferential* $\partial f(x) : X \to 2^{X^*}$, whose elements, called *subgradients*, capture the local variational behaviour of f. For example, one possible definition of $\partial f(x)$ for $f = |\cdot|$ might be

$$\partial f(x) = \begin{cases} -1, & x < 0 \\ [-1,1], & x = 0 \\ 1, & x > 0 \end{cases}$$

In fact, this is the convex subdifferential of $f = |\cdot|$, which we now define.

1.2.1 The Convex Subdifferential

Let f be a convex real-valued function on a Banach space X. Then the *convex subdifferential* is defined by

$$\partial f(\overline{x}) := \{x^* \in X^* \quad | \quad \langle x^*, x - \overline{x}\rangle + f(\overline{x}) \leq f(x) \quad \forall x \in X\}$$

That is, the subdifferential is the set of slopes of affine functions that meet f at $\overline{x}$ and minorize f everywhere (see Figure 1.1). f is *subdifferentiable* at x if $\partial f(x)$ is not empty, i.e. there exists a subgradient at x. The set of points of subdifferentiability of f is called the *domain* of ∂f, $\operatorname{dom}\partial f$. If f is convex and x is in the interior of $\operatorname{dom} f$, then $\operatorname{dom}\partial f(x)$ is nonempty, so $\operatorname{int}\operatorname{dom} f \subset \operatorname{dom}\partial f$ (see [BL00] or [Roc97]). The subdifferential has an intuitive characterization in terms of the directional derivative:

Proposition 1.3 *Let f be convex, and let $\overline{x} \in \operatorname{dom}(f)$. Then $x^* \in \partial f(\overline{x})$ if and only if $\langle x^*, d\rangle \leq f'(\overline{x}; d)$ for all $d \in X$.*

Suppose f is convex and Gâteaux differentiable at x. Then for any $x^* \in \partial f(x)$ and direction d we have

$$\langle x^*, d\rangle \leq f'(x;d) = \langle \nabla f(x), d\rangle$$

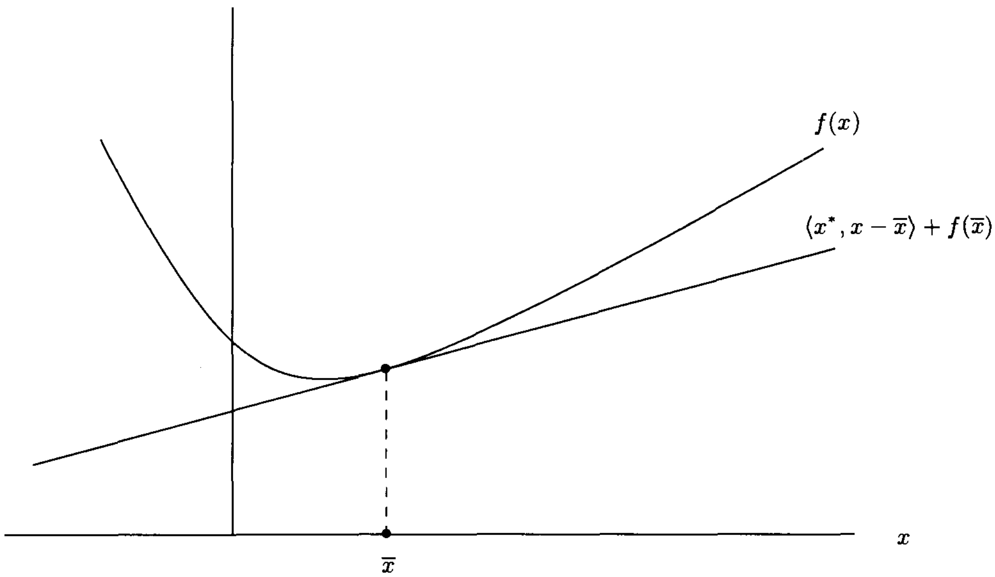

Figure 1.1: A convex subgradient, x^*

and

$$\langle x^*, -d\rangle \le f'(x; -d) = \langle \nabla f(x), -d\rangle$$

so the convex subdifferential reduces to the gradient in this case:

$$\partial f(x) = \{\nabla f(x)\}.$$

On the other hand, if $x \in \operatorname{core}\operatorname{dom} f$ and $\partial f(x) = \{x^*\}$ then $x^* = \nabla f(x)$, by Corollary 3.1.10 in [BL00].

We will primarily be interested in the convex subdifferential because of its interaction with the Fenchel conjugate f^*, discussed in section 1.3. It is the natural subdifferential to consider for convex functions, but if f is not convex, then ∂f is often not a useful object—it can be empty everywhere even for very well behaved functions, like $f = -\|\cdot\|^2$ on a normed space X. The alternate definition of the subdifferential supplied by Proposition 1.3 allows us to generalize the subdifferential by generalizing the definition of the directional derivative. Among the many choices of directional derivatives (see, e.g., [BL00]), we consider the Clarke directional derivative.

1.2.2 The Clarke Subdifferential

A function $f : X \to \mathbb{R}$ is said to be *Lipschitz* with constant K on a set $Y \subset X$ if there exists a $K \geq 0$ such that

$$|f(y) - f(y')| \leq K\|y - y'\|$$

for every $y, y' \in Y$. If f is Lipschitz on a neighborhood of $x \in X$, then we say that f is *locally Lipschitz around* x. f is locally Lipschitz on Y if it is locally Lipschitz at every point of Y. Locally Lipschitz functions are a broad class of functions, and a natural class to consider in analysis and optimization. For instance, a convex function on $\mathbb{R}^n$ is locally Lipschitz on the interior of its domain (see [BL00]).

The *Clarke directional derivative* $f^\circ(x; d)$ of a function $f : X \to \overline{\mathbb{R}}$ at a point x in the direction d is defined by

$$f^\circ(x; d) := \limsup_{y \to x, t \searrow 0} \frac{f(y + td) - f(y)}{t}.$$

Following Proposition 1.3, we define the *Clarke subdifferential*, ∂f, of f at x by

$$\partial f(x) := \{x^* \in X^* \mid \langle x^*, d\rangle \leq f^\circ(x; d) \text{ for all } d \in X\}.$$

Proposition 1.4 *Let $f : X \to \mathbb{R}$ be locally Lipschitz with constant K near x. Then*

(a) The function $d \to f^\circ(x; d)$ is finite and sublinear, *that is*

$$f^\circ(x; \alpha d_1 + \beta d_2) \leq \alpha f^\circ(x; d_1) + \beta f^\circ(d_2)$$

for all $\alpha, \beta \geq 0$ and $d_1, d_2 \in X$. Furthermore, $f^\circ(x; \cdot)$ is bounded by a multiple of the norm:

$$|f^\circ(x; d)| \leq K\|d\|.$$

(b) $f^\circ(x; -d) = (-f)^\circ(x; d)$.

(c) ∂f is closed (in fact w^*-closed) convex and nonempty, and $\partial f \subset KB^*$.

Proof: (a) If y is near x and $t > 0$ is small, then $y + td$ is near x, so

$$\left|\frac{f(y + td) - f(y)}{t}\right| \leq K\frac{\|y + td - y\|}{t} = K\|d\|.$$

Thus it follows that $|f^\circ(x; d)| \leq K\|d\|$.

To show sublinearity, we must prove that the function is subadditive and positively homogeneous. Positive homogeneity is easy: take $s > 0$ and note

$$\begin{aligned}
f^\circ(x; sd) &= \limsup_{y\to x, t\searrow 0} \frac{f(y+tsd)-f(y)}{t} \\
&= \limsup_{y\to x, t'\searrow 0} \frac{f(y+t'd)-f(y)}{(t'/s)} \\
&= s \limsup_{y\to x, t'\searrow 0} \frac{f(y+t'd)-f(y)}{t'} \\
&= sf^\circ(x; d).
\end{aligned}$$

For any sequence $x_n \to x$, $t_n \searrow 0$ and $\varepsilon > 0$, we have

$$\frac{f(x_n + t_n(d_1+d_2)) - f(x_n + t_n d_1)}{t_n} \le f^\circ(x; d_2) + \varepsilon$$

and

$$\frac{f(x_n + t_n d_1) - f(x_n)}{t_n} \le f^\circ(x; d_1) + \varepsilon$$

for all large n. Adding these two, and letting $n \to \infty$ we have

$$f^\circ(x; d_1 + d_2) \le f^\circ(x; d_1) + f^\circ(x; d_2) + 2\varepsilon.$$

Since ε was arbitrary, the result follows.

(b)

$$\begin{aligned}
f^\circ(x; -d) &= \limsup_{y\to x, t\searrow 0} \frac{f(y-td)-f(y)}{t} \\
&= \limsup_{u\to x, t\searrow 0} \frac{f(u+td-td)-f(u+td)}{t} \\
&= \limsup_{u\to x, t\searrow 0} \frac{(-f)(u+td)-(-f)(u)}{t} = (-f)^\circ(x; d)
\end{aligned}$$

(c) Since $f^\circ(x; \cdot)$ is finite and sublinear, there exists a linear functional $x^* \in X^*$ that minorizes it everywhere by the Hahn-Banach theorem; see, e.g., Theorem 4.14.5 in [FK70]. But then $x^* \in \partial f(x)$.

Closure and convexity are immediate, since ∂f is the intersection of w^*-closed halfspaces H_d defined by

$$H_d = \{x^* \in X^* \mid \langle x^*, d\rangle \le f^\circ(x; d)\}.$$

Finally, $x^* \in \partial f(x)$ implies

$$\langle x^*, d\rangle \le f^\circ(x; d) \le K\|d\|$$

for all d, so

$$\|x^*\| = \sup_{d\neq 0} \frac{\langle x^*, d\rangle}{\|d\|} \le K.$$

■

Since ∂f is bounded and w^*-closed when f is locally Lipschitz, ∂f is w^*-compact, by Alaoglu's theorem (see, for example, [FHH+01]). In particular, ∂f is norm compact if X is finite dimensional.

The following result shows that there is no real ambiguity to using ∂f to mean both the convex subdifferential and the Clarke subdifferential. The second part shows that the Clarke subdifferential is also compatible with differentiability :

Proposition 1.5 **([Cla83])** *Let $f : X \to \mathbb{R}$ be locally Lipschitz at $x \in X$.*

1. *If f is convex on an open set $U \subset X$ with $x \in U$, then the convex and Clarke subdifferentials coincide at x, and*

$$f^\circ(x; d) = f'(x; d) \text{ for all } d \in X. \tag{1.1}$$

2. *Suppose X is finite dimensional, and let Ω_f be the points where f is not differentiable. Then for any set S of Lebesgue measure 0 the following holds:*

$$\partial f(x) = \overline{\text{conv}}\{\lim \nabla f(x_n) \mid x_n \to x, x \notin \Omega_f \cup S\}.$$

In particular, if f is C^1 on a neighborhood of x, then

$$\partial f(x) = \{\nabla f(x)\}.$$

If f satisfies equation (1.1), then f is said to be *regular* at x.

The Clarke subdifferential obeys a number of useful calculus rules that we will need later. The proofs of these rules may be found in [Cla83]. For the Cartesian product rule below, we need to introduce the notion of a *partial Clarke subdifferential*. Given a point $(x, y) \in X \times Y$ and a real function G on $X \times Y$, we define $\partial_1 G(x, y)$ to be the subdifferential of the function $G(\cdot, y)$ at x. $\partial_2 G(x, y)$ is defined similarly.

Proposition 1.6 ([Cla83]) *Let $f, g : X \to \mathbb{R}$ be locally Lipschitz. Then the following calculus rules hold:*

> ***Scalar Multiplication*** $\partial(sf) = s\partial f$ *for s real.*
>
> ***Fermat's Rule*** *If f attains a local minimum at x, then $0 \in \partial f(x)$.*
>
> ***Sum Rule*** $\partial(f+g) \subset \partial f + \partial g$.
>
> ***Cartesian Product Rule*** *If G is regular at (x,y), then $\partial G(x,y) \subset \partial_1 G(x,y) \times \partial_2 G(x,y)$*
>
> ***Upper Semicontinuity*** *If X is finite dimensional, then for every $\varepsilon > 0$ there is a $\delta > 0$ such that*
> $$\partial f(x + \delta B) \subset \partial f(x) + \varepsilon B^*.$$

Note that upper semicontinuity implies the following sequential closedness property:

$$\left.\begin{array}{l} x_n \to x \\ x_n^* \to x^* \\ x_n^* \in \partial f(x_n) \end{array}\right\} \Longrightarrow x^* \in \partial f(x) \ .$$

In fact, if $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz at x then $\partial f(x)$ is bounded by a multiple of the ball, so the previous implication is equivalent to upper semicontinuity in this case.

1.3 Fenchel Conjugacy

Given a function $f : X \to \mathbb{R}$, we define the *Fenchel conjugate* $f^* : X^* \to \mathbb{R}$ of f by

$$f^*(x^*) = \sup_{x \in X}\{\langle x^*, x\rangle - f(x)\}.$$

Similarly, we define $f^{**} : X \to \mathbb{R}$ by

$$f^*(x) = \sup_{x^* \in X^*}\{\langle x, x^*\rangle - f^*(x^*)\},$$

so that $f^{**} = (f^*)^*$ for reflexive spaces. It is easily seen that f^* is convex and lower semicontinuous, as it is the supremum of a collection of affine functions. Furthermore, f^{**} is (pointwise) the greatest convex lower semicontinuous function that minorizes f. In particular, $f^{**} = f$ when f is convex and lower semicontinuous. f^* is also related to f and the convex subdifferential ∂f via Fenchel's inequality:

Theorem 1.7 (Fenchel's Inequality) *If $x \in X$ is in the domain of a function $f : X \to \mathbb{R}$, the following inequality holds for all $x^* \in X^*$:*

$$\langle x^*, x\rangle \leq f(x) + f^*(x^*).$$

Furthermore, the preceding holds with equality if and only if

$$x^* \in \partial f(x).$$

Proof: The inequality is immediate:

$$\begin{aligned} f^*(x^*) &= \sup_{x \in X}\{\langle x^*, x\rangle - f(x)\} \\ &\geq \langle x^*, x\rangle - f(x). \end{aligned}$$

Now $x^* \in \partial f(x)$ holds if and only if

$$\langle x^*, y - x\rangle + f(x) \leq f(y)$$

or

$$\langle x^*, y\rangle - f(y) + f(x) \leq \langle x^*, x\rangle$$

for all $y \in X$. Taking the supremum over all y, this is equivalent to

$$f^*(x^*) + f(x) \leq \langle x^*, x\rangle,$$

which gives the result. ■

Another important property of the Fenchel conjugate can be easily obtained from Fenchel's equation: for proper closed convex functions, the subdifferential of f^* is the inverse of the subdifferential of f.

Proposition 1.8 *For $f : X \to \overline{\mathbb{R}}$,*

$$x^* \in \partial f(x) \implies x \in \partial f^*(x^*). \tag{1.2}$$

Furthermore, the converse is true if f is convex and closed.

Proof: If $x^* \in \partial f(x)$ then by Theorem 1.7

$$f^*(x^*) + f(x) = \langle x^*, x\rangle,$$

so

$$f^*(x^*) + \langle y^* - x^*, x\rangle = \langle y^*, x\rangle - f(x)$$

for all $y^* \in X^*$. Taking the supremum over all x on the right hand side gives

$$f^*(x^*) + \langle y^* - x^*, x\rangle \leq f^*(y^*)$$

for all $y^* \in X^*$, so $x \in \partial f^*(x^*)$. If f is closed and convex, then $x \in \partial f^*(x^*)$ implies

$$\begin{aligned}\langle x, x^*\rangle &= f^*(x^*) + f^{**}(x)\\ &= f^*(x^*) + f(x),\end{aligned}$$

so $x^* \in \partial f(x)$. ■

The centerpiece of Fenchel conjugacy theory is *Fenchel duality.* As alluded to in the introduction, for real functions f and h we pair the *primal problem,*

$$p := \inf_{x\in X}\{f(x) + h(x)\},$$

with the following *dual problem*

$$d := \sup_{x^*\in X^*}\{-f^*(x^*) - h^*(-x^*)\}.$$

Fenchel duality asserts that these two problems are equivalent, provided a certain a constraint qualification holds:

Theorem 1.9 (Fenchel Duality [BL00]) *Let f and h be real functions on a Banach space X, with values in $\mathbb{R}\cup\{+\infty\}$. Then $p \geq d$.*

Furthermore, if f and h are convex and

$$0 \in \operatorname{core}(\operatorname{dom} h - \operatorname{dom} f) \tag{1.3}$$

then $p = d$ and the dual value is attained when finite.

Proof:

$$-f^*(x^*) - h^*(-x^*) \leq f(x) - \langle x^*, x\rangle + h(x) + \langle x^*, x\rangle = f(x) + h(x)$$

for all $x \in X$ and $x^* \in X^*$, so taking the supremum over x^* and the infimum over x gives $p \geq d$.

Suppose f and h are convex and that (1.3) holds. Define the *value function* $v : X \to \mathbb{R}$ by

$$v(u) := \inf_{x\in X}\big\{f(x) + h(x+u)\big\}.$$

v is convex, for if $r, s \in X$ and $\lambda \in [0,1]$ then

$$\begin{aligned}
v(\lambda r + (1-\lambda)s) &= \inf_{x\in X}\big\{f(x) + h(x + \lambda r + (1-\lambda)s)\big\} \\
&= \inf_{x,y\in X}\big\{f(\lambda x + (1-\lambda)y) + h(\lambda(x+r) + (1-\lambda)(y+r))\big\} \\
&\leq \inf_{x,y\in X}\big\{\lambda(f(x) + h(x+r)) + (1-\lambda)(f(y) + h(y+s))\big\} \\
&= \lambda v(r) + (1-\lambda)v(s)
\end{aligned}$$

Now we claim that $\operatorname{dom} v = \operatorname{dom} h - \operatorname{dom} f$. To see this, suppose $u \in \operatorname{dom} v$. This is true if and only if there exists $x \in X$ such that $f(x) + h(x+u) < \infty$. But this is equivalent to $x \in \operatorname{dom} f$ and $w = x + u \in \operatorname{dom} g$. That is, $u = w - x$ for some $w \in \operatorname{dom} h$ and $x \in \operatorname{dom} f$. If $v(0) = p$ is infinite, there is nothing to prove; if it is finite then v is proper (see [BL00], Lemma 3.2.6) and the constraint qualification (1.3) implies that v has a subgradient $-x^*$ at 0:

$$\begin{aligned}
v(0) &\leq v(u) + \langle -x^*, u\rangle \\
&\leq f(x) + h(x+u) - \langle x^*, u\rangle \\
&\leq \big[f(x) - \langle x^*, x\rangle\big] + \big[h(x+u) - \langle -x^*, x+u\rangle\big]
\end{aligned}$$

for all $x, u \in X$. Taking the infimum over x and then over u, we get

$$p = v(0) \leq -f^*(x^*) - h^*(-x^*) \leq d \leq p$$

so $p = d$ and the dual attains its supremum at x^*. ■

1.4 Essential Smoothness and Essentially Strict Convexity

We say a real-valued function f on $\mathbb{R}^n$ is *essentially smooth* if ∂f is single-valued when it is non-empty; that is, f it is Gâteaux differentiable on $\operatorname{dom}\partial f$. A function f on $\mathbb{R}^n$ is said to be *essentially strictly convex* if it is strictly convex on any convex subset of ∂f. There is a striking relationship between these two concepts:

Theorem 1.10 *Let f be a proper closed convex function on $\mathbb{R}^n$. Then f is essentially smooth if and only if f^* is essentially strictly convex, and f is essentially strictly convex if and only if f^* is essentially smooth.*

Proof: Let f be Gâteaux differentiable at a point $x \in \mathbb{R}^n$, so that $\partial f(x) = \{x^*\}$, where $x^* = \nabla f(x)$. First we wish to show that

$$f^*(z^*) > f^*(x^*) + \langle x, z^* - x^* \rangle \tag{1.4}$$

for all points z^* in $\mathbb{R}^n$ distinct from x^*. Since $x^* \in \partial f(x)$, we have $x \in \partial f^*(x^*)$, so that

$$f^*(z^*) \geq f^*(x^*) + \langle x, z^* - x^* \rangle$$

by the subgradient inequality. Now suppose

$$f^*(z^*) = f^*(x^*) + \langle x, z^* - x^* \rangle$$

for some z^*. Then

$$\begin{aligned}
\langle z^*, x' \rangle - f(x') &\leq f^*(z^*) = f^*(x^*) + \langle x, z^* - x^* \rangle \\
\langle z^*, x' \rangle - f(x') &\leq -f(x) + \langle x^*, x \rangle + \langle x, z^* - x^* \rangle \\
\langle z^*, x' \rangle - f(x') &\leq -f(x) + \langle z^*, x \rangle \\
\langle z^*, x' - x \rangle + f(x) &\leq f(x')
\end{aligned}$$

for every $x' \in \mathbb{R}^n$. But then $z^* \in \partial f(x) = \{x^*\}$, i.e. $z^* = x^*$, contrary to assumption. So we have proven 1.4.

