NONSMOOTH OPTIMIZATION WITH SMOOTH SUBSTRUCTURE

by

Warren L. Hare
B.Sc., Brock University, 1998.

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in the Department
of
Mathematics

© Warren L. Hare 2004
SIMON FRASER UNIVERSITY
April 2004

All rights reserved. This work may not be
reproduced in whole or in part, by photocopy
or other means, without the permission of the author.
APPROVAL

Name: Warren L. Hare
Degree: Doctor of Philosophy
Title of Thesis: NONSMOOTH OPTIMIZATION WITH SMOOTH SUBSTRUCTURE

Examining Committee: Dr. D. Muraki
Chair

___________________________
Dr. A. Lewis
Senior Supervisor

___________________________
Dr. J. Borwein
Supervisory Committee

___________________________
Dr. P. Loewen
Supervisory Committee

___________________________
Dr. S. Choi
Internal Examiner

___________________________
Dr. J. Burke
External Examiner

Date of Defense: April 2, 2004
Partial Copyright Licence

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author’s written permission.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Bennett Library
Simon Fraser University
Burnaby, BC, Canada
Abstract

Over the past decade, several independent researchers studying nonsmooth optimization have developed classes of functions and sets that, though nonsmooth themselves, have underlying smooth substructure. These substructures are exploited for algorithmic purposes, to create calculus rules, and develop sensitivity analysis. In this thesis we explore nonsmooth optimization with smooth substructure.

The majority of this thesis focuses on the notion of partial smoothness (Lewis). Our goal is to show partial smoothness is a natural and powerful tool in the study of nonsmooth optimization.

To achieve this goal, we begin by comparing partial smoothness to other recently developed notions of smooth substructure. After some preliminary examples and basic results about partial smoothness, Chapter 2 shows that partial smoothness is a nonconvex extension of the convex notions of $C^p$-identifiability (Wright) and fast tracks (Mifflin and Sagastizábal). In Chapter 4 we compare partial smoothness to primal-dual gradient structures (Mifflin and Sagastizábal) and $g \circ F$ decompositions (Shapiro). Both notions are shown to be highly related to partial smoothness. This makes it clear that partial smoothness arises naturally in nonsmooth optimization.

Chapter 3 studies the calculus of partial smoothness. To begin, partly smooth functions are compared with their epigraphs (Theorems 2.18 and 3.5). Focusing on convex partly smooth functions, a new inf-convolution rule is created, and some sensitivity results developed (Section 3.2). In order to move beyond the convex case the notion of prox-regularity is required.

Prox-regularity (Poliquin and Rockafellar) extends some convexity properties to a nonconvex setting. Adding prox-regularity to partial smoothness ensures that the projection mapping is well behaved (Section 3.3), allows for the identification of active constraints (Sections 3.4, 5.2 and 5.3), and creates simplified optimality conditions (Section 3.5). These results show the power of prox-regular partial smoothness in nonsmooth optimization.

We conclude this work with a chapter on algorithm design. Chapter 5 provides concrete examples for when the active constraint identification results of Section 3.4 hold. It further discusses a new algorithm for approximating normal cones to oracle-based sets. Various results suggesting the algorithm's success are given. The chapter concludes with some numerical testing of the algorithm.
Dedication

To everyone who has ever taken me hiking, skiing, or snow-shoeing.
Especially to the few who have done all three.
Acknowledgements

Like any extensive work, this thesis is not the result of one man’s labour. Without the support of my supervisor, committee, family, and friends it may have never been completed. I lack the space, (and energy) to thank all of you appropriately; for which I hope you forgive me. Some special thanks go out to:

Adrian Lewis for teaching me more than just math;

Jon Borwein for guiding me through my first year at S.F.U.;

Philip Loewen for providing input even before he was on my committee;

Jim Burke for saying the right thing (prox-regularity) at the right time (Nov 3rd, 2001);

S. Choi, B. Mifflin, C. Sagastizábal, and H. Wiersma who gave many hours with nothing in return; and

N.S.E.R.C., Jon Borwein, S.F.U., and Adrian Lewis for funding me throughout this thesis.
Contents

Approval .......................................................... ii
Abstract .................................................................. iii
Dedication ........................................................... iv
Acknowledgements ................................................ v
Contents ............................................................. vi
List of Tables ........................................................ viii
List of Figures ....................................................... ix
1 Introduction ....................................................... 1
  1.1 Outline ......................................................... 2
  1.2 Historical Background ....................................... 4
    1.2.1 Smooth Substructure .................................... 4
    1.2.2 Constraint Identification ............................... 9
    1.2.3 Projections and Prox-regularity ....................... 13
    1.2.4 Algorithm Design ....................................... 15
  1.3 Definitions .................................................... 15
    1.3.1 Basic Notation ........................................... 16
    1.3.2 Nonsmooth Analysis .................................... 17
  1.4 Manifolds ...................................................... 23
    1.4.1 Normal and Tangent Cones ............................ 24
2 Partial Smoothness and Related Concepts ..................... 27
  2.1 Definition of Partial Smoothness ......................... 27
  2.2 Examples ...................................................... 30
  2.3 Epigraph comparisons ....................................... 35
  2.4 $C^p$-Identifiability ......................................... 38
2.5 UV-decompositions and Fast Tracks ........................................ 40

3 Calculus of Partial Smoothness ............................................... 46
  3.1 Classic Calculus of Partial Smoothness ................................. 47
  3.2 Convex Partial Smoothness .............................................. 49
  3.3 The Smooth Projection Theorem ......................................... 55
  3.4 Active Constraint Identification ....................................... 59
  3.5 Critical Point Analysis .................................................. 63
  3.6 Necessity of Prox-regularity .......................................... 66

4 Related Notions ....................................................................... 67
  4.1 Primal-dual Gradient Structures ....................................... 68
  4.2 $g \circ F$ Decompositions ................................................. 75

5 Algorithmic Results .................................................................. 80
  5.1 Algorithm Design ............................................................. 81
  5.2 Constrained Minimization .................................................. 83
    5.2.1 Gradient Projection Methods ...................................... 85
    5.2.2 Newton-like Methods ............................................... 87
  5.3 Unconstrained Minimization .............................................. 89
    5.3.1 Proximal Point Methods ............................................ 89
  5.4 Random Normal Minimization ........................................... 93
  5.5 Numerical Results ........................................................... 100
    5.5.1 Effect of Search Radius on Normal Convergence .......... 100
    5.5.2 Effects of Dimension and Number of Points ............... 104
    5.5.3 Testing Problem Solving ............................................ 106

6 Conclusions and Open Questions ............................................ 110

Appendix ................................................................................. 114
  A Index of Notation ............................................................. 114

Bibliography ........................................................................... 118
List of Tables

5.1 Mean and Standard Deviations for RNG on the Unit Ball .......... 102
5.2 Mean and Standard Deviations for RNG on the Complement Ball .... 102
5.3 Mean and Standard Deviations for RNG on the Half Plane .......... 104
5.4 Mean and Standard Deviations for RNG on the Positive Orthant ........ 105
5.5 Mean and Standard Deviations for RNM on the Unit Square .......... 108
5.6 Mean and Standard Deviations for RNM on the Donut ............ 109
## List of Figures

2.1 A Function Lacking Sub-continuity: \( \max\{3x^2 - 2y, 3x^2, 2y + x^2\} \) .................. 33
2.2 Non-unique Active Manifold ........................................... 35
2.3 A Function with a Fast Track: \( |x - y^2| + y^2 \) .................................. 42

3.1 The Smooth Projection Theorem ................................... 57
3.2 Need for \( \bar{n} \in \text{rint} \mathcal{N}_S(x) \) .......................... 60

4.1 Relating \( C^p \)-identifiable Surfaces, Fast Tracks, \( g \circ F \) Decompositions, Partial Smoothness, and PDG Structures. .................. 79

5.1 Effect of \( \mu \) on Random Normal Generation ..................... 97
5.2 Necessity of Interior Tangent Vectors for Random Normal Generation ........ 99
5.3 Effect of \( \varepsilon \) on RNG for the Unit Ball ......................... 101
5.4 Effect of \( \varepsilon \) on RNG for the Complement Ball .............. 103
5.5 Improved \( \tilde{N} \) via Convexification .............................. 103
Chapter 1

Introduction

Optimization is a flourishing field of mathematical research. Its focus is based around finding optimal values (minima or maxima) of a function. If the function is smooth much of the theory behind this has been developed. However, many functions are not smooth. *Nonsmooth optimization* studies optimization without the assumption that the function to be optimized is smooth.

Nonsmooth optimization is often split into two fields of research: theoretical and applied. However, the distinction between these fields is often unclear.

Theoretical research discusses the behaviour of functions near optimal values. This leads to analytical objects such as 'the subdifferential' and 'the normal cone'. Many texts have been written on the theory of nonsmooth optimization, such as [RW98], which is used heavily in this thesis.

Applied research proceeds by asking the question of how optimal values can be found. Such research often focuses on designing algorithms and proving their convergence. Like theoretical research, many texts have been written on this field, such as [Ber95], also used throughout this thesis.

As mentioned, the distinction between theoretical and applied research is often hazy. For example, one classic result of nonsmooth optimization,

\[ \text{the subdifferential at an optimal point must contain 0,} \]

can be thought of both in a theoretical and an applied manner. As a theoretical result, it makes use of the analytical object 'the subdifferential'. The proof, though straightforward, is also quite theoretical. As an applied result it suggests a method of finding optimal values: design an algorithm which forces the subdifferential towards zero. The strong interaction between theoretical and applied research can be seen in the many authors who find it unreasonable to focus solely the theoretical or applied side of nonsmooth optimization. See [HUL93a] and [HUL93b] for example, which are also heavily-used references for this work.
Another example of the overlap between theoretical and applied research, is the study of active set identification. The active set of an optimization problem describes the location where optimal values must occur. If the active set can be identified, the dimension of the problem can often be reduced. Thus, much research has been put into the development of algorithms which identify active sets, and into classifying functions which have active sets to be identified.

One way for a function to have an active set which can be identified, is for the function to have some sort of underlying smooth substructure. Depending on how this smooth substructure is defined, one might also be able to exploit it for other purposes. Recently a variety of definitions have surfaced which try to capture the notion of a smooth substructure in a manner that can be exploited in nonsmooth optimization. The primary goal of this thesis is to examine the recent idea of smooth substructure defined by partial smoothness.

This thesis begins by arguing that partial smoothness is a natural notion of smooth substructure. Partial smoothness occurs in many examples in nonsmooth optimization, and extends several convex notions of smooth substructure to a nonconvex setting. Partial smoothness also provides a broad framework into which many previous results of nonlinear programming naturally extend.

The thesis further argues partial smoothness provides a powerful tool in the study of nonsmooth optimization. On its own, partial smoothness has a rich calculus. When combined with the assumption of prox-regularity, partial smoothness allows us to ensure the projection mapping is well behaved, identify active sets of minimization problems, and create simplified optimality conditions.

A secondary focus of this thesis is on a method of generating normal cone information for an oracle-based set. We outline the suggested algorithm, and provide both theoretical and experimental evidence of its effectiveness.

### 1.1 Outline

This thesis is divided into six Chapters, one Appendix, and the "Bibliography".

Chapter 1, "Introduction", covers the background and basic material needed to read this thesis. Section 1.2 contains a comprehensive background of past research related to this thesis. The section is split into four subsections, each discussing one aspect of what this thesis entails. Section 1.3 contains the basic definitions required in reading this work. Of great importance is the definition of prox-regularity (Definition 1.12) which is a recurring theme in this work. Chapter 1 ends with a section on manifolds. Section 1.4 contains the definition of a manifold, and several useful formulae for computing tangent and normal cones to manifolds.

Chapter 2, "Partial Smoothness and Related Concepts", introduces the definition of partial smoothness (Definition 2.1). Section 2.1 discusses the definition and gives some equivalent tests for its various requirements. Section 2.2 provides a large collection of examples to help the reader.
become familiar with the notion of partial smoothness. One important example, Example 2.16, reappears in Section 3.6 to show the necessity of prox-regularity in this thesis. Section 2.3 shows the natural relation between a partly smooth function and its epigraph. Specifically, a function is partly smooth if and only if its epigraph is (Theorem 2.18). This theorem also appears in Chapter 3 where we give a more subtle proof of the same result. Section 2.4 discusses Wright’s notion of \( C^p \)-identifiable surfaces (Definition 2.19). (As in Wright’s work, partial smoothness is examined in a \( C^p \) setting, instead of the \( C^2 \) setting Lewis developed.) The section culminates with Theorem 2.21, which shows the correspondence between \( C^p \)-identifiable surfaces and convex partly smooth sets. Chapter 2 ends with Section 2.5, where the notion of a fast track (Definition 2.25) is examined. Theorem 2.27 shows the correspondence between fast tracks and partly smooth functions. Together, the results of Sections 2.4 and 2.5 make the correspondence of fast tracks and \( C^p \)-identifiable surfaces clear.

Chapter 3, "Calculus of Partial Smoothness", largely focuses on the effects of prox-regularity on partly smooth functions and sets. The chapter begins (Section 3.1) by discussing calculus rules previously developed in [Lew02]. These rules provide a more elegant approach to comparing partly smooth functions to their epigraphs (Theorem 3.5). Section 3.2 develops a new calculus rule for convex partly smooth functions: the Infimal Convolution Rule (Theorem 3.7). The section also contains some results on nondegenerate critical points of convex partly smooth functions (Theorem 3.10). Section 3.3 contains one of the most important theorems of this thesis, the Smooth Projection Theorem (Theorem 3.13). The Smooth Projection Theorem describes when the projection mapping onto a prox-regular partly smooth set "identifies" the active manifold of partial smoothness. This provides the foundation required to show the uniqueness of active manifolds (Theorem 3.14 and Corollary 3.15), and the many finite identification results of Section 3.4 (Theorems 3.16, 3.18 and 3.21). The Smooth Projection Theorem also makes use of the \( C^p \) setting of this thesis to provide bounds on the level of smoothness for the projection mapping. Chapter 3 continues with Section 3.5, which considers the question of when a critical point of a prox-regular partly smooth function is a local minimum. Theorem 3.22 shows that locally all critical points for a prox-regular partly smooth function must lie on the same active manifold, while Theorem 3.23 shows that if the critical point minimizes the function along the active manifold, then it (locally) minimizes the function off of the active manifold as well. Chapter 3 ends with an example showing the necessity of prox-regularity throughout the chapter (Section 3.6).

Chapter 4, "Related Notions", discusses the relationships between partial smoothness and two other ideas of smooth substructure. Section 4.1 provides the definitions required for primal-dual gradient structures and their related index sets, along with several examples to aid the reader. The section reproduces a theorem of Mifflin and Sagastizábal which shows when a primal–dual gradient structure implies the underlying function is partly smooth (see [MS03, Thm 7.4] and Theorem 4.6). The section ends by showing that the broad class of functions with primal-dual gradient structure
subsumes the class of partly smooth functions (Theorem 4.8). Section 4.2 examines Shapiro's definition of $g \circ F$-decomposable functions. It shows that a $g \circ F$-decomposable function is the composition of a partly smooth function with a smooth function (Proposition 4.14). By comparing this with the transversality condition and the amenability condition the relationship between $g \circ F$-decomposable functions, partly smooth functions and amenable functions becomes clear (Theorem 4.15 and Corollary 4.16).

Chapter 5, “Algorithmic Results”, discusses how the results of this thesis may affect algorithm design. The chapter begins with a brief overview of optimization algorithms (Section 5.1). Section 5.2 examines algorithms for minimizing over a constraint set. Finite constraint identification is shown for the Gradient Projection method and Newton-like methods (Theorems 5.8 and 5.10). Section 5.3 examines identifying active manifolds of an unconstrained minimization problem. Theorem 5.14 shows the Proximal Point method satisfies the conditions required for finite constraint identification. The last two sections of Chapter 5 focus on the idea of creating normal vectors via random sampling. Section 5.4 outlines a new algorithm called “Random Normal Generation”. Theorems 5.23 and 5.25 provide theoretical evidence that the algorithm will converge. Section 5.5 supports this evidence with numeric results.

Chapter 6, “Conclusions and Open Questions”, highlights the main theorems of the thesis and lists several open questions which remain.

The thesis ends with the “Index of Notation” (Appendix A) and “Bibliography”, which contain a list of the most used symbols of this thesis, and a list of works cited in this thesis.

1.2 Historical Background

In this section we give some historical background to the problems approached in this thesis. The problems break down into four categories: smooth substructure, constraint identification, projections and prox-regularity, and algorithm design. Smooth substructure studies how the idea of nonsmooth functions having smooth substructure can aid in optimization. Constraint identification studies how to simplify an optimization problem by solving the active constraints. Projections and prox-regularity studies what conditions can be placed on a set to ensure a well behaved projection mapping. Prox-regularity is key to successful research in this area. Algorithm design studies how the results of optimization research can be used to improve performance of current algorithms, and design new algorithms.

1.2.1 Smooth Substructure

A good portion of optimization research has been focused on finding classes of functions (or sets) which are narrow enough to be useful, but broad enough to be of interest. This search often focuses
CHAPTER 1. INTRODUCTION

around functions with some sort of underlying smooth substructure. This substructure is then
exploited for algorithmic purposes, to create calculus rules, or develop stability analysis. The notion
of smooth substructure could be anything, from the simple idea that a function be smooth on a
subspace, to complicated notions such as primal-dual gradient structures, or full amenability. In this
subsection we discuss some of the history of smooth substructure.

We begin our background on smooth substructure in 1982 when Rockafellar developed the class
of lower-$C^p$ functions\footnote{This is not to say the history of smooth substructure begins in 1982, simply that we find it a convenient starting
point.}. The notion of a lower-$C^p$ function was created to extend the class of finite
max functions to include infinite index sets. This provides a broad class of functions that are both
locally Lipschitz and regular. Though this thesis only deals with lower-$C^p$ functions in passing, for
the sake of completeness, we include the definition here.

**Definition 1.1 (Lower-$C^p$)** A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be lower-$C^p$ on an open set $O$, if
on $O$ there is a representation

$$f(x) := \max_{t \in T} f_t(x)$$

where $T$ is a compact set, and the functions $f_t$ and all of their partial derivatives through order $p$
depend continuously in $(x, t)$ jointly.

Rockafellar's desire to find a class of locally Lipschitz regular functions stemmed from subdif-
ferential calculus; if two locally Lipschitz regular functions have the same subdifferential mapping,
the functions differ only by a constant [Roc82, Cor 3]. Using this result Rockafellar shows that
lower-$C^2$ functions are the difference of two convex functions [Roc82, Thm 6]. He further develops
the somewhat surprising result:

*For all $p > 2$ the class of lower-$C^p$ functions is indistinguishable from the class of lower-$C^2$
functions.* [Roc82, Thm 6]

Rockafellar ends his 1982 paper with an example showing the classes of lower-$C^1$ and lower-$C^2$
functions are not the same. In 1983 Rockafellar published a survey work using lower-$C^p$ functions as
an example to display much of the calculus known for the subdifferential mapping [Roc83].

Another common approach to creating smooth substructure is to compose a convex function with
a smooth function. This approach has been used by many authors including Ioffe ([Iof79a], [Iof79b]
and [Iof79c]), Rockafellar ([Roc88]) and Wright ([Wri89]). We restrict ourselves to the approach
developed by Rockafellar in 1988.

Rockafellar's work focused on the composition of a convex function with a piecewise linear-
quadratic function. The addition of a basic constraint qualification forced the resulting function to
have a well behaved calculus [Roc88, Def 4.1]. The term *amenable* was not used until 1992 [PR92],
but the resulting class described in [Roc88] is exactly the class of fully amenable functions. Though this thesis only skims the topic of amenability, we include the definition here. (For definitions of the normal cone, null space, etc... please refer to Section 1.3.)

**Definition 1.2 (Amenable Functions)** A function $f$ is amenable at a point $\bar{x} \in \text{dom } f$ if $f$ can be expressed locally as $f := g \circ F$ for some proper lsc convex function $g$, and a $C^1$ mapping $F$, such that the constraint condition

$$N_{\text{cl dom } g}(F(\bar{x})) \cap \text{null } \nabla F(\bar{x}) = \{0\}$$

holds. Further, $f$ is strongly amenable if the function $F$ is $C^2$, while if $F$ is piecewise linear-quadratic then we say $f$ is fully amenable.

Equation (1.1) was introduced in [Roc88] under the name basic constraint qualification. Similar conditions have appeared elsewhere under the names transversality and nondegeneracy (see [Lew02] or [Sha03] for example). In this thesis we shall refer to equations (1.1) as the amenability condition, and refer to equations of this style as transversality conditions.

Rockafellar's 1988 work is focused on the development of the epi-derivative. He uses fully amenable functions as an example of functions which are twice epi-differentiable [Roc88, Thm 4.5]. The importance of this is, if a function is twice epi-differentiable, it has a second order expansion [Roc88, Prop 2.8]. Rockafellar furthered this result by providing formulae for the epi-derivative and epi-Hessian for fully amenable functions [Roc88, Thm 3.1 & Thm 4.5].

In 1992 Poliquin joined with Rockafellar to continue the work on amenability. Their joint paper [PR92] first introduced the term amenable, and provided the definitions of "amenable" and fully amenable (strongly amenable was defined later in [RW98]). They further provide examples showing amenable and fully amenable are distinct classes of functions [PR92, pp. 339-340]. Using results from their 1993 paper, Poliquin and Rockafellar developed optimality conditions and sensitivity analysis for amenable and fully amenable functions [PR92, Thm 3.5 & Thm 4.2].

Poliquin and Rockafellar's 1993 work introduces amenable sets by way of the indicator function [PR93, Def 2.6]. The work then focuses on the calculus of amenability, providing a sum rule, chain rule, set intersection rule, and set composition rule [PR93, Thm 3.1, Thm 3.5, Cor 3.2, & Cor 3.6].

On the surface the classes of lower-$C^2$ functions and amenable functions seem entirely different. For example, amenability is a pointwise property while lower-$C^2$ is defined over an open set. Moreover, they were developed by different approaches and with different goals in mind. Nonetheless there is a very strong correlation between the two. This correlation did not become evident until 1998 when Rockafellar and Wets defined strong amenability [RW98, Def 10.23]. Using this intermediate class between amenable and fully amenable, they showed the following remarkable result:

*A function is lower-$C^2$ about a point if and only if it is strongly amenable at that point and the point is in the interior of its domain.* [RW98, Ex 10.36]
This result holds the same surprise as Rockafellar's relation between lower-$C^2$ and lower-$C^\infty$ functions, showing a relationship between what would appear to be two completely different classes of functions. The style of this result is echoed throughout this thesis as we show many of the new notions of smooth substructure are either the same, or closely related (see Sections 2.4, 2.5, 4.1, and 4.2).

More recent notions of smooth substructure have hinged on studying the identification of active constraints. This history of this question is detailed in the next subsection; here we focus on its connection with smooth substructure.

In 1987 Dunn published an inspirational paper on identifying active constraints via the projected gradient algorithm [Dun87]. The paper's work on identifying active constraints is discussed in Subsection 1.2.2; here we focus on its contribution to smooth substructures. In [Dun87], Dunn introduced the notion of an open facet. An open facet (or quasi-polyhedral surface), is any surface of a set that locally appears "flat". In 1988 Burke and Moré used the notion of an open facet to create a new approach to the study of active constraint identification [BM88]. (Again, we forgo discussion on active constraint identification until Subsection 1.2.2.) Their work hinges on their proof that, on an open facet the normal cone is constant [BM88, Thm 2.3].

In 1993 Wright furthered the work of Burke and Moré by showing the normal cone being constant was not necessary, but only that the normal cone be continuous along some surface. Wright defines a surface to be identifiable if it is a manifold contained in the set along which the normal cone is continuous [Wri93, Def 2]. (Although Wright never uses the term “manifold”, his definition clearly makes use of one.) An identifiable surface therefore has two types of smooth substructure: the smooth surface (manifold) and the continuity of the normal cone. This may be the fundamental difference between the current notions studied in this thesis, and the notions of lower-$C^2$ and amenability (both of which only ask for one type of smooth substructure).

In 2002 Lewis took the next step in extending Burke and Moré's work [Lew02]. Though Lewis's main goal was sensitivity analysis, he was also interested in the identification of active constraints. His paper begins by listing several examples of nonsmooth optimization problems, and noting that if the “activity” for the solution is known, then the problem can be reduced to a smooth one [Lew02, p. 4]. His definition, partly smooth, is the focus of much of this thesis (see Definition 2.1). Unlike previous definitions, partial smoothness does not include convexity. However, like identifiable surfaces, partial smoothness uses two types of smooth substructure: the function must be smooth along a manifold and the subdifferential must be continuous along the same manifold. The connection between partly smooth functions, and identifiable surfaces is developed in [Lew02, Thm 6.3] and discussed later in this thesis (Theorem 2.21).

After defining partly smooth functions, Lewis proceeded by developing alternate definitions for partial smoothness [Lew02, Note 2.9]. We continue this in Lemma 2.3 of this thesis. Lewis also created a rich calculus for partial smoothness, and used it to develop several sensitivity analysis results. We discuss his calculus, and extend it somewhat in Section 3.1 and 3.2 of this work. Lewis
next related partial smoothness with other existing notions of smooth substructure. As mentioned, Sections 2.4, 2.5, 4.1, and 4.2 of this work focus on that task. Lewis ended his work with an example showing, even under favourable conditions, that strong critical points of a partly smooth function need not be minimums of the function [Lew02, Sec 7]. This example is reproduced in Example 2.16 and used throughout this thesis. We further show the condition lacking in his example and provide a theorem stating when a strong critical point of a partly smooth function must be a local minimum of the function (Theorem 3.23). The missing condition, prox-regularity, and its history, are discussed in Subsection 1.2.3 below.

Another approach to sensitivity analysis was put forth by Shapiro in 2003 [Sha03]. His idea stems from the notion of cone reducible sets [BS00, Def 3.135]. In [Sha03], Shapiro extends these sets to $g \circ F$ decomposable functions. His approach is similar to amenability, in that he writes a function as the composition of a smooth function and a convex function. His constraints on the convex function further narrow the class. Besides his work on sensitivity analysis [Sha03, Sec 5], Shapiro’s work also provides a plethora of examples and rich calculus for the class of $g \circ F$ decomposable functions [Sha03, Sec 2]. The relationship between $g \circ F$ decomposable functions and partly smooth function is discussed in Section 4.2 of this thesis.

In 1996 Mifflin noticed that the subdifferential of the simple max function, $\max\{f_1, f_2\}$, could be used find a subspace on which an “approximate Newton” step for minimization could be derived [Mif96]. This prompted the “Quasi-second-order Proximal Bundle” algorithm. The algorithm itself is discussed in the next subsection, while we end this subsection by discussing the notions of smooth substructure it helped develop. To do this we must first discuss work of Lemaréchal, Oustry, and Sagastizábal.

Seeking to create second order expansions for nondifferentiable functions, Lemaréchal, Oustry, and Sagastizábal, developed the idea of a $UV$-decomposition [LOS00]. Splitting the space $\mathbb{R}^n$ into two orthogonal subspaces $U$ and $V$, [LOS00] studies the directions in which a function is smooth ($U$) and in which it is sharp ($V$). Using these directions Lemaréchal, Oustry, and Sagastizábal define the $U$-Lagrangian, an approximation of the original function, and show along certain manifolds it can be used to create a second order expansion for a nondifferentiable function [LOS00, Sec 3.1].

Mifflin noticed that the $UV$-decomposition was exactly the space developed for the “Quasi-second-order Proximal Bundle” algorithm. This prompted the joint work with Sagastizábal [MS99], which shows how [Mif96] and [LOS00] relate. Focusing on finite max functions Mifflin and Sagastizábal calculated the $UV$-decomposition for such functions and discussed their $U$-Lagrangians. They continued by discussing the algorithm first suggested in [Mif96], rewriting it to a more concise form involving a “$V$-step” followed by a “$U$-step”.

In a related work Oustry studied the $U$-Lagrangian for the maximum eigenvalue function [Ous99]. His work finds a manifold on which the solution mapping for the $U$-Lagrangian lies, and shows that
the addition of a transversality condition forces the $U$-Lagrangian to be infinitely differentiable. He concludes by discussing the $UV$-algorithm of Mifflin and Sagastizábal in this setting.

When studying the $U$-Lagrangian, Lemaréchal, Oustry and Sagastizábal noted that the $U$-Lagrangian was often the restriction of the function to a “thick surface” [LOS00, p. 717]. This “thick surface” was later formalized in the definition of a fast track [MS02a]. A convex function is said to have a fast track if its $U$-Lagrangian is smooth and the solution mapping lies on a unique manifold [MS02a] (or Definition 2.25 of this thesis). The behaviour of the $UV$-algorithm when a function has a fast track is discussed in [MS02b], and shown to be favourable. Links between fast tracks and partly smooth functions are discussed in Section 2.5 of this thesis.

Also arising from $UV$-decomposition theory is the notion of a primal dual gradient (PDG) structure. Seeking to generalize the results of [MS99] to infinite max functions, Mifflin and Sagastizábal defined PDG structures. The definition of PDG structures has evolved over several papers ([MS00a] [MS00b] [MS03]); in this thesis we use the nonconvex definition, [MS03]. The links of PDG structures to partly smooth functions is discussed in [MS03], and reiterated in this thesis (Section 4.1).

The connections between identifiable surfaces, partial smoothness, $g \circ F$ decompositions, fast tracks and primal-dual gradient structures are discussed in a recent work by Hare [Har03] and in Sections 2.4, 2.5, 4.1, and 4.2 of this thesis.

1.2.2 Constraint Identification

As reflected in the last section, notions of smooth substructure are often inspired by the study of algorithmic convergence. As early as 1976 authors noted that certain algorithms would converge after a finite number of iterations, provided the problem had a favourable structure [Ber76] [Roc76]. By the late 1980's it was found that even if the algorithm did not converge in a finite number of steps, the active constraints could be determined finitely [Dun87] [BM88] [Lem89]. In this subsection we outline the progress in the theory behind identification of active constraints. We do not detail the actual algorithms here, instead leaving that for Chapter 5.

Often, when minimizing over a constraint set, the constraint set is defined in such a way that it can be partitioned into subsets of lower dimension. For example a polyhedron (in $\mathbb{R}^3$) can be partitioned into its interior, faces, edges, vertices. (Higher dimensions lead to analogous partitions.) When minimizing unconstrained problems, one can rewrite the problem to involve constraint sets by noting that $\min_x f = \min_{(x,t)} \{ t : f(x) \leq t \}$. The goal of constraint identification is twofold: to describe the partitions of the constraint set and to discover which partition the solution to a constrained minimization problem lies on. If these goals can be met, the dimension of the problem is decreased, as only the partition on which the solution lies needs to be considered.

One classical approach to constrained minimization is the gradient projection method. In this method one uses steepest descent, then solves a quadratic sub-problem to maintain feasibility [Ber95,
pp. 203-204]. This approach is most effective when the constraint set is linear, thus making the quadratic sub-problem easy [NW99, p. 455]. Good overviews can be found in [Ber95, Sec 2.3] and [NW99, Sec 16.6], while the complete algorithm is detailed in Subsection 5.2.1 of this thesis.

According to Bertsekas the gradient projection algorithm first appeared in the mid 1960's through the works of Goldstein (1964) and Levitin and Polyak (1965) [Ber76, p. 174]. However, our interest begins in 1976 when Bertsekas first discussed finite convergence of the algorithm [Ber76]. In this work Bertsekas shows that with several assumptions on the directional derivatives and Hessian mapping of a function, the active constraints of a minimization problem over a "box set" can be identified in a finite number of iterations [Ber76, Prop 3].

In 1987 Dunn replaced the assumptions Bertsekas laid on the directional derivative and Hessian mapping with the restriction that the minimum be a "uniformly isolated zero" and a nondegeneracy condition [Dun87, Thm 2.1]. This nondegeneracy condition has resurfaced in many pieces of literature on active constraint identification, including this thesis. Example 5.15 shows the necessity of this condition in the study of active constraint identification. In 1987 Dunn showed that if the minimum is a uniformly isolated zero and the nondegeneracy condition holds, that the active constraints for a linearly constrained minimization problem could be finitely identified [Dun87, Thm 2.1]. Dunn furthered his work by defining the idea of an open facet (see Subsection 1.2.1). Dunn then showed, that under the same conditions, the open facet on which a solution to a constrained minimization problem lies could be identified in a finite number of iterations [Dun87, Thm 3.1].

Also in 1987, Calamai and Moré published an important paper on identifying active constraints via the gradient projection algorithm [CM87]. Their major contribution was in defining the projected gradient [CM87, eq (3.1)] (also see Definition 5.3 of this thesis). Most of their work focuses around the properties of the projected gradient, and its uses in optimization. One important result shows how the projected gradient is a measure of optimality for a constrained minimization problem [CM87, Lem 3.1]. Calamai and Moré also make use of the nondegeneracy condition of Dunn to show that active constraints of a linear constraint set are identified finitely if and only if the projected gradient converges to zero [CM87, Thm 4.1]. This theorem is expanded to the nonconvex, nonlinear case in Theorem 5.7 of this thesis.

The ideas of the gradient projection method has also been extended to nondifferentiable functions. The subgradient projection method replaces the gradient map with the projection of zero onto the subdifferential [Flå92]. In 1992 Flåm showed how the nondegeneracy condition of Dunn resulted in finite identification of active constraints via the subgradient projection method [Flå92, Thm 3.1 & 4.1]. This thesis suggests Flåm's results might be expandable to a broader setting (see Note 5.9); however, this is left for later research.

We now turn our attention to Newton-like methods. A good overview on Newton-like methods can be found in [NW99, Sec 6, & Sec 8], while the complete algorithm can be found in Subsection 5.2.2 of this thesis.
In 1988 Burke and Moré showed that, under the nondegeneracy condition of Dunn, Newton-like methods identify the open facet on which a solution lies in a finite number of iterations [BM88, Thm 4.1]. This result appears as a side note however, as the focus of their work is a meta-algorithmic approach to constraint identification. (We discuss this approach shortly.) In Theorem 5.10, this thesis extends Burke and Moré's result on open facets to partly smooth sets.

In 1991 Al-Khayyal and Kyparisis developed some interesting results closely related to constraint identification via Newton-like methods [AKK91]. Al-Khayyal and Kyparisis considered the case of optimizing over a convex constraint set under the condition that the optimality point was a strict critical point. Under this condition they showed how any algorithm could be modified to ensure finite convergence to the solution [AKK91, pp. 329-330]. Their technique involved adding a Newton-like step after every iteration, and proving that the Newton-like method must converge finitely [AKK91, Thm 3.1]. These results can easily be rederived by combining Example 2.5, Corollary 3.20 and Theorem 5.10 of this thesis.

The identification of active constraints for unconstrained minimization problems is an equally rich field of research. One of the first examples lies in Rockafellar's work [Roc76], which discusses the proximal point method.

The proximal point method was first introduced by Martinet in 1970 [Mar70, Sec 4]. Later Martinet published a concise work focusing on the algorithm and the basic convergence results [Mar72]. Good overviews of the proximal point method can be found in [Lem89], [HUL93b, Sec XV.4] and [Ber95, Sec 5.4.6], while the complete algorithm appears in Section 5.3 of this thesis.

In 1976 Rockafellar showed the first finite convergence results for the proximal point algorithm [Roc76]. Specifically, if at some point zero is in the interior of the subdifferential of a convex function, then that point can be found via a finite number of iterations of the proximal point algorithm. He further shows that in the case when the function is polyhedral, the proximal point algorithm identifies the active constraints of the minimization problem regardless of the behaviour of the subdifferential [Roc76, Prop 8]. Theorem 5.14 of this thesis provides an alternate approach to both these results.

Also working with convex functions, Ferris further examined the finite convergence of the proximal point algorithm [Fer91]. His approach considered convex functions which grew sharply in directions away from the set of minima (see [Fer91, Def 1]). For such functions the proximal point method was shown to converge in a finite number of iterations [Fer91, Thm 6]. Ferris's ideas are captured in the sharpness conditions of partial smoothness.

Recent work by Mifflin and Sagastizábal on fast tracks has shown that the proximal point method identifies fast tracks for a convex function in a finite number of iterations [MS02a]. Simultaneous work by Gilbert, recently presented at the International Symposium on Mathematical Programming 2003, has shown similar results [Gil03]. Mifflin's "Quasi-second-order Proximal Bundle" algorithm, later called the "VU-proximal Point" algorithm, tries to make use of this fact by alternating proximal point steps and Newton-like steps [Mif96] [MS02b]. The algorithm is still conceptual as it requires
knowledge of the $U$ and $V$ subspaces discussed in Subsection 1.2.1 and Section 2.5 of this thesis. However, methods for approximating these subspaces can be found in [MS02b]. Convergence results for the algorithm can be found in [Mif96], [Ous99] and [MS02b] for various classes of functions.

In 1988 Burke and Moré commented that results on the identification of constraints had “only been established for a few algorithms, and then under restrictive hypothesis” [BM88, p. 1]. Their stated goal was to create a meta-algorithmic approach to constraint identification. Instead of showing that a specific algorithm identified active constraints, they focused on finding the characteristics an algorithm must have in order to identify active constraints. Specific algorithms could then be checked for these conditions, and new algorithms designed with these conditions in mind. In [BM88, Thm 3.4] Burke and Moré showed that, if the minimum of a smooth function over a constraint set occurs at a nondegenerate stationary point on an open facet, then any algorithm which forces the projected gradient to zero (and convergence to the minimum) identifies the open facet in a finite number of iterations. This statement subsumes the results of Calamai and Moré discussed above. Burke and Moré also gave examples to show this result was achievable [BM88, Sec 4]. This appears to mark the first time authors took this approach to the question. This meta-algorithmic style is echoed in Sections 3.3, 3.4, 5.2 and 5.3 of this thesis.

In 1993 Wright developed the idea of identifiable surfaces [Wri93]. Using these, he extended the work of Burke and Moré to a broader class of constraint sets. Identifiable surfaces are discussed in more detail in Subsection 1.2.1 and Section 2.4 of this thesis.

In 1990 Burke wrote a follow-up paper to his 1988 work with Moré [Bur90]. The paper considers the case of a nonconvex constraint set. By creating a convex linearization of the nonconvex constraint set, Burke applied the results of [BM88] to a nonconvex case. When successive linearizations of the constraint set become increasingly accurate and some nondegeneracy conditions hold, Burke shows the tangent cone to the constraint set at the optimal point can be finitely identified [Bur90, Thm 6.2]. If the user’s knowledge of the set is strong enough this can identify the active constraints; even if it is not, Burke’s results effectively identify the smooth subspace of a $UV$-decomposition.

Like Burke’s results, the results in Section 3.4 of this thesis make no assumptions on the convexity of constraint set. Instead prox-regularity is assumed. Prox-regularity subsumes the cases of Burke [Bur90], and also holds for many other examples (see Subsection 1.2.3 below or [PR96b]). The results of Section 3.4 also differ from Burke’s in that they apply directly to the identification of active constraints, and not the tangent cone to the constraint set.

The case of a polyhedral constraint set was reconsidered by Burke and Moré in [BM94]. This work shows that the nondegeneracy condition can be removed when the constraint set is polyhedral [BM94, Thm 4.2]. This is because in the polyhedral case the nondegeneracy condition must apply for some exposed face upon which the optimal point lies (combine Theorems 2.4 and 3.3 of [BM94] or see [BM94, p. 585]). Though new in study of meta-algorithmic constraint identification, this result was known to Rockafellar in 1976, when he used it to show finite convergence of the proximal point
algorithm [Roc76]. Burke and Moré's 1994 work, also discusses critical point analysis for the case of a polyhedral constraint set [BM94, Thm 6.1 & Thm 6.2]. Their work shows that, the behaviour of the problem on the exposed face is sufficient to discuss behaviour in a global sense. These results are extended to the partly smooth prox-regular case in Section 3.5 of this thesis.

In 1999 Facchinei, Fischer and Kanzow published another approach to identifying active constraints of optimization problems [FFK99]. Unlike previous techniques their approach uses duality theory. As a result the theory is more technical and less geometric than the style of this thesis. Because of this, the connections between their work and this thesis are unclear. Other algorithmic approaches to problems with smooth substructure include Osborne (see [Osb01] for example). We do not examine his approach in this thesis.

Most of the results on constraint identification in this thesis also appear in a recent paper by Hare and Lewis [HL03], based on this thesis (Sections 3.1 3.3, 3.5, and 3.6). Results showing constraint identification for specific algorithms (Sections 5.2 and 5.3) do not appear in the paper.

1.2.3 Projections and Prox-regularity

The 1993 work of Wright shows the projection mapping is closely related to the idea of identifying active constraints [Wri93]. His paper also shows that the smoothness of the projection mapping relies on the local behaviour of the boundary of the set [Wri93, Thm 2.6].

The question of how the projection mapping to a convex set relates to the set has been long standing in convex geometry [Hol73]. Classical results include the fact that the projection mapping to a convex set is Lipschitz continuous. This shows that the projection mapping is differentiable almost everywhere [Hol73]. In 1973 Holmes asked when is the projection mapping differentiable outside the set (the map is never differentiable along the boundary of the set) [Hol73]. This question is further inspired by well known connections of the projection map with the distance function [Hol73, eq (1.1)], and the inverse of the projection mapping with the normal cone [Hol73, eq (1.7)]. By using the implicit function theorem, Holmes shows if the convex set has a $C^p$ boundary then the projection mapping is $C^{p-1}$ [Hol73, Thm 2].

These results were extended in 1982 by Fitzpatrick and Phelps [FP82]. Their work shows the equivalence:

*The projection mapping to a convex set is $C^{p-1}$ everywhere outside the set if and only if the boundary of the set is $C^p$. [FP82, Thm 3.10]*

On the surface this seems to complete the study of relating the projection mapping and the convex set it projects onto. However, the globalness of the result leaves open the question, what if the set has a boundary that is $C^p$ at some points but not others?

One step towards answering this question might be to describe the projection mapping to a
convex set with smooth substructure. Wright’s 1993 work did exactly this, showing that if a convex set has a $C^p$-identifiable surface, then there exists a cone emanating from the surface on which the projection map is $C^{p-1}$ [Wri93, Thm 2.6]. Since a set is convex if and only if the projection mapping is single valued everywhere, extending Wright’s work to a nonconvex setting naturally contains some difficulties. However, Theorem 3.13 (the Smooth Projection Theorem) of this thesis extends Wright’s result to a nonconvex setting. The power to do this lies both in partial smoothness and a property defined in 1996 by Poliquin and Rockafellar: prox-regularity [PR96b].

Though prox-regularity was developed for functions [PR96b], it is somewhat easier to understand in the case of sets. It has been long known that the projection mapping is related to the idea of a normal cone. Specifically, the direction from a point to its projection onto a set is in the negative normal cone to the set at the projection [Loegl, Lem 3C.2]. If a normal vector has the opposite property (the projection returns small displacements in the normal direction to the point at which they are normal) it is called a called a proximal normal vector [Loegl, Rem 4A.2]. A concise overview of proximal normals can be found in [Loegl], while a more detailed study can be found in [W98]. The prox-regularity condition ensures that all normals are proximal normals [PR96b, Thm 1.31, the uses of which will be seen throughout this thesis. Before continuing our discussion of prox-regularity we discuss some related notions.

In 1995 Clarke, Stern, and Wolenski developed the notion of a proximally smooth set [CSW95]. A set is defined to be proximally smooth if the distance function to the set is $C^2$ in a “tube” about the set. This property allowed Clarke, Stern and Wolenski to study derivatives and regularity of the distance function in a nonconvex setting [CSW95, Sec 31. Due to the relations between distance functions and projections, a set being proximally smooth implies the projection mapping to the set is $C^1$ in a “tube” about the set [CSW95, Thm 4.11. This suggests that proximal smoothness is the property we require to extend Wright’s results to a nonconvex setting. However, proximal smoothness is a property applying to the whole set, not a local property about a point. This is stronger than we require. Prox-regularity distills this property down to the local setting we desire.

The definition of prox-regularity was first introduced in 1996 by Poliquin and Rockafellar [PR96b]. This work discussed the question of when the proximal envelope is convex [PR96b]. To approach the question Poliquin and Rockafellar defined prox-regularity [PR96b, Def 1.1] (Definition 1.12 of this thesis), and provided a collection of examples of functions which were prox-regular. These include smooth functions, convex functions, finite max functions, lower-$C^2$ functions, and strongly amenable functions [PR96b, pp. 1810-1811]. Poliquin and Rockafellar used prox-regularity to resolve the question of when the proximal envelope is convex [PR96b, Sec 4]. They further showed that prox-regularity has applications towards second order theory [PR96b, Sec 5]. These applications were explored in [PR96a]. Using the results from [PR96b], Poliquin and Rockafellar show a prox-regular function is twice epi-differentiable if and only if the proximal envelope has a second order expansion [PR96a, Thm 3.5].
Poliquin and Rockafellar's original paper also contains several alternate definitions of prox-regularity. For example, a function is prox-regular if and only if it has a (locally) premonotone subdifferential map \([PR96b, \text{Thm } 3.2 \& \text{Prop } 4.8]\). This shows prox-regularity is an extension of convexity, since all convex functions have monotone subdifferential maps.

By definition, a set is prox-regular whenever its indicator function is. Early work by Poliquin and Rockafellar showed, not unexpectedly, a function is prox-regular if and only if its epigraph is \([PR96b, \text{Thm } 3.5]\). This allows the results of prox-regularity developed in \([PR96b]\) and \([PR96a]\) to be shifted between functions and sets quite easily.

In 2000 Poliquin, Rockafellar and Thibault showed that prox-regularity was the correct method of changing proximal smoothness to a local setting \([PRTO0]\). Their main result shows that a set is prox-regular at a point if and only if the projection mapping is single valued near that point \([PRTO0, \text{Thm } 1.3]\). This result shows that prox-regularity is the exact extension of convexity needed for this thesis. Most of \([PRTO0]\) is dedicated to the proof of this result and its various equivalences. Lemma 1.14 of this thesis collects the tools which we find of use.

1.2.4 Algorithm Design

While predominantly theoretical, this thesis is substantially motivated by algorithm design. A brief overview of algorithm design is discussed in Section 5.1. For more information we refer the reader to the plethora of texts on the topic, \([HUL93a]\) [HUL93b] [Ber95] and \([NW99]\) amongst others.

Recent work by Burke, Lewis and Overton suggests an approach to creating and using approximations to the subdifferential map \([BLO02]\). In it they recall that, in applied optimization, most functions are differentiable almost everywhere \([BLO02, \text{pp. } 567, 568]\). In such circumstances the subdifferential map can be written as the convex closure of all limiting gradients \([BLO02, \text{eq. } (1.1)]\). This suggests a method of random gradient sampling to approximate subdifferential maps.

As a first step to studying basic algorithms for optimization over nonsmooth sets, in Section 5.4 we attempt to emulate this idea for the case of constrained optimization. The Random Normal Generation algorithm (p. 94) attempts to generate approximate normal cones to a set described only by an oracle via random sampling. Section 5.4 also contains several results suggesting that the approximate normal cones created by the algorithm will converge to the correct normal cone. Some preliminary numerical results suggesting the algorithm's promise are given in Section 5.5.

1.3 Definitions

We assume the reader is familiar with the basic notions of mathematical analysis, and in general we will follow the notation laid out by Rockafellar and Wets in \([RW98]\). To aid the reader, a list of the
CHAPTER 1. INTRODUCTION

Notation used throughout this thesis can be found in Appendix A.

1.3.1 Basic Notation

We denote the set of all real numbers by \( \mathbb{R} \), the set of all positive real numbers by \( \mathbb{R}_+ \), and the set of extended real numbers by \( \bar{\mathbb{R}} := \mathbb{R} \cup \{\infty\} \). Similarly \( \mathbb{R}^n \) is the real vector space of dimension \( n \), while \( \mathbb{R}_+^n \) is the positive orthant contained in \( \mathbb{R}^n \). Unless otherwise stated, functions map \( \mathbb{R}^n \) into \( \mathbb{R} \), and points and sets are contained in \( \mathbb{R}^n \). We define the open ball of radius \( \varepsilon \) about the point \( \bar{x} \) by
\[
B_\varepsilon(\bar{x}) := \{x : |x - \bar{x}| < \varepsilon\},
\]
where \(|\cdot|\) is the Euclidean \( l_2 \) norm of a vector. To distinguish norms, we use \( || \cdot || \) to denote the \( l_2 \) norm of a matrix.

Basic set and function operations will be denoted by abbreviations of their names. For a set \( S \) we denote the closure by \( \text{cl} S \), the interior by \( \text{int} S \), the relative interior by \( \text{rint} S \), and the convex hull by \( \text{conv} S \). The subspace parallel to a nonempty convex set \( C \) is created by selecting any point \( \bar{z} \in \text{rint} C \), shifting \( C \) by that point and taking all scalar multiples of the result:
\[
\text{par} C := \mathbb{R}_+(C - \bar{z}).
\]
(1.2)
It is easy to confirm that this subspace is unique and independent of the point selected. For a function \( f \) we define the domain by \( \text{dom} f := \{x : f(x) < \infty\} \), while for a set valued map \( G \) the domain is \( \text{dom} G := \{x : F(x) \neq \emptyset\} \). For a vector valued map \( F : \mathbb{R}^n \to \mathbb{R}^m \) we denote the range by \( \text{rng} F \), and the null set by \( \text{null} F \). For a sublinear function \( f \) we define the linearity space by
\[
\text{lin} f := \{x : f(x) = -f(-x)\}.
\]
(1.3)

For a function \( f \) we use \( \min\{f\} \) to denote the minimum of \( f \), \( \max\{f\} \) the maximum, \( \inf\{f\} \) the infimum, \( \sup\{f\} \) the supremum, \( \arg\min\{f\} \) the argument of the minimum (set of all minimizing points), and \( \arg\max\{f\} \) the argument of the maximum (set of all maximizing points). The minimum notation is augmented in the usual manner: \( \min_S\{f\} \) is the minimum of the function \( f \) over the set \( S \), \( \min_y\{f(x,y)\} \) is the minimum of \( f(x,\cdot) \) over all choices of \( y \), and \( \min\{f : A\} \) is the minimum of \( f \) such that condition \( A \) is true. The max, \( \inf, \sup, \arg\min, \text{and} \arg\max \) are likewise augmented.

We shift from sets to functions via the indicator function. A set \( S \) has indicator function
\[
\delta_S := \begin{cases} 0 & x \in S \\ \infty & x \notin S \end{cases}
\]
(1.4)
To move from a function \( f \) to a set we make use of both level sets and epigraphs, defined:
\[
\text{lev} f := \{x : f(x) \leq 0\}, \quad \text{and}
\]
(1.5)
\[
\text{epi} f := \{(x,\alpha) : \alpha \geq f(x)\}.
\]
(1.6)

We say a function \( f \) belongs to the class of \( C^p \) functions (abbreviated \( f \in C^p \)) if it is \( p \) times continuously differentiable. A function is \( C^p \) at a point \( \bar{x} \) if it is \( p \) times continuously differentiable
at \( \bar{x} \), and a function is \( C^p \) near a point \( \bar{x} \) if it is \( p \) times continuously differentiable on some open neighbourhood of \( \bar{x} \). We denote these derivatives by \( \nabla f, \nabla^2 f, \ldots \nabla^p f \). Throughout this thesis we will use the term smooth to mean \( C^p \) where \( p \) is arbitrary, but fixed for the remainder of the proof or definition. In cases where more detail is required, or more than one function or set is involved we use the more precise notation \( C^p \).

We say a function \( f \) is \( o(g) \) at a point \( \bar{x} \) if

\[
\lim_{x \to \bar{x}} \frac{f(x)}{g(x)} = 0, \tag{1.7}
\]

while \( f \) is \( O(g) \) if there exists a \( K > 0 \) such that

\[
|f(x)| \leq K|g(x)| \text{ for all small } x. \tag{1.8}
\]

Loosely translated these state, \( f \) converges to zero faster than \( g \) (little 'oh') and \( f \) converges to zero at least as fast as \( g \) (big 'oh'). As such, the definitions are only of interest if the function \( g \) is continuous with \( g(0) = 0 \).

This completes the basic notation used throughout this thesis. For easy reference we suggest the reader make Appendix A readily available.

### 1.3.2 Nonsmooth Analysis

In order to work with discontinuous functions semi-continuity is a useful notion. We include it here, in its function and set form. In order to state the definitions of semi-continuity we require the notions of lower, upper, inner, and outer limits.

**Definition 1.3 (Lower, Upper, Inner and Outer Limits)** For a function \( f \) and a point \( \bar{x} \), the lower limit at \( \bar{x} \) and upper limit at \( \bar{x} \) are defined

\[
\liminf_{x \to \bar{x}} f(x) := \lim_{\delta \to 0} \left[ \inf_{x \in B_\delta(\bar{x})} f(x) \right],
\]

and

\[
\limsup_{x \to \bar{x}} f(x) := \lim_{\delta \to 0} \left[ \sup_{x \in B_\delta(\bar{x})} f(x) \right].
\]

For a set valued map \( F \) and a point \( \bar{x} \) the inner limit at \( \bar{x} \) and outer limit at \( \bar{x} \) are defined

\[
\liminf_{x \to \bar{x}} F(x) := \{ \bar{u} : \text{ for all } x_k \to \bar{x}, \text{ there exists } u_k \to \bar{u} \text{ with } u_k \in F(x_k) \},
\]

and

\[
\limsup_{x \to \bar{x}} F(x) := \{ \bar{u} : \text{ there exists } x_k \to \bar{x}, \text{ and } u_k \to \bar{u} \text{ with } u_k \in F(x_k) \}.
\]
CHAPTER I. INTRODUCTION

From the definition it is clear that, for a function \( f \) and a set valued mapping \( F \), the following inclusions always hold.

\[
\liminf_{z \to \bar{x}} f(z) \leq f(\bar{x}) \leq \limsup_{z \to \bar{x}} f(z) \\
\liminf_{z \to \bar{x}} F(z) \subseteq F(\bar{x}) \subseteq \limsup_{z \to \bar{x}} F(z).
\]

With this in mind we now state the following definitions of semi-continuity.

**Definition 1.4 (Semi-continuity)** A function \( f \) is lower semi-continuous (Lsc) at a point \( \bar{x} \) if

\[
\liminf_{z \to \bar{x}} f(z) \geq f(\bar{x}),
\]

while it is upper semi-continuous (usc) at \( \bar{x} \) if

\[
\limsup_{z \to \bar{x}} f(z) \leq f(\bar{x}).
\]

A set valued map \( F \) is inner semi-continuous (isc) at a point \( \bar{x} \) if

\[
\liminf_{z \to \bar{x}} F(z) \supseteq F(\bar{x}),
\]

while it is outer semi-continuous (osc) at \( \bar{x} \) if

\[
\limsup_{z \to \bar{x}} F(z) \subseteq F(\bar{x}).
\]

The mapping is called continuous at \( \bar{x} \) if it is both isc and osc there.

All definitions may also be applied globally, in that a function or set valued map is lsc (usc, isc, osc) if it is lsc (usc, isc, osc) at all points in its domain.

The term "closed" is often used in place of lsc. This is because a function is lsc if and only if its epigraph is closed [RW98, Thm 1.6]. It is also not difficult to see that the classical notion of continuity corresponds with a function being both lsc and usc [RW98, Ex 1.12].

For a set \( S \) the polar cone to \( S \) is

\[
S^\circ := \{d : \langle d, x \rangle \leq 0 \text{ for all } x \in S\}.
\] (1.9)

Often we make use polar cones in the case when \( S \) is a subspace of \( \mathbb{R}^n \). In this case, the polar cone is the subspace perpendicular to \( S \). We say two cones are polar if the polar of each is the other.

Closely related to the notion of a polar cone are the definitions of the normal and tangent cones. Loosely speaking, the normal cone to a set is the set of all directions pointing "orthogonally" away from the set, whereas the tangent cone is the set of all directions pointing into the set. The exact definitions follow.
Definition 1.5 (Normal and Tangent Cones) For a set $S$ containing a point $\bar{x}$ the regular normal cone to $S$ at $\bar{x}$ is defined as

$$\tilde{N}_S(\bar{x}) := \{ n : \langle n, x - \bar{x} \rangle \leq o(|x - \bar{x}|) \text{ for } x \in S \}.$$  

From this we create the normal cone (limiting normal cone), defined

$$N_S(\bar{x}) := \limsup_{z \to \bar{x}} \tilde{N}_S(z).$$

For a set $S$ containing a point $\bar{x}$ the tangent cone to $S$ at $\bar{x}$ is defined

$$\limsup_{\tau \to 0} \frac{S - \bar{x}}{\tau}.$$  

From this we create the regular tangent cone, defined

$$\tilde{T}_S(\bar{x}) := \liminf_{x \to \bar{x}} T_S(x).$$

The regular normal cone, normal cone, regular tangent cone, and tangent cone are all defined to be empty whenever evaluated at a point outside of the set.

Sets for which the regular normal cone and normal cone agree are called regular. The regularity of a function is checked via its epigraph.

Definition 1.6 (Regularity) If a set is locally closed at a point $\bar{x}$, and the regular normal cone and normal cone coincide at $\bar{x}$, we say the set is regular (Clarke regular) at $\bar{x}$. If the set is regular at every point it contains we call the set regular.

A function is regular (at a point $\bar{x}$) if its epigraph is regular (at $\bar{x}$).

The normal and tangent cones share a simple polarity relationship. The polar of the normal cone is the regular tangent cone, while the polar of the tangent cone is the regular normal cone [RW98, Thm 6.28]. Since polars cones are always convex, the regular normal cone and regular tangent cone are always convex. A second corollary is that a set is regular at a point if and only if the tangent cone and regular tangent cone coincide at that point [RW98, Cor 6.29]. Thus regularity implies that the normal and tangent cones are both convex. Lastly, if a set is regular then the regular normal cone mapping is osc, while the tangent cone mapping is isc [RW98, Cor 6.29].

It is not difficult to create examples of regular functions. For example, all convex functions, finite max functions, smooth functions, and lower-$C^2$ functions are regular [RW98, Ex 7.27, Ex 7.28, (Ex 7.28) & Thm 10.31]. It is a simple exercise to show a set is regular if and only if its indicator function is regular [RW98, Ex 7.28].

Our next goal is to define subgradients and provide some insight into the relationship between normal cones and subgradients. To do this we first define the subderivative of a function.
Definition 1.7 (Subderivatives) For a function \( f \) and a point \( \bar{x} \) where \( f \) is finite, the subderivative of \( f \) at \( \bar{x} \) in the direction \( \bar{d} \) is defined as

\[
\frac{df(\bar{x}, \bar{d})}{d} := \liminf_{\tau \to 0, d \to \bar{d}} \frac{f(\bar{x} + \tau \bar{d}) - f(\bar{x})}{\tau}.
\]

Using the subderivative we can define subgradients, and the subdifferential map.