We now prove that f^* is essentially strictly convex. Let $C \subset \operatorname{dom}(\partial f^*)$ be convex and choose distinct x^*, y^* in C and $\lambda \in (0,1)$. Let $w^* = \lambda x^* + (1-\lambda) y^*$, and note that

$w^* \in \operatorname{dom} \partial f^*$, so there exists $z \in \partial f^*(w^*)$. Since f is essentially smooth, $w^* = \nabla f(z)$, and we may apply (1.4) to the points x^* and y^*, which are distinct from w^*:

$$\lambda f^*(x^*) > \lambda(f^*(w^*) + \langle z, x^* - w^*\rangle)$$
$$(1-\lambda)f^*(y^*) > (1-\lambda)(f^*(w^*) + \langle z, y^* - w^*\rangle).$$

Adding these two equations yields

$$\lambda f^*(x^*) + (1-\lambda)f^*(y^*) > f^*(w^*),$$

so f^* is essentially strictly convex.

Conversely, if f is not essentially smooth then there exists a point x in $\mathbb{R}^n$ such that there are $y^*, z^* \in \partial f(x)$ with $y^* \neq z^*$. But then $x \in \partial f^*(y^*) \cap \partial f^*(z^*)$. The subdifferential inequality gives

$$\langle x, u^* - y^*\rangle \leq f^*(u^*) - f^*(y^*)$$
$$\langle x, v^* - z^*\rangle \leq f^*(v^*) - f^*(z^*)$$

for all $u^*, v^* \in E^*$. Substituting $u^* = z^*$ and $v^* = y^*$ gives

$$\langle x, z^* - y^*\rangle = f^*(z^*) - f^*(y^*).$$

For $\lambda \in [0,1]$, we have

$$\begin{aligned}\lambda f^*(y^*) + (1-\lambda)f^*(z^*) &= \lambda f^*(y^*) + (1-\lambda)(\langle x, z^* - y^*\rangle + f^*(y^*)) \\ &= f^*(y^*) + \langle x, ((1-\lambda)z^* + \lambda y^*) - y^*\rangle \leq f^*(\lambda y^* + (1-\lambda)z^*).\end{aligned}$$

Thus f^* is not essentially strictly convex.

Finally, f^* is essentially smooth if and only if f^{**} is essentially strictly convex, and $f^{**} = f$, since f is closed and convex. ■

The concepts of essential smoothness and essentially strict convexity can be extended to general Banach spaces by adding a local boundedness condition to ∂f and $(\partial f)^{-1}$; for a thorough discussion in of these concepts in a general setting, see [BBC01]. We now use the preceding result to approximate a convex function f on a bounded domain with a function whose conjugate is continuously differentiable.

1.5 Smoothing Devices

Our technique in the Chapters 2 and 3 will be to prove results first for smooth functions, and then use a limiting argument to prove their nonsmooth analogs by smooth approximation. In particular, we use Proposition 1.11 to smoothly approximate the conjugate of a proper closed convex function, and Proposition 1.12 to smoothly approximate a Lipschitz function.

Proposition 1.11 *Let $f : \mathbb{R}^n \to [-\infty, \infty]$ be a proper convex lower semicontinuous function with* $\mathrm{dom}(f) \subset \mathbb{R}^n$ *bounded, and let $f_\varepsilon := f + \varepsilon\|\cdot\|^2$, $\varepsilon > 0$. Then f_ε^* is continuously differentiable on $\mathbb{R}^n$.*

Proof: Since f is convex and $\|\cdot\|^2$ is strictly convex, f_ε is strictly convex. By Theorem 1.10, then, f_ε^* is essentially smooth. We need only show that $\mathrm{dom}\, f_\varepsilon^* = \mathbb{R}^n$, since $\mathrm{dom}\, \partial f_\varepsilon^* \supset \mathrm{int}\,\mathrm{dom}\, f_\varepsilon^*$. To this end, take $x^* \in \mathbb{R}^n$, and let $M = \sup_{x \in \mathrm{dom} f} \|x\|$. Then

$$\begin{aligned} f_\varepsilon^*(x^*) &= \sup_{x \in \mathrm{dom} f} \{\langle x^*, x\rangle - f(x) - \varepsilon\|x\|^2\} \\ &\leq \sup_{x \in \mathrm{dom} f} \{\langle x^*, x\rangle - f(x)\} \\ &\leq M \cdot \|x^*\| - \inf_{x \in \mathrm{dom} f} f(x) < \infty \end{aligned}$$

since f is lower semicontinuous on the compact set $M \cdot B$. So f^* is everywhere finite, and therefore $\{\nabla f^*(x^*)\} = \partial f^*(x^*)$ exists everywhere.

Since f^* is convex, finite and differentiable on $\mathbb{R}^n$, ∇f^* is everywhere continuous, by Corollary 25.5.1 in [Roc97]. ■

Proposition 1.12 *Let g be Lipschitz and real valued on $B(x;\varepsilon) \subset \mathbb{R}^n$, and let ϕ_ε be a nonnegative continuously differentiable function with support contained in $B(0;\varepsilon)$ and integral equal to 1. Then $g_\varepsilon = \phi_\varepsilon * g$ is continuously differentiable, $|g_\varepsilon - g| \leq \varepsilon\,\mathrm{Lip}(g)$, and $\nabla g_\varepsilon(x) \in \overline{\mathrm{conv}}\partial g(B(x;\varepsilon))$.*

Proof: The continuous differentiability of the convolution is well-known. A suitably general proof may be found in [Eva98].

To show that g_ε approaches g uniformly as $\varepsilon \to \infty$, we compute

$$\begin{aligned}
|g_\varepsilon(x) - g(x)| &= \left| \int_{B(0;\varepsilon)} \phi_\varepsilon(y) g(x-y) \mathrm{d}\lambda^n(y) - g(y) \right| \\
&= \left| \int_{B(0;\varepsilon)} \phi_\varepsilon(y) \big(g(x-y) - g(x)\big) \mathrm{d}\lambda^n(y) \right| \\
&\le \int_{B(0;\varepsilon)} |\phi_\varepsilon(y)| |g(x-y) - g(x)| \mathrm{d}\lambda^n(y) \\
&\le \int_{B(0;\varepsilon)} \phi_\varepsilon(y) \operatorname{Lip}(g) |y| \mathrm{d}\lambda^n(y) \\
&\le \int_{B(0;\varepsilon)} \phi_\varepsilon(y) \operatorname{Lip}(g) \varepsilon \mathrm{d}\lambda^n(y) \\
&= \varepsilon \operatorname{Lip}(g).
\end{aligned}$$

To show that $\nabla g_\varepsilon(x) \in \overline{\text{conv}} \partial g(B(x;\varepsilon))$, we note that Theorem 2.7.2 in [Cla83] says that

$$\partial g_\varepsilon(x) \subset \int_{B(0;\varepsilon)} \partial g(x-y) \phi_\varepsilon(y) \mathrm{d}\lambda^n(y), \tag{1.5}$$

where 1.5 is to be interpreted as follows: For every $x^* \in \partial g_\varepsilon(x)$ there exists a mapping $\zeta : B(0;\varepsilon) \to \mathbb{R}^n$ such that

$$\zeta(y) \in \partial g(x-y) \subset \partial g(B(x;\varepsilon)) \tag{1.6}$$

almost everywhere in $B(0;\varepsilon)$. Furthermore, the composite mapping $\langle \zeta(y), v \rangle$ is integrable on $B(x;\varepsilon)$ and

$$\langle x^*, v \rangle = \int_{B(0;\varepsilon)} \langle \zeta(y), v \rangle \phi_\varepsilon(y) \mathrm{d}\lambda^n(y)$$

for every $v \in \mathbb{R}^n$. In particular, taking v to be the canonical unit basis vectors gives

$$x^* = \int_{B(0;\varepsilon)} \zeta(y) \phi_\varepsilon(y) \mathrm{d}\lambda^n(y).$$

Since $\zeta(y) \in \partial g(B(x;\varepsilon))$ and $\int_{B(0;\varepsilon)} \phi_\varepsilon(y) \mathrm{d}\lambda^n(y) = 1$, we have

$$\nabla g_\varepsilon(x) \subset \overline{\text{conv}} \partial g(B(x;\varepsilon)),$$

by 1.13. ■

1.6 Jensen's Inequality

We will need to use two forms of the well-known Jensen's inequality, one for sets and one for functions. Let C be a convex subset of X and choose $x_1, x_2, \ldots, x_n \in C$ and nonnegative $\lambda_1, \lambda_2, \ldots, \lambda_n$ with $\sum_{i=1}^n \lambda_i = 1$. The set version of Jensen's inequality, Proposition 1.13, is a finite-dimensional extension of the fact that

$$\sum_{i=1}^{n} \lambda_i x_i \in C.$$

For $f : C \to \overline{\mathbb{R}}$ convex the functional version, Proposition 1.14, generalizes the following fact:

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i f(x_i).$$

Proposition 1.13 (Jensen's inequality on sets) *Let μ be a positive measure on a σ-algebra in a set Ω with $\mu(\Omega) = 1$. Furthermore, let $u : \Omega \to C \subset \mathbb{R}^n$ have the property that $\langle v, u(\cdot)\rangle$ is integrable for every $v \in \mathbb{R}^n$. Then*

$$x' := \int_\Omega u \, d\mu \in \overline{\text{conv}} C.$$

Proof: Corollary 11.5.1 in [Roc97] says that the closed convex hull of a set C is the intersection of all closed halfspaces containing C:

$$\overline{\text{conv}} C = \bigcap_{\alpha \in \mathcal{A}} H_\alpha$$

Where

$$H_\alpha = \{x \in \mathbb{R}^n \mid \langle a_\alpha, x\rangle \leq b_\alpha\},$$

so that $\sup_{c \in C} \langle a_\alpha, c\rangle \leq b_\alpha$. For a given $\alpha \in \mathcal{A}$,

$$\langle a_\alpha, u(x)\rangle \leq b_\alpha$$

for all $x \in \Omega$. Integrating over Ω gives

$$\begin{aligned}
\langle a_\alpha, x'\rangle &= \left\langle a_\alpha, \int_\Omega u \, d\mu \right\rangle \\
&= \int_\Omega \langle a_\alpha, u\rangle \, d\mu \\
&\leq \int_\Omega b_\alpha \, d\mu = b_\alpha,
\end{aligned}$$

So $x' \in H_\alpha$. As α was arbitrary, we have

$$x' \in \bigcap_{\alpha \in \mathcal{A}} H_\alpha = \overline{\text{conv}} C.$$

■

In particular, we will apply Proposition 1.13 to the usual Lebesgue measure on $\mathbb{R}^n$.

Proposition 1.14 (Jensen's inequality on a function) *Let $F : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ be convex, with $u : [a,b] \subset \mathbb{R} \to \operatorname{dom}(F) \subset \mathbb{R}^m$ summable. Then*

$$F\left(\frac{1}{b-a}\int_a^b u(t)\,dt\right) \leq \frac{1}{b-a}\int_a^b (F \circ u)(t)\,dt.$$

Proof: $\operatorname{dom}(F)$ is convex, so

$$\overline{u} := \frac{1}{b-a}\int_a^b u(t)\mathrm{d}t \in \operatorname{cl}\operatorname{dom}(F)$$

by the previous proposition. We prove the result for the case $\partial F(\overline{u}) \neq \emptyset$, for instance if $\overline{u} \in \operatorname{int}\operatorname{dom} F$. Then there is a $z^* \in \mathbb{R}^m$ such that

$$\begin{aligned} &F\left(\frac{1}{b-a}\int_a^b u(t)\mathrm{d}t\right) \\ &\leq F(y) - \left\langle z^*, y - \frac{1}{b-a}\int_a^b u(t)\mathrm{d}t\right\rangle, \end{aligned} \tag{1.7}$$

for all $y \in \mathbb{R}^m$. Letting $y = u(t)$, we integrate (1.7) over $s \in [a,b]$ to get

$$\begin{aligned} &(b-a)F\left(\frac{1}{b-a}\int_a^b u(t)\mathrm{d}t\right) \\ &\leq \int_a^b (F \circ u)(t)\mathrm{d}t - \left\langle z^*, \int_a^b u(t)\mathrm{d}t - \frac{b-a}{b-a}\int_a^b u(t)\mathrm{d}t\right\rangle. \end{aligned} \tag{1.8}$$

Dividing by $(b-a)$ gives the result. ■

1.7 Schauder's Fixed Point Theorem

A mapping f from a set S into itself is said to have a *fixed point* x if $f(x) = x$. There is an extensive fixed point theory that seeks to determine which properties of sets and functions guarantee the existence of fixed points. In particular, Schauder's fixed point theorem is central in our development of duality inequalities. We follow the method of proof provided in [GK90] by proving Schauder's result from Brouwer's fixed point theorem, which we merely state. In the following, $\mathbb{B}^n$ denotes the unit ball in $\mathbb{R}^n$.

Theorem 1.15 (Brouwer) *Every continuous function $f : \mathbb{B}^n \to \mathbb{B}^n$ has a fixed point.*

First, we prove a slightly more general version of Theorem 1.15:

Corollary 1.16 *Let $C \subset \mathbb{R}^n$ be closed, bounded and convex. Then every continuous function $f : C \to C$ has a fixed point.*

Proof: Let C be as above. Then given $x \in \mathbb{R}^n$, there exists a unique $y \in C$ such that $\|x - y\| = \inf_{z \in C} \|x - z\|$. The existence of y is due to the continuity of $\|x - \cdot\|$ over the compact set C, and uniqueness comes from the convexity of C and the strict convexity of the square of the Euclidean norm. Indeed, if $\|x - z\|^2 = \|x - y\|^2$ for some $z \neq y$ in C, then

$$\begin{aligned}
\left\| x - \frac{y+z}{2} \right\|^2 &= \left\langle x - \frac{y+z}{2}, x - \frac{y+z}{2} \right\rangle \\
&= \frac{\langle x-y, x-y\rangle}{2} + \frac{\langle x-z, x-z\rangle}{2} - \frac{\langle y-z, y-z\rangle}{4} \\
&= \frac{\|x-y\|^2}{2} + \frac{\|x-z\|^2}{2} - \frac{\|y-z\|^2}{4} \qquad (1.9) \\
&< \frac{\|x-y\|^2}{2} + \frac{\|x-z\|^2}{2} \\
&= \|x-y\|^2,
\end{aligned}$$

a contradiction. Thus the map T defined by $T(x) = \operatorname{argmin}_{y \in C} \|x - y\|$ is well defined. It is easy to see that $\|T(x) - T(x')\| \leq \|x - x'\|$ for $x, x' \in \mathbb{R}^n$, so T is continuous (in fact, nonexpansive). Also, $T(x) = x$ for any $x \in C$.

Without loss of generality, we may suppose that $C \subset \mathbb{B}^n$. Let $f : C \to C$ be continuous, and define $\widetilde{T} : \mathbb{B}^n \to \mathbb{B}^n$ by $\widetilde{T} = f \circ T$. Since $\widetilde{T}$ is continuous, it has a fixed point x, by Theorem 1.15. Since $\widetilde{T}(\mathbb{B}^n) \subset C$, $x \in C$. Thus $f(x) = f(T(x)) = \widetilde{T}(x) = x$, so x is a fixed point of f. ■

Now we are ready to prove the main fixed point result:

Theorem 1.17 (Schauder) *Let $C \neq \emptyset$ be a compact convex subset of a Banach space X. Then every continuous map from C into C has a fixed point.*

Proof: Choose $\varepsilon > 0$ and let $A = \{a_1, a_2, \dots, a_p\}$ be an ε-net for C, that is $C \subset A + \varepsilon B$. Define real-valued functions $m_i, i = 1, 2, \dots, p$ by

$$m_i(x) = \begin{cases} \varepsilon - \|a_i - x\|, & \|a_i - x\| \leq \varepsilon \\ 0, & otherwise \end{cases}$$

Now define a mapping $\phi : C \to C_0 = \operatorname{span} A \cap C$ by

$$\phi(x) = \frac{\sum_{i=1}^p m_i(x) a_i}{\sum_{i=1}^p m_i(x)}.$$

Then ϕ is continuous, since the denominator in the definition of ϕ can never be zero, and for any $x \in C$ we have

$$\begin{aligned} \|x - \phi(x)\| &= \left\| \frac{\sum_{i=1}^p m_i(x)(a_i - x)}{\sum_{i=1}^p m_i(x)} \right\| \\ &\leq \frac{\sum_{i=1}^p m_i(x) \|a_i - x\|}{\sum_{i=1}^p m_i(x)} \qquad (1.10) \\ &\leq \varepsilon. \end{aligned}$$

Let $T : C \to C$ be continuous. Then the mapping $\widetilde{T} : C_0 \to C_0$ defined by $\widetilde{T} = \phi \circ T$ has a fixed point x_0, by Theorem 1.16. This yields

$$\begin{aligned} \|T(x) - x\| &\leq \|T(x) - \widetilde{T}(x)\| + \|\widetilde{T}(x) - x\| \\ &= \|T(x) - \phi(T(x))\| + 0 \qquad (1.11) \\ &\leq \varepsilon. \end{aligned}$$

Since ε was arbitrary, we can construct a sequence $\{x_n\}_{n \in \mathbb{N}}$ such that $T(x_n) - x_n \to 0$ as $n \to \infty$. C is compact, so x_n has a convergent subsequence $x_{n_k} \to x_0 \in C$, and we have

$$T(x_0) = \lim_{k \to \infty} T(x_{n_k}) = \lim_{k \to \infty} x_{n_k} = x_0.$$

■

In Chapters 2 and 3, we apply Schauder's fixed point theorem to an operator on a set of functions in the uniform norm topology. As the theorem requires a compact set, our next theorem provides a compactness condition (see [Rud73] and [FK70]). Consider subsets S_1 and S_2 of two Banach spaces X and Y. We denote the space of continuous functions x from S_1 to S_2 with norm

$$\|x\| = \max_{s \in S_1} \|x(s)\|$$

by $C(S_1, S_2)$. Let W be a collection of functions in $C(S_1, S_2)$. Then W is said to be *uniformly equicontinuous* if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\|s - s'\| \leq \delta \text{ implies } \|x(s) - x(s')\| \leq \varepsilon$$

for every $s, s' \in S_1$ and $x \in W$. W is *uniformly equibounded* if

$$\sup_{x \in W, s \in S} \|x(s)\| < \infty.$$

The compactness result we require is as follows:

Theorem 1.18 (Ascoli-Arzela) *Let S_1, S_2 be compact subsets of Banach spaces X and Y. Then a collection of functions $D \subset C(S_1, S_2)$ is compact if and only if it is uniformly equibounded and uniformly equicontinuous.*

Chapter 2

Duality Results on Three Functions

In this chapter, we prove a central duality inequality given in [BF01], which we demonstrate to be equivalent to a nonsmooth mean value inequality given in [CL94], and to a Fenchel-like sandwich theorem given in [LR96]. We initially prove the result for smooth functions, and then extend it to locally Lipschitz functions, by uniformly approximating these with smooth functions.

The main nonsmooth result is as follows:

Theorem 2.1 (Borwein, Fitzpatrick) *Let C be a nonempty compact convex subset of $\mathbb{R}^n$, and let f, h be proper convex lower semicontinuous with* $\mathrm{dom}(f) \cup \mathrm{dom}(h) \subset C$. *Then for any $g : C \to \mathbb{R}$, Lipschitz on a neighborhood of C, there is a $z^* \in \partial g(C)$ such that*

$$\min(f - g) + \min(h + g) \leq -f^*(z^*) - h^*(-z^*) \leq \min(f + h). \tag{2.1}$$

This rather formidable looking result has two interesting specializations: the Clarke-Ledyaev two set Mean Value Inequality (see [CL94]), and the Lewis-Ralph sandwich theorem found in [LR96]. We will discuss these further after we have proven the main result.