Definition 1.8 (Subgradients and the Subdifferential) For a function \( f \) and a point \( \bar{x} \) where \( f \) is finite, the regular subdifferential is defined as

\[
\hat{\partial}f(\bar{x}) := \{ w : (w, d) \leq df(\bar{x}, d) \text{ for all } d \}.
\]

Vectors in the regular subdifferential are called regular subgradients.

From the regular subdifferential we create the (limiting) subdifferential as follows:

\[
\partial f(\bar{x}) := \limsup_{\tau(x) \to f(x)} \hat{\partial}f(x).
\]

Vectors in the subdifferential are called (limiting) subgradients.

We also create the horizon subdifferential,

\[
\partial^\infty f(\bar{x}) := \{ w : \text{there exists } x_k \to \bar{x}, w_k \in \hat{\partial}f(x_k), \lambda_k \to 0 \text{ with } f(x_k) \to f(\bar{x}), \lambda_k w_k \to w \}.
\]

Vectors in the horizon subdifferential are called horizon subgradients.

Loosely speaking, regular subgradients are the slopes of affine maps which undercut the function, while subgradients are the limits of all such nearby slopes. It is reassuring to know that when a function is differentiable the regular subdifferential is the single gradient vector [RW98, Ex 8.8], while if the function is smooth then the subdifferential is the single gradient vector [RW98, Ex 8.8].

From the definition it is clear that the regular subdifferential is a convex set contained in the subdifferential [RW98, Thm 8.6]. If a function is regular then the regular subdifferential and subdifferential map agree [RW98, Cor 8.11]. Thus, when a function is regular the subdifferential is convex.

The subdifferential map is closely related to the normal cone map via the epigraph of a function. Specifically, for a function \( f \) and a point \( \bar{x} \) where \( f \) is finite, the subdifferential and normal cone are related by the equations

\[
\partial f(\bar{x}) = \{ v : (v, -1) \in N_{\text{epi}} f(\bar{x}, f(\bar{x})) \},
\]

and

\[
N_{\text{epi}} f(\bar{x}, f(\bar{x})) = \{ \lambda (v, -1) : v \in \partial f(\bar{x}), \lambda > 0 \} \cup \{(v, 0) : v \in \partial^\infty f(\bar{x}, f(\bar{x})) \},
\]

[RW98, Thm 8.9]. These relationships prompt the definition of the Clarke subdifferential, below.
Definition 1.9 (Clarke Subdifferentials) For a function $f$ and a point $x$ where $f$ is finite, the Clarke subdifferential of $f$ at $x$ is defined

$$\partial f(x) := \{ w : (w, -1) \in \text{conv} \mathcal{N}_{\text{epi}} f(x, f(x)) \}.$$ 

If a function is regular, the regular, limiting and Clarke subdifferentials agree.

The subdifferential is often used to check necessary conditions of optimality. If a point is a local minimizer for a function, the subdifferential at the point must contain zero [RW98, Thm 10.1]. This leads to the definition of a critical point. Since we will only use critical points when the function is regular, we restrict our definition to this case.

Definition 1.10 (Critical Points) Given a regular function $f$, a point $x$ is a critical point of $f$ if $0 \in \partial f(x)$, a nondegenerate critical point of $f$ if $0 \in \text{rint} \partial f(x)$ and a strict critical point of $f$ if $0 \in \text{int} \partial f(x)$.

In the case of a smooth function $f$ and a regular set $S$, a point $x$ is a critical point of $f$ restricted to $S$ if $-\nabla f(x) \in N_S(x)$, a nondegenerate critical point of $f$ restricted to $S$ if $-\nabla f(x) \in \text{rint} N_S(x)$ and a strict critical point of $f$ restricted to $S$ if $-\nabla f(x) \in \text{int} N_S(x)$.

As mentioned for a point to be a local minimizer of a function (over a set) it must be a critical point of the function. Strict critical points are much stronger in that they are not only a necessary condition of minimization, but a sufficient condition. Nondegenerate critical points lie somewhere between critical and strict critical points, and are used throughout this work.

The final notion we must introduce in this section is prox-regularity. In order to understand prox-regularity we formally define the distance function and projection mapping.

Definition 1.11 (Distances and Projections) The distance from a point $x$ to a set $S$ is defined

$$\text{dist}(x, S) = \inf \{|x - z| : z \in S\}.$$ 

The related set valued mapping

$$P_S(x) := \text{argmin} \{|x - z| : z \in S\},$$

is called the projection mapping for $S$.

It is clear that for any point in a set, the distance of that point to the set is 0, while the projection mapping is the single point. Therefore the interest in distances and projections is focused on points outside of the set. In some cases the projection mapping is not single valued (project the origin onto the unit sphere), while in others it is empty (project onto any open set). However, if the set is closed and convex, the projection mapping is always single valued. Indeed, it is a classic result that, the
projection mapping is single valued everywhere if and only if the set is closed and convex ([CSW95, Cor 4.12] amongst many others). In order to gain some control over the projection mapping without forcing convexity onto the set, we make use of prox-regularity. We define this next.

**Definition 1.12 (Prox-regularity)** A function $f$ is prox-regular at a point $\bar{x}$ for a subgradient $\bar{w} \in \partial f(\bar{x})$ if $f$ is finite and locally lsc at $\bar{x}$, and

there exists $\varepsilon > 0$ and $\rho \geq 0$ such that

$$f(x') - f(x) - \langle w, x' - x \rangle \geq -\frac{\rho}{2} |x' - x|^2$$

whenever $x', x \in B_\varepsilon(\bar{x})$, $w \in \partial f(x)$, $|w - \bar{w}| < \varepsilon$, and $|f(x) - f(\bar{x})| < \varepsilon$.

When this holds for all $\bar{w} \in \partial f(\bar{x})$ we say $f$ is prox-regular at $\bar{x}$, while if $f$ is prox-regular at all points in its domain we say $f$ is prox-regular.

A set is prox-regular (at a point, for a normal vector) if its indicator function maintains this property.

It is not difficult to show that, if a function is prox-regular at a point for a subgradient then the subgradient is a regular subgradient. Thus, if a function is prox-regular at a point, it is regular at that point [RW98, p. 610]. It is also easy to show, if a function is prox-regular at a point, then it is prox-regular at all nearby points. Analogous results hold for prox-regular sets.

Examples of prox-regularity are almost as abundant as examples of regularity. All convex functions, smooth functions, finite max functions, and lower-C² functions are prox-regular everywhere [RW98, Ex 13.30, p. 13.27, (Prop 13.33) & Prop 13.33].

The following alternate definition of prox-regular set can be proven by simply applying the prox-regularity definition to the indicator function [RW98, Ex 13.31].

**Lemma 1.13** A closed set $S$ is prox-regular at a point $\bar{x} \in S$ for a normal vector $\bar{n} \in N_S(\bar{x})$ if and only if there exists $\varepsilon > 0$ and $\rho \geq 0$ such that

$$\langle n, x' - x \rangle \leq \frac{\rho}{2} |x' - x|^2$$

whenever $x, x' \in S \cap B_\varepsilon(\bar{x})$, $n \in N_S(x)$, and $|n - \bar{n}| < \varepsilon$.

The next lemma explains why prox-regularity is so key in the study of projection mappings.

**Lemma 1.14 (Prox-normal Neighbourhood)** Suppose the set $S$ is closed. Then, $S$ is prox-regular at the point $\bar{x} \in S$ if and only if the projection mapping $P_S$ is single valued near $\bar{x}$.

In this case there exists an open neighbourhood $\mathcal{N}$ of $\bar{x}$ on which the following properties hold:
CHAPTER 1. INTRODUCTION

(1) $P_S(\cdot)$ is single valued and Lipschitz continuous on $\mathcal{N}$.

(2) $P_S(\cdot) = (I + N_S)^{-1}(\cdot)$ on $\mathcal{N}$.

(3) For any point $x \in \mathcal{N}$, all normal vectors to $S$ in $(\mathcal{N} - x)$ are "proximal normals": that is, for $x$ and $v$ in $\mathcal{N}$, $v - x \in N_S(x)$ implies $x = P_S(v)$.

Proof: The equivalence statement is shown in [PRT00, Thm 1.3]. Parts (i) and (ii) can be found in [RW98, Ex 13.38], while part (iii) is a simple consequence.

\[\square\]

1.4 Manifolds

The idea of a manifold is key to this thesis. Loosely speaking a manifold is a surface in $\mathbb{R}^n$ that locally behaves like $\mathbb{R}^m$. For example, though we all know the earth is round, we locally describe it by flat maps. This is because locally the surface of the earth behaves like $\mathbb{R}^2$. An excellent introduction to manifolds can be found in Chapter 13 of [Str95]. We provide the formal definition of a manifold below.

Definition 1.15 (Manifold) Let the point $\bar{x}$ be contained in the set $\mathcal{M} \subseteq \mathbb{R}^n$. We say $\mathcal{M}$ is a $C^p$-manifold about $\bar{x}$ if there exists a $C^p$ function $G : \mathbb{R}^m \rightarrow \mathbb{R}^n$, and a point $\bar{y}$ such that

(i) $G(\bar{y}) = \bar{x},$

(ii) $\nabla G(\bar{y})$ is one to one, and

(iii) locally $\mathcal{M}$ is $\{G(y) : y \in \mathbb{R}^m\}$ (i.e. locally $\mathcal{M}$ is the image of $G$).

If this is true for all $\bar{x} \in \mathcal{M}$ then we simply call $\mathcal{M}$ a $C^p$-manifold. In the case of $p = 2$ we simply use the term manifold.

In this thesis we will often restrict our consideration of a function to how it behaves on a manifold. This leads to two important definitions. The first allows a concise language for describing the smoothness of a function relative to a manifold, while the second provides the language for describing critical point analysis when restricting functions to manifolds.

Definition 1.16 (Smooth Restrictions) Let $\mathcal{M}$ be a $C^p$-manifold. We say a function $f$ is smooth at a point $\bar{x}$ relative to $\mathcal{M}$, if there exists a function $g$ such that $g$ is smooth at $\bar{x}$, and $f(\bar{x}) = g(\bar{x})$ for all $\bar{x} \in \mathcal{M}$. We say a function $f$ is smooth along $\mathcal{M}$ if there exists a smooth function $g$ such that $f(x) = g(x)$ for all $x \in \mathcal{M}$. 
Definition 1.17 (Strong Critical Point) Let the regular function \( f \) be smooth along the manifold \( M \). We call a point \( \bar{x} \) a strong critical point of \( f \) relative to \( M \) if

\[
0 \in \text{rint } \partial f(\bar{x})
\]

and there exists \( \epsilon > 0 \) such that

\[
f(x) \geq f(\bar{x}) + \epsilon|x - \bar{x}|^2
\]

for all points \( x \in M \) near \( \bar{x} \).

In Definition 1.15 we defined manifolds as the images of smooth maps. This may not be the intuitive approach as we said manifolds are sets that locally behave like surfaces of lower dimension. Instead we might define a manifold as the pre-image of a smooth function mapping to \( \mathbb{R}^m \).

Theorem 1.18 (Alternate Manifold Definition) Let the point \( \bar{x} \) be contained in the set \( M \subseteq \mathbb{R}^n \). Then the following are equivalent:

(i) \( M \) is a \( C^p \)-manifold about \( \bar{x} \);

(ii) there exists a \( C^p \) function \( F : \mathbb{R}^n \rightarrow \mathbb{R}^m \) such that locally

\[
M = \{ x : F(x) = F(\bar{x}) \},
\]

and \( \nabla F(\bar{x}) \) is onto.

Proof: Chapter 13 of [Str95] is devoted to this topic.

1.4.1 Normal and Tangent Cones

As locally a manifold behaves like \( \mathbb{R}^m \), one would expect the normal and tangent cones to manifold to be perpendicular subspaces. This indeed is true, a fact that will be important throughout this thesis. In this subsection we provide descriptions of the normal and tangent cones to various manifolds. We begin with a basic result showing how to find the normal and tangent cones to a manifold under our two equivalent definitions.

Theorem 1.19 (Tangents and Normals to Manifolds) Suppose the function \( G : \mathbb{R}^m \rightarrow \mathbb{R}^n \) is \( C^p \), and the point \( \bar{y} \) satisfies \( \nabla G(\bar{y}) \) is one-to-one. Define the manifold \( M := \{ G(y) : y \in \mathbb{R}^m \} \). Then the tangent and normal cones to \( M \) at \( \bar{x} := G(\bar{y}) \) can be found via the formulae

\[
T_M(\bar{x}) = \text{rng } \nabla G(\bar{y}) := \{ \nabla G(\bar{y})(y) : y \in \mathbb{R}^m \}, \quad \text{and}
\]

\[
N_M(\bar{x}) = \text{ker } \nabla G(\bar{y}) := \{ y : \nabla G(\bar{y})(y) = 0 \}, \quad \text{and}
\]
CHAPTER 1. INTRODUCTION

\[ N_M(\bar{z}) = \text{null } \nabla G(\bar{y})^* := \{d : \nabla G(\bar{y})^* (d) = 0\}. \]  

(1.14)

Alternately, suppose the function \( F : \mathbb{R}^n \to \mathbb{R}^m \) is \( C^p \), and the point \( \bar{x} \) satisfies \( \nabla F(\bar{x}) \) is onto. Define the manifold \( M := \{x : F(x) = F(\bar{x})\} \). Then the tangent and normal cones to \( M \) at \( \bar{x} \) can be found via the formulae

\[ T_M(\bar{x}) = \text{null } \nabla F(\bar{x}) := \{d : \nabla F(\bar{x}) (d) = 0\}, \quad \text{and} \]

(1.15)

\[ N_M(\bar{x}) = \text{rng } \nabla F(\bar{x})^* := \{\nabla F(\bar{x})^* (y) : y \in \mathbb{R}^m\}. \]

(1.16)

Proof: Chapter 13 of [Str95] gives an excellent overview of these ideas, while [RW98, Ex 6.8] gives the details required for our more general setting.

This thesis will make use of the normal and tangent cones to a manifold in various special settings. The first of these will be in Theorem 2.18 where we compare a partly smooth function to its epigraph. To do this we need a way to describe the normal and tangent cones to the graph of a smooth function. Though this theorem is certainly known, for completeness sake we include its proof here.

**Theorem 1.20** Suppose \( \mathcal{M} \) is a \( C^p \)-manifold about the point \( \bar{x} \), and the function \( f \) is \( C^p \) along \( \mathcal{M} \). Then the set

\( \mathcal{M} := \{(x, f(x)) : x \in \mathcal{M}\} \)

is a \( C^p \)-manifold about the point \( (\bar{x}, f(\bar{x})) \). Moreover if \( f_m \) is any smooth function agreeing with \( f \) on \( \mathcal{M} \), then the tangent and normal cones to \( \mathcal{M} \) at \( (\bar{x}, f(\bar{x})) \) can be found via the formulae

\[ T_{\mathcal{M}}(\bar{x}, f(\bar{x})) = \{(t, \nabla f_m(\bar{x}) (t)) : t \in T_M(\bar{x})\} \quad \text{and} \]

\[ N_{\mathcal{M}}(\bar{x}, f(\bar{x})) = \{(n + \lambda \nabla f_m(\bar{x})^* (1), -\lambda) : n \in N_M(\bar{x}), \lambda \in \mathbb{R}\}. \]

Proof: Since \( \mathcal{M} \) is a manifold, there exists a point \( \bar{y} \in \mathbb{R}^m \) and \( C^p \) function \( F : \mathbb{R}^m \to \mathbb{R}^n \) such that, \( F(\bar{y}) = \bar{x}, \nabla F(\bar{y}) \) is one to one, and \( \mathcal{M} = \{F(y) : y \in \mathbb{R}^m\} \). Select \( f_m \) a \( C^p \) function which agrees with \( f \) on \( \mathcal{M} \) and define the function \( G(u) = (F(u), f_m(F(u))) \). As the composition of \( C^p \) functions, \( G \) is \( C^p \). Moreover \( \mathcal{M} = \{G(y) : y \in \mathbb{R}^m\} \), and \( G(\bar{y}) = (\bar{x}, f_m(\bar{x})) \). Lastly \( \nabla G(\bar{y}) = (\nabla F(\bar{y}), \nabla f_m(F(\bar{y})) \nabla F(\bar{y})) = (\nabla F(\bar{y}), \nabla f_m(\bar{x}) \nabla F(\bar{y})) \) by the chain rule. Since, \( \nabla F(\bar{y}) \) is one to one, \( \nabla G(\bar{y}) \) is one to one. Therefore \( \mathcal{M} \) is a \( C^p \)-manifold about \( (\bar{x}, f(\bar{x})) \).

To see the tangent and normal cone formulae we apply Theorem 1.19,

\[ T_{\mathcal{M}}(\bar{x}, f(\bar{x})) = \text{rng } \nabla G(\bar{y}) \]

\[ = \{(\nabla F(\bar{y})(y), \nabla f_m(\bar{x}) \nabla F(\bar{y})(y)) : y \in \mathbb{R}^m\} \]

\[ = \{(t, \nabla f_m(\bar{x}) (t)) : t \in T_M(\bar{x})\}, \]
while
\[
N_{\tilde{\mathcal{L}}}(\bar{x}, f(\bar{x})) = \text{nul} \, \nabla G(\bar{g})^* \\
= \{(d, -\lambda) : (\nabla F(\bar{g}), \nabla f_m(\bar{x}) \nabla F(\bar{g}))^* (d, -\lambda) = 0\} \\
= \{(d, -\lambda) : \nabla F(\bar{g})^* (d) + (\nabla f_m(\bar{x}) \nabla F(\bar{g}))^* (-\lambda) = 0\} \\
= \{(d, -\lambda) : \nabla F(\bar{g})^* (d) = \nabla F(\bar{g})^* (\lambda \nabla f_m(\bar{x})^*)(1)\} \\
= \{(n + \lambda \nabla f_m(\bar{x})^*)(1), -\lambda) : n \in N_M(\bar{x})\}.
\]

A second circumstance that will arise in this thesis is the result of adding two manifolds together. We conclude this chapter with a result describing when the resulting set is a manifold, and if it is a manifold what its normal and tangent cones are. The classical result is easily checked via equation (1.13) and (1.14) and the inverse function theorem.

**Theorem 1.21 (Manifold Addition)** Let $M_u \subseteq \mathbb{R}^n$ be a manifold about the point $\bar{u}$, and $M_v \subseteq \mathbb{R}^n$ be a manifold about the point $\bar{v}$ such that

Then

$$T_{M_u}(\bar{u}) \cap T_{M_v}(\bar{v}) = \{0\}. \tag{1.17}$$

is a manifold about the point $\bar{u} + \bar{v}$. Moreover the normal and tangent cones to this manifold can be found at $\bar{u} + \bar{v}$ via the formulae

$$N_{M_u + M_v}(\bar{u} + \bar{v}) = N_{M_u}(\bar{u}) \cap N_{M_v}(\bar{v}), \tag{1.18}$$

and

$$T_{M_u + M_v}(\bar{u} + \bar{v}) = T_{M_u}(\bar{u}) + T_{M_v}(\bar{v}). \tag{1.19}$$

Lastly, for any point $w$ near $\bar{u} + \bar{v}$ in $M_u + M_v$ there exists unique points $w_u \in M_u$ and $w_v \in M_v$ depending continuously on $w$ such that $w = w_u + w_v$. 
Chapter 2

Partial Smoothness and Related Concepts

The desire for smooth substructure to a nonsmooth problem arises in many areas of nonsmooth optimization. Smooth substructures have been used successfully in algorithmic research, sensitivity analysis, and developing calculus rules (see Section 1.2 for more detail). This thesis focuses largely on a notion of smooth substructure introduced in [Lew02], partial smoothness. A brief history of partial smoothness can be found in Subsection 1.2.1.

After defining partial smoothness this chapter argues that partial smoothness is a naturally occurring tool in nonsmooth optimization. We begin by providing a large collection of examples of functions that both are and are not partly smooth. We then compare partial smoothness of functions and sets via the epigraph operation. We conclude the chapter by comparing partial smoothness to two previously known notions of smooth substructure: $C^p$-identifiable surfaces and fast tracks.

2.1 Definition of Partial Smoothness

Next we introduce the definition of partial smoothness. Although Lewis's work focuses on $C^2$-partial smoothness [Lew02], this thesis examines partial smoothness in a $C^p$ setting. As Section 3.1 shows, this extension causes little difficulty in its application.

Definition 2.1 (Partly Smooth) A function $f$ is $C^p$-partly smooth at a point $\bar{x}$ relative to a set $M$ containing $\bar{x}$ if $M$ is a $C^p$-manifold about $\bar{x}$ and:

(i) (smoothness) $f$ is $C^p$ along $M$ near $\bar{x}$;

(ii) (regularity) $f$ is regular at all points $x \in M$ near $\bar{x}$, with $\partial f(x) \neq \emptyset$;
(iii) (sharpness) $df(x, n) > -df(x, -n)$ for all $n \in N_M(x) \setminus \{0\}$; and

(iv) (sub-continuity) $\partial f$ restricted to $M$ is continuous at $x$.

A function shall be called $C^p$-partly smooth relative to $M$, if it is $C^p$-partly smooth at every point $x \in M$ relative to $M$. We use the term partly smooth to mean $C^2$-partly smooth.

Further, a set $S$ is $C^p$-partly smooth at $x \in S$ relative to $M$ if its indicator function maintains this property.

In all cases we refer to $M$ as the active manifold.

At first glance the definition of partly smooth functions (and sets) may be difficult to absorb. This prompts us to briefly discuss each portion of the definition in turn. The clearest part of the definition is condition (i) (smoothness). This simply states that, when restricted to the active manifold the function is smooth. Conditions (ii) and (iii) (regularity and sharpness) are somewhat linked. Together they say, in directions pointing orthogonally away from the active manifold the function has an upwards opening "kink". For example, the function $|x|$ has such a kink at the point 0, but $x^2$ does not. The function $-|x|$ also has a kink, but the kink opens downwards, so the function is not regular. The last condition of partial smoothness (sub-continuity) provides a second level of smoothness for the function. It states that, though the function may be nonsmooth off the manifold, it must be nonsmooth in a locally controlled manner. These notions will become clearer in Section 2.2 where examples of functions satisfying all but one condition of partial smoothness are given.

When dealing with partial smoothness of sets it is more convenient to use the following equivalent definition.

**Lemma 2.2 (Partly Smooth Sets)** A set $S$ is $C^p$-partly smooth at a point $x \in S$ relative to a manifold $M$ if and only if $M$ is a $C^p$-manifold about $x$ and the following hold:

(i) $S \cap M$ is a neighbourhood of $x$ in $M$;

(ii) $S$ is regular at all points in $M$ near $x$;

(iii) $N_M(x) \subseteq N_S(x) - N_S(x)$; and

(iv) the normal cone map $N_S(\cdot)$ is continuous at $x$ relative to $M$.

**Proof:** The proof of [Lew02, Prop 2.11] does not change at all for the case of $C^p$ instead of $C^2$.

Notice that condition (ii) (regularity) of partial smoothness appears weaker in the case of sets. In the case of functions the regularity condition also demands the function have at least one subgradient,
while the case of sets has no analogous demand. The reason for this is, the normal cone to a set is never empty (it always contains 0).

Lewis's 2002 work introduces several alternate definitions for the sharpness condition. We include them, and add some of our own, here.

**Lemma 2.3 (Sharpness Equivalences)** Let $M$ be a manifold about the point $\bar{z}$.

If the function $f$ is regular at every point $x \in M$, with nonempty subdifferentials, then the following are equivalent to the sharpness condition (Definition 2.1 (iii)):

(i) $\text{lin} df(\bar{z}, \cdot) \subseteq T_M(\bar{z})$,

(ii) $\text{lin} df(\bar{z}, \cdot) = T_M(\bar{z})$,

(iii) $N_M(\bar{z}) \subseteq \text{par} \partial f(\bar{z})$, and

(iv) $N_M(\bar{z}) = \text{par} \partial f(\bar{z})$.

Alternately, if the set $S$ is regular at every point $x \in M$, then the following are equivalent to the sharpness condition:

(v) $T_S(\bar{z}) \cap -T_S(\bar{z}) \subseteq T_M(\bar{z})$,

(vi) $T_S(\bar{z}) \cap -T_S(\bar{z}) = T_M(\bar{z})$, and

(vii) $N_M(\bar{z}) = N_S(\bar{z}) - N_S(\bar{z})$.

**Proof:** Statements (i) through (iv) can be found in [Lew02, Note 2.9]. To see statements (v) and (vi) consider a set $S$ which is regular on the manifold $M$, and the point $\bar{z} \in M$. The indicator function $\delta_S$ has the following subderivative at $\bar{z}$,

$$d\delta_S(\bar{z}, d) = \begin{cases} 0 & d \in T_S(\bar{z}) \\ \infty & d \notin T_S(\bar{z}) \end{cases}.$$ 

Thus the linearity space of $d\delta_S(\bar{z}, \cdot)$ is exactly the set of directions

$$\{d : d \in T_S(\bar{z}), -d \in T_S(\bar{z})\} = T_S(\bar{z}) \cap -T_S(\bar{z}).$$

Applying statements (i) and (ii) yield (v) and (vi) respectively. A similar argument is given in [Lew02, Prop 2.11] connecting (iii) and (iv) to 2.2 (iii) and statement (vii) above.

Like the sharpness condition, the sub-continuity condition can be simplified. Since the function (or set) is regular by condition (ii) of partial smoothness, the subdifferential (or normal cone) is osc [RW98, Prop 6.6 & Prop 8.7]. Thus the sub-continuity condition can be simplified as follows.
Lemma 2.4 (Sub-continuity Equivalence) Let \( M \) be a manifold containing the point \( \bar{z} \).

If the function \( f \) is regular at \( \bar{z} \), then the sub-continuity condition (Definition 2.1 (iv)) can be simplified to: the subdifferential \( \partial f \) restricted to \( M \) is isc at \( \bar{z} \).

Similarly, for a set \( S \) that is regular at \( \bar{z} \), the sub-continuity condition can be simplified to: the normal cone \( N_S \) restricted to \( M \) is isc at \( \bar{z} \).

2.2 Examples

In this section we provide a selection of examples demonstrating some properties of partly smooth functions and sets. These examples serve two purposes. First, they show the abundance of interesting partly smooth functions. Second, they help to clarify the definition and give the reader a feel for what a partly smooth function might look like. To aid in the second purpose we include four examples of functions that are almost, but not quite partly smooth (Examples 2.10, 2.11, 2.12, and 2.13). Each of these examples will have exactly three of the four conditions of partial smoothness fulfilled. The last series of examples (Examples 2.14, 2.15 and 2.16) provide several less obvious functions which are partly smooth. Example 2.16 will also appear later in this thesis to show the necessity of prox-regularity in many of the results of Chapter 3.

Our first example shows how partly smooth functions arise naturally in optimization. Recall that for a regular function \( f \) a strict critical point is a point where \( 0 \in \text{int} \partial f(x) \), while for a smooth function \( f \) and a regular constraint set \( S \), a strict critical point is a point where \(-\nabla f(x) \in \text{int} N_S(x) \) (Definition 1.10). In either case, a point being a strict critical point is a sufficient condition for strict local minimization. Strict critical points have been studied by many authors in optimization ([Roc76] or [AKK91] for example). Next we shows how strict critical points relate to partial smoothness.

Example 2.5 (Strict Critical Points) If \( \bar{x} \) is a strict critical point of the regular function \( f \), then \( f \) is partly smooth at \( \bar{x} \) relative to the manifold \( \{ \bar{x} \} \).

Indeed, as \( \{ \bar{x} \} \) is a singleton, the smoothness and sub-continuity conditions of partial smoothness hold true. The regularity condition is given, while the sharpness conditions can be seen from Lemma 2.3 and the fact \( 0 \in \text{int} \partial f(\bar{x}) \).

Similarly if the point \( \bar{x} \) is contained in the regular set \( S \) and the normal cone \( N_S(\bar{x}) \) has interior, then \( S \) is partly smooth at \( \bar{x} \) relative to the manifold \( \{ \bar{x} \} \). (The proof is equally simple.)

Our next example is due to Lewis [Lew02, Cor 4.8] and shows the abundance of partly smooth functions and sets.
CHAPTER 2. PARTIAL SMOOTHNESS AND RELATED CONCEPTS

Example 2.6 (Finite Max Functions) Select a finite number of $C^p$ functions $f_i$ and define the function $f(x) := \max\{f_i(x) : i \in 1, 2, \ldots, m\}$. For any point $\bar{x}$ define the active set $A := A(\bar{x}) = \{i : f_i(\bar{x}) = f(\bar{x})\}$. If the active gradients $\{\nabla f_i(\bar{x})\}_{i \in A}$ form a linearly independent set, then [Lew02, Cor 4.8] shows $f$ is $C^p$-partly smooth at $\bar{x}$ relative to the manifold

$$M := \{x : f(x) = f_i(x) \text{ for all } i \in A, f(x) \neq f_i(x) \text{ for } i \notin A\}.$$

Similarly, consider the set defined by $S := \{x : f_i(x) \leq 0 \text{ for all } i\}$. In this case the active set at a point $\bar{x}$ is defined $A := A(\bar{x}) = \{i : f_i(\bar{x}) = 0\}$. If the active gradients are linearly independent, the $S$ is $C^p$-partly smooth at $\bar{x}$ relative to the manifold

$$M := \{x : 0 = f_i(x) \text{ for all } i \in A, f(x) \neq f_i(x) \text{ for } i \notin A\}.$$

In the next series of examples we discuss two methods of building partly smooth functions from partly smooth sets. We begin with the distance function, showing the distance to a partly smooth set is partly smooth relative to the same manifold.

Example 2.7 (Distance to Partly Smooth Sets) Suppose the set $S$ is $C^p$-partly smooth at the point $\bar{x}$ relative to the manifold $M$. Then the distance to $S$ is $C^p$-partly smooth at $\bar{x}$ relative to $M$.

Indeed it is easy to confirm the smoothness condition, as the distance function is zero on the set. As $S$ is regular the distance to $S$ is regular with subdifferential $\partial_s \text{dist}(x, S) = N_S(x) \cap B_1(0)$ [RW98, Ex 8.53]. The nonempty subgradient, sharpness, and sub-continuity conditions now follow from the fact $S$ is partly smooth and Lemma 2.3 (iii).

The following corollary is immediate from the fact $C^p$-manifolds are partly smooth at any point on the manifold relative to the manifold.

Corollary 2.8 (Distance to a Manifold) The distance to a $C^p$-manifold is $C^p$-partly smooth relative to the manifold.

Corollary 2.8 shows the interesting fact that, given any manifold there is a real valued function that is partly smooth relative to that manifold. By a slightly different approach we can actually show that every manifold has a convex partly smooth function associated with it.

Example 2.9 (Arbitrary Active Manifold) Let $M$ be a $C^p$-manifold ($p \geq 2$) about the point $\bar{x}$. Then there exists a convex function which is partly smooth at $\bar{x}$ relative to $M$. 

CHAPTER 2. PARTIAL SMOOTHNESS AND RELATED CONCEPTS

To see this, notice there must exist a $C^p$ function $F$ such that $\mathcal{M} := \{x : F(x) = 0\}$. As $p \geq 2$, the function $f(x) := |F(x)| = \max\{|u, F(x)| : |u| \leq 1\}$ is lower-$C^2$. Therefore [RW98, Thm 10.33] shows that for sufficiently large $R$ the function $h_R(x) := f(x) + \frac{R}{2}|x - \bar{x}|^2$ is locally convex.

We claim $h_R$ is partly smooth at $\bar{x}$ relative to $\mathcal{M}$. Indeed, $h_R \equiv 0$ on $\mathcal{M}$ so smoothness holds.

As $h_R$ is (locally) convex, regularity and nonempty subgradients hold. Applying [RW98, Thm 10.31] we see

$$\partial h_R(x) = \text{conv} \left\{ \nabla F(x)^*(u) : |u| \leq 1 \right\}. \quad (2.1)$$

Comparing this to Theorem 1.19 and Lemma 2.3 yields sharpness. While sub-continuity follows directly from equation (2.1).

Lastly, to gain global convexity we note that partial smoothness is a local property, so we can add an indicator function to create global convexity. That is for some convex neighbourhood $U$ of $\bar{x}$ and $R$ sufficiently large, $h_R + \delta_U$ is a convex function which is partly smooth at $\bar{x}$ relative to $\mathcal{M}$.

We now turn our attention to functions that are not partly smooth. The next four examples fulfill two purposes. First, they show that the four conditions of partial smoothness are independent. That is, functions exist satisfying any three of the four conditions but not the fourth. Secondly, the examples should help the reader understand how the four conditions of partial smoothness interact. We approach the conditions of partial smoothness in order, beginning with the smoothness condition.

**Example 2.10 (Lacks Smoothness)** Let $f : \mathbb{R}^n \to \mathbb{R}$ be any $C^1$ function that is not $C^2$. Define $\mathcal{M}$ to be the manifold $\mathbb{R}^n$. Then $f$ satisfies the regularity, sharpness, and sub-continuity conditions of $C^2$-partial smoothness, but lacks the requisite degree of smoothness.

**Example 2.11 (Lacks Regularity)** Consider the function $-|x|$, the point $0$ and the manifold $\{0\}$. Since $\mathcal{M}$ is a singleton, the smoothness and sub-continuity conditions hold. Furthermore $df(0, n) = 1$ for any $n \neq 0$, so the sharpness condition holds. However the function is not regular at $0$.

**Example 2.12 (Lacks Sharpness)** Take $f$ to be any $C^1$ function. Select the point $\bar{x} = 0$ and the manifold $\mathcal{M} = \{0\}$. Then for any direction $w$ the subderivative satisfies $-df(0, w) = -(\nabla f(0), w) = (\nabla f(0), -w) = df(0, -w)$. Since $N_{\mathcal{M}}(\bar{x}) = \mathbb{R}^n$, sharpness fails. As $\mathcal{M}$ is a singleton, smoothness and sub-continuity relative to $\mathcal{M}$ hold trivially. The regularity condition holds as $f \in C^1$. 
Example 2.13 (Lacks Sub-Continuity) Consider the convex function

\[
f(x, y) := \max \begin{cases} 
3x^2 - 2y \\
3x^2 \\
2y + x^2
\end{cases}
\]

at the point \((0, 0)\), and the manifold \(M := \{(x, 0) : x \in \mathbb{R}\} \).

![Figure 2.1: A Function Lacking Sub-continuity: max\{3x^2 - 2y, 3x^2, 2y + x^2\}](image)

Clearly \(f\) is smooth on \(M\). Since \(f\) is convex, it is regular. The sharpness condition holds as the subdifferential at \((0, 0)\) is \(\partial f(0, 0) = \text{conv} \{(0, -2), (0, 2)\} = N_M(0, 0)\). However the sub-continuity condition fails as \(\partial f(x, 0) = \text{conv} \{(6x, -2), (6x, 0)\} \) for \(x \neq 0\).

Example 2.13 in many ways appears partly smooth. In Example 2.6 we see that finite max functions are partly smooth, and Example 2.13 is a finite max function. However, to ensure that a finite max is partly smooth we also need the active gradients (see Example 2.6) to be linearly independent. Example 2.13 fails this condition, and therefore also shows that the requirements of Example 2.6 cannot be weakened.

Reexamining the examples beginning this section one might notice that most of them are convex, or at least lower-C^2. Since all lower-C^2 functions are simply the difference of a convex function and a quadratic [RW98, Thm 10.33], this might lead one to asking if partial smoothness is related to convexity. The answer is no, as the next example shows.

Example 2.14 (Non-convex Partial Smoothness) Consider the function \(f(x) = \sqrt{|x|}\). It is easy to see \(f\) is partly smooth at 0 relative to the manifold \(\{0\}\), and cannot be transformed to a convex function by the addition of a smooth function.
CHAPTER 2. PARTIAL SMOOTHNESS AND RELATED CONCEPTS

The above example is not only nonconvex, it is not locally Lipschitz at 0. However even if we add the condition that the function be locally Lipschitz everywhere, a partly smooth function need not be lower-$C^2$. In the next example we produce a locally Lipschitz partly smooth function that is not prox-regular (all lower-$C^2$ functions are prox-regular [RW98, Prop 13.33]).

Example 2.15 (Non-prox-regular Partial Smoothness) Define the function

$$f(x) := \begin{cases} x & x \geq 0 \\ -|x|^\frac{3}{2} & x < 0 \end{cases}$$

and the manifold $\mathcal{M} := \{0\}$. Thus $f$ is lower-$C^1$, locally Lipschitz and partly smooth at 0 relative to $\mathcal{M}$, but $f$ is not prox-regular, and therefore not lower-$C^2$ [RW98, Prop 13.33].

To see that $f$ is partly smooth we examine the definition. As $\mathcal{M}$ is a singleton, $f$ is smooth on $\mathcal{M}$ and trivially satisfies the sub-continuity condition. Moreover, $f$ is regular with subdifferential

$$\partial f(x) = \begin{cases} \{1\} & x > 0 \\ [0,1] & x = 0 \\ \{\frac{3}{2}|x|^\frac{3}{2}\} & x < 0. \end{cases}$$

Since $df(0,-1) = 0$ and $df(0,1) = 1$, the sharpness condition is satisfied. Thus $f$ is partly smooth at 0 relative to $\mathcal{M}$.

To see $f$ is not prox-regular at 0, one can either apply a direct check, or notice the subgradient is not premonotone and apply [RW98, Thm 13.36].

We end this section with an example due to Lewis [Lew02, Sec 7]. The example was created to show the difficulty in critical point analysis for partly smooth functions. (Section 3.5 of this thesis addresses this problem in light of prox-regularity.) For now we use Lewis's example to show a function that is partly smooth relative to two distinct manifolds. The function also provides a second example of a partly smooth function which is not prox-regular. Later it will provide an example showing the necessity of prox-regularity in the results of Sections 3.3, 3.4 and 3.5 of this thesis.

Example 2.16 (Non-unique Active Manifold) Define the function

$$f(x,y) := \begin{cases} x^2 - y & (y \leq 0) \\ \sqrt{x^4 + 2x^2y - y^2} & (0 < y < 2x^2) \\ 3x^2 - y & (2x^2 \leq y \leq 4x^2) \\ y - 5x^2 & (4x^2 < y). \end{cases}$$

In [Lew02, Sec 7] it is shown that $f$ is regular, locally Lipschitz, continuous everywhere, and $C^2$ except on the manifolds

$$\mathcal{M}_1 := \{(x,y) : y = 0\}$$


\[ M_2 := \{(x, y) : y = 4x^2\}. \]

Furthermore Lewis shows \( f \) is partly smooth along either of these manifolds.

In Corollary 3.14 we will see that this means \( f \) cannot be prox-regular. For completeness we provide a direct proof below.

Since \( f \) is continuous, [RW98, Thm 13.36] shows that \( f \) is prox-regular at \((0, 0)\) if and only if \( \partial f + RI \) is locally monotone at \((0, 0)\) for some \( R > 0 \). Now consider points \((x, 0)\) and \((x, 4x^2)\) for \( x \) near 0. Referring again to [Lew02, Sec 7] we have

\[
(2x, 1) \in \partial f(x, 0) \quad \text{and} \\
(6x, -1) \in \partial f(x, 4x^2).
\]

For \( \partial f + RI \) to be monotone we would require

\[
\langle (2x - 6x, 2) + R(0, -4x^2), (0, -4x^2) \rangle \geq 0
\]

which reduces to \((2 - 4Rx^2)(-4x^2) \geq 0\). This is false when \( x \neq 0 \) is sufficiently small.

\[ \square \]

### 2.3 Epigraph comparisons

The goal of this section is to prove the expected relation between a partly smooth function and its epigraph. Specifically, a function is partly smooth if and only if its epigraph is a partly smooth. Theorem 2.18 states this in more detail, including how the active manifolds relate.
Besides the theorem’s intrinsic interest, it is extremely useful in applying results about partly smooth sets to partly smooth functions. The basic technique is to examine the epigraph of a partly smooth function; Theorem 2.18 shows the epigraph is partly smooth, therefore the result regarding sets holds. This technique is most prominently displayed in the proof of Theorem 3.21 in this thesis.

In Section 3.1 we will develop an alternate proof to Theorem 2.18 via calculus rules. This alternate approach can also be found in [HL03]. In this section however, we approach the theorem directly.

Before we continue we develop one simple lemma.

**Lemma 2.17** Let the function $f$ be regular at the point $\bar{x}$ with $\partial f(\bar{x}) \neq \emptyset$. Then $-df(\bar{x}, -t) \leq df(\bar{x}, t)$ for any direction $t$.

**Proof:** This is a consequence of [RW98, Thm 8.30].

The remainder of this section is devoted to Theorem 2.18 and its proof.

**Theorem 2.18 (Partial Smoothness and Epigraphs)** Given a function $f$, a point $\bar{x}$, and a manifold $M$, the following are equivalent:

1. $f$ is $C^p$-partly smooth at $\bar{x}$ relative to $M$, and
2. $f$ is $C^p$ along $M$ and $\text{epi } f$ is $C^p$-partly smooth at the point $(\bar{x}, f(\bar{x}))$ relative to the manifold $\widehat{M} := \{(x, f(x)) : x \in M\}$.

**Proof:** The proof proceeds by showing the equivalence of each of the four requirements for partly smooth functions and partly smooth sets.

Throughout the proof we suppose $f$ is a function that is finite at the point $\bar{x}$, $M$ is a manifold, and $\widehat{M} := \{(x, f(x)) : x \in M\}$.

**Smoothness** If $f$ is partly smooth at $\bar{x}$ relative to $M$, then clearly $f$ is smooth on $M$, while conversely if $f$ is smooth on $M$ then $f$ satisfies the smoothness condition (2.1 (i)).

**Regularity** By definition $f$ is regular if and only if $\text{epi } f$ is regular.

**Non-empty subdifferential** We only need to show this in the direction (ii) $\Rightarrow$ (i). By Lemma 2.3 (vii), if $\text{epi } f$ is $C^p$-partly smooth at $(\bar{x}, f(\bar{x}))$ relative to $\widehat{M}$, then $N_{\widehat{M}}(\bar{x}, f(\bar{x})) = N_{\text{epi } f}(\bar{x}, f(\bar{x})) - N_{\text{epi } f}(\bar{x}, f(\bar{x}))$. Theorem 1.20 therefore shows that

$$\{(n + \lambda \nabla g(\bar{x})^*, -\lambda) : n \in N_{\widehat{M}}(\bar{x}), \lambda \in \mathbb{R}\} = N_{\text{epi } f}(\bar{x}, f(\bar{x})) - N_{\text{epi } f}(\bar{x}, f(\bar{x})),$$

(2.2)

(where $g$ is any smooth function agreeing with $f$ on $M$). We claim that the set $\{v : (v, -1) \in N_{\text{epi } f}(\bar{x}, f(\bar{x}))\}$ is nonempty, which would imply $\partial f(\bar{x}) \neq \emptyset$ [RW98, Thm 8.9].
Indeed suppose \( \{v : (v, -1) \in N_{\text{epi}} f(\bar{x}, f(\bar{x})) \} \) is empty. Since \( N_{\text{epi}} f(\bar{x}, f(\bar{x})) \) is a cone this implies that \( w \geq 0 \) whenever \( (v, w) \in N_{\text{epi}} f(\bar{x}, f(\bar{x})) \). As we are taking the normal to a closed epigraph, \( (v, w) \in N_{\text{epi}} f(\bar{x}, f(\bar{x})) \) implies \( w \leq 0 \) [RW98, Thm 8.9]. Thus, if \( \{v : (v, -1) \in N_{\text{epi}} f(\bar{x}, f(\bar{x})) \} \) is empty then \( (v, w) \in N_{\text{epi}} f(\bar{x}, f(\bar{x})) \) implies \( w = 0 \). This contradicts equation (2.2) with \( \lambda = 1 \).

(Sharpness) To begin, note that [RW98, Thm 8.2] shows that \( T_{\text{epi}} f(\bar{x}, f(\bar{x})) = \text{epi} df(\bar{x}, \cdot) \), so

\[
T_{\mathcal{G}}(\bar{x}, f(\bar{x})) = -T_{\text{epi}} f(\bar{x}, f(\bar{x})) \cap T_{\text{epi}} f(\bar{x}, f(\bar{x}))
\]

\[
\iff T_{\mathcal{G}}(\bar{x}, f(\bar{x})) = -\text{epi} df(\bar{x}, \cdot) \cap \text{epi} df(\bar{x}, \cdot)
\]

\[
= \{ (-s, \alpha) : df(\bar{x}, s) \leq \alpha \} \cap \{ (t, \beta) : df(\bar{x}, t) \leq \beta \}
\]

\[
= \{ (t, \beta) : df(\bar{x}, -t) \leq -\beta, df(\bar{x}, t) \leq \beta \}
\]

\[
= \{ (t, \beta) : df(\bar{x}, t) \leq -\beta \leq -df(\bar{x}, -t) \}.
\]

In either direction Lemma 2.17 shows

\[
T_{\mathcal{G}}(\bar{x}, f(\bar{x})) = -T_{\text{epi}} f(\bar{x}, f(\bar{x})) \cap T_{\text{epi}} f(\bar{x}, f(\bar{x}))
\]

\[
\iff T_{\mathcal{G}}(\bar{x}, f(\bar{x})) = \{ (t, \beta) : df(\bar{x}, t) = \beta \leq -df(\bar{x}, -t) \}
\]

\[
= \{ (t, df(\bar{x}, t)) : t \in \text{lin} df(\bar{x}, \cdot) \}.
\]

Lastly, if \( g \) is any smooth function agreeing with \( f \) on \( \mathcal{M} \), Theorem 1.20 shows that

\[
T_{\mathcal{G}}(\bar{x}, f(\bar{x})) = \{ (t, \nabla g(\bar{x})(t)) : t \in T_{\mathcal{M}}(\bar{x}) \}.
\]

Thus \( T_{\mathcal{G}}(\bar{x}, f(\bar{x})) = -T_{\text{epi}} f(\bar{x}, f(\bar{x})) \cap T_{\text{epi}} f(\bar{x}, f(\bar{x})) \) if and only if

\[
T_{\mathcal{M}}(\bar{x}) = \text{lin} df(\bar{x}, \cdot).
\] (3.3)

Examining Lemma 2.3 we see this shows the two sharpness conditions are equivalent.

(Sub-continuity) By [RW98, Thm 8.9] we have the following relations between \( \partial f \) and \( N_{\text{epi}} f \):

\[
\partial f(\bar{x}) = \{ v : (v, -1) \in N_{\text{epi}} f(\bar{x}, f(\bar{x})) \},
\] (2.4)

\[
N_{\text{epi}} f(\bar{x}, f(\bar{x})) = \{ \lambda (v, -1) : v \in \partial f(\bar{x}), \lambda > 0 \} \cup \{ (v, 0) : v \in \partial_{\infty} f(\bar{x}) \}.
\] (2.5)

Using this we must show \( \partial f \) is continuous when restricted to \( \mathcal{M} \) if and only if \( N_{\text{epi}} f \) is continuous when restricted to \( \widehat{\mathcal{M}} \).

In either case, by Lemma 2.4 we only have to show inner semi-continuity.

Suppose \( \text{epi} f \) is partly smooth at \( (\bar{x}, f(\bar{x})) \) relative to \( \widehat{\mathcal{M}} \). Select any subgradient \( \bar{w} \in \partial f(\bar{x}) \) and sequence of points \( x_k \rightarrow \bar{x} \) in \( \mathcal{M} \). We must show there exists subgradients \( w_k \in \partial f(x_k) \) such that \( w_k \rightarrow \bar{w} \). By equation (2.4), \( (\bar{w}, -1) \in N_{\text{epi}} f(\bar{x}, f(\bar{x})) \). As \( N_{\text{epi}} f \) is continuous at \( (\bar{x}, f(\bar{x})) \), and \( f \) is continuous relative to \( \mathcal{M} \) there exists \( (w_n, v_n) \in N_{\text{epi}} f(x_n, f(x_n)) \) with \( (w_n, v_n) \rightarrow (\bar{w}, -1) \). Since \( N_{\text{epi}} f \) is a cone we can replace \( (w_n, v_n) \) by

\[
(w_n/|v_n|, -1) \in N_{\text{epi}} f(x_n, f(x_n)).
\]
As $|v_n| \to 1$, we have $w_n/|v_n| \to \bar{w}$. Lastly notice that $w_n/|v_n| \in \partial f(x_n)$ by equation (2.4), which proves the isc property.

Conversely, suppose $f$ is partly smooth at $\bar{x}$ relative to $M$. Using equation (2.5) we seek to show that for any sequence $(x_k, f(x_k)) \to (\bar{x}, f(\bar{x}))$ in $\tilde{M}$ and any normal vector $\bar{n} := (\bar{w}, -\lambda) \in N_{\text{epi}}(\bar{x}, f(\bar{x}))$, there exists normal vectors $n_k \in N_{\text{epi}}(x_k, f(x_k))$ such that $n_k \to \bar{n}$. We separate this into two cases: $\lambda \neq 0$ and $\lambda = 0$.

If $\lambda \neq 0$, then equation (2.5) shows $\frac{\lambda}{\bar{w}} \in \partial f(\bar{x})$. As $\partial f$ is continuous when restricted to $M$, there exists subgradients $w_k \in \partial f(x_k)$ such that $w_k \to \frac{\lambda}{\bar{w}}$. Therefore, $\lambda (w_k, -1) \to (\bar{w}, -\lambda)$. Equation (2.5) shows $\lambda (w_k, -1) \in N_{\text{epi}}(x_k, f(x_k))$, so we have the desired sequence.

If $\lambda = 0$, then equation (2.5) combined with [RW98, Cor 8.11] shows $\bar{w} \in \partial f(\bar{x}) = [\partial f(\bar{x})]^\infty$. Since $f$ is regular on $M$ near $\bar{x}$, $\partial f$ is convex on $M$, so

$$[\partial f(\bar{x})]^\infty = \{ w : \text{ for all } x_k \to \bar{x}, \text{ there exists } w_k \in \partial f(x_k), \lambda_k \downarrow 0 \text{ with } f(x_k) \to f(\bar{x}), \lambda_k w_k \to w \}$$

[RW98, Thm 4.25 (a) & 4.24]. Thus there exists subgradients $w_k \in \partial f(x_k)$ and scalars $\lambda_k \downarrow 0$ such that $\lambda_k w_k \to \bar{w}$. By equation (2.5), $(\lambda_k w_k, -\lambda_k) \in N_{\text{epi}}(x_k, f(x_k))$ for all $k$. As $(\lambda_k w_k, -\lambda_k) \to (\bar{w}, 0)$ we have the desired sequence.

$\blacksquare$

### 2.4 $C^p$-Identifiability

One of the goals of this thesis is to show partial smoothness is a natural notion of smooth substructure in nonsmooth optimization. So far we have approached this task by showing how strict critical points relate to partial smoothness (Example 2.5) and by displaying an abundance of partly smooth functions (and therefore sets). In the next two sections we approach this goal by comparing partial smoothness to two other notions of smooth substructure: $C^p$-identifiable surfaces and fast tracks.

The study of $C^p$-identifiable surfaces originated largely in research by Burke and Moré. In [BM88] Burke and Moré discuss algorithms which could identify the active constraints of a constrained optimization problem in a finite number of iterations. Their study focuses on open facets, a generalization of polyhedral faces in which the face need only be "flat" not polyhedral (see [BM88, Def 2.5]).

In 1993 Wright extended their research by studying convex sets which, although not defined by smooth equations, had subsets (or surfaces) defined via smooth equations [Wri93]. His work began by showing that the projection mapping onto a convex set is $C^{p-1}$ whenever the surface on which the solution lies is $C^p$-identifiable [Wri93, Thm 2.6]. Though this was interesting in its own right, and extended many results on the smoothness of projection mapping, it was not his main goal. Using this result Wright showed that certain algorithms will naturally identify the active constraints of an optimization problem in a finite number of iterations [Wri93, Thm 3.1].
The definition of $C^p$-identifiable follows.

**Definition 2.19 ($C^p$-identifiable)** Let $C$ be a closed convex set with interior. A connected set $M \subseteq C$ is $C^p$-identifiable if for any point $y \in \mathbb{R}^n \setminus C$ such that the projection $P_C(y)$ is in $M$ and $y - P_C(y) \in \text{rint} \ N_C(P_C(y))$, there exist functions $g_i \in C^p \ (i = 1, 2, \ldots, m)$, such that:

(i) $\{\nabla g_i(P_C(y))\}_{i=1}^m$ is linearly independent;

(ii) $\text{conv} \ {\nabla g_i(z)}_{i=1}^m \subseteq N_C(z)$ for all $z \in M$ near $P_C(y)$;

(iii) $y - P_C(y) \in \text{rint} [\text{conv} \ {\nabla g_i(P_C(y))}_{i=1}^m]$; and

(iv) $M = \{x : g_i(x) = 0 \text{ for } i = 1, 2, \ldots, m\}$ near $P_C(y)$.

In the original definition Wright stated, either $M$ is an open subset of $\text{int} C$ or $M$ satisfies conditions (i) through (iv). The interior case is vacuously included in the above definition.

To see the abundance of $C^p$-identifiable surfaces, and the types of sets they include we consider the following example due to Wright [Wri93, Thm 2.4].

**Example 2.20 (Abundance of $C^p$-identifiable surfaces)** Let $\{g_i\}_{i=1}^m$ be a collection of $C^p$ convex functions. Define the convex set

$$C := \{x : g_i(x) \leq 0, i = 1, 2, \ldots, m\},$$

and select a point $\bar{x} \in C$. Suppose the active set $A(\bar{x}) := \{i : g_i(\bar{x}) = 0\}$ is such that $\{\nabla g_i(\bar{x})\}_{i \in A(\bar{x})}$ is a linearly independent set. Then [Wri93, Thm 2.4] shows the set

$$M := \{x : g_i(x) = 0 \text{ for all } i \in A(\bar{x}), g_i(x) < 0 \text{ for all } i \notin A(\bar{x})\}$$

is a $C^p$-identifiable surface of $C$.

Comparing Examples 2.6 and 2.20 we see remarkable similarities. In fact, except for the word 'convex', the examples are identical. This suggests that there is a strong similarity between partly smooth sets and $C^p$-identifiable surfaces. Before discussing this similarity it is important to note one major difference. Partly smooth sets are defined relative to a point and a manifold, while $C^p$-identifiable surfaces are defined only in terms of a manifold. To connect the two, recall that a set is partly smooth relative to a manifold if and only if it is partly smooth at each point on the manifold.

The relation between $C^p$-identifiable surfaces and partly smooth sets was examined first in [Lew02]. Here Lewis shows $C^p$-identifiability and convex $C^p$-partial smoothness are the same for the case of $p = 2$ [Lew02, Thm 6.3]. The case of arbitrary $p$ is also true, and the proof shows no significant change. We include the statement of the result below, but omit the details of the proof.
Theorem 2.21 (Cp-identifiable sets and Cp-partial smoothness) Let $C$ be a closed convex set, and $M$ be a manifold contained in $C$. Then $C$ is Cp-partly smooth relative to $M$ if and only if $M$ is a Cp-identifiable surface of $C$.

Proof: The proof is identical to that of [Lew02, Thm 6.3] with the exception of '2' being replaced by 'p' throughout.

2.5 UV-decompositions and Fast Tracks

Arising from research on second order expansions of non-differentiable functions, fast tracks form another notion of smooth substructure. This section compares fast tracks and convex partly smooth functions. Before moving on to fast tracks we must first discuss UV-decompositions and the U-Lagrangian.

The idea behind a UV-decomposition was first given in [LOSO0] by Lemaréchal, Oustry, and Sagastizábal. By examining the directional derivative of a convex function one can attempt to predict in which directions the function will behave smoothly and in which directions it will not. Denoting the smooth subspace $U$ and the sharp subspace $V$ follows from the descriptive images of the letters $U$ and $V$ themselves. The exact definition follows.

Definition 2.22 (UV-Decomposition) Let $f$ be a convex function which is finite at the point $\bar{x}$. For a subgradient $\bar{w} \in \text{rint}\partial f(\bar{x})$, we define the UV-decomposition as the subspaces

$$U := N_{\partial f(\bar{x})}(\bar{w}) \quad \text{and} \quad V = R_+ (\partial f(\bar{x}) - \bar{w}).$$

One should immediately note that $U$ and $V$ are perpendicular subspaces. Therefore, for any point $x \in \mathbb{R}^n$ we may express $x$ via its projections onto $U$ and $V$; $x = P_U(x) + P_V(x) := x_u + x_v$. An extremely useful lemma on alternate definitions of UV-decompositions follows.

Lemma 2.23 [LOSO0, Prop 2.2] Let $f$ be a convex function, which is finite at the point $\bar{x}$. Then the UV-decomposition of $f$ is independent of the subgradient $\bar{w} \in \text{rint}\partial f(\bar{x})$ chosen. Moreover the following subspaces are equal to $U$:

(i) $\{d \in \mathbb{R}^n : \sup_{w \in \partial f(\bar{x})} \langle w, d \rangle = \inf_{w \in \partial f(\bar{x})} \langle w, d \rangle\}$,

(ii) $\text{lin}\partial f(\bar{x}, \cdot)$, and

(iii) $\{d \in \mathbb{R}^n : \langle w, d \rangle = \langle \bar{w}, d \rangle \quad \text{for all} \quad w \in \partial f(\bar{x})\}$. 


It is worth comparing Lemma 2.23 with Lemma 2.3. Together these state that a convex function is sharp relative to a manifold if and only if the tangent space to the manifold is the smooth space, $U$ (convex functions are regular everywhere). This suggests that any work using $UV$-decompositions is somehow related to partial smoothness.

The original purpose of $UV$-decompositions was to create second order expansions for non-differentiable functions [LOS00]. By separating the domain into the smooth and sharp subspaces Lemaréchal, Oustry and Sagastizábal could focus on the smooth substructure of a function. To study the smooth substructure better, the authors needed a way to smooth the function along $U$. This was accomplished via the $U$-Lagrangian. Behaving much like a normal Lagrangian, the $U$-Lagrangian is $C^1$ and provides useful derivative information for the original function. Moreover, if the $U$-Lagrangian is $C^2$ then the Hessian information creates a second order expansion of the original function [LOS00, Thm 3.9].

Though we will not actually use the $U$-Lagrangian in this work, its solution mapping is pivotal in defining fast tracks.

**Definition 2.24 (U-Lagrangian)** Let $f$ be a convex function which is finite at the point $\bar{x}$, with subgradient $\bar{\omega} \in \text{rint } \partial f(\bar{x})$. For $u \in U$, the $U$-Lagrangian is defined by

$$L_U(u, \bar{\omega}) := \min_{v \in V} \{ f(\bar{x} + (u + v)) - \langle \bar{\omega}, v \rangle \},$$

where $U$ and $V$ are the $UV$-decomposition subspaces. Its related solution mapping is defined via

$$W_U(u, \bar{\omega}) := \arg \min_{v \in V} \{ f(\bar{x} + (u + v)) - \langle \bar{\omega}, v \rangle \}.$$  

In 2002, Mifflin joined Sagastizábal in researching how the $U$-Lagrangian could be used for algorithmic purposes [MS02a] [MS02b]. This research found many functions had “$C^2$ trajectories” for the solution mapping. That is, locally there exists a $C^2$ function which provides a solution mapping for the $U$-Lagrangian. Mifflin and Sagastizábal called these trajectories fast tracks, and formalized them as follows.