First, we introduce a smooth version of the left hand inequality in Theorem 2.1:

Theorem 2.2 (Borwein, Fitzpatrick) *Let C be a nonempty compact convex subset of $\mathbb{R}^n$, and let f, h be proper convex lower semicontinuous with f^* and h^* continuously differentiable, and* $\mathrm{dom}(f) \cup \mathrm{dom}(h) \subset C$. *Then for any function g that is continuously differentiable on a neighborhood of C, there is a $z \in C$ such that*

$$\max(g - f) + \max(-g - h) \geq f^*(\nabla g(z)) + h^*(-\nabla g(z)). \tag{2.2}$$

To prove this result, we introduce a continuous operator T on a compact convex subset W of a certain function space. We then deduce the existence of a fixed point for this operator using Schauder's fixed point theorem, Theorem 1.17. The bulk of our efforts are devoted to proving that T and W meet the criteria required by Theorem 1.17; once this is proven, Theorem 2.2 is obtained more easily.

In the setting of Theorem 2.2, define the set W by

$$W = \{x : [0,1] \to C \mid \text{Lip}(x) \le M\}, \tag{2.3}$$

where $M = 2\sup\{\|c\| \mid c \in C\}$, and endow W with the topology induced by the uniform norm, so that

$$\|x\| = \sup_{t\in[0,1]} \|x(t)\|$$

for $x \in W$. Thus W consists of arcs in C whose "speed" is limited by M. Next we introduce the following nonlinear operator $T : W \to W$:

$$Tx(t) = \int_0^t \nabla f^* \circ \nabla g \circ x + \int_t^1 \nabla h^* \circ (-\nabla g) \circ x. \tag{2.4}$$

2.1 Properties of T and W

Proposition 2.3 *For T given by (2.4) and W given by (2.3), the following are true:*

1. *W is convex and nonempty, and it is compact in the uniform norm topology,*

2. *T is a continuous self-map on W.*

Proof: W is clearly nonempty, since any constant trajectory, $x(t) = c$ for some $c \in C$, is in W. Turning to convexity, take $x, y \in W$ with $z = \lambda x + (1-\lambda)y$ for some $\lambda \in [0,1]$. Then for $t \in [0,1]$ we have $x(t), y(t) \in C$, and C is convex, so

$$z(t) = \lambda x(t) + (1-\lambda)y(t) \in C.$$

For $s \in [0,1]$, we also have

$$\begin{aligned}
\|z(s) - z(t)\| &= \|\lambda x(s) + (1-\lambda)y(s) - \lambda x(t) + (1-\lambda)y(t)\| \\
&\le \lambda\|x(s) - x(t)\| + (1-\lambda)\|y(s) - y(t)\| \\
&\le \lambda M + (1-\lambda)M = M,
\end{aligned}$$

so $\mathrm{Lip}(z) \leq M$, and thus W is convex.

Endowing W with the uniform norm topology

$$x \in W \implies \|x\| = \sup_{t\in[0,1]} \|x(t)\|,$$

the Ascoli-Arzela theorem, Theorem 1.18, says that W is compact if it is *uniformly equicontinuous* and *uniformly equibounded.* Since W is a collection of functions sharing the same Lipschitz constant, equicontinuity is easy: Let $x \in W, \varepsilon > 0$. Then $\delta = \varepsilon/M$ gives a constant independent of x such that

$$|s-t| \leq \delta, s,t \in [0,1] \implies \|x(s) - x(t)\| \leq M|s-t| \leq \varepsilon.$$

On the other hand, equiboundedness is guaranteed by the compactness of C:

$$\|x(t)\| \leq \sup_{c\in C} \|c\| = M/2$$

for all $x \in W$ and $t \in [0,1]$, since $x(t) \in C$. So W is compact.

Turning now to the operator T, again let $x \in W$. We wish to show that $Tx \in W$, so we must show $Tx(t) \in C$ for $t \in [0,1]$, and $\mathrm{Lip}(Tx) \leq M$. Fixing $t \in [0,1]$, let

$$c(s) = \begin{cases} \nabla f^* \circ \nabla g \circ x(s) & 0 \leq s < t \\ \nabla h^* \circ (-\nabla g) \circ x(s) & t \leq s \leq 1 \end{cases}$$

To show that $Tx(t) \in C$, we notice that $c(s) \in C$ for all $s \in [0,1]$, since $\nabla f^*(y), \nabla h^*(y) \in C$ for all y. So

$$\begin{aligned} Tx(t) &= \int_0^t \nabla f^* \circ \nabla g \circ x(u)du + \int_t^1 \nabla h^* \circ (-\nabla g) \circ x(u)du \\ &= \int_0^1 c(u)du, \end{aligned}$$

and $Tx(t) \in \overline{\mathrm{conv}} C = C$ by Proposition 1.13.

Now let $s,t \in [0,1]$, and denote

$$G(u) = \nabla g(x(u)).$$

Then

$$\begin{aligned}\|Tx(s) - Tx(t)\| &= \left\| \int_0^s \nabla f^* \circ \nabla g \circ x(u)du + \int_s^1 \nabla h^* \circ (-\nabla g) \circ x(u)du \right. \\ &\quad \left. - \int_0^t \nabla f^* \circ \nabla g \circ x(u)du + \int_t^1 \nabla h^* \circ (-\nabla g) \circ x(u)du \right\| \\ &\leq \left\| \int_t^s \nabla f^*(G(u))du \right\| + \left\| \int_t^s \nabla h^*(-G(u))du \right\| \\ &\leq \int_t^s \|\nabla f^*(G(u))du\| + \int_t^s \|\nabla h^*(-G(u))du\|,\end{aligned}$$

but $\nabla f^*(y), \nabla h^*(y) \in C$ for all y, so

$$\begin{aligned}\|Tx(s) - Tx(t)\| &\leq |t-s| 2 \max_{c \in C} \|c\| \\ &= M|t-s|.\end{aligned}$$

Since $Tx(t) \in C$ for all $t \in [0,1]$ and $\operatorname{Lip}(Tx) \leq M$, we have obtained $Tx \in W$, or $T(W) \subset W$.

Let $\varepsilon > 0$. To show that T is continuous on W, we must show that there exists a $\delta > 0$ such that

$$\|x - y\| < \delta \implies \|Tx - Ty\| < \varepsilon$$

for x, y in W. Now define $F : C \to C$ and $H : C \to C$ by

$$\begin{aligned}F(x) &= \nabla f^*(\nabla g(x)) \\ H(x) &= \nabla h^*(-\nabla g(x)).\end{aligned}$$

Since f^*, h^* and g^* are continuously differentiable, both F and H are continuous. In fact, F and H are uniformly continuous on C, since C is compact. Thus there exists $\delta > 0$ such that $\|x - y\| < \delta$ and $x, y \in C$ imply

$$\|F(x) - F(y)\| < \varepsilon$$

and

$$\|H(x) - H(y)\| < \varepsilon.$$

Let $x, y \in W$ with $\|x - y\| < \delta$, so that

$$\|x(t) - y(t)\| < \delta$$

for all $t \in [0,1]$. Then

$$\begin{aligned}
\|Tx(t) - Ty(t)\| &= \left\| \int_0^t F(x(s))ds + \int_t^1 H(x(s))ds \right. \\
&\quad \left. - \int_0^t F(y(s))ds + \int_t^1 H(y(s))ds \right\| \\
&\le \left\| \int_0^t F(x(s)) - F(y(s))ds \right\| + \left\| \int_t^1 H(x(s)) - H(y(s))ds \right\| \\
&\le \int_0^t \|F(x(s)) - F(y(s))\|ds + \int_t^1 \|H(x(s)) - H(y(s))\|ds \\
&\le t\varepsilon + (1-t)\varepsilon = \varepsilon
\end{aligned}$$

for all $t \in [0,1]$. Therefore $\|Tx - Ty\| < \varepsilon$ and T is continuous. ■

We are now ready to prove our smooth three-function duality inequality:

Proof of Theorem 2.2: Since T and W satisfy the requirements of Schauder's fixed point theorem, we conclude that there is an $x \in W$ such that $x = Tx$, that is

$$x(t) = \int_0^t \nabla f^* \circ \nabla g \circ x + \int_t^1 \nabla h^* \circ (-\nabla g) \circ x$$

for all $t \in [0,1]$. Then, since $\nabla x = \nabla f^* \circ \nabla g \circ x - \nabla h^* \circ (-\nabla g) \circ x$, we have the following:

$$\begin{aligned}
g(x(1)) - g(x(0)) &= \int_0^1 (g \circ x)' \\
&= \int_0^1 \langle \nabla g \circ x, \nabla x \rangle \\
&= \int_0^1 \langle \nabla f^* \circ \nabla g \circ x, \nabla g \circ x \rangle \\
&\quad + \int_0^1 \langle \nabla h^* \circ (-\nabla g) \circ x, -\nabla g \circ x \rangle && (2.5) \\
&= \int_0^1 \Big(f^*(\nabla g \circ x) + f(\nabla f^* \circ \nabla g \circ x) \Big) \\
&\quad + \int_0^1 \Big(h^*(-\nabla g) \circ x + h(\nabla h^* \circ (-\nabla g) \circ x) \Big), && (2.6)
\end{aligned}$$

where (2.6) follows from (2.5) by Fenchel's equation, Theorem 1.7 as follows. For Fenchel's Equation to hold for f, we require that

$$\nabla g \circ x \in \partial f(\nabla f^* \circ \nabla g \circ x). \tag{2.7}$$

Since f is proper convex lower semicontinuous, we know that

$$x^* \in \partial f(x) \text{ iff } x \in \partial f^*(x^*).$$

Since f^* is continuously differentiable, we have

$$\{\nabla f^* \circ \nabla g \circ x\} = \partial f^*(\nabla g \circ x),$$

and therefore (2.7) holds, by Proposition 1.8, and the same argument holds for h. Continuing from (2.6), we have

$$\begin{aligned}
g(x(1)) - g(x(0)) &\geq \int_0^1 \Big(f^*(\nabla g \circ x) + f(\nabla f^* \circ \nabla g \circ x)\Big) \\
&\quad + \int_0^1 \Big(h^*(-\nabla g) \circ x + h(\nabla h^* \circ (-\nabla g) \circ x)\Big) \qquad (2.8) \\
&\geq \int_0^1 \nabla f^* \circ \nabla g \circ x + f\left(\int_0^1 \nabla f^* \circ \nabla g \circ x\right) \\
&\quad + \int_0^1 \nabla h^* \circ (-\nabla g) \circ x + h\left(\int_0^1 \nabla h^* \circ (-\nabla g) \circ x\right) \qquad (2.9) \\
&= \int_0^1 \Big(\nabla f^* \circ \nabla g \circ x + \nabla h^* \circ (-\nabla g) \circ x\Big) \\
&\quad + f(x(1)) + h(x(0)). \qquad (2.10)
\end{aligned}$$

Equation (2.9) follows from (2.8) by Jensen's inequality, Proposition 1.14, and using (2.10) we know there is a $z = x(t) \in C$ such that

$$g(x(1)) - f(x(1)) - h(x(0)) - g(x(0)) \geq f^*(\nabla g(z)) + h^*(-\nabla g(z)).$$

By taking the maxima over C we get

$$\max(g - f) + \max(-g - h) \geq f^*(\nabla g(z)) + h^*(-\nabla g(z)).$$

■

2.2 A Nonsmooth Duality Result on Three Functions

We are now ready to prove Theorem 2.1:

Proof of Theorem 2.1: The right hand equality is just weak Fenchel duality, as follows:

$$\begin{aligned}-f^*(z^*) - h^*(-z^*) &= \inf\{f(x) - \langle z^*, x\rangle\} + \inf\{h(y) - \langle -z^*, y\rangle\}\\ &\leq \inf\{f(x) + h(x)\}\\ &= \min(f+h),\end{aligned}$$

where the infimal attainment is due to f and h being lower semicontinuous on a compact set C (see, e.g., [FK70]). To prove the other inequality, first set

$$\begin{aligned}f_m &= f + (1/m)\|\cdot\|^2\\ h_m &= h + (1/m)\|\cdot\|^2\\ g_n &= g * \phi_{1/n}.\end{aligned}$$

From Theorem 2.2, we know that there exists $z \in C$ such that

$$\max(f_m - g_n) + \max(-h_m - g_n) \geq f_m^*(\nabla g_n(z)) + h_m^*(-\nabla g_n(z)).$$

That is, there exists $z_n^* \in \nabla g(C)$ such that

$$\max(f_m - g_n) + \max(-h_m - g_n) \geq f_m^*(z_n^*) + h_m^*(-z_n^*).$$

Remembering that $\nabla g_n(x) \in \operatorname{conv} \partial g(B(x;1/n))$, we note the following inclusions:

$$\begin{aligned}z_n^* \in \nabla g_n(C) &= \bigcup_{c\in C}\{\nabla g_n(c)\}\\ &\subset \bigcup_{c\in C}\overline{\operatorname{conv}}\ \partial g(B(c;1/n))\\ &\subset \overline{\operatorname{conv}} \bigcup_{c\in C}\partial g(B(c;1/n))\\ &= \overline{\operatorname{conv}}\ \partial g(C_n),\end{aligned}$$

where $C_n = C + B(x;1/n)$.

Since g is Lipschitz on a neighborhood of C, g must be Lipschitz on C_n for n large, which implies that $\partial g(x) \subset B(0;\mathrm{Lip}(g))$ for all $x \in C_n$.

Of course, this implies

$$\overline{\operatorname{conv}}\partial g(C_n) \subset B(0;\mathrm{Lip}(g)),$$

so that $\overline{\operatorname{conv}}\partial g(C_n)$ is compact for n large. Therefore $\{z_n^*\}$ has a subsequence converging to some z^*. We continue to denote this subsequence by $\{z_n^*\}$.

Now $z_n^* \in \overline{\text{conv}}\partial g(x_n + (1/n)B)$ for some $x_n \in C$. Without loss of generality, we may assume x_n converges to $x \in C$. Then

$$x_n \in x + \lambda_n B$$

for some $\lambda_n \to 0$. Thus

$$z_n^* \in \overline{\text{conv}}\partial g(x + (1/n + \lambda_n)B).$$

By upper semicontinuity, for any $\varepsilon > 0$ we can take N large enough so that

$$z_n^* \in \overline{\text{conv}}(\partial g(x) + \varepsilon B)$$

whenever $n \geq N$. Since ε was arbitrary,

$$z^* \in \overline{\text{conv}}(\partial g(x)) = \partial g(x).$$

We have established that there is $z_m^* \in \partial g(C)$ such that

$$\max(g - f_m) + \max(-g - h_m) \geq f_m^*(z_m^*) + h_m^*(-z_m^*) \tag{2.11}$$

Thus there exists $x_m \in C$ such that $z_m^* \in \partial g(x_m)$. $\text{Lip}(g) \cdot B^* \times C$ is compact in the product topology and $z_m^* \in \text{Lip}(g) \cdot B^*$, so (z_m^*, x_m) has a subsequence converging to some (z^*, x). Consider (z_m^*, x_m) to be this subsequence. Then $z_m^* \to z^*$, $x_m \to x$ and $z_m^* \in \partial g(x_m)$, which implies that $z^* \in \partial g(x)$.

Note that f_m converges uniformly to f on C, since

$$\sup_C \{f - f_m\} = (1/m) \sup_{c \in C} \|c\|^2,$$

and C is compact. But then f_m^* converges uniformly to f^*, since

$$\begin{aligned} f_m^*(y^*) - f^*(y^*) &= \sup_{x \in C}\{\langle y^*, x\rangle - f_m(x)\} + \inf_{x' \in C}\{f(x') - \langle y^*, x'\rangle\} \\ &\leq \sup_{x \in C}\{f(x) - f_m(x)\}. \end{aligned}$$

Thus both sides of (2.11) must converge to the respective sides of (2.1). ■

2.3 The Lewis-Ralph and Clarke-Ledyaev Inequalities

From Theorem 2.1, we obtain the following well-known duality results as corollaries:

Theorem 2.4 (Lewis-Ralph Sandwich Theorem) *Let C be a nonempty compact convex subset of $\mathbb{R}^n$, and let f, h be proper convex lower semicontinuous with* $\mathrm{dom}(f) \cup \mathrm{dom}(h) \subset C$. *Then for any Lipschitz function g such that $-h \le g \le f$ there is a $z^* \in \partial g(C)$ such that*

$$0 \ge f^*(z^*) + h^*(-z^*).$$

Proof: By Theorem 2.1, there is a $z^* \in \partial g(C)$ such that

$$f^*(z^*) + h^*(-z^*) \le -\min(f-g) - \min(h+g) \le 0.$$

■

Theorem 2.1 is actually equivalent to Theorem 2.4, since we can recover the former from the latter: Setting $\overline{f} = f - \min(f-g)$ and $\overline{h} = h - \min(h+g)$, we have

$$\overline{f} \ge g \ge -\overline{h},$$

since

$$f - g \ge \min(f-g) \text{ and } h + g \ge \min(h+g).$$

By Theorem 2.4, there exists a $z^* \in \partial g(C)$ such that

$$\begin{aligned} 0 &\ge \overline{f}^*(z^*) + \overline{h}^*(-z^*) \\ &= f^*(z^*) - \min(f-g) + h^*(-z^*) - \min(h+g) \end{aligned}$$

which is the result in Theorem 2.1.

As noted in [LR96] and [LL00], the Lewis-Ralph sandwich theorem has the following interpretation: If g is a Lipschitz function between $-h$ and f, then there is an affine function $k(x) = \langle x^*, x\rangle + b$ separating $-h$ from f (see Figure 2.1). Furthermore, this affine function is parallel to g at some point x:

$$x^* \in \partial g(x).$$

To see this, suppose f and $-h$ are affinely separated by x^*. Then there exists a real b such that

$$f(x) \ge \langle x^*, x\rangle + b \ge -h(x) \tag{2.12}$$

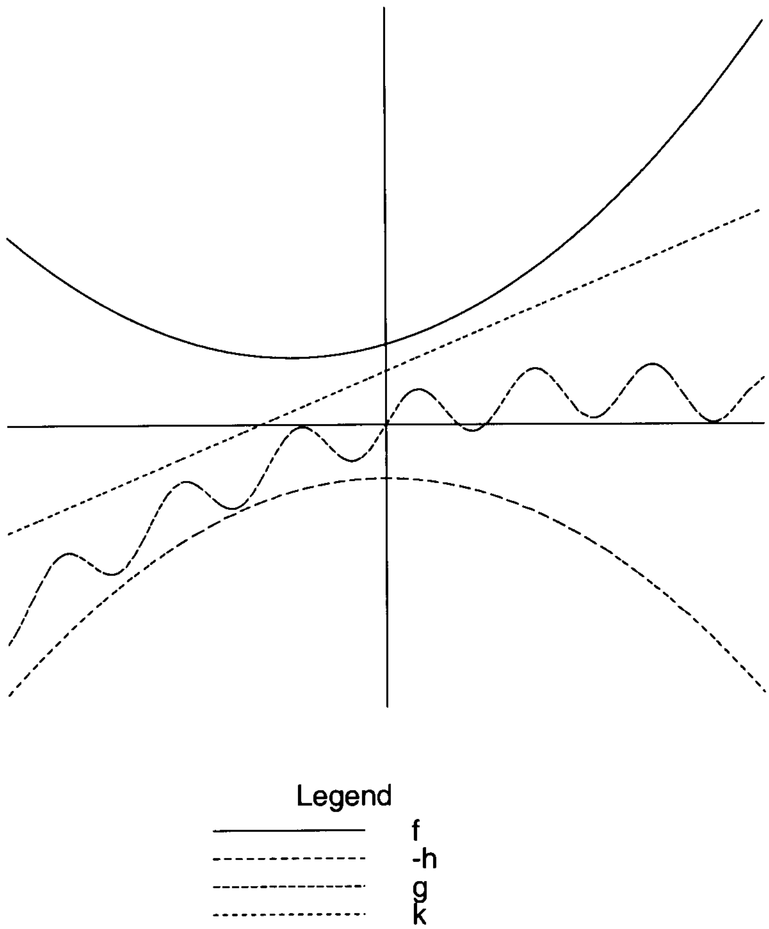

Figure 2.1: The Lewis-Ralph sandwich theorem

for all x. Subtracting $\langle x^*, x\rangle$ from both sides, this is clearly equivalent to

$$-f^*(x^*) = \inf_x\{f(x) - \langle x^*, x\rangle\} \geq b \geq \sup_x\{\langle -x^*, x\rangle - h(x)\} = h^*(-x^*)$$

for some real b. Such a b exists if and only if

$$f^*(x^*) + h^*(-x^*) \leq 0.$$

The Lewis-Ralph sandwich theorem guarantees this, with the added information that

$$x^* \in \partial g(C).$$

It is not known whether this conclusion can be strengthened so that the affine separator is tangent to g rather than merely parallel. If this were so, we could take b in equation (2.12) to be

$$b = g(x^*) - \langle x^*, x'\rangle,$$

where $x^* \in \partial g(x')$.

For f and h as above the Fenchel duality theorem, Theorem 1.9, says that there is a point $x^* \in \mathbb{R}^n$ such that

$$-f^*(x^*) - h^*(-x^*) \geq \inf_{x\in\mathbb{R}^n}\{f(x) - (-h(x))\} \geq 0,$$

provided

$$0 \in \operatorname{core}(\operatorname{dom} h - \operatorname{dom} f). \tag{2.13}$$

By the above discussion, for any $b \in [h^*(-x^*), -f^*(x^*)] \neq \emptyset$ the affine function $k(x) := \langle x^*, x\rangle + b$ separates f and $-h$. The Lewis-Ralph sandwich theorem can therefore be viewed as an extension of Fenchel duality, where the existence of a globally Lipschitz separator g takes the place of the constraint qualification (2.13).

Theorem 2.1 also gives us the following striking multidirectional mean value inequality, due to Clarke and Ledyaev ([CL94]):

Theorem 2.5 (Clarke-Ledyaev Mean Value Inequality) *Let X and Y be compact convex subsets of $\mathbb{R}^n$ and $g : [X, Y] \to \mathbb{R}$ be Lipschitz. Then there is $z^* \in \partial g([X, Y])$ such that*

$$\langle z^*, x - y\rangle \leq \max_X g - \min_Y g \tag{2.14}$$

for all $x \in X$ and $y \in Y$.

Proof: Define the *indicator function* I_S of a set S by

$$I_S(x) := \begin{cases} 0 & x \in S \\ +\infty & x \notin S \end{cases}$$

and let $f = I_X$ and $h = I_Y$ in Theorem 2.1. Then we have the following:

$$\begin{aligned} \max_X g - \min_Y g &= -\min(f - g) - \min(-h - g) \\ &\geq f^*(z^*) + h^*(-z^*) \\ &= \sup_{x \in X}\langle z^*, x\rangle + \sup_{y \in Y}\langle -z^*, y\rangle \\ &\geq \langle z^*, x - y\rangle, \end{aligned}$$

for all $x \in X$ and $y \in Y$. ■

Note that by setting $X = \{x\}$ and $Y = \{y\}$ we recover the classical two point Mean Value theorem, since Theorem 2.5 gives us $z \in [x, y]$ and $z^* \in \partial g(z)$ such that

$$\langle z^*, x - y\rangle \leq \max_X g - \min_Y g = g(x) - g(y).$$

In particular, when g is differentiable we have

$$\langle \nabla g(z), x - y\rangle \leq \max_X g - \min_Y g = g(x) - g(y). \tag{2.15}$$

Reversing the roles of x and y, we see there is $z' \in [x, y]$ with

$$\langle \nabla g(z'), x - y\rangle \geq g(x) - g(y). \tag{2.16}$$

When g is differentiable, the function $g_{[x,y]}(t) = g\big(tx + (1-t)y\big)$ is differentiable on $[0, 1]$ with derivative

$$g'_{[x,y]}(t) = \langle \nabla g\big(tx + (1-t)y\big), x - y\rangle.$$

Furthermore, the derivative has the *Darboux property* (see [Spr70], p. 197): if

$$g'_{[x,y]}(t_1) < \alpha < g'_{[x,y]}(t_2),$$

then there is a point t_3 between t_1 and t_2 with

$$g'_{[x,y]}(t_3) = \alpha.$$

For $\alpha = g(x) - g(y)$, equations (2.15) and (2.16) and the Darboux property show that there is a point $z'' \in [x, y]$ with

$$\langle \nabla g(z''), x - y \rangle = g(x) - g(y).$$

The following example from [CL94] illustrates the improvement that the Clarke-Ledyaev mean value inequality provides over the classical Mean Value theorem. The classical theorem is essentially linear in its conclusion: it doesn't allow us to specify the behaviour of the gradient of a function in multiple directions simultaneously at a single point. As the example shows, with Theorem 2.5 we can obtain multidimensional control on the gradient.

Example 1.

Let

$$X = \{(0, x) \in \mathbb{R}^2 \mid x \in [0, 1]\},$$
$$Y = \{(1, y) \in \mathbb{R}^2 \mid y \in [0, 1]\}$$

and let $f : [X, Y] \to \mathbb{R}$ be continuously differentiable, with $f|_X = 0$ and $f|_Y = 1$ (see, e.g. Figure 2.2). Then there is a point $z \in [X, Y]$ such that

$$f'_x(z) = 1 + |f'_y(z)|,$$

where $\nabla f(z) = (f'_x(z), f'_y(z))$.

Proof: By Theorem 2.5 there are points $u, v \in [X, Y]$ such that

$$\langle \nabla f(u), y - x \rangle \geq \min_Y f - \max_X f = 1 \tag{2.17}$$

and

$$\langle \nabla f(v), x - y \rangle \geq \min_X f - \max_Y f = -1 \tag{2.18}$$

for all $x \in X$ and $y \in Y$. Define the function $g : \mathbb{R}^2 \to \mathbb{R}$ by

$$g(v) = \min_{x \in X, y \in Y} \langle v, y - x \rangle.$$

Then

$$\begin{aligned} g(v) = g(v_1, v_2) &= \min\{v_1(1 - 0) + v_2(y_2 - x_2) \mid y_2, x_2 \in [0, 1]\} \\ &= \begin{cases} v_1 - v_2 & v_2 \geq 0 \\ v_1 + v_2 & v_2 \leq 0 \end{cases} \\ &= v_1 - |v_2|. \end{aligned}$$

Recalling Equations (2.17) and (2.18), we have

$$g(\nabla f(u)) = \min\{\langle \nabla f(u), y - x\rangle \mid x \in X, y \in Y\} \geq 1$$
$$g(\nabla f(v)) \leq \max\{\langle \nabla f(v), y - x\rangle \mid x \in X, y \in Y\} \leq 1$$

Since g and ∇f are continuous on $[X, Y]$, there is a point z in $[u, v]$ with

$$g(\nabla f(z)) = f'_x(z) - |f'_y(z)| = 1$$

by the Intermediate Value theorem. ■

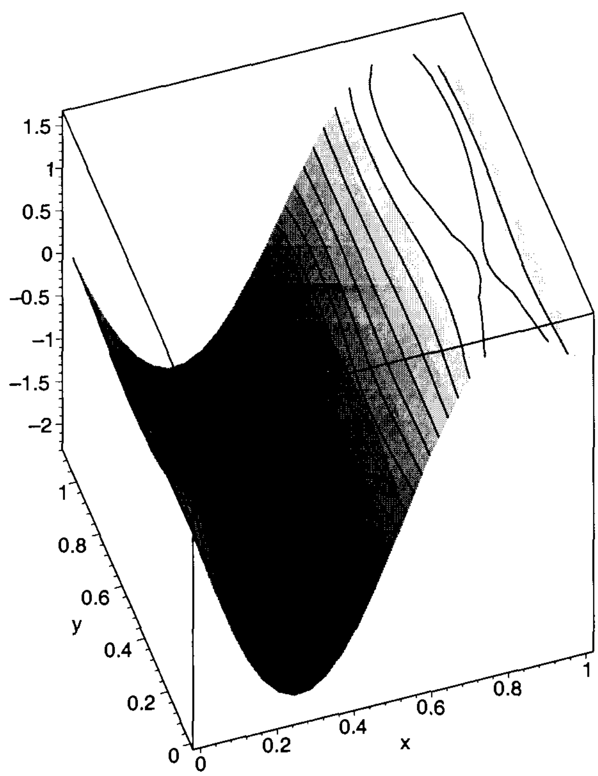

Figure 2.2: A possible function f.

Following the development in [LR96], we conclude this chapter by showing that the Lewis-Ralph sandwich theorem can be obtained from (2.14) as well, so these two theorems and Theorem 2.1 are all equivalent. First we introduce a technical result that, given a closed convex function f and $k \geq 0$, provides a globally k-Lipschitz minorant to f. This minorant

f_k, called the *Lipschitz regularization* of f, is defined by the infimal convolution of f with a multiple of the norm:

$$f_k(x) := f\Box k\|\cdot\| = \inf_{y\in X}\{f(y) + k\|x-y\|\}.$$

Intuitively, f_k is the function whose epigraph is the set sum of the epigraph of f and the epigraph of $k\|\cdot\|$.

Proposition 2.6 (Lewis, Ralph) *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be closed, proper and convex with bounded domain. Then f_k is convex and everywhere finite with global Lipschitz constant k, and $f_k(x) \le f(x)$ for every x. Furthermore, suppose $g : \mathbb{R}^n \to \mathbb{R}$ is globally k-Lipschitz on a set C containing* $\operatorname{dom} f$*, and that $g \le f$ on C. Then $g \le f_k$ on C as well.*

Proof: For $x, y \in \mathbb{R}^n$ and $\lambda \in [0,1]$ we have

$$\begin{aligned}
f_k(\lambda x + (1-\lambda)y) &= \inf_{z\in\mathbb{R}^n}\big\{f(z) + k\|\lambda x + (1-\lambda)y - z\|\big\} \\
&= \inf_{z_1,z_2\in\mathbb{R}^n}\big\{f(\lambda z_1 + (1-\lambda)z_2) + k\|\lambda(x-z_1) + (1-\lambda)(y-z_2)\|\big\} \\
&\le \inf_{z_1,z_2\in\mathbb{R}^n}\big\{\lambda\big(f(z_1) + k\|x-z_1\|\big) + (1-\lambda)\big(f(z_2) + k\|y-z_2\|\big)\big\} \\
&= \lambda \inf_{z_1\in\mathbb{R}^n}\big\{f(z_1) + k\|x-z_1\|\big\} + (1-\lambda)\inf_{z_2\in\mathbb{R}^n}\big\{f(z_2) + k\|y-z_2\|\big\} \\
&= \lambda f_k(x) + (1-\lambda)f_k(y)
\end{aligned}$$

So f_k is convex. Furthermore,

$$f_k(x) = \inf_{y\in\mathbb{R}^n}\{f(y) + k\|x-y\|\} \le f(x) + k\|x-x\| = f(x)$$

for all x. Since there exists an $x' \in \operatorname{dom} f$, this gives us $f_k(x') < +\infty$. On the other hand, f is bounded below, say $f \ge m$, so

$$f_k(x') \ge \inf y \in \mathbb{R}^n\{m + k\|x-y\|\} \ge m,$$

and therefore $x' \in \operatorname{dom} f_k$.

For $x, y \in \mathbb{R}^n$ we have

$$\begin{aligned}
f_k(x) - f_k(y) &= \inf_w\big\{f(w) + k\|x-w\|\big\} - \inf_z\big\{f(z) + k\|y-z\|\big\} \\
&= \sup_z\big\{\inf_w\{f(w) + k\|x-w\|\} - f(z) - k\|y-z\|\big\} \\
&\le \sup_z\big\{f(z) + k\|x-z\| - f(z) - k\|y-z\|\big\} \\
&= k\sup_z\big\{\|x-z\| - k\|y-z\|\big\} \le k\|x-y\|,
\end{aligned}$$

and interchanging x and y gives the desired Lipschitz bound. Since f_k is globally Lipschitz and finite at x', it is everywhere finite.

Finally, if $g(x) > f_k(x)$ then there exists a $y \in \operatorname{dom} f$ such that

$$g(x) > f(y) + k\|x - y\| \geq g(y) + k\|x - y\|,$$

which contradicts the Lipschitz bound on g. ∎

In the setting of Theorem 2.4, then, we may find continuous functions f_k and h_k such that

$$f(x) \geq f_k(x) \geq g(x) \geq -h_k(x) \geq -h(x)$$

for all x. Applying the theorem to f_k, g and h_k, and noting the order-reversing property of conjugation, we see that there is a $x^* \in \partial g(C)$ such that

$$0 \geq f_k^*(x^*) + h_k^*(-x^*) \geq f^(x^*) + h^*(-x^*).$$

We may therefore assume without loss of generality that the f and h are continuous on C.
Proof that Theorem 2.5 implies Theorem 2.4: As in the previous dicussion, assume f and h are continuous on C. Then $\alpha = \sup_C h$ and $\beta = \sup_C f$ are both finite, since C is compact. Define sets X and Y by:

$$X = \{(y,t) \in \mathbb{R}^{n+1} \mid -\alpha \leq t \leq -h(x)\}$$
$$Y = \{(x,s) \in \mathbb{R}^{n+1} \mid f(x) \leq s \leq \beta\}.$$

That is, X is the truncated hypograph of $-h$ on C and Y is the truncated epigraph of f on C. Y is closed and convex, since it is the intersection of a closed halfspace and the epigraph of a closed convex function. It is also bounded, as it is contained in the cylinder $C \times [\inf_C f, \beta]$, and it is clearly nonempty. Similarly, X is also compact, convex and nonempty.

Now define $G : X \times \mathbb{R} \to \mathbb{R}$ by $G(x,t) = -g(x) + t$ for x near C. The scalar multiplication and cartesian product rules of Proposition 1.6 apply to G as follows:

$$\partial G(x,t) \subset \partial_1 G(x,y) \times \partial_2 G(x,y) = \partial(-g)(x) \times \partial(t \to t) = -\partial g(x) \times \{1\}.$$

Then

$$\begin{aligned}\inf_Y G &= \inf\{-g(x) + t \mid f(x) \leq t \leq \beta\} \\ &= \inf_C\{f(x) - g(x)\} \geq 0\end{aligned}$$

and

$$\begin{aligned}\sup_X G &= \sup\{-g(x) + t \mid -\alpha \leq t \leq -h(x)\} \\ &= \sup_C\{-h(x) - g(x)\} \leq 0.\end{aligned}$$

By Theorem 2.5 and the above computation of ∂G, there exists $(c, t) \in C \times \mathbb{R}$ with $-c^* \in \partial g(c)$ such that

$$\langle(-c^*, 1), (y, s) - (x, r)\rangle \geq \inf_Y G - \sup_X G \geq 0$$

for all $x, y \in C$ and s, t with $f(y) \leq s \leq \beta$ and $h(x) \leq -r \leq \alpha$. But then

$$\langle -c^*, y\rangle + f(y) + \langle c, x\rangle - h(x) \leq 0$$

for all $x, y \in C$, so

$$f^*(c^*) + h^*(-c^*) \geq 0.$$

■

Chapter 3

Duality Inequalities on Two Functions

Modifying the operator T and set W given in the previous chapter, Borwein and Fitzpatrick proved a two-function nonsmooth duality result, given by Theorem 3.3. In this chapter, we provide a proof for this result and then examine several Rolle-type corollaries. We then introduce Ekeland's variational principle, and use it to obtain an approximate maximum principle for Lipschitz functions on the unit ball. Applying this variational result to the setting of one of our Rolle-type corollaries provides a much stronger result. However, not all of these results are as amenable to a variational treatment, as a conjectured improvement on another Rolle-type result turns out to be false.

As in the previous chapter, we first prove a smooth result, and then obtain the nonsmooth result by smooth approximation.

Theorem 3.1 (Borwein, Fitzpatrick) *Let C be a nonempty compact convex subset of $\mathbb{R}^n$, and let f be proper convex lower semicontinuous with f^* continuously differentiable, and* $\mathrm{dom}(f) \subset C$. *If $\alpha \neq 1$ and $g : [C, \alpha C] \to \mathbb{R}$ is continuously differentiable then there are $z \in [C, \alpha C]$ and $a \in C$ such that*

$$(g(\alpha a) - g(a))/(\alpha - 1) - f(a) \geq f^*(\nabla g(z)).$$

To prove this result, we again consider a nonlinear operator T on a function space W. Let $M = (1 + |\alpha|) \sup_{x \in C} \|x\|$, define W by

$$W = \{x : [0, 1] \to [C, \alpha C] \mid \mathrm{Lip}(x) \leq M\}, \tag{3.1}$$

and define an operator T by

$$Tx(t) = \alpha \int_0^t \nabla f^* \circ \nabla g \circ x + \int_t^1 \nabla f^* \circ \nabla g \circ x. \tag{3.2}$$

Then these modified T and W still satisfy the requirements of Schauder's fixed point theorem, Theorem 1.17:

Proposition 3.2 *For T given by* (3.2) *and W given by* (3.1), *the following are true:*

1. *W is convex and nonempty, and it is compact in the uniform norm topology,*

2. *T is a continuous self-map on W.*

Proof: The proof is entirely parallel to the proof of Proposition 2.3. ■

Proof of Theorem 3.1: From Schauder's theorem we know that there is a path $x \in W$ such that $x = Tx$. Parallel to the development of Theorem 2.2, we have the following:

$$\begin{aligned} g(x(1)) - g(x(0)) &= \int_0^1 (g \circ x)' \\ &= \int_0^1 \langle \nabla g \circ x, \nabla x \rangle \\ &= (\alpha - 1) \int_0^1 \langle \nabla f^* \circ \nabla g \circ x, \nabla g \circ x \rangle \\ &= (\alpha - 1) \int_0^1 \Big(f(\nabla f^* \circ \nabla g \circ x) + f^*(\nabla g \circ x) \Big). \end{aligned}$$

Again, we use Fenchel's equation here. Jensen's inequality, Proposition 1.14 then yields

$$\begin{aligned} \Big(g(x(1)) - g(x(0)) \Big) / (\alpha - 1) &\geq \int_0^1 f^*(\nabla g \circ x) + f \left(\int_0^1 \nabla f^* \circ \nabla g \circ x \right) \\ &= \int_0^1 f^*(\nabla g \circ x) + f(x(0)). \end{aligned}$$

Letting $a = x(0)$, there must be a $z = x(t) \in [C, \alpha C]$ such that

$$\Big(g(\alpha a) - g(a) \Big) / (\alpha - 1) - f(a) \geq f^*(\nabla g(z)).$$

■

3.1 A Nonsmooth Duality Inequality on Two Functions

As before, the smooth result gives rise to an analogous nonsmooth inequality.

Theorem 3.3 (Borwein, Fitzpatrick) *Let C be a nonempty compact convex subset of $\mathbb{R}^n$, and let f be proper convex lower semicontinuous with* $\mathrm{dom}(f) \subset C$. *If $\alpha \neq 1$ and $g : [C, \alpha C] \to \mathbb{R}$ is Lipschitz then there are $z^* \in \partial g([C, \alpha C])$ and $a \in C$ such that*

$$(g(\alpha a) - g(a))/(\alpha - 1) - f(a) \geq f^*(z^*)$$

Proof: By setting

$$g_n = g * \phi_{1/n}$$
$$f_m = f + (1/m)\|\cdot\|^2,$$

this follows from the smooth version in the same way that Theorem 2.1 follows from Theorem 2.2. ■

3.2 Corollaries to the Two Function Result

Theorem 3.3 gives rise to a host of interesting specializations:

Corollary 3.4 *Let C be a nonempty compact convex subset of $\mathbb{R}^n$, and let f be proper convex lower semicontinuous with* $\mathrm{dom}(f) \subset C$. *If $g : [C, -C] \to \mathbb{R}$ is Lipschitz then there are $z^* \in \partial g([C, -C])$ and $a \in C$ such that*

$$(g(a) - g(-a))/2 - f(a) \geq f^*(z^*).$$

Proof: Set $\alpha = -1$. ■

Note that, in particular, $f^*(z^*) \leq 0$ if f dominates the odd part of g on C.

Corollary 3.5 *Let C, f be as above. If $g : [C, 0] \to \mathbb{R}$ is Lipschitz then there are $z^* \in \partial g([C, 0])$ and $a \in C$ such that*

$$f(a) + f^*(z^*) \leq g(a) - g(0).$$

Proof: Set $\alpha = 0$. ■
In particular, $f^*(z^*) \leq 0$ if f dominates $g - g(0)$ on C.

3.3 Unit Ball Corollaries

With $C = B$ and $f = I_B$, we get particularly interesting inequalities relating g and the size of some $z^* \in \partial g(B)$:

Corollary 3.6 *Let B be the closed unit ball of $\mathbb{R}^n$ and $g : B \to \mathbb{R}$ be Lipschitz. Then for $\alpha \in [-1, 1)$ there is $x^* \in \partial g(B)$ such that*

$$\|x^*\|_* \leq \max_{a \in B} (g(\alpha a) - g(a))/(\alpha - 1).$$

Proof: If $\alpha \in [-1, 1)$, then $[C, \alpha C] = B$. Since $f = I_B$, f is proper convex lower semicontinuous and $f^*(x^*) = \sup\{\langle x^*, x\rangle \mid x \in B\} = \|x^*\|_*$. Then there are $x^* \in \partial g(B)$ and $a \in B$ such that

$$\begin{aligned} \|x^*\|_* &\leq \Big(g(\alpha a) - g(a)\Big)/(\alpha - 1) - I_B(a) \\ &= \Big(g(\alpha a) - g(a)\Big)/(\alpha - 1). \end{aligned}$$

The maximum must be still greater. ■

Two specializations of this last corollary are immediate upon setting $\alpha = -1$ and $\alpha = 0$, respectively.