**Definition 2.25 (Fast Tracks)** Let $f$ be a convex function, minimized at the point $\bar{x}$. We call $v(\cdot)$ a fast track if, for all subgradients $\bar{\omega} \in \text{rint } \partial f(\bar{x})$

(i) $v : U \rightarrow V$ is a $C^2$ selection of $W_U(u, \bar{\omega})$, (i.e. $v(u) \in W_U(u, \bar{\omega})$ for all $u \in U$) and

(ii) $L_U(u, \bar{\omega})$ is $C^2$ with respect to $u$.

A simple example helps show what these definitions are capturing.

**Example 2.26** Consider the convex function $f(x, y) := |x - y|^2 + y^2$. Clearly $f$ is minimized at $(0, 0)$, and the subdifferential takes the form $\partial f(0, 0) = \text{conv } \{(1,0), (-1,0)\}$. 


It is not hard to see that the $UV$-decomposition at $\bar{x} = (0,0)$ yields

$$U = \{(0,y) : y \in \mathbb{R}\}, \text{ and } V = \{(x,0) : x \in \mathbb{R}\}.$$ 

Selecting any $\bar{w} \in \text{rint} \partial f(0,0)$, the $U$-Lagrangian for $f$ is

$$L_U((0,y), \bar{w}) = \min_{z \in \mathbb{R}} \{|x - y^2| + y^2 - \langle \bar{w}, (x,0) \rangle\}.$$ 

Notice $\bar{w} = (\alpha, 0)$ for some $\alpha \in (-1,1)$. Therefore,

$$L_U((0,y), \bar{w}) = \min_{x \in \mathbb{R}} \{\max\{ (1 - \alpha)x, 2y^2 - (1 + \alpha)x \} \}.$$ 

Since, $(1 - \alpha) > 0$ and $-(1 + \alpha) < 0$, we know $(1 - \alpha)x$ and $2y^2 - (1 + \alpha)x$ are linear functions in $x$ with slopes of opposite signs. The minimax problem is therefore solved when

$$(1 - \alpha)x = 2y^2 - (1 + \alpha)x.$$ 

This simplifies to $x = y^2$. Therefore, $W_U((0,y), \bar{w}) = \{(y^2,0)\}$ regardless of $\bar{w}$, so $v(0,y) = (y^2,0)$ is a fast track for $f$. 

It is easy to confirm that the above example is partly smooth at $\bar{x} := (0,0)$ relative to the manifold $\mathcal{M} := \{(y^2,y) : y \in \mathbb{R}\}$. This manifold also results from the fast track found in Example 2.26 as $\mathcal{M} = \{(0,0) + (0,y) + v(0,y) : y \in \mathbb{R}\}$, or in $UV$ notation, $\mathcal{M} = \{(\bar{x} + u + v(u)) : u \in U\}$. This result is not a coincidence. Theorem 2.27 focuses on making precise the relationship of fast tracks to partial smoothness. The theorem, and its consequences, also appears in [Har03]. (Recall we use the term partly smooth to refer to $C^2$-partly smooth.)

**Theorem 2.27 (Fast Tracks and Partial Smoothness)** Let $f$ be a convex function minimized at the point $\bar{x}$. Then
CHAPTER 2. PARTIAL SMOOTHNESS AND RELATED CONCEPTS

(i) if $f$ is partly smooth at $\bar{x}$ relative to the manifold $\mathcal{M}$, then $\mathcal{M}$ defines a fast track, and

(ii) if $v(u)$ is a fast track for $f$ at $\bar{x}$, then $\mathcal{M} := \{\bar{x} + (u + v(u)) : u \in U\}$ is a manifold and $f$ is partly smooth at $\bar{x}$ relative to $\mathcal{M}$.

Theorem 2.27 involves the condition that the function examined must be minimized at the point of interest. This condition arises from the definition of fast tracks, and not any mathematical restrictions. By tilting a convex function by a linear function it is always possible to force a zero subgradient, and therefore force the point of interest to be a minimizer for the function. Using this idea, Theorem 2.27 could be written without the minimizer condition.

The first half of the proof to Theorem 2.27 requires the next result of Lewis.

**Theorem 2.28** [Lew02, Thm 6.1] If the function $f$ is partly smooth at the point $\bar{x}$ relative to the manifold $\mathcal{M}$, then there exists a $C^2$ function, $v : T_M(\bar{x}) \to N_M(\bar{x})$, such that locally:

(i) $v = O(|u|^2)$ (as $u \to 0$),

(ii) $\bar{x} + (u + w) \in \mathcal{M}$ if and only if $w = v(u)$,

(iii) for small $u$ and any subgradient $\bar{w} \in \text{rint} \partial f(\bar{x})$ the function

$$h(v) : V \to \mathbb{R}$$

$$w \mapsto f(\bar{x} + (u + w)) - \langle \bar{w}, (u + w) \rangle$$

is a minimized at $w = v(u)$.

Lewis's result yields the proof of part (i) of Theorem 2.28.

**Proof of 2.27** (i): By applying Theorem 2.28 to a partly smooth function, we find the required function, $v(u)$, for a fast track.

First note that the $UV$-decomposition yields $U = T_M(\bar{x})$ and $V = N_M(\bar{x})$. This can be seen as $T_M(\bar{x}) = \text{lin} \{df(\bar{x}, \cdot)\} = U$ by partial smoothness, Lemma 2.3 and Lemma 2.23.

Applying Theorem 2.28 we find $v(u)$ is $C^2$ by construction. Furthermore, part (iii) shows, for any subgradient $\bar{w} \in \text{rint} \partial f(\bar{x})$, that $v(u)$ is a selection of $W_U(u, \bar{w})$. Lastly, as $\bar{x} + (u + v(u)) \in \mathcal{M}$ and $v(u)$ is $C^2$, the $U$-Lagrangian is the sum of two $C^2$ functions:

$$L_U(u, \bar{w}) = f(\bar{x} + (u + v(u))) + \langle v(u), \bar{w} \rangle.$$

Therefore $L_U \in C^2$ as required.

To make the remainder of the proof of Theorem 2.28 more readable, we separate out two lemmas.
Lemma 2.29 (Fast Tracks and Sharpness) Let $f$ be a convex function, minimized at the point $\bar{x}$. If $f$ has a fast track $v(u)$ then

$$
\mathcal{M} := \{\bar{x} + (u + v(u)) : u \in U\}
$$

is a manifold with

$$T_{\mathcal{M}}(\bar{x}) = U = \text{lin} \, df(\bar{x}, \cdot).
$$

(Thus $f$ satisfies the sharpness condition of partial smoothness relative to $\mathcal{M}$ at $\bar{x}$.)

Proof: Define

$$G : U \to \mathbb{R}^n,
\quad u \mapsto \bar{x} + (u + v(u)).$$

As $G$ is $C^2$, $\mathcal{M} = \{G(u) : u \in U\}$, is a manifold about $\bar{x}$ provided $\nabla G(0)$ is one-to-one. Note,

$$\nabla G(0)(u) = u + \nabla v(0)(u).$$

As $\nabla v(0) = 0$ by [LOS00, Cor 3.5], we have $\nabla G(0)(u) = u$. Thus $\mathcal{M}$ is a manifold.

Since $T_{\mathcal{M}}(\bar{x}) = \text{rng} \, \nabla G(0)$ (Theorem 1.19) and $\nabla G(0)(u) = u$, we have $T_{\mathcal{M}}(\bar{x}) = U$. The second equality, $U = \text{lin} \, df(\bar{x}, \cdot)$, follows from Lemma 2.23, while the sharpness statement results from Lemma 2.3.

The above lemma shows the sharpness condition of partial smoothness holds. To prove the subcontinuity condition we use the next lemma, due essentially to Lemaréchal, Oustry, and Sagastizábal.

Lemma 2.30 Let $f$ be a convex function, minimized at the point $\bar{x}$. Suppose $v(u)$ is a fast track for $f$. Then for any subgradient $\bar{w} \in \text{rint} \, \partial f(\bar{x})$, and point $u \in U$ sufficiently small,

$$(\nabla_U L_U(u, \bar{w}), 0) + \bar{w} \in \partial f(\bar{x} + (u + v(u))).$$

Proof: As the $U$-Lagrangian is $C^2$ and $v(u)$ is a selection of $W_U$, [LOS00, Thm 3.3 (i)] rewrites to the above.

We now complete the proof of Theorem 2.27.

Proof of 2.27 (ii): By Lemma 2.29 we know if the function $f$ has a fast track $v(u)$ at the point $\bar{x}$, then $f$ is sharp relative to the manifold $\mathcal{M} := \{\bar{x} + (u + v(u)) : u \in U\}$ at $\bar{x}$. Furthermore, on this manifold, $f$ is the difference of two $C^2$ functions,

$$f(\bar{x} + (u + v(u))) = L_U(u, \bar{w}) - \langle v(u), \bar{w} \rangle.$$
Thus $f$ is $C^2$ on $M$. As convexity yields the regularity condition, we have only left to show that $\partial f$ is continuous along $M$ at $\bar{x}$.

By Lemma 2.4 it suffices to show that, for any sequence $x_k \in \mathcal{M} = \{\bar{x} + (u + v(u)) : u \in U\}$ converging to $\bar{x}$ and any subgradient $\bar{w} \in \text{rint} \partial f(\bar{x})$, there exists a sequence of subgradients $w_k \in \partial f(x_k)$ with $w_k \to \bar{w}$. Let $x_k = \bar{x} + u_k + v(u_k)$ and note that $x_k \to \bar{x}$ implies $u_k \to 0$ and $v(u_k) \to 0$. By Lemma 2.30, selecting $w_k = (\nabla L_U(u_k, \bar{w}), 0) + \bar{w}_v$ gives us $w_k \in \partial f(\bar{x} + (u_k + v(u_k))) = \partial f(x_k)$. As $L_U \in C^2$ with $(\nabla L_U(0, \bar{w}), 0) = \bar{w}_u$ [LOS00, Thm 3.3], we have $w_k \to \bar{w}_u + \bar{w}_v = \bar{w}$ as required.

With the comparison between partial smoothness and fast tracks complete, it becomes clear that $C^p$-identifiable surfaces and fast tracks are the same concept applied respectively to sets and functions. Furthermore partial smoothness is a nonconvex extension of both these concepts.

The concept of $C^p$-identifiable surfaces arose through the study of active constraint identification, fast tracks find their origin in the $U$-Lagrangian (a concept created to study second order expansions of non-differentiable functions), and partial smoothness began in studying sensitivity analysis. Despite the fact that each concept arose through different research they are the same. This shows partial smoothness is indeed a natural concept in nonsmooth optimization.
Chapter 3

Calculus of Partial Smoothness

In Chapter 2 we introduced the definition of partial smoothness. To understand partial smoothness better we provided several examples. We then compared partial smoothness to two convex notions of smooth substructure: $C^p$-identifiable surfaces and fast tracks. All three ideas arose naturally in the study of nonsmooth optimization, although through different research. Partial smoothness was shown to generalize both $C^p$-identifiable surfaces and fast tracks to a nonconvex setting. Therefore we have developed a strong argument showing that partial smoothness is a natural idea in nonsmooth optimization.

In this chapter we turn our attention to showing that partial smoothness is a powerful tool in nonsmooth optimization. To do this we examine the calculus of partial smoothness. (We use the term calculus as a broad notion including any rule which aids in calculation.) We begin by restating several calculus rules developed in [Lew02]. Although Lewis focused on $C^2$-partial smoothness, we see his results remain true in our broader framework of $C^p$-partial smoothness. In fact the proofs are essentially unchanged. As an application of these rules we provide an alternate proof to Theorem 2.18 (Partial Smoothness and Epigraphs). The chapter continues by developing a rule regarding the infimal convolution of two convex partly smooth functions. Theorem 3.10 continues the study of convex partly smooth functions by developing some interesting critical point analysis. In Section 3.3 we develop one of the most powerful results of this thesis, the Smooth Projection Theorem (Theorem 3.13). The Smooth Projection Theorem describes when the projection onto a prox-regular partly smooth set lies on the active manifold. This forces the active manifold of a prox-regular partly smooth function to be unique. The Smooth Projection Theorem further underlies many active constraint identification results which appear in Section 3.4 of this thesis. Section 3.5 develops some critical point analysis for prox-regular partly smooth functions. Theorem 3.23 shows that if a point minimizes a prox-regular partly smooth function along the active manifold, then it (locally) minimizes the function. We end Chapter 3 with an example, originally from Lewis [Lew02, Sec 7], showing the necessity of prox-regularity in the results of Sections 3.3, 3.4, and 3.5 of this thesis.
Combined, these results give a strong argument that partial smoothness (especially when combined with prox-regularity) is a powerful tool in nonsmooth optimization.

3.1 Classic Calculus of Partial Smoothness

The calculus of partial smoothness was richly developed in [Lew02]. This section extends some of the results of Lewis from $C^2$-partial smoothness to $C^p$-partial smoothness. Actually, all of the proofs of Lewis (not just the ones presented below) remain essentially unchanged when moving from '2' to 'p'. The goal of this section is not to show this fact (it is easy to see when reading the paper), but to demonstrate one application of the calculus of partial smoothness. Specifically we develop enough calculus to establish Theorem 2.18 (Partial Smoothness and Epigraphs) in a more elegant manner.

We begin with the classical notion of a chain rule. To understand the chain rule we first require a transversality condition from [Lew02, eq (4.1)]. We say a function $f$ and a set $M$ satisfy the transversality condition at the point $\bar{x}$ if

$$\text{nul}(\nabla f(\bar{x}))^* \cap N_M(f(\bar{x})) = \{0\}. \quad (3.1)$$

The transversality rule stated above shows strong similarity to the amenability condition (equation (1.1)). Both require that the normal cone of a set and the null set of a function intersect only at the trivial point. However, as the normal cones refer to two different sets it is difficult to compare these rules. In Section 4.2 of this thesis we will see one case where these two rules coincide. But now we return our attention to the chain rule.

**Theorem 3.1 (Chain Rule)** Let the $C^p$ function $G: \mathbb{R}^m \to \mathbb{R}^n$ and the manifold $M$ satisfy the transversality condition at the point $\bar{x}$. If the function $f$ is $C^p$-partly smooth at the point $G(\bar{x})$ relative to the manifold $M$ then the composition $f \circ G$ is $C^p$-partly smooth at $\bar{x}$ relative to the manifold $G^{-1}(M) := \{x : G(x) \in M\}$.

**Proof:** In the proof [Lew02, Thm 4.2] simply consider the word "smooth" to mean $C^p$ (instead of $C^2$ as Lewis intends).

Also proven in [Lew02] is the sum rule. Equation (3.2) (below) writes the transversality condition (equation (3.1)) to the form resulting from summing manifolds and functions.

**Theorem 3.2 (Sum Rule)** Consider the manifolds $M_1, M_2, \ldots, M_k$, the functions $f_1, f_2, \ldots, f_k$, and a point $\bar{x} \in \cap_i M_i$. Suppose that each function is $C^p$-partly smooth at the point $\bar{x}$ relative to its
respective manifold. If
\[
\left\{(y_1, y_2, \ldots, y_k) : \sum_{i=1}^{k} y_i = 0, y_i \in N_{M_i}(\bar{x}) \ \text{for} \ i = 1, 2, \ldots, k\right\} = \{0\},
\]
then the function \(\sum_i f_i\) is partly smooth at \(\bar{x}\) relative to the manifold \(\cap_i M_i\).

**Proof:** [Lew02, Cor 4.6] relies on the chain rule and [RW98, Prop 6.41], both of which hold in the \(C^p\) setting.

Since a \(C^p\) function is always partly smooth relative to the entire space, we have the following corollary.

**Corollary 3.3 (Smooth Perturbation Rule)** If the function \(f\) is \(C^p\)-partly smooth at the point \(\bar{x}\) relative to the manifold \(M\) and the function \(g\) is \(C^p\) on a neighbourhood of \(\bar{x}\), then the function \(f + g\) is \(C^p\)-partly smooth at \(\bar{x}\) relative to \(M\).

The last calculus rule we require before reproving Theorem 2.18 is the level set rule. Like the previous rules, it was developed in [Lew02], and the proof remains unchanged.

**Theorem 3.4 (Level Set Rule)** Suppose the function \(f\) is \(C^p\)-partly smooth at the point \(\bar{x}\) relative to the manifold \(M\), and \(\bar{x}\) is not a critical point of the problem \(\inf\{f(x) : x \in M\}\). Then the level set
\[
\text{lev } f := \{x \in \mathbb{R}^n : f(x) \leq 0\}
\]
is \(C^p\)-partly smooth at \(\bar{x}\) relative to the set
\[
M_0 := \{x \in M : f(x) = 0\}.
\]

**Proof:** Like the proof for the chain rule, we simply have to interpret the word "smooth" in [Lew02, Thm 4.10] as \(C^p\).

By combining the chain rule, sum rule, and level set rule together we can devise an alternate proof to the first direction of Theorem 2.18 (Partial Smoothness and Epigraphs). We end this section by showing if a function is partly smooth then its epigraph is also partly smooth via calculus rules. This approach to the proof of Theorem 2.18 also appears in [HL03].

**Theorem 3.5 (Epigraphs of Partly Smooth Functions)** If the function \(f\) is \(C^p\)-partly smooth at the point \(\bar{x}\) relative to the manifold \(M\) then the epigraph \(\text{epi } f\) is \(C^p\)-partly smooth at the point \((\bar{x}, f(\bar{x}))\) relative to the manifold \(\widehat{M} := \{(x, f(x)) : x \in M\}\).
Proof: Let the function \( f \) be \( C^{P} \)-partly smooth at the point \( \bar{x} \) relative to the manifold \( \mathcal{M} \). We next construct the epigraph via basic calculus rules.

Let \( P : \mathbb{R}^{n+1} \to \mathbb{R}^{n} \) be the restriction of \( \mathbb{R}^{n+1} \) onto the first \( n \) coordinates. Next define the function \( h \) via

\[
h : \mathbb{R}^{n+1} \to \mathbb{R}
\]

\[
(x, r) \mapsto f(P(x, r)) - r.
\]

Thus \( \text{lev} h := \{ (x, r) : h(x, r) \leq 0 \} = \text{epi} f. \)

Examining \( P \) we see

\[
\nabla P(\bar{x}, f(\bar{x})) = \begin{bmatrix}
100 \ldots 00 \\
010 \ldots 00 \\
\vdots \\
000 \ldots 10
\end{bmatrix},
\]

so \( \text{nul}(\nabla P(\bar{x}, f(\bar{x}))) = \{0\} \). Thus the transversality condition must hold, and the Chain Rule (Theorem 3.1) shows \( f(P(x, r)) \) is partly smooth at \( (\bar{x}, f(\bar{x})) \) relative to \( P^{-1}(\mathcal{M}) = \{(x, r) : P(x, r) \in \mathcal{M}\} \).

The Smooth Perturbation Rule (Corollary 3.3) shows \( h(x, r) \) is partly smooth at \( (\bar{x}, f(\bar{x})) \) relative to \( \{(x, r) : x \in \mathcal{M}\} \).

Lastly we apply the Level Set Rule (Theorem 3.4). Note that \( \partial h(\bar{x}, f(\bar{x})) = (\partial_{x} f(\bar{x}), -1) \), so \( (\bar{x}, f(\bar{x})) \) is not a critical point of \( \inf\{ h(x, r) : x \in \mathcal{M}, r \in \mathbb{R}\} \). Thus, \( \text{lev} f \) is partly smooth at \( (\bar{x}, f(\bar{x})) \) relative to \( \{(x, r) : x \in \mathcal{M}, h(x, r) = 0\} \). Simplifying we see \( \text{epi} f \) is partly smooth at \( (\bar{x}, f(\bar{x})) \) relative to \( \{(x, r) : x \in \mathcal{M}, r = f(x)\} \).

3.2 Convex Partial Smoothness

In this section we develop several new results for convex partly smooth functions.

First we turn our attention to the infimal convolution operation: for functions \( f \) and \( g \) the infimal convolution of \( f \) and \( g \) is denoted by the symbol \( \# \) and defined

\[
(f \# g)(y) := \inf_{\bar{x}} \{ f(x - y) + g(x) \}. \tag{3.3}
\]

The infimal convolution is useful in that it subsumes many operations used in nonsmooth analysis. For example, the distance to a set and the proximal envelope (see Definition 5.11 of this thesis) can both be phrased as infimal convolutions. Some researchers refer to infimal convolution as epi-addition, since the operation is equivalent to summing the epigraphs [RW98, Ex 1.28].
CHAPTER 3. CALCULUS OF PARTIAL SMOOTHNESS

After examining infimal convolutions of convex partly smooth functions, we examine nondegenerate critical points of convex partly smooth functions. The nondegeneracy condition was first studied by Dunn in the case of sets [Dun87]. In this section we examine the functional version: \( 0 \in \text{rint} \partial f(x) \). In Theorem 3.10 we show if a point is nondegenerate minimum of a convex partly smooth function, then all minima of the function lie on the tangent space to the active manifold.

In order to prove the Infimal Convolution Rule we require one simple lemma.

**Lemma 3.6** If convex sets \( C \) and \( D \) satisfy \( \text{rint} C \cap \text{rint} D \neq \emptyset \), then the parallel subspaces of \( C \) and \( D \) satisfy

\[
\text{par} C \cap \text{par} D = \text{par} (C \cap D).
\]

**Proof:** Select a point \( x \in \text{rint} C \cap \text{rint} D \). By definition \( \text{par} C = \mathbb{R}_+(C - x) \) and \( \text{par} D = \mathbb{R}_+(D - x) \). Therefore,

\[
\text{par} C \cap \text{par} D = \{ y : \lambda y \in C - x \text{ for some } \lambda > 0 \} \cap \{ z : \nu z \in D - x \text{ for some } \nu > 0 \}
\]

\[
= \{ y : \lambda y \in C - x \text{ for some } \lambda > 0 \text{ and } \nu y \in D - x \text{ for some } \nu > 0 \}
\]

\[
= \{ y : \lambda y \in C - x \text{ and } \lambda y \in D - z \text{ for some } \lambda > 0 \}
\]

\[
= \{ y : \lambda y + x \in C \text{ and } \lambda y + x \in D \text{ for some } \lambda > 0 \}
\]

\[
= \{ y : \lambda y + x \in (C \cap D) \text{ for some } \lambda > 0 \}
\]

\[
= \{ y : \lambda y \in (C \cap D) - x \text{ for some } \lambda > 0 \}
\]

\[
= \mathbb{R}_+((C \cap D) - x).
\]

As \( x \in \text{rint} C \cap \text{rint} D \), we have \( x \in \text{rint} (C \cap D) \) [RW98, Prop 2.42]. Thus, \( \mathbb{R}_+((C \cap D) - x) = \text{par} (C \cap D) \) and the proof is complete.

Next we state and prove the Infimal Convolution Rule.

**Theorem 3.7 (Infimal Convolution Rule)** Suppose \( f \) and \( g \) are convex functions, where \( f \) is \( C^p \)-partly smooth at the point \( \bar{x}_f \) relative to the manifold \( M_f \) and \( g \) is \( C^p \)-partly smooth at the point \( \bar{x}_g \) relative to the manifold \( M_g \). Further suppose that \( M_f \subseteq \text{argmin} f \), \( M_g \subseteq \text{argmin} g \),

\[
T_{M_f}(\bar{x}_f) \cap T_{M_g}(\bar{x}_g) = \{0\}, \tag{3.4}
\]

and

\[
\text{rint} \partial f(\bar{x}_f) \cap \text{rint} \partial g(\bar{x}_g) \neq \emptyset. \tag{3.5}
\]

Then the infimal convolution of \( f \) and \( g \),

\[
h(y) = (f \# g)(y) := \inf_x \{ f(y - x) + g(x) \}, \tag{3.6}
\]

is partly smooth at \( \bar{y} := \bar{x}_f + \bar{x}_g \) relative to the manifold \( M := M_f + M_g \).
Proof: Since \( f \) and \( g \) are bounded below (\( M_f \subseteq \text{argmin } f \) and \( M_g \subseteq \text{argmin } g \)), \( h \) is finite valued.

By Theorem 1.21, \( M \) is a manifold containing \( \bar{y} \). Theorem 1.21 further shows that there exists a neighbourhood \( V \) of \( \bar{y} \) such that each \( y \in M \cap V \) can be represented as a unique sum of \( x_f(y) \in M_f \) and \( x_g(y) \in M_g \). Moreover, as \( y \to_M \bar{y} \) one has \( x_f(y) \to_M \bar{x}_f \) and \( x_g(y) \to_M \bar{x}_g \). To simplify notation we shall write \( y = x_f + x_g \) instead of \( y = x_f(y) + x_g(y) \), but it should be remembered that \( x_f \) and \( x_g \) are defined uniquely by \( y \).

Next we examine the four conditions of partial smoothness.

(Smoothness) Let \( y = x_f + x_g \in M \) where \( x_f \in M_f \), and \( x_g \in M_g \). Then
\[
h(y) = \inf_{x_1 + x_2 = y} \{ f(x_1) + g(x_2) \} \\
\leq f(x_f) + g(x_g) \\
= \inf f + \inf g \\
\leq h(y),
\]
so we have equality. Thus \( h \) is constant on \( M \), therefore smooth.

(Regularity) As \( f \) and \( g \) are convex, \( h \) is convex, and therefore regular.

(Nonempty subdifferential and sub-continuity) Given \( y = x_f + x_g \), equation (3.6) attains its minimum at \( z = x_g \), so [RW98, Ex 10.18] yields
\[
\partial h(y) = \partial f(x_f) \cap \partial g(x_g).
\]
Equation (3.5) shows \( \partial h(\bar{y}) \neq \emptyset \).

In order to apply [RW98, Thm 4.32], we next show that \( \partial f(\bar{x}_f) \) and \( \partial g(\bar{x}_g) \) cannot be separated. Select any \( \bar{w} \in \text{rint } \partial f(\bar{x}_f) \cap \text{rint } \partial g(\bar{x}_g) \). To say \( \partial f(\bar{x}_f) \) and \( \partial g(\bar{x}_g) \) cannot be separated is to say there does not exist nonzero \( \alpha \in \mathbb{R}^n \), such that
\[
\partial f(\bar{x}_f) \subseteq \{ z : \langle z, \alpha \rangle \leq \langle \bar{w}, \alpha \rangle \} \quad \text{and} \quad \partial g(\bar{x}_g) \subseteq \{ z : \langle z, \alpha \rangle \geq \langle \bar{w}, \alpha \rangle \},
\]
(see [RW98, p. 62]).

Suppose such an \( \alpha \) exists. Then
\[
\partial f(\bar{x}_f) \subseteq \{ z : \langle z - \bar{w}, \alpha \rangle \leq 0 \} \\
\partial f(\bar{x}_f) - \bar{w} \subseteq \{ (z - \bar{w}) : \langle z - \bar{w}, \alpha \rangle \leq 0 \} \\
\mathbf{R}_+(\partial f(\bar{x}_f) - \bar{w}) \subseteq \{ z : \langle z, \alpha \rangle \leq 0 \} \\
\text{par } \partial f(\bar{x}_f) \subseteq \{ z : \langle z, \alpha \rangle \leq 0 \}.
\]
Applying Lemma 2.3 tells us \( N_{M_f}(\bar{x}_f) \subseteq \{ z : \langle z, \alpha \rangle \leq 0 \} \), so \( \alpha \in T_{M_f}(\bar{x}_f) \). The same argument applied to \( \partial g(\bar{x}_g) \subseteq \{ z : \langle z - \bar{w}, -\alpha \rangle \leq 0 \} \) yields \( -\alpha \in T_{M_g}(\bar{x}_g) \). As \( T_{M_g}(\bar{x}_g) \) is a subspace, this implies \( \alpha \in T_{M_g} \), which contradicts equation (3.4). Thus, \( \partial f(\bar{x}_f) \) and \( \partial g(\bar{x}_g) \) cannot be separated.
As \( \partial f \) and \( \partial g \) have closed convex images, the conditions of [RW98, Thm 4.2] are fulfilled. Thus
\[
\lim_{x_f \to \bar{x}_f, x_g \to \bar{x}_g} (\partial f(x_f) \cap \partial g(x_g)) = \partial f(\bar{x}_f) \cap \partial g(\bar{x}_g),
\]
which shows \( \lim_{y \to \mathcal{M}} \partial h(y) = \partial h(\bar{y}) \) and the sub-continuity condition is satisfied. As \( \partial h(\bar{y}) \neq \emptyset \) we also see that \( \partial h \) is nonempty on \( \mathcal{M} \).

(Sharpness) Writing \( \bar{y} = \bar{x}_f + \bar{x}_g \) as before and applying Theorem 1.21 we have
\[
N_M(\bar{y}) = N_{M_f}(\bar{x}_f) \cap N_{M_g}(\bar{x}_g).
\]

Lemma 2.3 (iv) and Lemma 3.6 combine to show
\[
N_{M_f}(\bar{x}_f) \cap N_{M_g}(\bar{x}_g) = \text{par} (\partial f(\bar{x}_f) \cap \partial g(\bar{x}_g)).
\]
As \( \partial f(\bar{x}_f) \cap \partial g(\bar{x}_g) = \partial h(\bar{y}) \), Lemma 2.3 (iv) shows the sharpness condition.

The infimal convolution of an indicator function and the norm creates the distance to a set. Using this, Theorem 3.7 can be used as an alternate approach to show the distance to a partly smooth set is partly smooth. Corollary 3.8 was shown in a broader nonconvex framework in Example 2.7, but it is nonetheless a nice application.

Corollary 3.8 (Distances to Partly Smooth Sets) Suppose the convex set \( C \) is partly smooth at the point \( \bar{x} \) relative to the manifold \( \mathcal{M} \). Then the distance to \( C \) is partly smooth at \( \bar{x} \) relative to \( \mathcal{M} \).

Proof: Write the distance function as the infimal convolution of the norm function and the indicator function of \( C \), \( \text{dist}(\cdot, C) = |\cdot||\delta_C \), then apply Theorem 3.7.

Examining Theorem 3.7 we might ask ourselves whether the rule could be generalized. In light of Corollary 3.8 versus Example 2.7 it is clear that in some cases the answer is yes. However, it is easy to see that without the convexity and the argmin conditions the infimum may occur away from the active manifolds. In this case the partial smoothness of the function is essentially ignored, so no interesting result would follow. The tangent cone condition is required to ensure \( M \) is a manifold (see Theorem 1.21), so it too would be difficult to remove. Next we provide an example showing the necessity of the relative interior of the two subdifferentials intersecting.

Example 3.9 (Necessity of \( \text{rint} \partial f \cap \text{rint} \partial g \neq \emptyset \)) Consider the functions \( f := \max\{-x, 0\} \) and \( g := \max\{0, x\} \). Notice that \( f \) and \( g \) are both partly smooth at the point 0 relative to the manifold.
\[ \{0\} = \arg\min f \cap \arg\min g. \] However, the infimal convolution \( f \# g \) is the constant function 0. This is not partly smooth relative to the manifold \( \{0\} + \{0\} \), so the result of Theorem 3.7 fails.

This completes our examination of the the conditions for applying the Infimal Convolution Rule. We now turn our attention to the nondegeneracy condition introduced by Dunn in [Dun87]. The next theorem examines a nondegenerate critical point of a partly smooth function. In Theorem 3.10 we see that for a convex partly smooth function nondegenerate critical points in some way locate all minimal points for the function.

**Theorem 3.10 (Nondegeneracy and Convex Partial Smoothness)** If the function \( f \) is partly smooth at the point \( \bar{x} \) relative to the manifold \( M \), then \( 0 \in \rint \partial f(\bar{x}) \) if and only if the three conditions,

\[
\begin{align*}
df(\bar{x}, t) &= 0 \text{ for all } t \in T_M(\bar{x}), \\
df(\bar{x}, n) &> 0 \text{ for all nonzero } n \in N_M(\bar{x}), \text{ and} \\
\partial f(\bar{x}) &\subseteq N_M(\bar{x})
\end{align*}
\]

hold.

Furthermore, if the function is convex then \( 0 \in \rint \partial f(\bar{x}) \) implies

\[
\bar{x} \in \arg\min f \subseteq T_M(\bar{x}) + \{\bar{x}\}.
\]

**Proof:** Suppose \( 0 \in \rint \partial f(\bar{x}) \). By [RW98, Thm 8.30] this implies

\[
\text{df}(\bar{x}, w) \geq 0
\]

for all \( w \in \mathbb{R}^n \). Therefore partial smoothness and Theorem 2.3 yield equation (3.7). Theorem 2.3 also shows that \( \text{par } \partial f(\bar{x}) \subseteq N_M(\bar{x}). \) As \( 0 \in \rint \partial f(\bar{x}) \) this implies \( \partial f(\bar{x}) \subseteq N_M(\bar{x}) \) (equation (3.9)).

If \( n \in N_M(\bar{x}) \), equation (3.11) shows

\[
df(\bar{x}, -n) \geq 0 \geq -df(\bar{x}, n).
\]

The sharpness condition on \( f \) then guarantees that at least one of these is nonzero. We seek to show that \( df(\bar{x}, n) > 0 \). If it is not, then \( df(\bar{x}, -n) > 0 \), so [RW98, Thm 8.30] shows that there must exist some \( w \in \partial f(\bar{x}) \) such that \( \langle w, -n \rangle > 0 \). As \( 0 \in \rint \partial f(\bar{x}) \) we know that \( -\tau w \in \partial f(\bar{x}) \) for some \( \tau > 0 \). Thus

\[
df(\bar{x}, n) = \sup \partial f(\bar{x}), n) \geq \langle -\tau w, n \rangle > 0,
\]
a contradiction.
Suppose now that equations (3.7) to (3.9) hold. As $\partial f(\bar{x}) \subseteq N_M(\bar{x})$, equation (3.8) combined with [RW98, Thm 8.30] shows that for all nonzero $n \in N_M(\bar{x})$ there exists $w \in \partial f(\bar{x}) \cap N_M(\bar{x})$ such that $\langle w, n \rangle > 0$. As $N_M(\bar{x})$ is a subspace, and $\partial f(\bar{x})$ is convex this implies $0 \in \text{int}_{N_M(\bar{x})} \partial f(\bar{x})$, (where $\text{int}_{N_M(\bar{x})}$ is the interior taken with respect to the subspace $N_M(\bar{x})$). Since $\partial f(\bar{x}) \subseteq N_M(\bar{x})$ we have $\text{rint} \partial f(\bar{x}) \supseteq \text{int}_{N_M(\bar{x})} \partial f(\bar{x})$, so $0 \in \text{rint} \partial f(\bar{x})$ as desired.

Before considering the added condition that $f$ be convex, we set up some notation. As $T_M(\bar{x})$ and $N_M(\bar{x})$ are perpendicular subspaces whose union span $\mathbb{R}^n$ we know that any $x \in \mathbb{R}^n$ can be written $x = P_{N_M(\bar{x})}(x) + P_{T_M(\bar{x})}(x)$. To simplify notation we shall write $x := x_N + x_T$ where $x_N := P_{N_M(\bar{x})}(x)$ and $x_T := P_{T_M(\bar{x})}(x)$.

Now suppose $f$ is convex, and equations (3.7) to (3.9) hold. Combining equations (3.8) and (3.9) shows that $x = \bar{x}$. Given any $x \in \text{argmin} f$ we need to show that $x \in T_M(\bar{x}) + \bar{x}$. As $f$ is convex, argmin $f$ is a closed convex set, thus $\text{conv} \{ \bar{x}, x \} \subseteq \text{argmin} f$. This implies that $df(\bar{x}, x - \bar{x}) = 0$. (This does not show that $df(\bar{x}, x - \bar{x}) = 0$, a statement which would complete the proof.)

We claim $P_{N_M(\bar{x})}(x) =: x_N = \bar{x}_N$. This would imply $x - \bar{x} = x_T - \bar{x}_T \in T_M(\bar{x})$, or $x \in T_M(\bar{x}) + \{ x \}$. 

If $x_N \neq \bar{x}_N$ then we would have $x_N - \bar{x}_N \in N_M(\bar{x}) \setminus \{ 0 \}$ which, by equation (3.8), would imply $df(\bar{x}, x_N - \bar{x}_N) > 0$. Theorem 8.30 of [RW98] would then show the existence of some $w \in \partial f(\bar{x})$ such that $\langle w, x_N - \bar{x}_N \rangle > 0$. As $\partial f(\bar{x}) \subseteq N_M(\bar{x})$ we have $w = w_N$. By [RW98, Thm 8.30] this would show $df(\bar{x}, x - \bar{x}) = \sup(\partial f(\bar{x}), x - \bar{x}) \geq \langle w_N, x - \bar{x} \rangle = \langle w_N, x_N - \bar{x}_N \rangle > 0$. This contradicts $df(\bar{x}, x - \bar{x}) = 0$, so $x_N$ must equal $\bar{x}_N$ as required.

The interest of Theorem 3.10 is, for a convex partly smooth function, the existence of a nondegenerate critical point implies that the active manifold locates all minima of the function. The convex function $f(x) := \max\{ x, 0 \}$ provides an example of a partly smooth function ($f$ is partly smooth at the minimal point $0$ relative to the manifold $\{ 0 \}$) for which not all minima lie in the tangent space to the active manifold. Conversely, the convex function $g(x) = \begin{cases} x^2 & x \leq 0 \\ x & x > 0. \end{cases}$ is partly smooth at the minimal point $0$ relative to the manifold $\{ 0 \}$ and all minima lie in the tangent space to the active manifold, but $0$ is not a nondegenerate critical point of $g$.

In Section 3.5 we revisit the relationship of nondegenerate critical points to optimality. There we show the nonconvex result that a nondegenerate critical point forces all minima (and indeed all critical points) to lie on the active manifold.
3.3 The Smooth Projection Theorem

We now turn our attention to a class of nonconvex partly smooth functions. In order to maintain some control over the function we impose the prox-regularity condition. Recall that prox-regularity is a generalization of convexity developed by Poliquin and Rockafellar [PR96b]. Recall also that all finite max functions are prox-regular, so Example 2.6 shows the abundance of prox-regular partly smooth functions.

In this section we use prox-regularity to develop one of the most powerful results of this thesis, the Smooth Projection Theorem. Loosely speaking, the Smooth Projection Theorem says the projection mapping for a prox-regular partly smooth function is well behaved in some open set. Specifically, it shows two things. To begin, the Smooth Projection Theorem shows that, for a prox-regular partly smooth set, the projection mapping is smooth. The level of smoothness is guaranteed as one degree less than than the level of partial smoothness. This extends many previously known results on the smoothness of the projection mapping (see [Hol73] [FP82] and [Wri93] for example). The Smooth Projection Theorem also shows, for a prox-regular partly smooth set, the projection mapping (locally) maps points onto the active manifold of partial smoothness. This is an extremely useful tool in the study of active constraint identification (a concept which we study in the next section). It also provides us with a proof that the active manifold of a prox-regular partly smooth set is unique. We finish the section by relating this result to prox-regular partly smooth functions, $C^p$-identifiable surfaces, and fast tracks.

Before stating the Smooth Projection Theorem we begin with two preliminary results. Lemma 3.11 is a technical result which also appears in [HL03].

**Lemma 3.11** If $M$ is a $C^2$ manifold about 0 and the normal vector $\bar{g} \in N_M(0)$ is sufficiently small then

\[
x \in M \implies |x - \bar{g}|^2 \geq |\bar{g}|^2 + \frac{1}{2} |x|^2.
\]

**Proof:** The required inequality is equivalent to

\[
|x - 2\bar{g}|^2 \geq |2\bar{g}|^2.
\]

As $M$ is $C^2$ it is prox-regular at 0. Selecting $\bar{g}$ sufficiently small ensures that the projection of $2\bar{g}$ onto $M$ is 0 (Lemma 1.14, part (iii)). Now notice that

\[
|x - 2\bar{g}|^2 \geq \min\{|x - 2\bar{g}|^2 : x \in M\} = |P_M(2\bar{g}) - 2\bar{g}|^2 = |2\bar{g}|^2,
\]

as required.
The next theorem is essentially due to Lewis who proved it for the case of $p = 2$. Like the results in Section 3.1 the proof is essentially unchanged for the general case, so omitted. Theorem 3.12 makes use of the notion of a strong critical point, which is defined in Definition 1.17 of this work. To aid readers we remind them here that we call a point $\bar{x}$ a strong critical point of $f$ relative to $\mathcal{M}$ if $0 \in \text{rint } \partial f(\bar{x})$ and there exists a constant $\varepsilon > 0$ such that

$$f(x) \geq f(\bar{x}) + \varepsilon |x - \bar{x}|^2$$

for all points $x \in \mathcal{M}$ near $\bar{x}$. With this in mind we present the second preliminary result needed to prove the Smooth Projection Theorem.

**Theorem 3.12 (Parametric Minimization)** Suppose the function $\rho : \mathbb{R}^k \times \mathbb{R}^m \to [-\infty, \infty]$ is $CP$-partly smooth at the point $(\bar{y}, \bar{z})$ relative to the manifold $\mathbb{R}^k \times \mathcal{M}$. If $\bar{z}$ is a strong critical point of $\rho_\bar{y}(\cdot) = \rho(\bar{y}, \cdot)$ relative to $\mathcal{M}$ then there exists neighbourhoods, $U$ of $\bar{z}$ and $V$ of $\bar{y}$ and a function $\Phi \in C^{p-1}$ such that for all parameters $y \in V$,

(i) $\Phi(\bar{y}) = \bar{z}$;

(ii) $\rho_y$ restricted to $\mathcal{M} \cap U$ has a unique critical point at $\Phi(y)$; and

(iii) $\Phi(y)$ is a strong critical point of $\rho_y$ restricted to $\mathcal{M} \cap U$.

**Proof:** Lewis shows this in [Lew02, Thm 5.7] for the case of $p = 2$. The proof is essentially unchanged for the broader case.

The Smooth Projection Theorem is an application of Theorem 3.12. The theorem's basic setting is a prox-regular partly smooth set. Example 2.6 shows such sets exist in abundance. The second condition of Theorem 3.13, the normal vector of interest should be contained in the relative interior of the normal cone, is also not strenuous. The normal cone to a prox-regular set is always convex, so has nonempty relative interior.

The Smooth Projection Theorem also appears in [HL03]. We forgo further discussion until after the statement and proof of Theorem 3.13.

**Theorem 3.13 (Smooth Projection Theorem)** Let the set $S$ be prox-regular at the point $\bar{z}$ and $CP$-partly smooth ($p \geq 2$) there relative to the manifold $\mathcal{M}$. Then for any normal vector $\bar{n} \in \text{rint } N_S(\bar{z})$

sufficiently small, there exists a neighbourhood of $\bar{z} + \bar{n}$ on which the projection mappings satisfy

$$P_{\mathcal{M}} \equiv P_S \in C^{p-1}. \quad (3.12)$$
Proof: By shifting the set we may assume without loss of generality that \( \bar{x} = 0 \).

Now define the function \( \rho \) via

\[
\rho : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}, \quad (n, x) \mapsto \frac{1}{2} |x - n|^2 + \delta_S(x),
\]

where \( \delta_S \) is the indicator function of \( S \), and (as in Theorem 3.12) define \( \rho_n = \rho(\bar{n}, \cdot) \). Assume \( \bar{n} \in \text{rint} \, N_S(0) \) is sufficiently small for Lemma 1.14 and Lemma 3.11 to hold on an open neighbourhood of \( \bar{n} \). We claim \( \rho \) satisfies the conditions of Theorem 3.12 at \( (\bar{n}, 0) \).

By [Lew02, Prop 4.51], \( \rho \) is partly smooth at \( (n, 0) \) relative to \( \mathbb{R}^m \times M \) for all vectors \( n \in \mathbb{R}^m \). Since \( \partial \rho_n(0) = -\bar{n} + N_S(0) \), our choice of \( \bar{n} \) ensures \( 0 \in \text{rint} \partial \rho_n(0) \). Lastly, to see that \( (\bar{n}, 0) \) is a strong critical point, note that equation (1.12) is equivalent to

\[
|x - \bar{n}|^2 \geq |\bar{n}|^2 + 2\varepsilon |x|^2 \quad \forall x \in M \text{ near } \bar{x}.
\]

This holds for \( \varepsilon = \frac{1}{4} \) by Lemma 3.11.

Therefore there exists a function \( \Phi \in C^{p-1} \) such that, for \( n \) near \( \bar{n} \), \( \Phi(n) \) is a strong critical point of \( \rho_n \) relative to \( M \) near 0. By (ii), \( \Phi(n) \in M \), while (iii) shows

\[
0 \in \text{rint} \partial \rho_n(\Phi(n)) = \text{rint} \{(\Phi(n) - n) + N_S(\Phi(n))\}.
\]

Therefore \( n - \Phi(n) \in \text{rint} \, N_S(\Phi(n)) \).

As \( \Phi \in C^{p-1} \) near \( \bar{n} \), we do not leave the prox-normal neighbourhood guaranteed by Lemma 1.14. Thus \( n - \Phi(n) \in \text{rint} \, N_S(\Phi(n)) \) implies \( P_S(n) \equiv \Phi(n) \equiv P_M(n) \in C^{p-1} \), completing the proof.

As mentioned before, the Smooth Projection Theorem actually tells us two things regarding projections onto prox-regular partly smooth sets. First, the projection is smooth, and second the projection lies on the active manifold.
CHAPTER 3. CALCULUS OF PARTIAL SMOOTHNESS

The first result (the projection is smooth) is of interest for several reasons. First it extends results of Holmes [Hol73], Fitzpatrick and Phelps [FP82], and Wright [Wri93] on the smoothness of the projection mapping. This extension is notable as it discusses the smoothness of projection mappings for nonconvex sets, something unconsidered in the above references. The first result further shows that prox-regularity is key in Theorem 3.13. Recent work of Poliquin, Rockafellar, and Thibault show that the projection mapping is single valued if and only if the set is prox-regular [PRT00, Thm 1.3] (or see Theorem 1.14 of this thesis). Since the Smooth Projection Theorem involves the projection mapping being smooth, the projection must be single valued, and therefore the set must be prox-regular. An example further demonstrating the necessity of prox-regularity appears in Section 3.6.

The second result of the Smooth Projection Theorem is that the projection onto a prox-regular partly smooth set must lie on the active manifold. If there existed a prox-regular set which was partly smooth relative to two distinct manifolds this quickly leads to a contradiction. The next corollary explores this idea. Like the Smooth Projection Theorem, the next corollary also appears in [HL03].

Corollary 3.14 (Uniqueness of Manifolds) Consider a set $S$ that is prox-regular at the point $\bar{x}$ and $C^p$-partly smooth ($p \geq 2$) there relative to both manifolds $M_1$ and $M_2$. Then locally $M_1 \equiv M_2$.

Similarly consider a function $f$ that is prox-regular at the point $\bar{x}$ and $C^p$-partly smooth ($p \geq 2$) there relative to both manifolds $M_1$ and $M_2$. Then locally $M_1 \equiv M_2$.

Proof: Suppose the set $S$ is prox-regular at the point $\bar{x}$ and $C^p$-partly smooth ($p \geq 2$) there relative to both manifolds $M_1$ and $M_2$. Select any normal vector $\bar{n} \in \text{rint} N_S(\bar{x})$ sufficiently small that the Smooth Projection Theorem (Theorem 3.13) and Lemma 1.14 both hold. Now select any sequence of points $x_k \in M_1$ converging to $\bar{x}$. Since $S$ is partly smooth relative to $M_1$, the normal cones $N_S(x_k)$ converge to $N_S(\bar{x})$. Therefore there exists normal vectors $n_k \in N_S(x_k)$ such that $n_k \rightarrow \bar{n}$.

We now apply the Smooth Projection Theorem with regards to $M_2$. As $x_k \rightarrow \bar{x}$ and $n_k \rightarrow \bar{n}$, eventually $x_k + n_k$ is sufficiently close to $\bar{x} + \bar{n}$ that $P_S(x_k + n_k) = P_{M_2}(x_k + n_k)$, and $n_k$ is in the proximal normal neighbourhood developed in Lemma 1.14. Thus the normals $n_k$ are proximal normals, and therefore $P_S(x_k + n_k) = x_k$. Therefore $x_k \in M_2$ as desired.

For the case of functions consider a function $f$ that is prox-regular at the point $\bar{x}$ and $C^p$-partly smooth ($p \geq 2$) there relative to both manifolds $M_1$, and $M_2$. Then the epigraph of $f$ is partly smooth at $(\bar{x}, f(\bar{x}))$ relative to both $\bar{M}_1 := \{(x, f(x)) : x \in M_1\}$ and $\bar{M}_2 := \{(x, f(x)) : x \in M_2\}$ (Theorem 2.18). Applying the above we have $\bar{M}_1 \equiv \bar{M}_2$ (locally), so we must also have $M_1 \equiv M_2$ (locally).
CHAPTER 3. CALCULUS OF PARTIAL SMOOTHNESS

We end this section by considering the uniqueness of the active manifold in the setting of $C^p$-identifiable surfaces, and fast tracks. These results can also be found in [Har03].

Corollary 3.15 (Uniqueness of Identifiable Surfaces and Fast Tracks) Let the convex set $C$ contain the point $\bar{x}$. If $M_1$ and $M_2$ are $C^p$-identifiable surfaces ($p \geq 2$) of $C$ at $\bar{x}$, then locally $M_1 \equiv M_2$.

Similarly, let $f$ be a convex function minimized at the point $\bar{x}$. If $v_1(u)$ and $v_2(u)$ are fast tracks for $f$, then locally $v_1 \equiv v_2$.

Proof: Apply Theorem 2.21, Theorem 2.27 and the fact all convex functions are prox-regular to Corollary 3.14.

3.4 Active Constraint Identification

In the last section we developed the Smooth Projection Theorem and several of its consequences on prox-regular partial smoothness. We discussed how the Smooth Projection Theorem extended results of Holmes, Fitzpatrick and Phelps, and Wright on the smoothness of the projection mapping. In this section we turn our attention to the other result of the Smooth Projection Theorem, that the projection must lie on the active manifold. Using this we will develop several results on the finite identification of active constraints. Though we will focus mainly on the case of identifying active constraints of a constraint set, we will also discuss the identification of active constraints in the functional setting. Almost all of the results of this section can also be found in [HL03] (the exception is Example 3.17), although Corollary 3.20 is slightly more general in the setting we have developed here.

We begin with the basic theorem describing active constraint identification.

Theorem 3.16 (Finite Identification for Sets) Consider a set $S$ that is $C^p$-partly smooth ($p \geq 2$) at the point $\bar{x}$ relative to the manifold $M$ and prox-regular at $\bar{x}$. If the normal vector $\bar{n}$ is in $\text{rint} \ N_S(\bar{x})$, and the sequences $\{x_k\}$ and $\{d_k\}$ satisfy

$$
\begin{align*}
    x_k &\to \bar{x}, \quad d_k \to \bar{n}, \quad \text{and } \text{dist}(d_k, N_S(x_k)) \to 0, \\
    &\quad \quad \quad \quad \quad (3.13)
\end{align*}
$$

then

$$
    x_k \in M \quad \text{for all large } k.
$$

Proof: Select a sequence of normal vectors $n_k \in N_S(x_k)$ such that $|d_k - n_k| \to 0$. Noting that $\bar{n} \in N_S(\bar{x})$ implies $\lambda \bar{n} \in N_S(\bar{x})$ for any $\lambda > 0$ we may select $\lambda > 0$ such that Lemma 1.14 and the
Smooth Projection Theorem (Theorem 3.13) both hold for $\lambda \bar{n}$. Hence there exists a neighbourhood $V$ of $\bar{x} + \lambda \bar{n}$ on which the projection mappings satisfy
\[
P_S \equiv p_M \in C^{p-1}.
\] (3.14)

As $d_k \to \bar{n}$ and $|d_k - n_k| \to 0$, we must have $\lambda n_k \to \lambda \bar{n}$. As $x_k \to \bar{x}$ for large $k$, $\lambda n_k + x_k$ will be in $V$. As $\lambda n_k \in N_S(x_k)$, the prox-normal property (Lemma 1.14 part (iii)) combined with equation (3.14) imply that $x_k = P_S(\lambda n_k + x_k) = P_M(\lambda n_k + x_k)$. Thus $x_k \in M$ for all large $k$.

Before continuing with our investigation on identifying active constraints, we briefly consider the demand that the normal vector be in the relative interior of the normal cone. It is easy to see prox-regularity implies that the normal cone is closed and convex, and therefore has nonempty relative interior. However, one might feel that by some sort of limiting argument the relative interior condition on the normal vector might be removed. If the vector were not in the relative interior of the normal cone, it could be approximated by a limit of vectors in the relative interior. Applying Theorem 3.16 to each of these we might, in the limit, create results which bypass the relative interior condition. The next example shows that the relative interior condition cannot be removed.

**Example 3.17 (Necessity of $\bar{n} \in \text{rint} N_S(\bar{x})$)** Consider the convex set $C := \{(x,y) : x \leq 0, y \geq 0\}$. As $C$ is convex it is prox-regular. Furthermore $C$ is partly smooth at the point $\bar{x} := (0,0)$ relative to the manifold $M := \{(0,0)\}$.

Consider now the normal vector $\bar{n} := (0,-1) \in N_C(0,0)$. This vector is not in the relative interior of the normal cone. The sequence of points $x_k := (\frac{-1}{k},0)$ and normal vectors $n_k := (-1,0)$ satisfy $x_k \to \bar{x}$, $n_k \to \bar{n}$, and $\text{dist}(n_k, N_S(x_k)) \to 0$, but $x_k \notin M$ for all $k$.

![Figure 3.2: Need for $\bar{n} \in \text{rint} N_S(\bar{x})$](image)
When thinking on Theorem 3.16 it is helpful to consider the minimization of a $C^1$ function over a prox-regular partly smooth constraint set (or indeed any constraint set). A well known necessary condition for a minimum is that the negative gradient of the function be in the normal cone to the set. In [Dun87], Dunn defined a point to be a nondegenerate critical point if the negative gradient lies in the relative interior to the normal cone. Using this he showed that the active constraints of a linearly constrained set could be finitely identified. His work was extended by Burke and Moré in 1988, who also made use of the nondegeneracy condition [BM88]. In 1993 Wright’s development of $C^p$-identifiable surfaces further extended the notion of finitely identifying active constraints. He too made use of the nondegeneracy condition [Wri93]. (Many others were also involved in this research: for more details please refer to Section 1.2 of this thesis.)

In all cases the results worked along the same lines. An algorithm (or collection of algorithms) which forced normal vectors to converge to a nondegenerate critical point was presented. The result was that the active constraints of a convex set were identified in a finite number of iterations. Our next theorem unifies, clarifies and extends these results to the setting of prox-regular partly smooth sets. It does not specify an algorithm, only the basic properties the algorithm must have. Remarkably, the key property the algorithm must have is the natural condition that it drives the the error in the first order optimality conditions to zero.

**Theorem 3.18 (Identification with Constraints)** Consider a $C^1$ function $f$. Let the set $S$ be $C^p$-partly smooth ($p \geq 2$) at the point $\bar{z}$ relative to the manifold $M$, and prox-regular there. Suppose $x_k \to \bar{z}$ and $-\nabla f(\bar{z}) \in \text{rint } N_S(\bar{z})$. Then

$$x_k \in M \text{ for all large } k$$

if and only if

$$\text{dist} (-\nabla f(x_k), N_S(x_k)) \to 0.$$ 

**Proof:** ($\Rightarrow$) If $x_k \in M$ for all $k$ sufficiently large, then $S$ being partly smooth implies that $S$ is regular at $x_k$ and $N_S(x_k) \to N_S(\bar{z})$ (Lemma 2.2 (ii) and (iv)). As $f \in C^1$ we know $-\nabla f(x_k) \to -\nabla f(\bar{z})$. Therefore $N_S(x_k) + \nabla f(x_k)$ converges to $N_S(\bar{z}) + \nabla f(\bar{z})$ [RW98, Thm 4.25 & Ex 4.29]. Combining this with [RW98, Cor 4.7] and $-\nabla f(\bar{z}) \in \text{rint } N_S(\bar{z})$ yields

$$\text{dist} (0, N_S(x_k) + \nabla f(x_k)) \to \text{dist} (0, N_S(\bar{z}) + \nabla f(\bar{z})) = 0,$$

which is equivalent to $\text{dist} (-\nabla f(x_k), N_S(x_k)) \to 0$.

($\Leftarrow$) Set $d_k = -\nabla f(x_k)$ and apply Theorem 3.16, noting that $f \in C^1$.

Example 2.6 describes a large class of $C^p$-partly smooth sets. Sets defined by a finite number of smooth inequality constraints are always prox-regular, so we have a large collection of sets which
satisfy the conditions of Theorem 3.18. Hence the following example, which also appears in [Bur90, Thm 6.1].

Example 3.19 (Identification with Inequalities) Consider minimizing a $C^1$ function $f$ over the set

$$S := \{ z : g_i(z) \leq 0, \ i = 1, 2, \ldots, m \}$$

for $C^2$ functions $g_i$. Suppose at the point $\bar{z}$ that the active gradients $\{ \nabla g_i(\bar{z}) : i \in A(\bar{z}) \}$ are linearly independent, where $A(x)$ is the active set $\{ i : g_i(x) = 0 \}$ (see Example 2.6). Suppose $x_k \to \bar{z}$ and $-\nabla f(\bar{z}) \in \text{rint} \ N_S(\bar{z})$, or equivalently strict complementarity holds in the first order conditions:

$$-\nabla f(\bar{z}) = \sum_{i \in A(\bar{z})} \lambda_i \nabla g_i(\bar{z}), \text{ where } \lambda_i > 0 \text{ for all } i \in A(\bar{z}).$$

Then by Theorem 3.18,

$$\text{dist}(-\nabla f(x_k), N_S(x_k)) \to 0$$

if and only if

$$A(x_k) = A(\bar{z}) \text{ for all large } k.$$

An interesting corollary arises when we combine Example 2.5 and Theorem 3.18.

Corollary 3.20 (Locally Sharp Minimizer) Suppose the set $S$ is prox-regular at the point $\bar{x}$. If $f$ is a $C^1$ function with $-\nabla f(\bar{z}) \in \text{int} \ N_S(\bar{z})$ and $x_k$ is a sequence of points converging to $\bar{x}$ then

$$x_k = \bar{x} \text{ for all large } k$$

if and only if

$$\text{dist}(-\nabla f(x_k), N_M(x_k)) \to 0.$$

Proof: Notice $S$ is regular and the normal cone $N_S(\bar{x})$ has interior, and then consider Example 2.5 and Theorem 3.18.

We end this section by seeing what Theorem 3.18 means when considering the minimization of functions. We could likewise switch Example 3.19 and Corollary 3.20 to the functional case, but we feel the results are clear enough without further expansion.