Corollary 3.7 *Let B be the closed unit ball of $\mathbb{R}^n$ and $g : B \to \mathbb{R}$ be Lipschitz. Then there is $x^* \in \partial g(B)$ such that*

$$\|x^*\|_* \leq \max_{a \in B} (g(a) - g(-a))/2.$$

The right hand side of this corollary is a measure of g's evenness; one consequence of this result is the well-known fact that if g is even on the unit ball then $0 \in \partial g(B)$, i.e. g has a critical point on B.

Corollary 3.8 *Let B be the closed unit ball of $\mathbb{R}^n$ and $g : B \to \mathbb{R}$ be Lipschitz. Then there is $x^* \in \partial g(B)$ such that*

$$\|x^*\|_* \leq \max_{a \in B} g(a) - g(0).$$

This result guarantees the existence of small Clarke subgradients, when g does not greatly exceed $g(0)$ on B.

It is perhaps instructive to examine the fixed point arc x of the operator T given by (3.2) in the setting of Corollaries 3.7 and 3.8: suppose f is the indicator function of $C = B$, g is continuously differentiable, and $\alpha = 0$ or $\alpha = -1$. In this case

$$M := (1 + |\alpha|) \sup\{\|c\| \mid c \in B\} = 1 + |\alpha|$$

and

$$W := \{x : [0, 1] \to B \mid \mathrm{Lip}(x) \leq 1 + |\alpha|\}.$$

The conjugate of the indicator function of the ball is the dual norm; if we consider B to be the Euclidean ball, then $f^*(x) = \|x\|_* = \|x\|$, and $\nabla f^*(x) = x/\|x\|$ for $x \neq 0$. If $\nabla g(x) = 0$ at some point x in the ball, then Corollaries 3.7 and 3.8 hold automatically, so let us assume $\nabla g(x) \neq 0$ for any $x \in B$. The fixed point arc of T satisfies

$$\begin{aligned} x(t) = Tx(t) &= \alpha \int_0^t \nabla f^* \circ \nabla g \circ x + \int_t^1 \nabla f^* \circ \nabla g \circ x \\ &= \alpha \int_0^t \frac{\nabla g(x(s))}{\|\nabla g(x(s))\|} ds + \int_t^1 \frac{\nabla g(x(s))}{\|\nabla g(x(s))\|} ds. \end{aligned}$$

For $\alpha = -1$, the fixed point arc satisfies

$$x(t) = -\int_0^t \frac{\nabla g(x(s))}{\|\nabla g(x(s))\|} ds + \int_t^1 \frac{\nabla g(x(s))}{\|\nabla g(x(s))\|} ds. \tag{3.3}$$

Evaluating at 0 and 1 gives

$$x(0) = \int_0^1 \frac{\nabla g(x(s))}{\|\nabla g(x(s))\|} ds = -x(1),$$

so the fixed point arc stops at the opposite point on the ball to where it started. Furthermore, the condition $M = 1 + |\alpha| = 2$ shows that the arc length of x is less than 2. In particular, if $\|x(0)\| = 1$ then $x(1)$ is diametrically opposite on the sphere, and thus x is a straight line connecting the two points. If $\|x(0)\| < 1$, then the strict convexity of the Euclidean ball shows that $\|x(t)\| < 1$ for all $t \in [0, 1]$. Differentiating x with respect to t, we see that x satisfies the following differential equation:

$$x'(t) = \frac{-\nabla g(x(t))}{\|\nabla g(x(t))\|} + \frac{-\nabla g(x(t))}{\|\nabla g(x(t))\|} = -2\frac{\nabla g(x(t))}{\|\nabla g(x(t))\|}.$$

In summary, the fixed point arc starts at a point a, moves with speed 2 in the direction of greatest decrease of g, and stops at the opposite point $-a$.

On the other hand, for $\alpha = 0$ the fixed point arc satisfies

$$x(t) = \int_t^1 \frac{\nabla g(x(s))}{\|\nabla g(x(s))\|} ds. \tag{3.4}$$

It is clear that $x(1) = 0$, and $M = 1 + |\alpha| = 1$ implies that the arc length of x is less than 1. Differentiating equation (3.4) with respect to t gives

$$x'(t) = -\frac{\nabla g(x(t))}{\|\nabla g(x(t))\|}$$

as well. So x is a smooth path of length 1 that moves with unit speed in the direction of greatest decrease of g, and stops at the origin.

In the next section, we give a result strengthening Corollary 3.8 whose proof employs Ekeland's variational principle.

3.4 Ekeland's Variational Principle

Given a function F and a point that is nearly infimal for F, a *variational principle* guarantees the existence of a nearby point that is the strict minimum of a slightly perturbed function. In the case of Ekeland's theorem, we perturb the function by a scaled translate of the norm. There are other useful variational principles that perturb F by smoother functions, like sums of powers of the norm, or Fréchet bump functions—see [DGZ93] and [BP87] for details.

Ekeland's variational principle is a powerful tool in optimization and analysis, with applications to fixed point theory, partial differential equations and nonlinear analysis. Although Ekeland's theorem holds for any complete metric space, for clarity we will consider the case of interest to us, Banach space (see [Eke74]).

Theorem 3.9 (Ekeland) *Let $F : X \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function on a Banach space X, with F bounded below on X. Given $\varepsilon, \lambda > 0$ and a point u such that*

$$F(u) \leq \inf F + \varepsilon,$$

there exists a point $v \in X$ such that:

1. $F(v) \leq F(u)$,

2. $\|u - v\| \leq \lambda$,

3. and

$$F(w) > F(v) - \varepsilon/\lambda\|v - w\|$$

for all $w \neq v$.

To prove this, consider a relation $\prec$ on $X \times \mathbb{R}$ by

$$(x_1, t_1) \prec (x_2, t_2) \text{ iff } (t_2 - t_1) + \alpha\|x_2 - x_1\| \leq 0$$

for some fixed $\alpha > 0$. Then $\prec$ is reflexive, since

$$(t_1 - t_1) + \alpha\|x_1 - x_1\| = 0.$$

The relation $\prec$ is also antisymmetric: if $(x_1, t_1) \prec (x_2, t_2)$ and $(x_2, t_2) \prec (x_1, t_1)$ then

$$(t_2 - t_1) + \alpha\|x_2 - x_1\| \leq 0$$

and

$$(t_1 - t_2) + \alpha\|x_1 - x_2\| \leq 0,$$

and adding these two inequalities yields

$$\|x_1 - x_2\| \leq 0,$$

so $x_1 = x_2$. But then $0 \leq t_2 - t_1 \leq 0$, so $t_1 = t_2$. Finally, $\prec$ is transitive: if $(x_1, t_1) \prec (x_2, t_2)$ and $(x_2, t_2) \prec (x_3, t_3)$ then

$$t_3 - t_1 + \alpha\|x_3 - x_1\| \leq (t_3 - t_2) + (t_2 - t_1) + \alpha\|x_3 - x_2\| + \alpha\|x_2 - x_1\| \leq 0$$

so $(x_1, t_1) \prec (x_3, t_3)$.

Note also that $\prec$ is continuous in the sense that sets of the form

$$S = \{(x, t) \mid (x, t) \succ (y, s)\}$$

are closed, where (y, s) is a fixed element of X. For if $(x_n, t_n) \in S$ for all n, with $x_n \to x$ and $t_n \to t$, then

$$t_n - s + \alpha\|x_n - y\| \leq 0$$

for all n. Taking limits gives

$$t - s + \alpha\|x - y\| \leq 0.$$

Before we prove Theorem 3.9, we introduce a technical lemma:

Lemma 3.10 (Ekeland) *Suppose $S \subset X \times \mathbb{R}$ is closed and bounded below: there is an $m \in \mathbb{R}$ such that*

$$(x, t) \in S \implies t \geq m.$$

Then for every $(x, t) \in S$ there is a $(\overline{x}, \overline{t}) \in S$ that is maximal in S with respect to $\prec$, and $(\overline{x}, \overline{t}) \succ (x, t)$.

Proof: We will recursively construct a sequence (x_n, t_n) as follows. Given (x_n, t_n), define S_n and m_n by

$$\begin{aligned} S_n &= \{(x, t) \in S | (x, t) \succ (x_n, t_n)\}, \\ m_n &= \inf_{(x,t) \in S_n} t. \end{aligned}$$

Note that $m_n \geq m$, since $t \geq m$ for all $(x, t) \in S_n \subset S$. Now pick $(x_{n+1}, t_{n+1}) \in S_n$ to satisfy

$$(t_n - t_{n+1}) \geq 1/2(t_n - m_n).$$

This yields

$$\begin{aligned} |t_{n+1} - m_{n+1}| &= t_{n+1} - m_{n+1} \leq t_{n+1} - m_n \\ &\leq 1/2(t_n + m_n) - m_n = 1/2(t_n - m_n) \leq 1/2^n(t_1 - m). \end{aligned}$$

For any $(x, t) \in S_{n+1}$ we have

$$m_{n+1} \leq t \leq t_{n+1}$$

so

$$\begin{aligned} |t - m_{n+1}| &= t - m_{n+1} \leq t_{m+1} - m_{n+1} \leq 1/2^n(t_1 - m) \\ \|x - x_{n+1}\| &\leq 1/\alpha|t - t_{n+1}| \leq 1/\alpha|t - m_{n+1}| \\ &\leq 1/2^n 1/\alpha(t_1 - m). \end{aligned}$$

This shows that the diameter of S_n goes to 0, but $S_{n+1} \subset S_n$ and each S_n is closed and nonempty, so there is a unique element $(\overline{x}, \overline{t})$ in

$$\bigcap_n S_n.$$

Furthermore, by definition

$$(\overline{x}, \overline{t}) \succ (x_n, t_n)$$

for all n; in particular, this is true for $n = 1$. Now suppose $(\tilde{x}, \tilde{t}) \succ (\overline{x}, \overline{t})$. Then $(\tilde{x}, \tilde{t}) \succ (x_n, t_n)$ for all n, by transitivity, and

$$(\tilde{x}, \tilde{t}) \in \bigcap_n S_n = \{(\overline{x}, \overline{t})\}.$$

So $(\overline{x}, \overline{t})$ is maximal. ■

To prove Theorem 3.9, we now apply Lemma 3.10 to

$$S = \operatorname{epi}(F) = \{(x, t) \in X \mid F(x) \leq t\},$$

$$(x_1, t_1) = (u, F(u))$$

and

$$\alpha = \varepsilon/\lambda.$$

S is closed, since F is lower semicontinuous. We conclude that there is a maximal $(v, t) \in \operatorname{epi}(F)$ with $(v, t) \succ (u, F(u))$. Since (v, t) is maximal and $(v, F(v)) \succ (v, t)$ for any $(v, t) \in \operatorname{epi}(F)$, it must be that $t = F(v)$. Furthermore, $(v, F(v)) \succ (u, F(u))$ implies that

$$F(v) \leq F(u) - \varepsilon/\lambda\|u - v\| \leq F(u),$$

and

$$F(v) \geq \inf_X F \geq F(u) - \varepsilon$$

implies

$$\|u - v\| \leq \lambda/\varepsilon(F(u) - F(v)) \leq \lambda.$$

Finally, if $w \neq v$ then $(w, F(w)) \not\succ (v, F(v))$ by the maximality of $(v, F(v)$:

$$F(w) - F(v) + \varepsilon/\lambda\|w - v\| > 0,$$

or

$$F(v) < F(w) + \varepsilon/\lambda\|w - v\|.$$

3.5 A Nonsmooth Maximum Principle

Our next result uses Ekeland's variational principle to link the growth of a Lipschitz function on the boundary of a set S to the size of its smallest subgradient on the interior of S.

Theorem 3.11 (Borwein, Fitzpatrick) *Let S be the closure of a nonempty open bounded set in a Banach space X and let $g : S \to \mathbb{R}$ be Lipschitz. If $x \in \operatorname{int} S$ and*

$$t := \inf \{\|z^*\|_* \mid z^* \in \partial g(z), z \in \operatorname{int} S\} > 0$$

then

$$\sup_{u \in \partial S} (g(u) - t\|u - x\|) \geq g(x).$$

Proof: Let $0 < \alpha < t$ and $0 < \varepsilon < t - \alpha$. Setting $h(y) = \alpha\|y - x\| - g(y)$, we see

$$\begin{aligned} h(y) &= \alpha\|y - x\| - g(y) \\ &\geq -g(y) \geq -g(x) - \operatorname{Lip}(g)\|x - y\| \\ &\geq -g(x) - 2 \cdot D \cdot \operatorname{Lip}(g), \end{aligned}$$

where $D = \sup_{x \in S} \|x\|$. Since h is bounded below, we may apply Ekeland's Principle. Thus, there is $u \in S$ such that

$$h(u) \leq h(y) + \varepsilon\|u - y\| \text{ for all } y \in S.$$

Suppose $u \in \operatorname{int} S$. u minimizes $h + \varepsilon\|u - \cdot\|$ over S, so

$$\begin{aligned} 0 &\in \partial(h + \varepsilon\|u - \cdot\|)(u) \\ &\subset \partial h(u) + \partial(\varepsilon\|u - \cdot\|)(u) \\ &= \partial h(u) + \varepsilon B^* \\ &= \partial(\alpha\| \cdot - x\| - g)(u) + \varepsilon B^* \\ &\subset -\partial g(u) + \alpha B^* + \varepsilon B^* \\ &= -\partial g(u) + (\alpha + \varepsilon) B^*, \end{aligned}$$

using the properties of the Clarke subdifferential (Proposition 1.6). Thus, $0 \in -\partial g(u) + (\alpha + \varepsilon)B^*$, so there is $u^* \in \partial g(u)$ with $\|u^*\|_* \leq (\alpha + \varepsilon) < t$. This contradicts the definition of t, so $u \in \partial S$.

Since u minimizes $h + \varepsilon\|u - \cdot\|$ we have $h(u) \leq h(x) + \varepsilon\|u - x\|$, or

$$\alpha\|u - x\| - g(u) \leq \alpha\|x - x\| - g(x) + \varepsilon\|u - x\|.$$

This implies that

$$\sup_{u \in \partial X} \{g(u) - (\alpha - \varepsilon)\|u - x\|\} \geq g(x).$$

Letting $\alpha \to t$ will make $\varepsilon \to 0$, and the result follows by uniform convergence. ∎

The condition $t > 0$ means that g has no approximate critical points on the interior of S, and the conclusion says that for every $\varepsilon > 0$ and $x \in \operatorname{int} S$ there is a point u on the boundary of S such that

$$g(u) - g(x) \geq t\|u - x\| - \varepsilon$$

so Theorem 3.11 imposes a kind of minimum growth condition on g. Among other things, it is immediate that

$$\sup_S g = \sup_{\partial S} g,$$

hence the term "nonsmooth maximum principle". When g is constant on the boundary of S, we can use Theorem 3.11 to show that g has an approximate critical point:

Corollary 3.12 *In the setting of Theorem 3.11, suppose that*

$$g(x) = a \quad \forall x \in \partial S.$$

Then

$$t := \inf\{\|z^*\|_* \mid z^* \in \partial g(z), z \in \operatorname{int} S\} = 0. \tag{3.5}$$

Proof: Note that if $f := -g$ then $\partial f = -\partial g$, so

$$\inf\{\|z^*\|_* \mid z^* \in \partial f(z), z \in \operatorname{int} S\} = t.$$

Now suppose that $t > 0$. Then we may apply Theorem 3.11 to f, g and some $x \in \operatorname{int} S$ to get

$$\sup_{u \in \partial S}\{g(u) - t\|u - x\|\} \geq g(x)$$

and

$$\sup_{u \in \partial S}\{-g(u) - t\|u - x\|\} \geq -g(x)$$

Adding these two together, we see

$$\sup_{u \in \partial S}\{g(u) - t\|u - x\|\} \geq \inf_{u \in \partial S}\{g(u) + t\|u - x\|\}.$$

Now $x \in \operatorname{int} S$, so there is an $M > 0$ such that $\|u - x\| > M$ for all $u \in \partial S$:

$$\begin{aligned} a - tM = \sup_{u \in \partial S} g(u) - tM &\geq \sup_{u \in \partial S} \{g(u) - t\|u - x\|\} \\ &\geq \inf_{u \in \partial S} \{g(u) + t\|u - x\|\} \geq \inf_{u \in \partial S} g(u) + tM = a + tM \end{aligned}$$

So $t \leq 0$, contradicting our assumption. ■

Several similar variationally derived theorems treating Gâteaux differentiable functions can be found in [AGJ97].

3.5.1 A Variational Improvement to Corollary 3.8

We can also use Theorem 3.11 to get the following corollary, upon setting $S = B$ and $x = 0$:

Corollary 3.13 *Let B be the closed unit ball in a Banach space X and let $g : B \to \mathbb{R}$ be Lipschitz. Then*

$$\inf\{\|z^*\|_* \mid z^* \in \partial g(z), \|z\| < 1\} \leq \max\left(\sup_{\partial B} g - g(0), 0\right). \tag{3.6}$$

Proof: The left hand side of (3.6) equals t, since $S = B$. If $t > 0$ then Theorem 3.11 gives us:

$$\begin{aligned} \sup_{\partial B} g - t &= \sup_{\partial B}\{g - t\} \\ &= \sup_{\partial B}\{g - t\| \cdot - 0\|\} \\ &\geq g(0) \end{aligned}$$

so that $t \leq \sup_{\partial B}(g - g(0))$. ■

Note that this result substantially strengthens Corollary 3.8, since the supremum is taken over the boundary of the ball rather than over the whole of the unit ball. In [BF01], Borwein and Fitzpatrick noted this refinement made possible by Ekeland's theorem, and wondered if Corollary 3.7 could be strengthened in a similar way. That is, they asked whether or not the following conjecture is true:

Conjecture 1 *Let B be the closed unit ball of $\mathbb{R}^n$ and $g : B \to \mathbb{R}$ be Lipschitz. Then there is $x^* \in \partial g(B)$ such that*

$$\|x^*\|_* \leq \max_{a \in \partial B} (g(a) - g(-a))/2.$$

Such a result would be a striking multidimensional version of Rolle's Theorem. Indeed, for $n = 1$ the conjecture actually *is* Rolle's Theorem, so it is true in that case. It is well known, however, that this version of Rolle's Theorem is not true in infinite dimensions. In [Fer96], for example, Ferrer constructed a continuously differentiable globally Lipschitz operator T on the Hilbert space l_2 of square summable real sequences such that $f = 0$ on the unit sphere of l^2, but $\nabla f \neq 0$ anywhere on the unit ball. We reproduce his construction in the next section. Then, in the next chapter, we construct a very strong counterexample to Conjecture 1, showing it to be false even in $\mathbb{R}^2$. This suggests that a variational proof of Corollary 3.7 is unlikely to be discovered.

3.6 Ferrer's Construction

Let $L, R : l_2 \to l_2$ be the linear operators on l_2 defined by

$$Lx = (x_2, x_3, x_4, \ldots)$$
$$Rx = (0, x_1, x_2, x_3, \ldots)$$

for sequences $x \in l_2$. That is, L and R are the left and right shift operators, respectively. They are clearly bounded, since $\|Lx\| \leq \|x\|$ and $\|Rx\| = \|x\|$, so they are continuous. They are also adjoint to each other, since

$$\langle x, Ru \rangle = \langle Lx, u \rangle$$

for all $x, u \in l_2$. Now define the continuous function $T : l_2 \to l_2$ by

$$T(x) = (1/2 - \|x\|^2)e_1 + Rx,$$

where $e_1 = (1, 0, 0, 0, \ldots) \in l_2$. If $T(x) = x$ for $x \in l_2$, then

$$1/2 - \|x\|^2 = x_1 = x_2 = x_3 = x_4 = \ldots, \tag{3.7}$$

so $\|x\|^2 - 1/2 \neq 0$ implies $\lim_{n\to\infty} \neq 0$, contrary to the assumption that $x \in l^2$. On the other hand, if $\|x\|^2 - 1/2 = 0$, then Equation (3.7) implies $x = 0$, so

$$0 = \|x\|^2 - 1/2 = 0 - 1/2 = -1/2,$$

a contradiction. Therefore T has no fixed points in l_2.

The function $f : l_2 \to \mathbb{R}$ referred to at the end of the previous section is defined by

$$f(x) = \frac{1 - \|x\|^2}{\|x - T(x)\|^2}. \tag{3.8}$$

Then f is continuous, since it is the quotient of continuous functions, and the denominator on the left hand side of Equation (3.8) is never zero. Clearly $f|_S = 0$, where S is the unit sphere in l_2. We wish to show that f is Fréchet differentiable, and to determine the gradient ∇f of f. The function $x \to \|x\|^2$ is Fréchet differentiable, with gradient $2x$, and a continuous linear operator $A : l_2 \to l_2$ has Fréchet derivative A, so both the numerator and denominator of f are differentiable, since they sums and compositions of these two kinds of functions. But then f is clearly Fréchet differentiable, being the quotient of two Fréchet functions whose denominator is never zero. To compute the gradient, first note that

$$T'(x)u = Ru - 2\langle x, u\rangle e_1.$$

The quotient rule for f gives:

$$\langle \nabla f(x), u\rangle = \frac{1}{\|x - T(x)\|^4} \times \left[-2\langle x, u\rangle \|x - T(x)\|^2 - 2\langle x - T(x), u - T'(x)u\rangle\right]$$

for any $x, u \in l_2$. Since L is the adjoint of R, $\langle T(x), e_1\rangle = 1/2 - \|x\|^2$ and $L(T(x)) = x$, we have

$$\begin{aligned}
\langle x - T(x), u - T'(x)u\rangle &= \langle x - T(x), u\rangle + \langle x - T(x), 2\langle x, u\rangle e_1 - Ru\rangle \\
&= \langle x - T(x), u\rangle + 2\langle x, u\rangle x_1 - 2\langle x, u\rangle(1/2 - \|x\|^2) - \langle x - T(x), Ru\rangle \\
&= \langle x - T(x) + 2x_1 x - x(1 - 2\|x\|^2) - L(x - T(x)), u\rangle \\
&= \langle (1 + 2x_1 + 2\|x\|^2)x - T(x) - Lx, u\rangle.
\end{aligned}$$

Thus

$$\begin{aligned}
\langle \nabla f(x), u\rangle = \frac{-2}{\|x - T(x)\|^4} \times \langle (\|x - T(x)\|^2 + (1 - \|x\|^2)(1 + 2x_1 + 2\|x\|^2))x \\
- (1 - \|x\|^2)(Lx + T(x)), u\rangle,
\end{aligned}$$

so that

$$\nabla f(x) = \frac{-2}{\|x - T(x)\|^4} \times \left(\|x - T(x)\|^2 + (1 - \|x\|^2)(1 + 2x_1 + 2\|x\|^2)\right)x - (1 - \|x\|^2)(Lx + T(x)).$$

Suppose there is a sequence x in the interior of the unit ball of l_2 such that $\nabla f(x) = 0$. Then

$$Lx + T(x) = sx,$$

where

$$s = \frac{\|x - T(x)\|^2}{1 - \|x\|^2} + 1 + 2x_1 + 2\|x\|^2. \tag{3.9}$$

Applying L to both sides gives

$$L^2x - sLx + x = 0,$$

so x satisfies the second-order linear recurrence relation

$$x_{n+2} - sx_{n+1} + x_i = 0, \quad n \geq 1,$$

whose characteristic equation is

$$t^2 - st + 1 = 0.$$

There are three types of solutions to the recurrence, depending on the discriminant of the characteristic equation:

Case 1: $|s| = 2$.

Then the sequences $u_n = (s/2)^{n-1}$ and $v_n = (n-1)(s/2)^{n-1}$ are basic elements of $\mathrm{Ker}(L^2x - sLx + x)$. So there are real constants C_1, C_2 such that

$$x = C_1u + C_2v.$$

But $x \in l_2$ implies $\lim_{n\to\infty} x_n = 0$, so $C_1 = C_2 = 0$. Then $x = 0$, and yet $\nabla f(0) = 16e_1$, contrary to assumption.

Case 2: $|s| < 2$

In this case, the roots of the characteristic equation come in conjugate pairs

$$\alpha = \cos\theta + i\sin\theta, \quad \beta = \cos\theta - i\sin\theta$$

where $\sin\theta \neq 0$. Then for $x \in \mathrm{Ker}(L^2x - sLx + x)$ there are complex constants C_1 and C_2 such that

$$x_n = C_1\alpha^{n-1} + C_2\beta^{n-1}, \quad n \geq 1.$$

Since x is a real sequence, there exist real constants C_3 and C_4 such that

$$x_n = C_3 \cos(n-1)\theta + C_4 \sin(n-1)\theta, \quad n \geq 1,$$

using DeMoivre's theorem. But $\sin\theta \neq 0$ implies this sequence has no limit, contrary to the assumption that $x \in l_2$.

Case 3: $|s| > 2$.

In this case, the characteristic equation has two distinct real roots α and β whose product is 1. Without loss of generality, let $|\alpha| > 1$, $|\beta| < 1$. Then

$$x_n = C_1\alpha^{n-1} + C_2\beta^{n-1}, \quad n \geq 1$$

for some real C_1 and C_2. But $\lim_{n\to\infty} x_n = 0$ implies $C_1 = 0$, so x is the geometric sequence

$$x_n = x_1\beta^{n-1}, \quad n \geq 1.$$

Also, the familiar formula for the geometric series gives

$$\|x\|^2 = \sum_{n=1}^{\infty} x_1^2(\beta^2)^{n-1} = \frac{x_1^2}{1-\beta^2}$$

and

$$\begin{aligned}\|x - T(x)\|^2 &= \left(x_1 + \frac{x_1^2}{1-\beta^2} - 1/2\right)^2 + \sum_{n=1}^{\infty}(x_1\beta^{n-1} - x_1\beta^n)^2 \\ &= \left(x_1 + \frac{x_1^2}{1-\beta^2} - 1/2\right)^2 + \frac{x_1^2(1-\beta)}{1+\beta}.\end{aligned}$$

From $Lx + T(x) - sx = 0$ and $\beta + \frac{1}{\beta} = s$ we have the following quadratic equation in x_1:

$$x_1^2 + \frac{1-\beta^2}{\beta}x_1 - \frac{1}{2}(1-\beta^2) = 0. \tag{3.10}$$

But then

$$x_1 + \frac{x_1^2}{1-\beta^2} - 1/2 = \left(1 - \frac{1}{\beta}\right)x_1$$

so

$$\|x - T(x)\|^2 = \left(1 - \frac{1}{\beta}\right)^2 x_1^2 + \frac{x_1^2(1-\beta)}{1+\beta} = \frac{x_1^2(1-\beta)}{\beta^2(1+\beta)}.$$

Substituting these values into the definition of s, Equation (3.9), we get

$$\beta + \frac{1}{\beta} = s = \frac{x_1^2(1-\beta)}{\beta^2(1+\beta)} \cdot \frac{1-\beta^2}{1-\beta^2-x_1^2} + 1 + 2x_1 + 2\frac{x_1^2}{1-\beta^2};$$

upon collecting terms we get

$$1 = (\beta - 2x_1)\left(1 + \frac{(1-\beta^2)(1-\beta)}{2\beta^2(1-x_1^2-\beta^2)}\right). \tag{3.11}$$

There are two solutions to the quadratic equation (3.10). If

$$x_1 = \frac{-1+\beta^2-\sqrt{1-\beta^4}}{2\beta},$$

then Equation (3.11) gives $0 < \beta - 2x_1 < 1$, since $\|x\| < 1$ implies $x_1^2 + \beta^2 < 1$. But then $0 < \big(1+\sqrt{1-\beta^4}\big)/\beta < 1$, which is impossible, since $|\beta| < 1$. On the other hand, suppose

$$x_1 = \frac{-1+\beta^2+\sqrt{1-\beta^4}}{2\beta}.$$

Rearranging Equation (3.10) gives

$$1 - x_1^2 - \beta^2 = \frac{1}{2\beta}(1-\beta^2)(\beta+2x_1),$$

so putting these two values into Equation (3.11) gives

$$\begin{aligned}
1 &= \frac{1}{\beta}\big(1-\sqrt{1-\beta^4}\big)\frac{2\beta^2-\beta+\sqrt{1-\beta^4}}{2\beta^2-1+\sqrt{1-\beta^4}} \\
\beta\big(2\beta^2-1+\sqrt{1-\beta^4}\big) &= \big(1-\sqrt{1-\beta^4}\big)\big(2\beta^2-\beta+\sqrt{1-\beta^4}\big) \\
2\beta^3 &= \big(1-\sqrt{1-\beta^4}\big)\big(2\beta^2+\sqrt{1-\beta^4}\big) \\
2\big(1+\sqrt{1-\beta^4}\big) &= \beta\big(2\beta^2+\sqrt{1-\beta^4}\big) \\
2(1-\beta^3) &= (\beta-2)\sqrt{1-\beta^4},
\end{aligned}$$

which is a contradiction, since the left side of the last line above is positive, but the right side is negative.

Chapter 4

A counterexample to a conjecture by Borwein and Fitzpatrick

At the end of the previous chapter we considered a problem posed by Borwein and Fitzpatrick in [BF01]; namely, can we strengthen the Rolle-type duality inequality:

Corollary 3.7 *Let B be the closed unit ball of $\mathbb{R}^n$ and $g : B \to \mathbb{R}$ be Lipschitz. Then there is $x^* \in \partial g(B)$ such that*

$$\|x^*\|_* \leq \max_{a \in B} (g(a) - g(-a))/2 \tag{4.1}$$

so that the maximum is taken over the unit sphere, ∂B, instead of the unit ball? Suppose that such a restriction is possible and let $G : B \to \mathbb{R}$ be Lipschitz and even on the the sphere:

$$G(x) = G(-x) \text{ for all } x \in \partial B.$$

Then

$$\max_{a \in \partial B} (G(a) - G(-a))/2 = 0,$$

so there would have to be $x \in B$ with $0 \in \partial G(x)$. However, in [BKW02] Borwein, Kortezov and Wiersma construct a C^1 function G on the unit ball in $\mathbb{R}^2$ such that G is even on the unit circle, but ∇G is nowhere 0 on the ball. Thus, the answer to the question is strongly negative: We cannot restrict the maximum to the sphere, even when G is continuously differentiable and $n = 2$. Since our conjecture is false, it seems likely that no variational proof of Corollary 3.7 will be found: the result seems to rely strongly on a fixed point argument. Recall that Corollary 1 compares the norm of a subgradient of g to the odd part of g, which suggests a link with the *Borsuk-Ulam* theorem (see [BL00]):

Theorem 4.1 (Borsuk-Ulam) *For any positive integers $m < n$, if the function $f : S_n \to \mathbb{R}^m$ is continuous then there is a point x in S_n satisfying $f(x) = f(-x)$.*

In the above, S_n is the Euclidean unit sphere in $\mathbb{R}^n$. Since the Borsuk-Ulam theorem implies Brouwer's fixed point theorem, we have further evidence of the inherently fixed-point theoretic nature of Corollary 3.7

The remainder of this chapter is devoted to constructing the counter-example function G alluded to before.

4.1 Notation

Throughout this chapter $\|\cdot\|$ refers to the 2-norm in $\mathbb{R}^2$. We denote by B the closed unit 2-ball centered at 0, by O the interior of B, and $S = B \backslash O$. Whenever $x \in \mathbb{R}^2$ is expressed as $x = (\cdot, \cdot)$, we mean its polar coordinates, whereas $x = (\cdot, \cdot)_*$ will be used for Cartesian coordinates. For $\phi \in \mathbb{R}$, let $s(\phi) \in S$ denote the point of argument ϕ:

$$s(\phi) = (\sin(\phi), \cos(\phi))_* .$$

As in previous chapters, for $x, y \in \mathbb{R}^2$, $[x, y]$ denotes the closed line-segment with endpoints x and y. Finally, χ_S denotes the characteristic function of the set S:

$$\chi_S(x) := \begin{cases} 1 & x \in S \\ 0 & \text{otherwise} \end{cases}$$

4.2 Construction

The desired function G is constructed in two stages: first we define a function G_0 with no critical points on the sphere, and then we apply Lemma 4.2 to G_0 in order to even it out on S without introducing any new critical points on B. The construction is quite technical and so we accompany all steps with pictures.

4.2.1 Defining Calderas

We start by defining a function F_θ that generates a caldera-shaped graph, and we center translates of F_θ at three different points in the plane: at $\mathbf{A} = (\sqrt{2}, \pi/4)$, $\mathbf{B} = (\sqrt{2}, \pi)$, and $\mathbf{C} = (\sqrt{2}, 3\pi/2)$ (Figure 4.2). F_θ is defined as follows.

Let $f : [0, +\infty) \to \mathbb{R}$ be defined by

$$f(\rho) := \begin{cases} \cos^2(\frac{5\pi}{6}(\rho - 1)) & \rho \in [0.4, 1.6] \\ 0 & \rho \in [0, 0.4] \cup [1.6, +\infty) \end{cases}$$

Note that:

1. $f \in C^1([0, +\infty))$,
2. $f'|_{(0.4,1)} > 0$,
3. $f'|_{(1,1.6)} < 0$,
4. $f(1 + \rho) = f(1 - \rho)$ for $\rho \in [0, 1]$.

Given some $\theta \in \mathbb{R}$, define $F_\theta \in C^1(\mathbb{R}^2)$ by

$$F_\theta((\sqrt{2}, \theta) + (\rho, \phi)) := f(\rho);$$

This defines a caldera centered at $(\sqrt{2}, \theta)$ (Figure 4.1). Let

$$F := F_{\frac{\pi}{4}} - F_\pi - F_{\frac{3\pi}{2}};$$

this defines one upright caldera centered at $\mathbf{A}$ and two inverted calderas at $\mathbf{B}$ and $\mathbf{C}$ (Figure 4.3). Since calderas $\mathbf{B}$ and $\mathbf{C}$ are inverted, the critical points of F in the first quadrant are exactly the critical points of $F_{\frac{\pi}{4}}$, i.e. the rim of caldera $\mathbf{A}$.

In the second quadrant, the critical points of F are a subset of the critical points of F_π, since caldera $\mathbf{C}$ removes those critical points of F_π that are within the support of $F_{\frac{3\pi}{2}}$. By symmetry, the critical points of F in the fourth quadrant are exactly those critical points of $F_{\frac{3\pi}{2}}$ not in the support of F_π.

In the third quadrant, we have a critical point only at $(1, \frac{5\pi}{4})$.

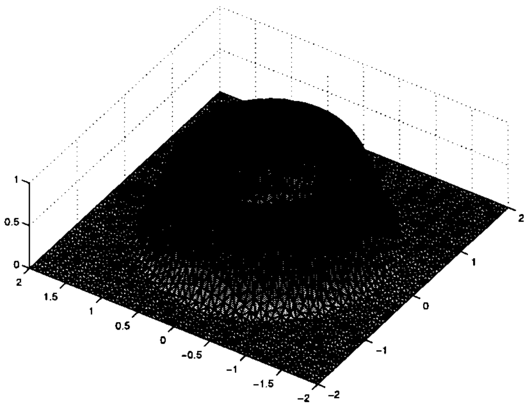

Figure 4.1: The function F_θ, translated to the origin

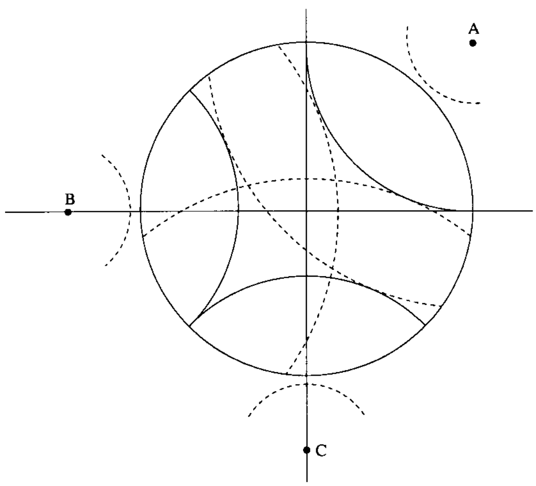

Figure 4.2: The three calderas are centered at **A**, **B** and **C**, and the rim of each caldera is indicated with solid arcs. The support of each caldera is delimited by two dashed arcs.

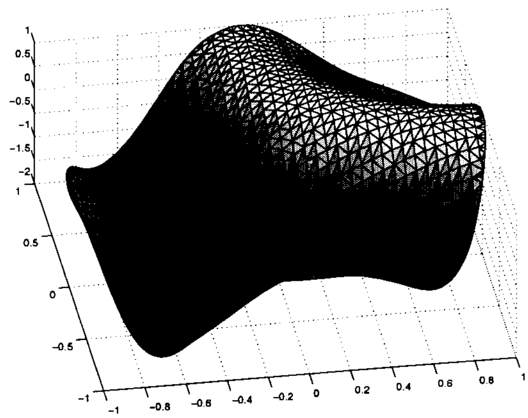

Figure 4.3: F, the sum of three calderas.

4.2.2 Removing Critical Points

Each of the four quadrants contain critical points that we wish to remove. In the first quadrant, we add a thin rising ridge to the top of caldera **A**. In the second and fourth quadrants, we add a narrow gorge to the bottom of calderas **B** and **C**. The following function f_0 will serve as the radial component of these topographical features.

Define $f_0 : [0, +\infty) \to \mathbb{R}$ by

$$f_0(\rho) := \begin{cases} \frac{\cos^2(50\pi(\rho-1))}{5} & \rho \in [0.99, 1.01] \\ 0 & \rho \in [0, 0.99] \cup [1.01, +\infty) \end{cases}$$

Then we have:

1. $f_0 \in C^1([0, +\infty))$,
2. $f_0'(\rho) > 0$ for $\rho \in (0.99, 1)$,
3. $f_0'(\rho) < 0$ for $\rho \in (1, 1.01)$,
4. $f_0(1 + \rho) = f(1 - \rho)$ for $\rho \in [0, 1]$.

In the second quadrant, there is a flat section on a region of $\mathbf{B} + S$. We add a narrow circular valley to the rim of caldera **B** to remove these critical points. The angular component of this valley is defined by the following function.

Let $p_1 \in \mathbf{C}^1(\mathbb{R})$ be a 2π-periodic function such that:

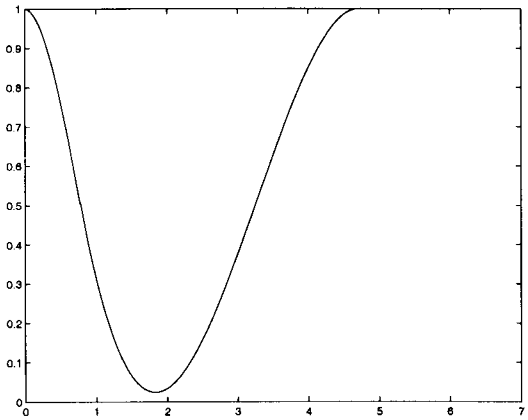

Figure 4.4: The function p_1.

1. $p_1|_{[-\frac{\pi}{4},\frac{\pi}{4}]} \geq \frac{1}{2}$,

2. $p_1|_{[-\frac{\pi}{2},0]} = 1$,

3. $p_1'|_{(0,\frac{\pi}{4}]} < 0$,

4. $p_1(\phi) = \frac{2-\pi}{4} - \phi$ for $\phi \in [\frac{\pi}{4} - 0.01, \frac{\pi}{4}]$

(see Figure 4.4). Define $G_1 \in \mathbf{C}^1(\mathbb{R}^2)$ by

$$G_1(\mathbf{B} + (\rho, \phi)) := -f_0(\rho)p_1(\phi).$$

This determines a gorge on the bottom of caldera $\mathbf{B}$, which slopes down from a height of $-\frac{1}{2}$ at $\mathbf{B} + (1, \frac{\pi}{4})$ to a height of -1 at $\mathbf{B} + (1, -\frac{\pi}{4})$ (Figure 4.5). Note that we only care about the values of p_1 on the interval $[-\frac{\pi}{4}, \frac{\pi}{4}]$, since this is the angular range of the region $(\mathbf{B} + S) \cap B$. In the fourth quadrant, we define a gorge on the flat spots of caldera $\mathbf{C}$ by taking the mirror image of G_1 along the $y = x$ line of the plane. Let $p_2(\phi) := p_1\left(\frac{\pi}{2} - \phi\right)$ and define $G_2 \in \mathbf{C}^1(\mathbb{R}^2)$ by

$$G_2(\mathbf{C} + (\rho, \phi)) := -f_0(\rho)p_2(\phi).$$

By symmetry, this defines a gorge on the bottom of caldera $\mathbf{C}$ that takes a value of $-\frac{1}{2}$ at $\mathbf{C} + (1, \frac{\pi}{4})$ and decreases to -1 at $\mathbf{C} + (1, \frac{3\pi}{4})$.