Theorem 3.21 (Identification for Functions) Let the function $f$ be $C^p$-partly smooth $(p \geq 2)$ at the point $\bar{x}$ relative to the manifold $M$, and prox-regular there, with $0 \in \text{rint} \ \partial f(\bar{z})$. Suppose $x_k \to \bar{x}$ and $f(x_k) \to f(\bar{x})$. Then

$$x_k \in M \text{ for all large } k$$
CHAPTER 3. CALCULUS OF PARTIAL SMOOTHNESS

if and only if

\[ \text{dist}(0, \partial f(x_k)) \to 0. \]

Proof: Let

\[ S := \text{epi } f, \]

and

\[ g : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R} \]

\[ (x, r) \mapsto r. \]

Define \( z_k := (x_k, f(x_k)) \) and \( \bar{x} := (\bar{x}, f(\bar{x})) \), and notice that

\[ -\nabla g(\bar{x}) = (0, -1) \in \text{rint } N_{\bar{x}}(\bar{x}), \]

by [RW98, Thm 8.9]. Theorem 2.18 shows that \( S \) is partly smooth at \( \bar{x} \) relative to \( \bar{M} := \{(x, f(x)) : x \in M \} \), and [PR96b, Thm 3.5] gives us prox-regularity at \( \bar{x} \), so we may apply Theorem 3.18. Thus we have

\[ x_k \in M \text{ for all large } k \]
\[ \iff z_k \in \bar{M} \text{ for all large } k \]
\[ \iff \text{dist}(-\nabla g(z_k), N_S(x_k, f(x_k))) \to 0 \]
\[ \iff \text{dist}((0, -1), N_S(x_k, f(x_k))) \to 0 \]
\[ \iff \text{dist}((0, -1), \{\lambda(u, -1) : u \in \partial f(x_k), \lambda > 0\}) \to 0 \]
\[ \iff \text{dist}(0, \partial f(x_k)) \to 0, \]

where the fourth equivalence follows from [RW98, Thm 8.9] (equation (1.11) in this work).

As a last note for this section, we mention that we could use Theorem 3.21 to reconstruct Theorem 3.18 for a function \( f \) and a set \( S \) by applying it to \( f + \delta_S \). However, this yields a slightly weaker result, as \( f \) must be \( C^2 \) rather than \( C^1 \).

3.5 Critical Point Analysis

We have seen many results so far that support the statement that prox-regular partly smooth functions (and sets) are a powerful tool in nonsmooth optimization. Many of these results suggest that if a function is prox-regular and partly smooth at the location of the minimum, the active manifold plays a key role in how the function behaves near the minimizer. The results of this section further support this, showing that the function need only be examined along the active manifold. The results of this section can also be found in [HL03].
We begin with a result showing that if a prox-regular partly smooth set or function has a non-degenerate critical point then locally all critical points lie on the active manifold.

**Corollary 3.22** Let the set $S$ be $C^2$-partly smooth at the point $\hat{x}$ relative to the manifold $\mathcal{M}$ and prox-regular there. Consider a function $f \in C^1$ with $-\nabla f(\hat{x}) \in \text{rint} N_S(\hat{x})$. Then locally all critical points of the constrained problem $\inf \{ f(x) : x \in S \}$ lie on $\mathcal{M}$.

Similarly, suppose the function $g$ is $C^2$-partly smooth at the point $\hat{x}$ relative to the manifold $\mathcal{M}$ and prox-regular there, with $0 \in \text{rint} \partial g(\hat{x})$. If $g$ is continuous at $\hat{x}$ then locally all critical points of the unconstrained problem $\inf \{ g(x) \}$ lie on $\mathcal{M}$.

**Proof:** Suppose $x_k$ is a sequence of critical points for the problem $\min_S \{ f(x) \}$ converging to $\hat{x}$. This sequence must satisfy Theorem 3.18. Thus $x_k$ must lie on the active manifold for all $k$ large. (The same argument applied to Theorem 3.21 shows the result for $\min \{ g(x) \}$.)

Like Theorem 3.10, Corollary 3.22 examines nondegenerate critical points of partly smooth functions. Theorem 3.10 shows that, in the convex case, if a function is partly smooth at a nondegenerate critical point $\hat{x}$ relative to a manifold $\mathcal{M}$ then all minima for the function lie on a shift of the tangent space to the active manifold $(\hat{x} + T_M(\hat{x}))$. Corollary 3.22 subsumes this result. Indeed, suppose a function $f$ is convex and partly smooth at a nondegenerate critical point $\hat{x}$ relative to a manifold $\mathcal{M}$, and $x$ is a minimal point of $f$. By convexity the line segment $[x, \hat{x}]$ consists of minimal points of $f$. Therefore, each point on $[x, \hat{x}]$ is a critical point of $f$, so Corollary 3.22 shows $[x, \hat{x}] \subseteq \mathcal{M}$.

In the above argument Corollary 3.22 shows that, if a convex function is partly smooth at a nondegenerate critical point, then all minima of the function lie on the active manifold. This cannot be true in general, as without convexity global control cannot be established. Our next result examines nondegenerate critical points for prox-regular partly smooth functions. In this case, it is sufficient to consider optimality only on the active manifold to determine local optimality for the function. This answers a question posed by Lewis in [Lew02, Sec 7] as to what extra conditions to partial smoothness are required to ensure this.

**Theorem 3.23 (Sufficient Optimality Conditions)** Consider a function $f$ that is prox-regular at the point $\hat{x}$ and $C^2$-partly smooth there relative to the manifold $\mathcal{M}$.

(i) If $\hat{x}$ is a strict local minimizer of $f$ restricted to $\mathcal{M}$, and satisfies $0 \in \text{rint} \partial f(\hat{x})$, then $\hat{x}$ is in fact an unconstrained strict local minimizer of $f$.

(ii) If $\hat{x}$ is a strong critical point of $f$ relative to $\mathcal{M}$, then $f$ grows at least quadratically near $\hat{x}$.  

CHAPTER 3. CALCULUS OF PARTIAL SMOOTHNESS

Proof: Considering the two cases, we see that both imply the existence of some smooth nonnegative function $g$ such that

$$f(x) > f(\bar{x}) + g(x - \bar{x})$$

(3.15)

for all $x \in \mathcal{M}$ near $\bar{x}$, with $x \neq \bar{x}$. Furthermore, in both cases it suffices to show

$$f(z) > f(\bar{z}) + g(z - \bar{z})$$

for all points $z \in \mathbb{R}^m$ near but not equal to $\bar{z}$. Indeed, for case (i) $g \equiv 0$, while for case (ii) $g = \varepsilon |\cdot|^2$ for some $\varepsilon > 0$.

For the sake of eventual contradiction, suppose there exists a sequence $z_k \to \bar{z}$ with

$$f(z_k) \leq f(\bar{z}) + g(z_k - \bar{z}) \text{ for all } k.$$  

(3.16)

Equation (3.15) shows $z_k \notin \mathcal{M}$, so the projection of $z_k$ onto $\mathcal{M}$, $x_k := P_{\mathcal{M}}(z_k)$ must differ from $z_k$. Thus the normal vectors

$$n_k := \frac{z_k - x_k}{|z_k - x_k|} \in N_\mathcal{M}(z_k)$$

(3.17)

are well defined, with $|n_k| = 1$. Dropping to a subsequence as necessary, and noting $\mathcal{M}$ is a $C^2$ manifold, we may further assume $n_k \to \bar{n}$ for some normal vector $\bar{n} \in N_\mathcal{M}(\bar{z})$, as $k \to \infty$.

As $0 \in \text{rint } \partial f(\bar{z})$ and $f$ is partly smooth we know $N_{\mathcal{M}}(\bar{z}) = \mathbb{R}_+ \partial f(\bar{z})$. Thus there exists some $\lambda > 0$ so $\lambda \bar{n} \in \partial f(\bar{z})$. Moreover, $\partial f$ is continuous along $\mathcal{M}$, so there exists $w_k \in \partial f(x_k)$ with $w_k \to \lambda \bar{n}$.

Since $f$ is prox-regular at $\bar{z}$ for $\lambda \bar{n}$, there exists a constant $R > 0$ so for all large $k$ we have

$$f(x_k) \geq f(x_k) + \langle w_k, x_k - x_k \rangle - R|x_k - x_k|^2,$$

(3.18)

and from equation (3.15) we also have

$$f(x_k) - g(x_k - \bar{z}) \geq f(\bar{z}).$$

(3.19)

Combining these with equation (3.16) we find

$$f(x_k) - g(x_k - \bar{z}) + g(x_k - \bar{z}) \geq f(x_k) + \langle w_k, x_k - x_k \rangle - R|x_k - x_k|^2,$$

which simplifies to

$$\frac{g(x_k - \bar{z}) - g(x_k - \bar{z})}{|x_k - x_k|} \geq \langle w_k, n_k \rangle - R|x_k - x_k|.$$

In both cases (i) and (ii), we note $g$ is differentiable at the origin with gradient zero. Thus taking the limit as $k \to \infty$ yields

$$0 \geq \lambda \langle \bar{n}, \bar{n} \rangle = \lambda,$$

a contradiction. Thus the existence of $z_k$ is impossible and the proof is complete.
CHAPTER 3. CALCULUS OF PARTIAL SMOOTHNESS

3.6 Necessity of Prox-regularity

Examining the results of Sections 3.3, 3.4, and 3.5 it is natural to ask, to what level is prox-regularity required? As discussed, Lemma 1.14 argues that prox-regularity is key in the Smooth Projection Theorem. However, perhaps some of the other results could be approached in a different manner. We end this chapter by attempting to show prox-regularity is necessary for the results of Sections 3.3, 3.4, and 3.5. To do this we reexamine Example 2.16. Recall that this example of Lewis [Lew02, Sec 7] originally provided a partly smooth function with non-unique active manifold. Below we see how the example also demonstrates the necessity of prox-regularity for the remaining results of Sections 3.3, 3.4, and 3.5.

Example 3.24 (Necessity of Prox-Regularity) Define the function

\[
f(x, y) := \begin{cases}
x^2 - y & (y \leq 0) \\
\sqrt{x^4 + 2x^2y - y^2} & (0 < y < 2x^2) \\
3x^2 - y & (2x^2 \leq y \leq 4x^2) \\
y - 5x^2 & (4x^2 < y)
\end{cases}
\]

(see Figure 2.16). Recall that \(f\) is partly smooth at \((0, 0)\) relative to both

\[
M_1 := \{(x, y) : y = 0\} \\
M_2 := \{(x, y) : y = 4x^2\}.
\]

Immediately we see the conclusion of Corollary 3.14 (Uniqueness of Manifolds) fails. In [Lew02, Sec 7] it is shown \(f\) has a strong critical point at \((0, 0)\) with respect to \(M_1\), but \((0, 0)\) is not a local minimum of \(f\). Thus the conclusion of Theorem 3.23 (Sufficient Optimality Conditions) fails.

Furthermore despite the fact \(0 \in \text{rint} \partial f(0, 0)\) the conclusions of Theorem 3.21 (Identification for Functions) do not hold. To see this consider the sequence of points \(x_k = (1/k, 0) \to (0, 0)\) in \(M_1\). The subdifferential of \(f\) is explicitly found in [Lew02, Sec 7], but it suffices to note that for all real \(x\),

\[
\partial f(x, 0) = \text{conv} \{(2x, -1), (2x, 1)\}.
\]

Thus we see \(\text{dist}(0, \partial f(x_k)) \to 0\), but \(x_k \notin M_2\) for all \(k\). (We could similarly swap the roles of \(M_1\) and \(M_2\).)

The problem of course is that \(f\) is not prox-regular at \((0, 0)\), as detailed in Example 2.16.
Chapter 4

Related Notions

In Chapter 2 of this thesis we introduced the notion of a partly smooth function or set. We examined the definition, and compared it to other previously defined notions of smooth substructure, showing it was a natural tool in the study of nonsmooth optimization.

Chapter 3 focused on showing the power of partial smoothness in nonsmooth optimization. To do this we discussed calculus rules, active constraint identification, and critical point analysis. We began Chapter 3 by showing partial smoothness had a well developed calculus. To demonstrate this we used calculus rules to reprove the relationship of a partly smooth function and its epigraph. Chapter 3 continued by creating an infimal convolution rule and developing some critical point analysis for convex partly smooth functions. In Section 3.3 we developed the Smooth Projection Theorem (Theorem 3.13), which extended many previously known results. We then showed how the Smooth Projection Theorem entailed uniqueness of the active manifold, and allowed for the identification of active constraints. This led to the critical point analysis of Section 3.5. Together these results strongly suggest that partial smoothness is a powerful tool in the study of nonsmooth optimization, especially when combined with prox-regularity.

In Chapter 4 we discuss two recent notions of smooth substructure: primal-dual gradient (PDG) structures and $g \circ F$ decompositions. Both are broad classes of nonconvex functions with underlying smooth substructure. As in Sections 2.4 and 2.5 we focus largely on how these notions are related to partial smoothness. Unlike $C^p$-identifiable surfaces and fast tracks, the relationships of PDG structures and $g \circ F$ decompositions to partial smoothness are not easily defined. In Section 4.1 we see that PDG structures are a much broader class than partly smooth functions (see Theorem 4.8 for example). However, with the addition of some structure on certain index sets, a PDG structure implies partial smoothness. Section 4.2 discusses the relationship of $g \circ F$ decompositions with partial smoothness. Specifically it shows that $g \circ F$ decomposable functions are compositions of partly smooth functions with smooth functions. This leads to some additional properties of $g \circ F$
decomposable functions.

Most of the results from this chapter can also be found in [Har03].

4.1 Primal-dual Gradient Structures

Like fast tracks, primal-dual gradient (PDG) structures are closely related to UV-decompositions. When Mifflin and Sagastizábal starting studying how UV-decompositions could be used for algorithmic purposes they sought a class of functions for which the UV-decomposition could be easily determined. The obvious answer was the class of finite max functions, which Mifflin and Sagastizábal study in [MS99]. They end this paper with the goal of extending their results to a class of infinite max functions. To this end PDG structures were defined [MS00a]. It was soon seen that this broad class of functions, with some additional structure, also served as an example of functions with fast tracks [MS00b].

The definition of PDG structures has evolved over several papers, from the class of convex functions in [MS00a] to a much broader class of nonconvex functions in [MS03] (a third definition somewhat between these two appears in [MS00b]). In this section we will examine the most recent definition from [MS03]. We provide this in Definition 4.1 below.

It is useful to recall that the Clarke subdifferential to a function $f$ at a point $\bar{x}$ is defined

$$\partial f(\bar{x}) := \{w : (w, -1) \in \text{conv } N_{\text{epi } f}(\bar{x}, f(\bar{x}))\},$$

(Definition 1.9).

**Definition 4.1 (Primal-Dual Gradient Structures)** A lsc function $f$ has primal-dual gradient structure (PDG structure) at a point $\bar{x}$ relative to a set $P$ if there exists smooth primal functions $\{f_i\}_{i=0}^{m_1}$ and smooth dual functions $\{\phi_j(x)\}_{j=1}^{m_2}$, and a closed convex set $\Delta \subseteq \mathbb{R}^{m_1+1+m_2}$ locally satisfying:

(i) $\bar{x} \in \{x : f_i(x) = f(x) \text{ for all } i, \phi_j(x) = 0 \text{ for all } j\} \subseteq P$;

(ii) if $(\alpha, \beta) := (\alpha_0, \alpha_1, ..., \alpha_{m_1}, \beta_1, \beta_2, ..., \beta_{m_2}) \in \Delta$ then $\alpha$ is an element of the set $\{(\alpha_0, ..., \alpha_{m_1}) : \sum_{i=0}^{m_1} \alpha_i = 1, \alpha_i \geq 0\}$;

(iii) the $i$th unit vector is in $\Delta$ for $i = 0, 1, ..., m_1$;

(iv) for each $\bar{j} = 1...m_2$, there exists $(\alpha, \beta) \in \Delta$ such that $\beta_j \neq 0$ and $\beta_j = 0$ for $j \neq \bar{j}$;

(v) for each $x \in P$,

(a) $f(x) = f_i(x)$ for some $i$ and
(b) \( w \in \partial f(x) \) if and only if

\[
    w = \sum_{i=0}^{m_1} \alpha_i \nabla f_i(x) + \sum_{j=1}^{m_2} \beta_j \nabla \phi_j(x),
\]

where \((\alpha, \beta) \in \Delta\) satisfies

\[
    \begin{align*}
    \alpha_i &= 0 \quad \text{if } f_i(x) \neq f(x) \\
    \beta_j &= 0 \quad \text{if } \phi_j(x) \neq 0.
    \end{align*}
\]

With the number of conditions required for a PDG structure to exist, one might assume PDG structures are a rather restrictive class of functions. The next example both gives the reader a feel for what PDG structures are capturing, and shows that they exist in abundance.

**Example 4.2 (Abundance of PDG Structures)** Consider any finite collection of \(C^2\) functions \(\{f_i\}_{i=0}^m\) and define the finite max function \(f(x) = \max\{f_i(x) : i = 0, 1, \ldots, m\}\). For any point \(\bar{x}\), a PDG structure can be created as follows.

Select the original \(f_i\) as the primal functions, \(\phi_1 = 0\) as the dual function, \(P = \mathbb{R}^n\), and \(\Delta := \{(a_0, a_1, \ldots, a_m, \beta_1) : \sum_{i=0}^m a_i = 1, a_i \geq 0, \beta_1 \in [0, 1]\}\). These clearly satisfy conditions (i) to (iv), and (v)(a) of Definition 4.1. Condition (v)(b) is true by [RW98, Ex 8.31].

Let us now attempt to explain each of the conditions of a PDG structure. The definition begins by asking for a set \(P\) over which the PDG structure is defined. Condition (i) states that this set must contain all the points at which all the primal and dual functions are active. In conditions (ii) to (iv) a second set, \(\Delta\), is constructed. This set is used in condition (v) to define the Clarke subdifferential for the function. In this manner PDG structures are demanding both control over the active constraints of the structure, and control over the subdifferential of the function.

Before discussing the similarities of this to partial smoothness we will require another definition (Definition 4.4 (PDG Index sets), below). To inspire this, we examine another example on the abundance of PDG structures. In the next example we show how a function can be associated with multiple PDG structures. The example further shows that regularity is not a necessary condition for a function to have a PDG structure.

**Example 4.3 (Nonregular Nonunique PDG Structure)** It is easy to confirm that the primal function \(f_0 = 0\), the dual function \(\phi_1(x) = x\), the set \(P = \{0\}\), and the set \(\Delta = \{(1, \beta) : \beta \in [-1, 1]\}\) provide a PDG structure for the nonregular function \(-|x|\) at the point 0.

Alternately, primal functions \(f_0(x) = x\) and \(f_1(x) = -x\), dual function \(\phi_1(x) = 0\), \(P\) any subset of \(\mathbb{R}\) containing 0, and \(\Delta = \{(a_0, a_1, \beta) : a_0 + a_1 = 1, a_i \geq 0, \beta \in [0, 1]\}\) also defines a PDG structure for \(-|x|\) at 0. Moreover the addition of the function \(g(x) = x^n\) for \(n \geq 2\) to either the
primal or dual function list has no effect on the PDG structure (except to alter the set \( \Delta \)). Thus a single function can have several (indeed infinite) different PDG structures.

Furthermore, all of the above lists also provide PDG structures for the convex function \(|x|\).

The above example shows both a strength and weakness of PDG structures. As a strength we see that PDG structures include functions that are not regular. The weakness lies in the nonuniqueness of the PDG definition. In fact, no function has a unique PDG structure.

To see this consider a PDG structure \( (\{f_i\}_{i=1}^{m_1}, \{\phi_j\}_{j=1}^{m_2}, P, \Delta) \) at a point \( \bar{x} \) for a function \( f \). Adding the smooth function \( \phi_{m_2+1}(x) := (x - \bar{x})^2 \) to the list of dual functions and adjusting the \( P \) set to \( \{\bar{x}\} \) creates a new PDG structure for the function. Indeed all that needs to be confirmed is a new \( \Delta \) set can be created. This can be accomplished by the set

\[
\Delta_{\text{new}} := \{(\alpha, \beta, \beta_{m_2+1}) : (\alpha, \beta) \in \Delta, \beta_{m_2+1} \in [0,1]\} = \Delta \times [0,1].
\]

In order to gain better control over PDG structures, Mifflin and Sagastizábal introduced several additional definitions. The concept of an index set, a list of the primal and dual functions which locally are of interest, is natural. By focusing on index sets with favourable properties one can gain further control over the PDG structure. We provide Mifflin and Sagastizábal's definitions from [MS03] next.

**Definition 4.4 (PDG Index sets)** Let the function \( f \) have PDG structure at the point \( \bar{x} \) relative to the set \( P \). Let \( \{f_i\}_{i=1}^{m_1} \) be the collection of primal functions, \( \{\phi_j\}_{j=1}^{m_2} \) be the collection of dual functions, and \( \Delta \) be the closed convex set as described in Definition 4.1.

An ordered pair of sets \( K := (K_f, K_\phi) \subseteq (\{0,1,\ldots,m_1\}, \{1,2,\ldots,m_2\}) \) is a basic index set if \( 0 \in K_f \) and

\[
\left\{ \begin{bmatrix} \nabla f_i(\bar{x}) \\ 1 \end{bmatrix} \right\}_{i \in K_f} \cup \left\{ \begin{bmatrix} \nabla \phi_j \\ 0 \end{bmatrix} \right\}_{j \in K_\phi}
\]

is a linearly independent set.

A basic index set \( K \) is UV-transversal if

\[
V_K := \conv \mathbb{R}_+(\{ \nabla f_i(\bar{x}) - \nabla f_0(\bar{x}) \}_{i \in K_f} \cup \{ \nabla \phi_j(\bar{x}) \}_{j \in K_\phi})
\]

is equal to the "sharp" subspace \( V = \operatorname{par} \tilde{\partial}f(\bar{x}) \).

A nonsingleton basic index set \( K \) is primal feasible if for all sufficiently small points \( u \) in the "smooth" subspace \( U = N_{\tilde{\partial}f(\bar{x})}(\bar{w}) \) (where \( \bar{w} \in \operatorname{rint} \tilde{\partial}f(\bar{x}) \)), the unique solution (see [MS03, Thm 4.2])

\[
v = v_K(u)
\]

is equal to

\[
\begin{align*}
f_i(\bar{x} + u + v) &= f_0(\bar{x} + u + v) \\
\phi_j(\bar{x} + u + v) &= 0
\end{align*}
\]

(4.1)
satisfies
\[ x(u) := \bar{x} + u + v_K(u) \in P \text{ and } f(x(u)) = f_i(x(u)) \text{ for all } i \in K_f. \]

A basic index set \( K \) is dual feasible relative to a Clarke subgradient \( w \in \partial f(\bar{x}) \) if there exists an \( (\alpha, \beta) \in \Delta \) such that
\[
w = \sum_{i \in K_f} \alpha_i \nabla f_i(\bar{x}) + \sum_{j \in K_\phi} \beta_j \nabla \psi_j(\bar{x}) \quad \text{with} \quad \alpha_i = 0 \text{ for all } i \notin K_f
\]
\[
\beta_j = 0 \text{ for all } i \notin K_\phi.
\]

A basic index set is dual feasible if it is dual feasible relative to all Clarke subgradients \( w \in \partial f(\bar{x}) \).

Lastly, we say \( f \) has a UV-transversal (primal feasible, dual feasible) PDG structure if it has a PDG structure with a UV-transversal (primal feasible, dual feasible) index set.

The first definition presented, basic index set, immediately denies trivial functions (such as \(|-\bar{x}|^2\)) from entering an index set. This puts a considerable restriction on the index set.

The concept of a UV-transversal index set is extremely useful in PDG analysis. By yielding another description of the UV-decomposition it allows many results regarding the smooth space \((U)\) to be developed. Using this property, first and second order derivative information for the \(U\)-Lagrangian can be obtained [MS03].

A primal feasible index is a condition on both the index set, and on the selection of the set \( P \). To satisfy primal feasibility, \( P \) must be sufficiently large to contain solutions to the system of equations (4.1). This puts more control on the trajectories created by equations (4.1), thus allowing results on the smoothness of the \(U\)-Lagrangian to be developed [MS03, Thm 5.1 & Thm 5.2].

Lastly, a dual feasible index says the indexed functions are the only functions required to describe the Clarke subdifferential. Dual feasibility is examined extensively in [MS00b] and [MS03] where many of the results for UV-transversality are shown to have stronger consequences when dual feasible index sets are present.

Notice UV-transversal, primal feasible, and dual feasible index sets all begin with the condition that the index set be a basic index set. As basic index sets \( \{K_f, K_\phi\} \) must satisfy \( 0 \in K_f \), none of the more restrictive index sets may be created by the empty set. To further see the connections between the various types of index sets, and to simply gain a feel for index sets, we provide the following example.

Example 4.5 (Index Sets) In Example 4.3 we explained several different approaches to defining a PDG structure for the convex function \(|x|\).

The first approach is to select the primal function \( f_0 = 0 \), the dual function \( \phi_1(x) = x \), the set \( P = \{0\} \), and the set \( \Delta = \{(1, \beta) : \beta \in [-1, 1]\} \). There are only two permissible basic index sets for this
CHAPTER 4. RELATED NOTIONS

PDG structure: \(\{K_f, K_\phi\} = \{\{0\}, \emptyset\}\) and \(\{K_f, K_\phi\} = \{\{0\}, \{1\}\}\). If \(K_\phi = \emptyset\) then the resulting basic index set is not \(UV\)-transversal, as \(V_K = \text{conv} \mathbb{R}_+\{\nabla f_0(0)\} = \{0\}\) while \(\text{par} \partial f(0) = \mathbb{R}\). Alternately if \(K_\phi = \{0\}\) the resulting basic index set is \(UV\)-transversal, as \(V_K = \text{conv} \mathbb{R}_+\{\nabla f_0(0), \nabla \phi_1(0)\} = \mathbb{R}\). In this case, it is easy to check that the basic index set is also primal and dual feasible.

The second approach is to use primal functions \(f_0(x) = x\) and \(f_1(x) = -x\), dual function \(\phi_1(x) = 0\), \(P\) any subset of \(\mathbb{R}\) containing 0, and \(\Delta = \{(a_0, a_1, \beta) : a_0 + a_1 = 1, a_i \geq 0, \beta \in [0, 1]\}\). As \(\nabla \phi_1 = 0\), the only permissible basic index sets for this decomposition are \(\{K_f, K_\phi\} = \{\{0\}, \emptyset\}\), and \(\{K_f, K_\phi\} = \{\{0, 1\}, \emptyset\}\). The first is not \(UV\)-transversal, while the second is \(UV\)-transversal, primal feasible and dual feasible.

As a third approach we could select primal functions \(f_0(x) = \frac{x}{2}\) and \(f_1(x) = -x\), dual function \(\phi_1 = \frac{x}{2}\), \(P = \{0\}\), and \(\Delta := \{(a_0, a_1, \beta) : a_0 + a_1 = 1, a_i \geq 0, \beta = a_0\}\). In this case the only permissible index sets are \(K = \{\{0\}, \{1\}\}\) and \(K = \{\{0, 1\}, \emptyset\}\). Both of these index sets are \(UV\)-transversal and primal feasible, but not dual feasible.

The above example contains many types of basic index sets, ranging from not \(UV\)-transversal, to \(UV\)-transversal and both primal and dual feasible. The final approach yields an example of a \(UV\)-transversal primal feasible index set that is not dual feasible. One of our next tasks is to create a PDG structure that is \(UV\)-transversal and dual feasible but not primal feasible. To do this we must discuss the connection between PDG structures and partly smooth functions. The next theorem, due to Mifflin and Sagastizábal, shows that if a function has a primal dual feasible PDG structure then the function is partly smooth.

**Theorem 4.6 (PDG Structures and Partial Smoothness)** [MS03, Thm 7.4] Suppose the regular function \(f\) has PDG structure at \(\bar{x}\) relative to \(P\). If \(K = \{K_f, K_\phi\}\) is an index set that is both primal and dual feasible, then \(f\) is partly smooth at \(\bar{x}\) relative to the manifold

\[
M_K := \{x : f_i(x) - f_0(x) = 0 \text{ for each } i \in K_f, \phi_j(x) = 0 \text{ for each } j \in K_\phi\}.
\]

Theorem 4.6 shows that a nicely behaved PDG structure is actually a partly smooth function. It therefore also shows a connection between convex PDG structures and fast tracks. This connection is explored in [MS02a], where convex primal dual feasible PDG structures were used as an example of functions with fast tracks.

Theorem 4.6 also allows us to develop an otherwise difficult example. The next example shows a dual feasible PDG structure with no primal feasible index set.
Example 4.7 (Non-primal feasible PDG) Consider the convex function

\[
f(x, y) := \max \begin{cases} 
3x^2 - 2y \\
3x^2 \\
2y + x^2 
\end{cases}
\]

at the point (0, 0). (See Figure 2.13.) As \( f \) is a finite max function it has a PDG structure with a dual feasible index set at (0, 0) [MS00b, Lem 4.3]. However, as \( f \) is not partly smooth (Example 2.13), Theorem 4.6 shows that \( f \) has no primal feasible index set.

There are still several open questions regarding how index sets for PDG structures interact. It is known that dual feasible index sets must be UV-transversal [MS03]. We have shown dual feasible index sets need not be primal feasible. We have also shown primal feasible index sets need not be dual feasible. However, we have no example of a primal feasible index set which is not UV-transversal, or of a UV-transversal index set which is neither primal nor dual feasible.