The first quadrant has critical points along the length of the rim of caldera $\mathbf{A}$, that is, on $(\mathbf{A} + S) \cap B$. To remove these points, we put a ridge on $\mathbf{A} + S$ rising upward along the arc from $\mathbf{A} + (1, \pi)$ to $\mathbf{A} + (1, \frac{3\pi}{2})$. Let $p_3 \in \mathbf{C}^1(\mathbb{R})$ be a 2π-periodic function such that

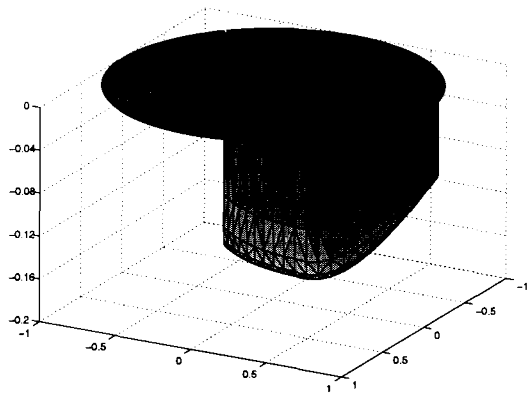

Figure 4.5: The function G_1

1. $p_3|_{[\pi,\frac{3\pi}{2}]} \geq \frac{1}{2}$,
2. $p_3'|_{[\pi,\frac{3\pi}{2}]} > 0$,
3. $p_3(\phi) = \frac{1}{2} - \pi + \phi$ for $\phi \in [\pi, \pi + 0.01]$

(see Figure 4.6).

We create the desired ridge with the function $G_3 \in \mathbf{C}^1(\mathbb{R}^2)$ defined by

$$G_3(\mathbf{A} + (\rho, \phi)) := f_0(\rho)p_3(\phi) \quad \text{(Figure 4.8.}$$

In the third quadrant, we must remove the critical point at $(1, \frac{5\pi}{4})$. To accomplish this we add a function G_4 that has positive slope at $(1, \frac{5\pi}{4})$ in the direction $(1, \frac{\pi}{4})$. The definition of G_4 is as follows.

Let $p_4 \in \mathbf{C}^1(\mathbb{R})$ be a 2π-periodic function such that $p_4|_{[0,\frac{\pi}{8}]} \geq 0$, $p_4'|_{[0,\frac{\pi}{8})} < 0$, and $p_4|_{[\frac{\pi}{8},\frac{15\pi}{8}]} = 0$. We use p_4 to generate the angular component of G_4 (Figure 4.6).

We define $G_4 \in \mathbf{C}^1(\mathbb{R}^2)$ piecewise on a partition of B. On $(\mathbf{B} + B) \cap B$, G_4 is defined by

$$G_4(\mathbf{B} + (\rho, \phi)) := -f_0(\rho)p_4\left(\phi + \frac{\pi}{4}\right) \text{ for } \rho \leq 1.$$

On $(\mathbf{C} + B) \cap B$, G_4 is defined by

$$G_4(\mathbf{C} + (\rho, \phi)) := -f_0(\rho)p_4\left(\frac{3\pi}{4} - \phi\right) \text{ for } \rho \leq 1.$$

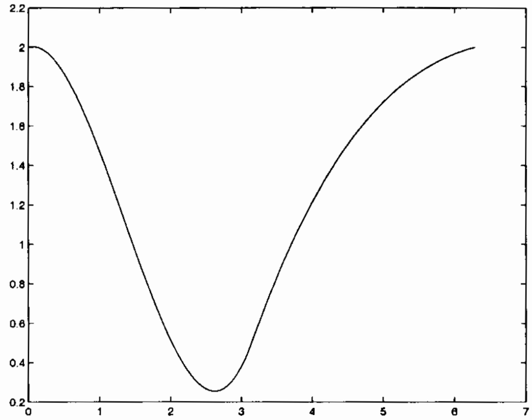

Figure 4.6: The function p_3.

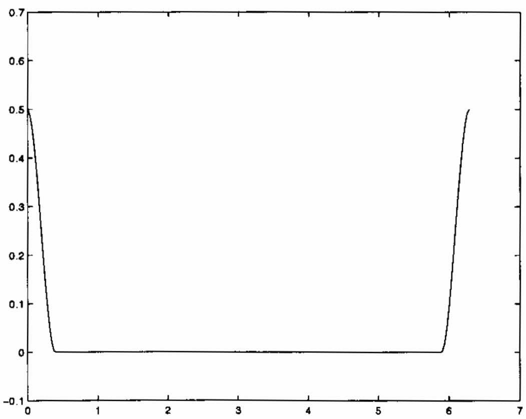

Figure 4.7: The function p_4.

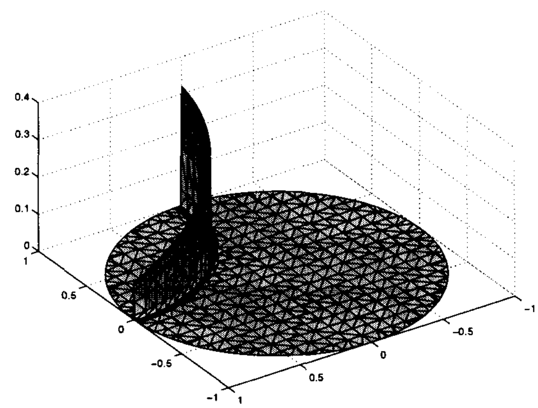

Figure 4.8: The function G_3.

Finally, for $(x,y)_* \notin (\{\mathbf{B},\mathbf{C}\}+B)$, we define G_4 by

$$G_4((x,y)_*) := -f_0(1)p_4\left(2\arctan\frac{(x\sqrt{2}+1)^2+(y\sqrt{2}+1)^2}{2(x+y+\sqrt{2})\sqrt{2}}\right). \tag{4.2}$$

If C is a circle tangent to the line segment $[\mathbf{B},\mathbf{C}]$ at the midpoint $\mathbf{M}=\left(-\sqrt{2}/2,-\sqrt{2}/2\right)_*$ of $\mathbf{B}$ and $\mathbf{C}$, then (4.2) makes G_4 constant on the arc D defined by the intersection of C with $B\setminus(\{\mathbf{B},\mathbf{C}\}+O)$ (Figure 4.9). To see why this is so, consider Figure 4.10. For any point $\mathbf{P}=(x,y)_*\in D$, there is a unique circle C as above. The center of this circle, $\mathbf{Z}=(z,z)_*$, satisfies $\|\mathbf{Z}-\mathbf{M}\|=\|\mathbf{Z}-\mathbf{P}\|$, which means that

$$z = 1/2\left(\frac{x^2+y^2-1}{x+y+\sqrt{2}}\right).$$

We wish for G_4 to be constant on D, and we require that $G_4(D)=G_4(\mathbf{P}')$ where $\mathbf{P}'$ is the point of intersection of D and $\mathbf{B}+S$. Since $\triangle\mathbf{BMZ}$ is right angled, the angle ϕ satisfies

$$\begin{aligned}\phi &= \arctan\|\mathbf{Z}-\mathbf{M}\|/\|\mathbf{M}-\mathbf{B}\| && (4.3)\\ &= \arctan\frac{(x\sqrt{2}+1)^2+(y\sqrt{2}+1)^2}{2(x+y+\sqrt{2})\sqrt{2}}. && (4.4)\end{aligned}$$

Since the angle $\angle\ \mathbf{MBP}'$ is double the angle ϕ, the argument of p_4 in (4.2) is the same as that of $\mathbf{P}'$. The moduli also coincide, since the modulus of $\mathbf{P}'$ is 1. Of course, the definition of G_4 guarantees that $G_4(\mathbf{P}')=G(\mathbf{P}'')$, where $\mathbf{P}''$ is the point of intersection of D and $\mathbf{C}+S$.

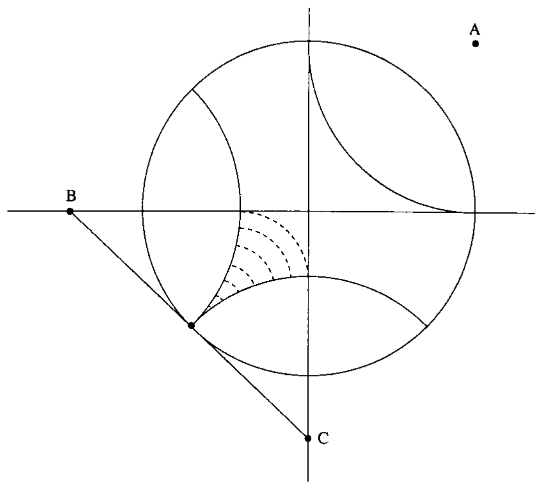

Figure 4.9: The dashed curves are arcs of circles tangent to the line **BC** at $(1, \frac{5\pi}{4})$. G_4 is constant along these arcs.

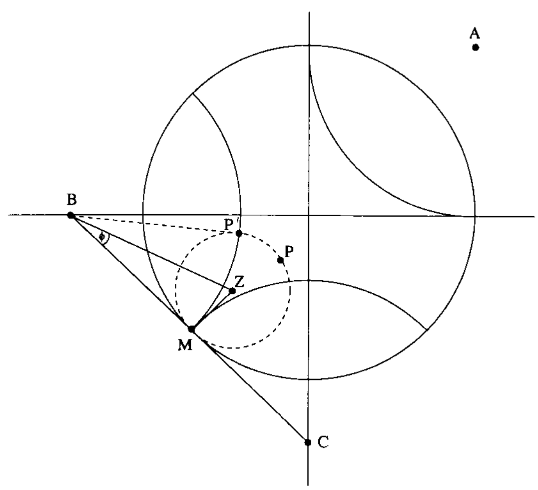

Figure 4.10: Geometry for G_4

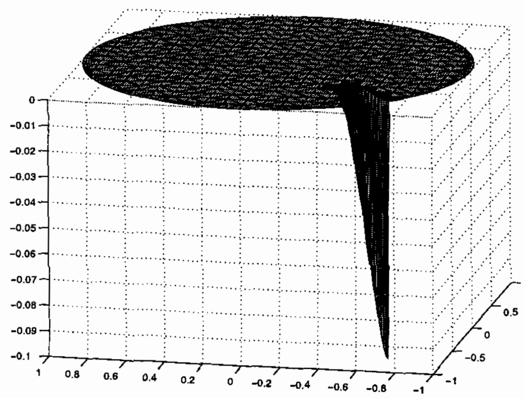

Figure 4.11: The function G_4.

We have shown that G_4 is continuous; it remains to show that it is continuously differentiable. To see that G_4 remains a C^1 function after the gluing of the pieces, we observe that:

- for each $x = \mathbf{B} + (1, \phi)$, the directional derivative in direction $(1, \phi)$ is zero on both sides and is continuous there, while that in direction $(1, \phi + \frac{\pi}{2})$ equals $f_0(1)p_4'(\phi + \frac{\pi}{4})$ on both sides and is continuous there;
- for each $x = \mathbf{C} + (1, \phi)$, a symmetric argument applies.

The definition also guarantees that supp G_4 is contained in the third quadrant.

Now let $G_0 \in \mathbf{C}^1(\mathbb{R}^2)$ be defined by

$$G_0 := F + G_1 + G_2 + G_3 + G_4.$$

To summarize, G_0 is composed of the sum F of the three calderas, with the gorges/ridges G_1, G_2 and G_3 added to remove the critical points on the rims of the three calderas, and the bump G_4 added to remove the critical point at $(1, \frac{5\pi}{4})$.

4.2.3 Verifying that G_0 has no Critical Points.

We begin at caldera $\mathbf{A}$. Note that $G_0(x) = F_{\frac{\pi}{4}}(x) + G_3(x)$ whenever $x = \mathbf{A} + (\rho, \phi)$ for any $\rho \in [0, 1.01]$ and $\phi \in [\pi, \frac{3\pi}{2}]$. The components G_1, G_2 and G_4 are identically zero in the first quadrant of B, and

$$\begin{aligned}\|\mathbf{A}-\mathbf{B}\| = \|\mathbf{A}-\mathbf{C}\| &= \sqrt{4+2\sqrt{2}} > 1.01+1.6 \\ &= 1.01 + \max\|\mathbf{B} - \partial \text{ supp } G_1\| \\ &= 1.01 + \max\|\mathbf{C} - \partial \text{ supp } G_2\|,\end{aligned} \tag{4.5}$$

so F_π and $F_{\frac{3\pi}{2}}$ are identically zero for x as above. Since $G_0(x) = F_{\frac{\pi}{4}}(x) + G_3(x)$, and we have:

(I-1) $\frac{d}{d\phi}G_0(\mathbf{A}+(1,\phi)) > 0$, for $\phi \in [0,1) \cup (1, 1.01]$, and

(I-2) $\frac{d}{d\rho}G_0(\mathbf{A}+(\rho,\phi)) < 0$ for $\rho \in (0.4, 1)$.

Near caldera $\mathbf{B}$, we have

$$G_0(x) = -F_\pi(x) - F_{\frac{3\pi}{2}}(x) + G_1(x) + G_2(x) + G_4(x)$$

whenever $x = \mathbf{B} + (\rho, \phi)$ with $\rho \in [0, 1.01]$ and $\phi \in [-\frac{\pi}{4}, \frac{\pi}{4}]$. Again, this is true by a similar argument to (4.5). On the rim of caldera $\mathbf{B}$, G_0 is strictly increasing in the clockwise direction, i.e. $\frac{d}{d\phi}G_0(\mathbf{B}+(1,\phi)) > 0$. To see this, note that

(II-1) $\frac{d}{d\phi}F_\pi(\mathbf{B}+(1,\phi)) = 0$;

(II-2) $\frac{d}{d\phi}F_{\frac{3\pi}{2}}(\mathbf{B}+(1,\phi)) \leq 0$ and is strictly positive whenever $\phi \in (-\frac{\pi}{4}, 0]$;

(II-3) $\frac{d}{d\phi}G_1(\mathbf{B}+(1,\phi)) \geq 0$ and is strictly positive whenever $\phi \in (0, \frac{\pi}{4}]$;

(II-4) $\frac{d}{d\phi}G_2(\mathbf{B}+(1,\phi)) \geq 0$;

(II-5) $\frac{d}{d\phi}G_4(\mathbf{B}+(1,\phi)) \geq 0$ and is strictly positive for $\phi = -\frac{\pi}{4}$.

For points less than one unit away from $\mathbf{B}$, that is, for points $\mathbf{B} + (\rho,\phi) \in B \cap (\mathbf{B} + O)$, the radial slope is strictly negative:

$$\frac{d}{d\rho}G_0(\mathbf{B}+(\rho,\phi)) < 0$$

To see this, note that for such points:

(III-1) $\frac{d}{d\rho}F_\pi((\sqrt{2},\pi)+(\rho,\phi)) > 0$;

(III-2) $\frac{d}{d\rho}F_{\frac{3\pi}{2}}((\sqrt{2},\pi)+(\rho,\phi)) \geq 0$;

(III-3) $\frac{d}{d\rho}G_1((\sqrt{2},\pi)+(\rho,\phi))\leq 0$;

(III-4) $\frac{d}{d\rho}G_2((\sqrt{2},\pi)+(\rho,\phi))\leq 0$;

(III-5) $\frac{d}{d\rho}G_4((\sqrt{2},\pi)+(\rho,\phi))\leq 0$.

Near caldera $\mathbf{C}$, we have a symmetric situation. At points of distance one from $\mathbf{C}$, G_0 is strictly increasing in the counterclockwise direction:

$$\frac{d}{d\phi}G_0(\mathbf{C}+(1,\phi))<0,$$

and for points less than one unit away from $\mathbf{C}$ the radial slope is strictly negative:

$$\frac{d}{d\rho}G_0(\mathbf{C}+(\rho,\phi))<0.$$

The final piece to check is the set of points whose distance from $\{\mathbf{A},\mathbf{B},\mathbf{C}\}$ is greater than one. Let $x\in R=B\setminus\big(\{\mathbf{A},\mathbf{B},\mathbf{C}\}+B\big)$. Then the slope at x in the direction $(1,1)_*$ is strictly positive:

$$\nabla G_0(x)\cdot(1,1)_*>0.$$

We verify this fact as follows:

(IV-1) $\nabla F_{\frac{\pi}{4}}(x)\cdot(1,1)_*\geq 0$ and is strictly positive on $R\cap(\mathbf{A}+1.5B)$;

(IV-2) $\nabla G_3(x)\cdot(1,1)_*\geq 0$ on $R\cap(\mathbf{A}+1.5B)$;

(IV-3) $\nabla(-F_\pi)(x)\cdot(1,1)_*\geq 0$ and is strictly positive on $R\cap(\mathbf{B}+1.5B)$;

(IV-4) $\nabla G_1(x)\cdot(1,1)_*\geq 0$, on $R\cap(\mathbf{B}+1.5B)$;

(IV-5) $\nabla(-F_{\frac{3\pi}{2}}+G_2)(x)\cdot(1,1)_*\geq 0$ and is strictly positive on $R\cap(\mathbf{C}+1.5B)$;

(IV-6) $\nabla G_2(x)\cdot(1,1)_*\geq 0$, on $R\cap(\mathbf{C}+1.5B)$; Since $R\subset\mathbf{A},\mathbf{B},\mathbf{C}+1.5B$, we have accounted for all of R.

4.2.4 Making G_0 Even on the Sphere

The constructed function G_0 has no critical points inside B, but it requires some additional work in order to be made even on S. We introduce a lemma that allows us to even out the boundary:

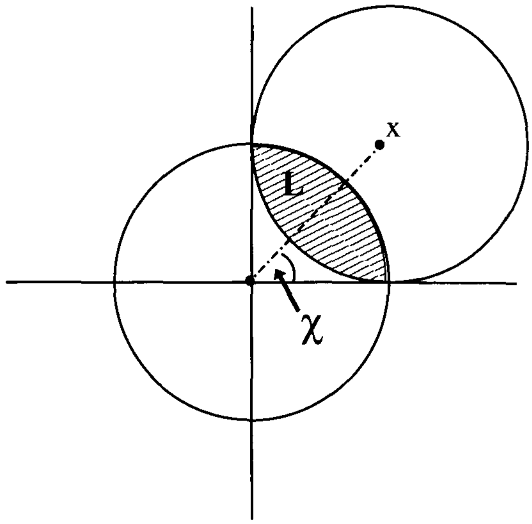

Figure 4.12: A lunette L

Lemma 4.2 *Let $\chi \in \mathbb{R}$, $x = (\sqrt{2}, \chi)$, and $L := B \cap (x+B)$ (see Figure 4.12). Let $h \in C^1(\mathbb{R})$ satisfy*

1. $\operatorname{supp} h \subset [\chi - \frac{\pi}{4}, \chi + \frac{\pi}{4}]$,
2. $h'(\chi - \frac{\pi}{4}) = h'(\chi + \frac{\pi}{4}) = 0$,
3. $h'(\phi) \geq 0$ *for* $\phi \in (\chi - \frac{\pi}{4}, \chi)$ *and*
4. $h'(\phi) \leq 0$ *for* $\phi \in (\chi, \chi + \frac{\pi}{4})$.

Then there exists a function $H \in C^1(\mathbb{R}^2)$ such that:

1. $H(s(\phi)) = h(\phi)$ *for all* $\phi \in [\chi - \frac{\pi}{4}, \chi + \frac{\pi}{4}]$;
2. $\frac{d}{d\rho} H(x + (\rho, \psi)) \leq 0$ *for* $\rho \leq 1$ *and* $\psi \in [-\chi - \frac{\pi}{4}, -\chi + \frac{\pi}{4}]$ *(that is, in the right sector of $x + B$ containing L);*
3. $H(x) = 0$ *for all* $x \notin (1, \infty)L$.

Proof: Let $f \in C^1[0, +\infty)$ satisfy $f(0) = f'(0) = f'(1) = 0$, $f|_{[1,+\infty)} = 1$ and $f' \geq 0$. For any $\phi \in [\chi - \frac{\pi}{4}, \chi + \frac{\pi}{4}]$, define

$$p(\phi) := \sqrt{2}\cos(\phi - \chi) - \sqrt{\cos 2(\phi - \chi)}$$

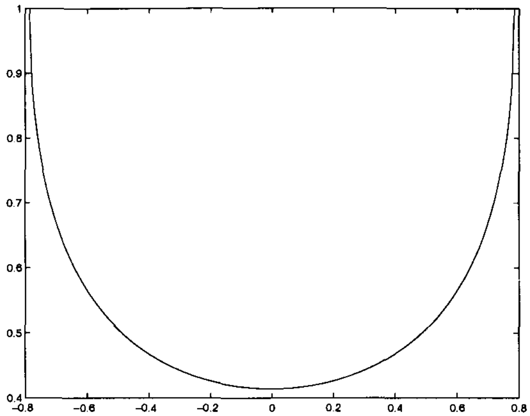

Figure 4.13: $p(\phi) = \min\{\rho \mid (\rho, \phi) \in L\}$

This defines $p(\phi)$ to equal $\min\{\rho > 0 : (\rho, \phi) \in L\}$; note that $p(\phi) \in [\sqrt{2} - 1, 1)$ whenever $\phi \in (\chi - \frac{\pi}{4}, \chi + \frac{\pi}{4})$ (see Figure 4.13). Now let:

$$H((\rho, \phi)) := \begin{cases} h(\phi) f(\frac{\rho - p(\phi)}{1 - p(\phi)}) & (\rho, \phi) \in (1, \infty)L \\ 0 & \text{otherwise} \end{cases}$$

Note that $H((p(\phi), \phi)) = 0$ and $H((1, \phi)) = h(\phi)$ for all $\phi \in [\chi - \frac{\pi}{4}, \chi + \frac{\pi}{4}]$. The definition guarantees that H is C^1 everywhere except possibly on the boundary of the two sets; on the other side, it is directly checked that H' equals zero at all points of the boundary and is continuous there (due to the facts that $f(0) = f'(0) = 0$ and $h(\chi) = h(\chi + \frac{\pi}{2}) = h'(\chi) = h'(\chi + \frac{\pi}{2}) = 0$). Then (1) and (3) follow from the definition; to see (2), note that, for a fixed ψ, $H(x + (\rho, \psi))$ is non-increasing in direction ρ since $f' \geq 0$, $h'(\phi) \geq 0$ for $\phi \in (\chi - \frac{\pi}{4}, \chi)$ and $h'(\phi) \leq 0$ for $\phi \in (\chi, \chi + \frac{\pi}{4})$. ■

As an example of how Lemma 4.2 works, Figure 4.14 shows the graph of H when $h(\phi) = \cos^2(2\phi)$ and $\chi = 0$. We will use H to smoothly reshape the function G_0 on one portion of the sphere at a time, leaving values outside the associated lunette unchanged.

Define g_0 by $g_0(\phi) := G_0((1, \phi))$. Then $g_0 \in C^1$ is a 2π-periodic function that traces G_0 over the sphere S (Figure 4.15). It is readily checked that

$$g_0(\phi) > 0 \text{ for } \phi \in [0, \frac{\pi}{2}],$$

$$g_0(\phi) < 0 \text{ for } \phi \in [\frac{3\pi}{4}, \frac{7\pi}{4}],$$

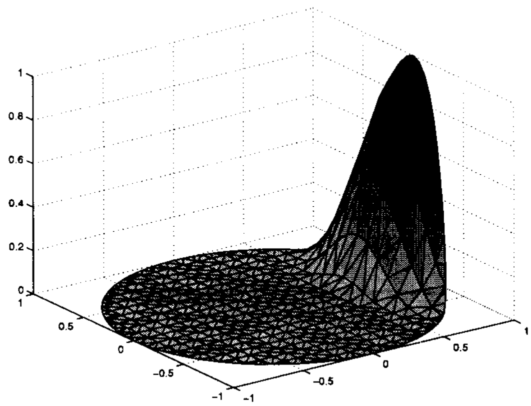

Figure 4.14: Graph of H, with $h(\phi) = cos^2(2\phi)$ and $\chi = 0$.

$$g_0'(\phi) > 0 \text{ for } \phi \in (\frac{\pi}{4}, \frac{\pi}{2}) \cup (\frac{3\pi}{4}, \pi) \cup (\frac{5\pi}{4}, \frac{3\pi}{2}) \cup (\frac{7\pi}{4}, 2\pi),$$

and

$$g_0'(\phi) < 0 \text{ for } \phi \in (0, \frac{\pi}{4}) \cup (\frac{\pi}{2}, \frac{3\pi}{4}) \cup (\pi, \frac{5\pi}{4}) \cup (\frac{3\pi}{2}, \frac{7\pi}{4}).$$

The reflective definition of G_2 and the overall symmetry guarantees that

$$g_0(\frac{5\pi}{4} - \phi) = g_0(\frac{5\pi}{4} + \phi) \text{ for } \phi \in [0, \frac{\pi}{2}]. \tag{4.6}$$

$$-\frac{d}{d\phi} g_0(\frac{3\pi}{4} - \phi) \geq \frac{d}{d\phi} g_0(\frac{3\pi}{4} + \phi) \text{ for } \phi \in [0, \frac{\pi}{4}]. \tag{4.7}$$

Similarly,

$$\frac{d}{d\phi} g_0(\frac{7\pi}{4} + \phi) \geq -\frac{d}{d\phi} g_0(\frac{7\pi}{4} - \phi) \text{ for } \phi \in [0, \frac{\pi}{4}]. \tag{4.8}$$

Denote

1. $M := g_0(0) > 0$,

2. $m := g_0(\frac{\pi}{2}) > 0$,

3. $l := g_0(\frac{3\pi}{4}) = g(\frac{7\pi}{4}) < 0$ and

4. $L := g_0(\frac{5\pi}{4}) < 0$.

Then M is the greatest height of the ridge on caldera **A**, m is the lowest height on that ridge, l is the depth of the shallowest point of the gorges on the inverted calderas **B** and **C**,

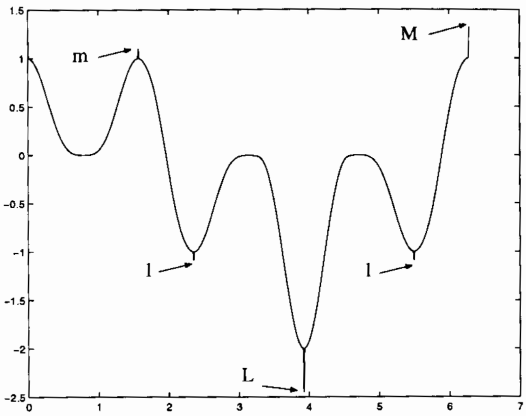

Figure 4.15: The function g_0.

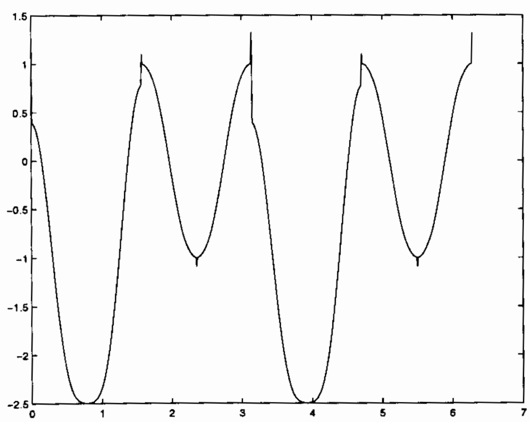

Figure 4.16: The function g

and L is the depth of the deepest point (Figure 4.15). We want to use Lemma 4.2 to remold G_0 along its rim, making the function π-periodic along the rim (i.e. even on the sphere), without adding any critical points to the function.

Choose some π-periodic $g \in C^1(\mathbb{R})$, such that:

$$g(0) = M,\ g(\frac{\pi}{4}) = L \text{ and } g(\phi) = g_0(\phi) \text{ for } \phi \in [-\frac{\pi}{4}, 0] \cup [\frac{\pi}{2}, \frac{3\pi}{4}],$$

$$g'(\phi) \leq g_0'(\phi) \text{ for } \phi \in (0, \frac{\pi}{4}) \cup (\pi, \frac{5\pi}{4}) \cup (\frac{3\pi}{2}, \frac{7\pi}{4}), \tag{4.9}$$

$$g'(\phi) \geq g_0'(\phi) \text{ for } \phi \in (\frac{\pi}{4}, \frac{\pi}{2}) \cup (\frac{3\pi}{4}, \pi) \cup (\frac{5\pi}{4}, \frac{3\pi}{2}) \tag{4.10}$$

(we use here (4.6), (4.7), (4.8)), and

$$g_0(\phi + \pi) \leq g(\phi) \leq g_0(\phi) \text{ for } \phi \in [0, \frac{\pi}{2}].$$

This is possible since here $g_0(\phi + \pi) < 0 < g_0(\phi)$ and

$$g_0'(\frac{j\pi}{4}) = 0 \text{ for all } j \in \mathbb{N}. \tag{4.11}$$

The last implies also that

$$g(\phi) \geq g_0(\phi) \text{ for } \phi \in [\pi, \frac{3\pi}{2}].$$

We now have a π-periodic function g that agrees with g_0 on $[-\frac{\pi}{4}, 0] \cup [\frac{\pi}{2}, \frac{3\pi}{4}]$, and we wish to make g_0 agree with g on the remainder of the unit circle. To this end, we use (4.9), (4.10) and apply Lemma 4.2 three times to get $H_1, H_2, H_3 \in C^1(\mathbb{R}^2)$, using

$$\chi_1 := \frac{\pi}{4},\ \chi_2 := \pi \text{ and } \chi_3 := \frac{3\pi}{2};$$

$$x_1 := \mathbf{A},\ x_2 := \mathbf{B} \text{ and } x_3 := \mathbf{C};$$

and

$$h_i := (g - g_0)\chi_{[\chi_i - \frac{\pi}{4}, \chi_i + \frac{\pi}{4}]},\ i = 1, 2, 3.$$

In each of the three cases, $h_i \in C^1(\mathbb{R})$ since (4.11) and the definition of g assure us that $h_i'(\chi_i - \frac{\pi}{4}) = h_i(\chi_i - \frac{\pi}{4}) = h_i'(\chi_i + \frac{\pi}{4}) = h_i(\chi_i + \frac{\pi}{4}) = 0$. As a result, we obtain functions $H_i \in C^1(\mathbb{R}^2)$, $i = 1, 2, 3$, such that:

(V-1) $H_1(s(\phi)) = (g - g_0)(\phi)$ for all $\phi \in [0, \frac{\pi}{2}]$;

(V-2) $H_2(s(\phi)) = (g - g_0)(\phi)$ for all $\phi \in [\frac{3\pi}{4}, \frac{5\pi}{4}]$;

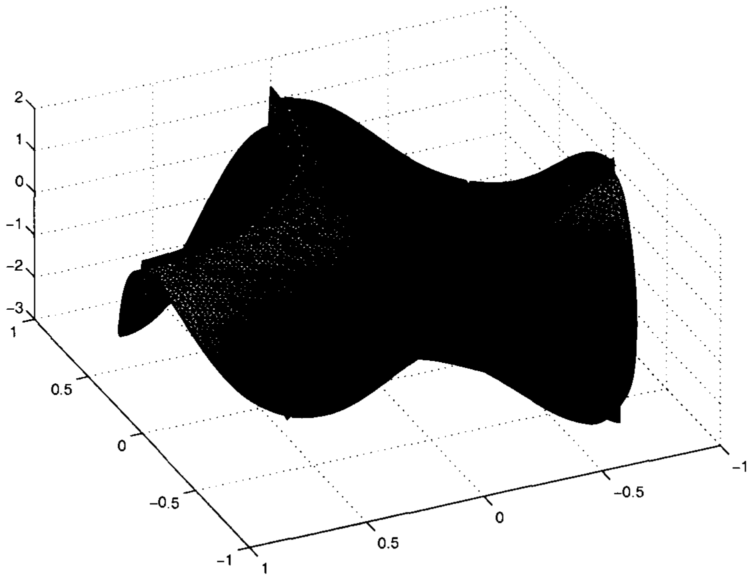

Figure 4.17: The composite function G.

(V-3) $H_3(s(\phi)) = (g - g_0)(\phi)$ for all $\phi \in [\frac{5\pi}{4}, \frac{7\pi}{4}]$;

(V-4) $\frac{d}{d\rho}H_1(x + (\rho, \psi)) \geq 0$ for $\rho \leq 1$ and $\psi \in [\pi, \frac{3\pi}{2}]$;

(V-5) $\frac{d}{d\rho}H_i(x + (\rho, \psi)) \leq 0$ for $\rho \leq 1$ and $\psi \in [-\chi_i - \frac{\pi}{4}, -\chi_i + \frac{\pi}{4}]$, $i = 2, 3$;

(V-6) $H_i(x) = 0$ for all $x \notin (1, \infty)(B \cap (x_i + B))$, $i = 1, 2, 3$.

Let $G := G_0 + H_1 + H_2 + H_3$. Then the previous calculations for G_0 and the properties of H_i above guarantee that G has no critical points.

Let $\tilde{g}(\phi) = G(1, \phi)$. Then

$$\begin{aligned}\tilde{g}(\phi) &= G_0(\phi) + H_1(\phi) + H_2(\phi) + H_3(\phi) \\ &= g_0(\phi) + (g - g_0)\chi_{[0,\frac{\pi}{2}]\cup[\frac{3\pi}{4},\frac{5\pi}{4}]\cup[\frac{5\pi}{4},\frac{7\pi}{4}]}(\phi) \\ &= g(\phi),\end{aligned}$$

so G is even on the sphere. Figure 4.17 provides a view of the composite function, viewed from the second quadrant.

4.3 Even-valuedness on a Neighborhood of S

Although the preceding construction shows that Conjecture 1 is false, it is very nearly true when $g : \mathbb{R}^2 \to \mathbb{R}$ is C^1: the next result from [BKW02] shows that if g is even on a neighborhood of the unit circle, then $\nabla g = 0$ for some point in the open unit ball O. For a continuous vector field $v : U \to \mathbb{R}^2$ on an open subset U of the plane, and a smooth oriented curve $C \subset \mathbb{R}^2$, define the *winding number* $\mathrm{rot}(v, C)$ of v through C to be the number of counterclockwise rotations performed by v over the curve C. The following fact may be found in [Dei85] and [Pri95]:

Proposition 4.3 *Let v and C be as above. Then*

1. *If C is closed then*

$$\mathrm{rot}(v, C) \in \mathbb{N}.$$

If $v \neq 0$ on U and C is closed and contractible, *that is, homeomorphic to a point, then*

$$\mathrm{rot}(v, C) = 0.$$

2. *If C is the* join *(see [Pri95]) of oriented curves C_1 and C_2 then*

$$\mathrm{rot}(v, C) = \mathrm{rot}(v, C_1) + \mathrm{rot}(v, C_2).$$

Roughly speaking, the join of two oriented curves C_1 and C_2 is constructed by attaching the end of the curve C_1 to the start of C_2.

Proposition 4.4 *Let $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ be a C^1-function defined on a neighborhood U of B. If f is even in some neighborhood W of S, then there is some $x_0 \in B$ such that $\nabla f(x_0) = 0$.*

Proof: Suppose $v = \nabla f$ does not vanish on the ball. Let C be the unit circle, oriented counterclockwise, let C_1 be the part of C in the upper halfspace, and let C_2 be the part in the lower halfspace. Since f is even on a W, for any $x \in S$ and $d \in \mathbb{R}^2$ we have

$$\begin{aligned}\langle \nabla f(x), d\rangle &= \lim_{t \searrow 0} \frac{f(x+td) - f(x)}{t} \\ &= \lim_{t \searrow 0} \frac{f(-x+t(-d)) - f(-x)}{t} = -\langle \nabla f(-x), d\rangle,\end{aligned}$$

so $v(-x) = -v(x)$. But then $v((-1,0)_*) = -v((1,0)_*)$, so clearly

$$\mathrm{rot}(v, C_1) = n + \frac{1}{2}$$

and

$$\mathrm{rot}(v, C_2) = m + \frac{1}{2}$$

for some $m, n \in \mathbb{N}$. But f is even on W, so $m = n$. By Proposition 4.3,

$$\mathrm{rot}(v, C) = \mathrm{rot}(v, C_1) + \mathrm{rot}(v, C_2) = 2n + 1 \neq 0.$$

But this contradicts part 2 of Proposition 4.3. So $\nabla f = 0$ at some point on the ball. ■

Bibliography

[AGJ97] D. Azagra, J. Gomez, and J.A. Jaramillo. Rolle's theorem and negligibility of points in infinite-dimensional Banach spaces. *J. Math. Anal. Appl.*, 213:487–495, 1997.

[BBC01] H. H. Bauschke, J.M. Borwein, and P. L. Combettes. Essential smoothness, essential strict convexity, and legendre functions in Banach spaces. *Commun. Contemp. Math.*, 4:615–647, 2001.

[BF01] J. M. Borwein and S.P. Fitzpatrick. Duality inequalities and sandwiched functions. *Nonlinear Analysis*, 46:365–380, 2001.

[BKW02] J.M. Borwein, I. Kortezov, and H. Wiersma. A C^1 function that is even on a sphere and has no critical points in the ball. *Journal of Nonlinear and Convex Analysis*, 3:1–16, 2002.

[BL00] Jonathan M. Borwein and Adrian S. Lewis. *Convex Analysis and Nonsmooth Optimization.* CMS Books in Mathematics. Springer-Verlag: New York, 2000.

[BP87] J. Borwein and D. Preiss. A smooth variational principle with applications to subdifferentiability and differentiability of convex functions. *Trans. Am. Math. Soc.*, 303:517–512, 1987.

[CL94] F.H. Clarke and Yu.S. Ledyaev. Mean value inequalities. *Proceedings of the American Mathematical Society*, 122:1075–1083, 1994.

[Cla83] Francis H. Clarke. *Optimization and Nonsmooth Analysis.* Canadian Mathematical Society Series of Monographs and Advanced Texts. John Wiley and Sons: New York, 1983.

[Dei85] K. Deimling. *Nonlinear Functional Analysis.* Springer, Berlin, Heidelberg, New York, Tokyo, 1985.

[DGZ93] R. Deville, G. Godefroy, and V. Zizler. *Smoothness and Renormings in Banach Spaces*, volume 64 of *Pitman Monographs.* Pitman, New York, 1993.

[Eke74] I. Ekeland. On the variational principle. *Journal of Mathematical Analysis and Applications*, 47:324–353, 1974.

[Eva98] Lawrence C. Evans. *Partial Differential Equations.* Graduate Studies in Mathematics. American Mathematical Society, 1998.

[Fer96] J. Ferrer. Rolle's theorem fails in l_2. *MAA Monthly*, 103:161–165, 1996.

[FHH$^+$01] M. Fabian, P. Habala, P. Hájek, V. M. Santalucia, J. Pelant, and V. Zizler. *Functional Analysis and Infinite Dimensional Geometry.* CMS Books in Mathematics. Springer-Verlag: New York, 2001.

[FK70] S.V. Fomin and A.N. Kolmogorov. *Introductory Real Analysis.* Dover Publications, Inc: New York, 1970.

[GK90] Kazimierz Goebel and W.A. Kirk. *Topics in Metric Fixed Point Theory.* Cambridge Studies in Advanced Mathematics. Cambridge University Press: Cambridge, New York, 1990.

[HUL93a] J.B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I*, volume 305 of *Grundlehren Math. Wiss.* Springer-Verlag, Berlin, 1993.

[HUL93b] J.B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms II*, volume 306 of *Grundlehren Math. Wiss.* Springer-Verlag, Berlin, 1993.

[LL00] A.S. Lewis and R.E. Lucchetti. Nonsmooth duality, sandwich and squeeze theorems. *SIAM J. Control. Optim.*, 38:613–626, 2000.

[LR96] A.S. Lewis and D.A. Ralph. A nonlinear duality result equivalent to the clarke-ledyaev mean value inequality. *Nonlinear Analysis TMA*, 26:343–350, 1996.

[Pri95] H.A. Priestley. *Introduction to Complex Analysis.* Oxford University Press, 1995.

[Roc97] R. Tyrrell Rockafellar. *Convex Analysis.* Princeton Landmarks in Mathematics. Princeton University Press: Princeton, New Jersey, 1997.

[Rud73] Walter Rudin. *Functional Analysis.* McGraw-Hill Series in Higher Mathematics. McGraw-Hill Book Company: New York, 1973.

[Spr70] David A. Sprecher. *Elements of Real Analysis.* Academic Press: New York and London, 1970.

[Wad95] W.R. Wade. *An Introduction to Analysis.* Prentice Hall, New Jersey, 1995.