Theorem 4.6 shows relations between well behaved PDG structures and partly smooth functions. The question naturally arises of when a partly smooth function must have a well behaved PDG structure. The next theorem answers this, but in a rather unsatisfactory way.

Although Theorem 4.8 shows that all partly smooth functions have a PDG structure with a dual feasible index set, it does not maintain the active manifold when creating the PDG structure. Instead it forces the PDG structure to hold only for the point of partial smoothness, and ignores the remaining smooth substructure. We discuss this further after the proof.

Theorem 4.8 (Partial Smoothness and PDGs) If a regular function \( f \) and a manifold \( \mathcal{M} \) satisfy the sharpness condition (Definition 2.1 (iii)) at the point \( \bar{x} \), then \( f \) has a PDG structure at \( \bar{x} \) relative to \( P = \{ \bar{x} \} \). Moreover the PDG structure has a dual feasible index set.

In particular, any partly smooth function has a dual feasible PDG structure.

Proof: First select a subgradient to the function \( f \) at the point \( \bar{x} \), such that \( \bar{w} \in \text{rint} \ \partial f(\bar{x}) \). Then define the affine function \( f_0(x) := \langle \bar{w}, x - \bar{x} \rangle + f(\bar{x}) \). Note that \( f_0 \) satisfies \( f_0(\bar{x}) = f(\bar{x}) \) and \( \nabla f_0(\bar{x}) = \bar{w} \).

As \( \mathcal{M} \) is a manifold about \( \bar{x} \), there exists functions \( \{ \phi_j \}_{j=1}^m \) such that \( \mathcal{M} := \{ x : \phi_j(x) = 0 \} \), and the set \( \{ \nabla \phi_j(\bar{x}) \}_{j=1}^m \) is linearly independent.

For our PDG structure we shall select the set \( P = \{ \bar{x} \} \), the primal function \( f_0 \), and the collection of dual functions \( \{ \phi_j \}_{j=1}^m \cup \{ \phi_{m+1} := | \cdot - \bar{x} |^2 \} \). Thus, we have only to create an appropriate \( \Delta \) set.
By Lemma 2.3 the sharpness condition implies that
\[
\partial f(\bar{x}) = N_M(\bar{x}) = \left\{ \sum_{j=1}^{m} \lambda_j \nabla \phi_j(\bar{x}) : \lambda_j \in \mathbb{R} \right\}. \tag{4.2}
\]
Since \( \nabla f_0(\bar{x}) \in \text{rint} \, \partial f(\bar{x}) \), we have \( \partial f(\bar{x}) = R_+(\partial f(\bar{x}) - \nabla f_0(\bar{x})) \). Therefore there exists a closed convex set \( \Delta_B \) such that
\[
\partial f(\bar{x}) - \nabla f_0(\bar{x}) = \left\{ \sum_{j=1}^{m} \beta_j \nabla \phi_j(\bar{x}) : \beta \in \Delta_B \right\}. \tag{4.3}
\]
Since \( \nabla \phi_{m+1}(\bar{x}) = 0 \) we may add any multiple of \( \nabla \phi_m(\bar{x}) \) without altering equation (4.3). That is
\[
\partial f(\bar{x}) = \{ \nabla f_0(\bar{x}) + \sum_{j=1}^{m} \beta_j \nabla \phi_j(\bar{x}) + \gamma \nabla \phi_{m+1}(\bar{x}) : \beta \in \Delta_B, \gamma \in \mathbb{R} \}.
\]
We therefore define \( \Delta := \{(1, \beta, \gamma) : \beta \in \Delta_B, \gamma \in \mathbb{R} \} \).

It is clear that the above satisfies parts (i), (ii), (iii) and (v) of Definition 4.1. To see (iv) simply note that \( \nabla \phi_{m+1}(\bar{x}) + \nabla f_0(\bar{x}) \in \partial f(\bar{x}) \), while equation (4.2) and the fact \( \nabla f_0(\bar{x}) \in \text{rint} \, \partial f(\bar{x}) \) imply that for any \( j \in \{1, 2, \ldots, m\} \) that \( \lambda \nabla \phi_j(\bar{x}) + \nabla f_0(\bar{x}) \in \partial f(\bar{x}) \) for some value \( \lambda \neq 0 \).

Lastly note that \( K_f = \{0\} \) and \( K_\phi = \{1, 2, \ldots, m\} \) provides a basic index set which is dual feasible.

Although Theorem 4.8 answers the question of when partly smooth functions have PDG structures, it is not a very satisfying answer. The main problem is that we select \( P \) to be the singleton \{\( \bar{x} \)\}. By doing this the only property of partial smoothness we actually need is sharpness relative to a manifold. The magnitude of this problem becomes clear in the following corollary.

**Corollary 4.9 (Regularity and PDGs)** *Any lsc function that is both regular at a point \( \bar{x} \) and has nonempty subdifferential there has a dual feasible PDG structure at \( \bar{x} \).*

*Thus all lower-C\(^2\) and strongly amenable functions have dual feasible PDG structures.*

**Proof:** Select any subgradient \( w \in \text{rint} \, \partial f(\bar{x}) \). Define the manifold \( M \) as the affine subspace \( \bar{x} + N_{\partial f(\bar{x})}(w) \). Then \( M \) satisfies
\[
N_M(\bar{x}) = T_{\partial f(\bar{x})}(w).
\]
Thus \( f \) is sharp relative to \( M \) at \( \bar{x} \).

Lower-C\(^2\) functions (and strongly amenable function) satisfy the requirements of this result by [RW98, Ex 10.26 & Thm 10.31].
The question one would really like to answer is: when can a PDG structure be created having the set $P$ the same as a given manifold, $\mathcal{M}$? Equivalently, when does a partly smooth function have a primal dual feasible PDG structure? This question appears much harder, as conditions ensuring the set $\Delta_B$ (from the proof of Theorem 4.8) remains constant while $x$ moves along $\mathcal{M}$ are not transparent.

### 4.2 $g \circ F$ Decompositions

Shapiro began studying $g \circ F$ decompositions in the form of cone-reducible sets [BS00, Def 3.135], a subset of the class of reducible sets which were developed to study constrained optimization problems. In [Sha03] Shapiro defined $g \circ F$ decomposable functions and demonstrated a one-to-one correspondence between them and cone-reducible sets [Sha03, Ex 2.4, and Ex 2.5]. We begin this section with the definition of $g \circ F$ decomposable functions.

**Definition 4.10 ($g \circ F$ Decomposable Functions)** A function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is $g \circ F$ decomposable at a point $\bar{x} \in \text{dom } f$ if $f$ can locally be written as $f(\bar{x}) + g(F(\bar{x}))$ such that the functions $F : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \bar{\mathbb{R}}$ satisfy:

(i) $F$ is $C^2$ with $F(\bar{x}) = 0$ and

(ii) $g$ is sublinear (i.e. $g$ is convex with $g(0) = 0$ and $g(\lambda x) = \lambda g(x)$ for all $x \in \mathbb{R}^n$, $\lambda > 0$).

Similar to amenability, a $g \circ F$ decomposable function is the composition of a convex function with a smooth function. The class differs in that it removes the amenability condition (see Definition 1.2), and forces the convex function to be sublinear. Shapiro describes the class of $g \circ F$ decomposable functions as "somewhat between the classes of amenable and fully amenable functions" [Sha03, p. 1]. He goes on to state that "the function $f(\cdot)$ is real valued . . . and hence $g(\cdot)$ is real valued" is a reasonable simplifying assumption for most of his research [Sha03, Remark 1.1]. The next proposition and its connected example explore this assumption.

**Proposition 4.11 (Finite $g \circ F$ Decompositions)** Let the sublinear function $g$ and the smooth function $F$ provide a $g \circ F$ decomposition for the function $f$ at the point $\bar{x}$. Suppose $g$ is finite valued near $\bar{x}$. Then $f$ is strongly amenable and lower-$C^2$ near $\bar{x}$, and prox-regular at $\bar{x}$.

**Proof:** Suppose the finite valued sublinear function $g$ and the smooth function $F$ provide a $g \circ F$ decomposition for the function $f$ at the point $\bar{x}$. This implies that $N_{\text{cl}(\text{dom } g)}(F(\bar{x})) = \{0\}$, so the amenability condition,

$$N_{\text{cl}(\text{dom } g)}(F(\bar{x})) \cap \text{nul } \nabla F(\bar{x}) = \{0\},$$
is satisfied. As $F$ is $C^2$, $f$ is strongly amenable and locally finite valued. Next, [RW98, Ex 10.36] states $f$ is lower-$C^2$ about $\bar{x}$ if and only if $f$ is strongly amenable at $\bar{x}$ and $\bar{x} \in \text{int}(\text{dom } f)$. Lastly, all strongly amenable functions are prox-regular by [RW98, Prop 13.32].

The next example shows that a function $f$ which is real valued and $g \circ F$ decomposable, but for which the $g$ function cannot be real valued.

**Example 4.12 (Non-continuous $g \circ F$ Decomposition)** Consider the well known example $g : \mathbb{R}^2 \to \mathbb{R}$ by

$$g(x, y) = \begin{cases} \frac{x^2}{y} & y > 0 \\ 0 & x = y = 0 \\ +\infty & \text{otherwise}. \end{cases}$$

It is easy to confirm $g$ is positive homogeneous. Define the set $S := \{(a, -\frac{1}{2}a^2) : a \in \mathbb{R}\}$, and notice, $g$ is twice the conjugate of the indicator function $\delta_S$:

$$g(x, y) = 2\sup_{a,b}\{ax + by - \delta_S(a, b)\}.$$ 

Therefore $g$ is convex and lsc [RW98, Thm 11.1].

Compose $g$ with the $C^2$ function

$$F : \mathbb{R}^2 \to \mathbb{R}^2$$

$$(u, v) \mapsto (u, u^2 + v^2).$$

As $F(0, 0) = (0, 0)$ the resulting function is $g \circ F$ decomposable. However the resulting function is

$$f(u, v) := g(F(u, v)) = \begin{cases} \frac{u^2}{v^4 + v^2} & (u, v) \neq (0, 0) \\ 0 & (u, v) = (0, 0), \end{cases}$$

which is not continuous at $(0, 0)$. As all lower-$C^2$ functions are continuous, $f$ cannot be lower-$C^2$. Proposition 4.11 therefore shows that no real valued $g$ can be used to create a $g \circ F$ decomposition for $f$, even though $f$ is real valued.

The remainder of this section is devoted to developing the connection between partly smooth functions and $g \circ F$ decomposable functions. A simple example shows that $g \circ F$ decomposable functions include functions that are not partly smooth. Indeed the following example shows a $g \circ F$ decomposable function that is not regular.
Example 4.13 (A Nonregular $g \circ F$ Decomposition) Let $K := \{(x, y) : y \geq |x|\}$, and consider the indicator function of $K \cup -K$, $\delta_{K \cup -K}$. As the set $K \cup -K$ is not regular at $(0, 0)$, its indicator function $\delta_{K \cup -K}$ is not regular there. Therefore, $\delta_{K \cup -K}$ is not partly smooth at $(0, 0)$. However, it is $g \circ F$ decomposable at $(0, 0)$. Indeed, selecting $g(x, y) := \delta_{K}(x, y)$ and $F(x, y) := (x^2, y^2)$ provide a decomposition.

Example 4.13 clearly shows that a function being $g \circ F$ decomposable does not imply that the function is partly smooth. However, Shapiro shows that the addition of a transversality condition to a $g \circ F$ decomposition forces the resulting function to be partly smooth. Shapiro used this transversality condition almost without exception throughout [Sha03]. Before discussing this condition we consider sublinear functions in light of partial smoothness. The next proposition shows that all sublinear functions are partly smooth at the origin, and describes their active manifolds. Before stating it we recall that the linearity space of a sublinear function $f$ is $\{x : f(x) = -f(-x)\}$.

Proposition 4.14 (Partly Smooth via Sublinearity) Sublinear functions are partly smooth at the origin relative to their linearity spaces.

Proof: This can be shown via Shapiro's work by selecting a sublinear function $g$ and considering its composition with the identity map in light of [Sha03, p. 11]. We include a direct proof here.

Let $g$ be a sublinear function, and define the subspace $U := \{x : -dg(0, x) = dg(0, -x)\}$. Applying Definition 1.7 we see that $U$ is the linearity space of $g$. Thus $g$ is linear on $U$, and therefore smooth along $U$. The regularity and nonempty subgradient condition follows from the convexity of $g$. The sharpness condition follows immediately from Theorem 2.23 part (ii).

To see sub-continuity is satisfied it suffices to show that $\partial g(0) \subseteq \partial g(u)$ for all $u \in U$. This is true, as $w \in \partial g(0)$ implies that

$$g(-u) \geq \langle w, -u \rangle \text{ for all } u \in U$$

and

$$g(y) \geq \langle w, y \rangle \text{ for all } y \in \mathbb{R}^n.$$ 

Adding these inequalities, and noting that $g(-u) = -g(u)$, yields

$$g(y) \geq g(u) + \langle w, y - u \rangle \text{ for all } y \in \mathbb{R}^n, u \in U.$$ 

This shows $w \in \partial g(u)$, so $\partial g(0) \subseteq \partial g(u)$ for all $u \in U$. 

$\blacksquare$
Proposition 4.14 shows that \( g \circ F \) decompositions are compositions of smooth functions with convex partly smooth functions. Reexamining the Chain Rule (Theorem 3.1) reminds us that the addition of a transversality condition ensures such compositions produce partly smooth functions. This result was discovered by Shapiro who showed that under the transversality condition,

\[
\text{rng } \nabla F(\bar{x}) + N_{\partial g(0)}(w) = \mathbb{R}^m,
\]

where \( w \in \text{rint } \partial g(0) \), \( g \circ F \) decompositions are partly smooth relative to \( \{ x : F(x) \in N_{\partial g(0)}(w) \} \) \cite[pp. 10]{Sha03}. Note that the subspace \( N_{\partial g(0)}(w) \) is exactly the \( U \) subspace of \( UV \)-decompositions (Definition 2.22), and therefore Lemma 2.23 shows that the subspace is independent of the subgradient used. Using Proposition 4.14 we reprove Shapiro's result by use of the Chain Rule.

**Theorem 4.15 (\( g \circ F \) Decompositions and Partial Smoothness)** Let the function \( f \) be \( g \circ F \) decomposable at \( \bar{x} \). Let \( U \) be the linearity space of \( g \) (see Proposition 4.14). Then the transversality condition, equation (4.4), is equivalent to

\[
\text{rng } \nabla F(\bar{x}) + T_U(F(\bar{x})) = \mathbb{R}^m,
\]

which is also equivalent to

\[
\text{nul } (\nabla F(\bar{x})^\top) \cap N_U(F(\bar{x})) = \{0\}.
\]

Furthermore, any of these conditions imply the amenability condition (equation (1.1)).

Moreover, if \( g \) and \( F \) satisfy any of these conditions then \( f \) is partly smooth at \( \bar{x} \) relative to \( \{ x : F(x) \in U \} \).

**Proof:** Theorem 2.23 shows that \( U = N_{\partial g(0)}(w) \) for any \( w \in \text{rint } \partial g(0) \). Moreover, as \( U \) is a subspace and \( F(\bar{x}) = 0, T_U(F(\bar{x})) = U \). Thus Shapiro's transversality condition is equivalent to equation (4.5). The equivalence of equation (4.5) to equation (4.6) is clear.

Notice that the linearity space, \( U \), must be a subset of \( \text{cl}(\text{dom } g) \). Therefore we have

\[
N_{\text{cl}(\text{dom } g)}(F(\bar{x})) \subseteq N_U(F(\bar{x})),
\]

showing that the transversality condition implies the amenability condition (equation (1.1)).

Theorem 3.1 shows that if \( g \) is partly smooth relative to \( M \), and \( F \) is a \( C^2 \) function such that transversality (equation (4.5)) holds, then \( g \circ F \) is partly smooth relative to \( F^{-1}(M) \). Applying Proposition 4.14 and rewriting this in terms of \( U \) yields the final result.

Since the transversality condition implies the amenability condition we immediately gain the following corollary.
Corollary 4.16 (Transversal $g \circ F$ Decompositions) Let the function $f$ be $g \circ F$ decomposable at the point $\bar{x}$, and the transversality condition (equation (4.4)) hold. Then $f$ is partly smooth, strongly amenable, and prox-regular at $\bar{x}$. Moreover the active manifold of partial smoothness is $M := F^{-1}(\text{N}_\partial g(0)(w))$, which is unique and independent of the choice of representing functions $g$ and $F$, and of the subgradient $w \in \text{rint} \partial g(0)$ used.

**Proof:** As mentioned previously, strongly amenable functions are prox-regular [RW98, Prop 13.32], and prox-regularity implies that the active manifold is unique (Corollary 3.14). The alternate description of the active manifold follows from Lemmas 2.3 and 2.23.

Corollary 4.16 shows that, at least under Shapiro's transversality condition, the study of $g \circ F$ decomposable functions is a subset of the study of amenable functions and a subset of the study of partly smooth functions. Notice that Example 2.16 is partly smooth and locally Lipschitz, but cannot be $g \circ F$ decomposable with the transversality condition, as it is not prox-regular. Therefore, this second inclusion is strict.

Corollary 4.16 also shows how the transversality condition forces some uniqueness onto the $g \circ F$ decomposition. Even though $g \circ F$ decompositions are never unique, the interaction between the $g$ and $F$ function is.

Having now compared partial smoothness to $C^p$-identifiable surfaces, fast tracks, PDG structure, and $g \circ F$ decompositions, we end this chapter with a figure describing the relationships we have investigated.

![Figure 4.1: Relating $C^p$-identifiable Surfaces, Fast Tracks, $g \circ F$ Decompositions, Partial Smoothness, and PDG Structures.](image-url)
Chapter 5

Algorithmic Results

Chapter 2 of this thesis introduced partial smoothness, then provided several examples of partial smoothness and showed that partial smoothness is an extension of two convex notions of smooth substructure. This argues partial smoothness is a natural tool in the study of nonsmooth optimization.

This argument is reinforced in Chapter 4, where partial smoothness is compared to primal-dual gradient (PDG) structures and $g \circ F$ decompositions. Although neither notion is subsumed by partial smoothness, both show strong connections. Results of Mifflin and Sagastizábal show that if a function has a well behaved PDG structure that it must be partly smooth. On the other hand, if a function is partly smooth, or indeed simply regular, Theorems 4.8 and 4.9 show that a PDG structure exists. Unfortunately the created PDG structure is not well behaved. In Section 4.2 we saw $g \circ F$ decompositions are the composition of partly smooth functions and smooth functions (Proposition 4.14). By an application of the Chain Rule we showed that a transversality condition guarantees the resulting function is partly smooth, fully amenable, and prox-regular (Corollary 4.16). These results show partial smoothness’s intimate relation to these two developments in nonsmooth optimization.

In Chapter 3 we argued partial smoothness is also a powerful tool in nonsmooth optimization, especially when coupled with prox-regularity. The chapter developed several calculus rules, including the Infimal Convolution Rule, the Smooth Projection Theorem, and some critical point analysis.

In this chapter we discuss how the results of this thesis may affect algorithm design. The first half of the chapter focuses on how the active constraint identification results of Section 3.4 behave for existing algorithms. Section 5.2 examines the identification of active manifolds when minimizing a function over a prox-regular partly smooth set. To do this we rewrite the constraint identification results in terms the projected gradient. The projected gradient was introduced by Calamai and Moré [CM87], where many of its applications in nonsmooth optimization are explored. This thesis uses the projected gradient to describe the finite identification of active constraints for two algorithmic
CHAPTER 5. ALGORITHMIC RESULTS

styles: Gradient Projection and Newton-like Methods. These results extend many results on the finite identification of active constraints ([Dun87] [CM87] [BM88] [Lem89] [Bur90] [AKK91] [Wri93] and [BM94] for example).

In Section 5.3 we consider the problem of identifying the active manifold in the unconstrained minimization of a prox-regular partly smooth function. We focus on the classical Proximal Point method. Theorem 5.14 shows that the Proximal Point method identifies the active manifold in a finite number of iterations. This extends many results on identification of active constraints via the proximal point algorithm ([Roc76] and [MS02a] for example).

The last half of Chapter 5 introduces an algorithm for determining an approximate normal cone to an oracle-based set. Section 5.4 begins by outlining the algorithm and its inspiration. It continues with some theoretical results on describing the normal cone to a set without access to first order information (Theorems 5.23 and 5.25). These results support the likely success of the algorithm. In Section 5.5 numerical testing is used to further argue the algorithm’s success.

We next provide a brief discussion on the terminology and layout of optimization algorithms.

5.1 Algorithm Design

In this section we discuss algorithm design. Our goal here is not to provide a complete overview of the subject, but instead provide a sketch. Descent methods for optimization can be broken down into two broad classes: line search methods and trust region methods. As this thesis examines only line search methods, we do not discuss trust region methods here. Readers interested in trust region methods, or wishing a larger discussion on algorithm design, should consider [HUL93a] [HUL93b] [Ber95] and [NW99].

Line search algorithms can be broken down into 5 stages: initialization, selecting a search direction, selecting a step size, updating the data, and checking the stopping criterion. We discuss each stage in turn.

Algorithms begin with an initialization stage. This is the stage in which the user must tell the algorithm what problem it is trying to solve. The user typically supplies a point where the search for a minimum begins. This point is termed $x_0$, while successive iterations are labeled $x_1$, $x_2$, and so on. Depending on the algorithm, the user may also supply several additional parameters. These parameters may describe a maximum step size or penalty term.

The next two stages, search direction and step size, are connected. In line search algorithms a search direction is chosen first, followed by the step size. The search direction represents a direction in which a decrease in function value is likely to be found (while maintaining feasibility). The step size is selected in a manner that ensures a decrease is found, and eventual convergence is guaranteed.
CHAPTER 5. ALGORITHMIC RESULTS

One classical line search algorithm is Steepest Descent. In this case the gradient (or subdifferential) of the function is examined to determine which direction causes the most rapid decrease in local function values. In the case of a smooth function this direction is the negative of the gradient vector at the iteration point, while in the nondifferentiable case the negative of the smallest subgradient vector might be used. A line search is then performed in the direction of steepest descent.

Many other methods of selecting search directions exist, some of which are mentioned in Sections 5.2 and 5.3 of this thesis. We do not discuss this further, but instead refer the reader to the collection of literature on the topic.

The selection of a good step size is crucial in algorithm design. If a step size is too small convergence may never occur. If the step is too large, the algorithm may “jump” back and forth over a minimum. (Consider for example the iterates \( x_k = (-1)^k/k \) as a sequence approaching the minimum of the function \( f(x) = x^2 \).)

One obvious way to select a step size is to try a step of unit length. This is used in Newton’s method and the Proximal Point methods where second order information is used to incorporate the step size into the selection of the search direction. In first order methods however, the search direction contains no information about good step sizes. Hence, more complicated approaches to selecting a step length are required. We discuss two such approaches: the Limited Minimization Rule [Ber95, p. 24] and the Armijo Rule [Ber95, p. 25].

**Definition 5.1 (Limited Minimization Rule)** Given a function \( f \), a point \( x \), and a search direction \( d \) the Limited Minimization Rule states: select step size \( s \) as a solution to the one-dimensional problem

\[
s \in \arg\min \{ f(x + sd) : s \in [0,1], x + sd \text{ feasible} \}.
\]

The Limited Minimization Rule is not a practical rule. Occasionally it may be possible to solve the sub-problem, \( \arg\min_{s \in [0,1]} \{ f(x + sd) \} \), but usually this is not the case. The rule’s main application is in proofs of algorithmic convergence. The more practical Armijo Rule follows.

**Definition 5.2 (Armijo Rule)** Given a smooth function \( f \), a point \( x \), a search direction \( d \), a “shrinking parameter” \( \beta \in (0,1) \), and a parameter \( \sigma > 0 \) the Armijo Rule states: determine the smallest positive integer \( m \) such that

\[
f(x) - f(x + \beta^m d) \geq -\sigma \beta^m \nabla f(x)(d) \quad \text{and} \quad f(x + \beta^m d) \text{ is feasible},
\]

then set the step size to \( s = \beta^m \).

The Armijo Rule relies on the fact that a function (locally) looks like the linear function described by its gradient. Decreasing step lengths are tested until the decrease found is reasonable compared
to the predicted decrease. For an unconstrained function, it is easy to check that the Armijo Rule must be satisfied for sufficiently large integers $m$. Typically convergence results for algorithms using the Armijo rule can be found in [NW99, Chpt 3].

Many other rules for selecting step sizes exist. Instead of the continuing discussion we refer the reader to the literature on the subject.

The final stage in any algorithm is checking if the stopping criteria is fulfilled. If it is, the algorithm stops, while otherwise the algorithm performs another iteration. Like choosing search directions and step sizes the stopping criterion can be simple or complicated. Simple criterion include, repeat for $N$ iterations (where the positive integer $N$ is fixed in the initialization stage), or repeat until successive iterations are less than a distance $\varepsilon$ apart (again, $\varepsilon > 0$ is fixed in the initialization stage). More complicated stopping criterion are beyond the scope of this thesis: interested readers should consult the literature on the topic.

5.2 Constrained Minimization

In this section we examine algorithms for constrained minimization. A convenient tool for examining convergence in constrained minimization is the projected gradient. The projected gradient is a notion combining gradient and feasibility information into a single vector. We provide the definition as written in Calamai and Moré [CM87] below.

**Definition 5.3 (Projected Gradient)** For a differentiable function $f$ and a regular set $S$ containing the point $\bar{z}$ the projected gradient of $f$ at $\bar{z}$ relative to $S$ is

$$\nabla_S f(\bar{z}) := P_T S(\bar{z})(-\nabla f(\bar{z})).$$

**Note 5.4** The projected gradient described above is occasionally called the negative projected gradient because of the negative sign inside the projection mapping. Since the positive projected gradient is seldom used in literature (and never in this thesis) we avoid this cumbersome language.

In unconstrained optimization the negative gradient is the direction of steepest descent. The projected gradient extends this idea to constrained minimization, attempting to describe the steepest descent which maintains feasibility. Of course, as we are projecting onto the tangent cone, feasibility is only maintained in a limiting sense. That is, the projected gradient may produce a direction which is not feasible, but for which some nearby direction maintains feasibility.

To understand the projected gradient it is useful to recall two pieces of information. First, for a regular set the normal and tangent cones are polar. Second, for optimality one requires the negative gradient be in the normal cone to the constraint set. The next proposition shows how the projected gradient combines these conditions to give a sort of measure of the distance to optimality.
Proposition 5.5 (Approximate First Order Conditions) If the function \( f \) is \( C^1 \) and the set \( S \) is regular at the point \( \bar{x} \in S \), then

\[
|\nabla_S f(\bar{x})| = \text{dist}(\nabla f(\bar{x}), N_S(\bar{x})),
\]

\[
\leq |v + \nabla f(\bar{x})| \text{ for all } v \in N_S(\bar{x}).
\]

In particular

\[
\nabla_S f(\bar{x}) = 0 \iff \bar{x} \text{ is a critical point for } \min_S f.
\]

Proof: Since \( S \) is regular at \( \bar{x} \), the tangent and normal cones are closed convex and polar to each other [RW98, Ex 6.24]. Thus \( |P_{T_S}(\nabla f(\bar{x}))| = \text{dist}(\nabla f(\bar{x}), N_S(\bar{x})) \) [RW98, Ex 12.22]. Furthermore,

\[
\text{dist}(\nabla f(\bar{x}), N_S(\bar{x})) = \inf\{v - (\nabla f(\bar{x})) : v \in N_S(\bar{x})\}
\]

\[
\leq |v + \nabla f(\bar{x})| \text{ for all } v \in N_S(\bar{x}).
\]

Lastly, \( \bar{x} \) is a critical point if and only if \( \text{dist}(\nabla f(\bar{x}), N_S(\bar{x})) = 0 \).

Proposition 5.5 suggests driving the projected gradient to zero might be a good way to minimize a differentiable function over a regular set. Thus, when constructing an algorithm one might demand, each step cause a reduction in the projected gradient. Unfortunately, even if the function is \( C^\infty \) and the set convex, this is not always reasonable, as shown by the following very easy example.

Example 5.6 (Projected Gradient Nonzero) Consider the \( C^\infty \) function \( f(x, y) := y \) and the convex set \( S := \{(x, y) : y \geq |x|\} \). Then the minimum of \( f \) restricted to \( C \) is 0 and is obtained only at \((0,0)\). The projected gradient is

\[
\nabla_C f(x, y) := \begin{cases} 
(0, -1) & y > |x| \\
\left(\frac{1}{2}, \frac{-1}{2}\right) & y = -x \neq 0 \\
\left(-\frac{1}{2}, \frac{-1}{2}\right) & y = x \neq 0 \\
(0, 0) & y = x = 0.
\end{cases}
\]

Thus the projected gradient is bounded away from zero except at the optimal point \((0,0)\).

The main use of projected gradients in this work arises from rewriting Theorem 3.18 in the language of projected gradients.

Theorem 5.7 (Identification with Constraints) Consider a \( C^1 \) function \( f \). Let the set \( S \) be \( C^p \)-partly smooth \((p \geq 2)\) at the point \( \bar{x} \) relative to the manifold \( M \), and prox-regular there. Suppose \( x_k \to \bar{x} \) and \( -\nabla f(\bar{x}) \in \text{rint} N_S(\bar{x}) \). Then \( x_k \in M \) for all large \( k \) if and only if \( |\nabla_S f(x_k)| \to 0 \).
5.2.1 Gradient Projection Methods

The Gradient Projection method attempts to use the idea of steepest descent in a constrained minimization setting. By solving a quadratic sub-problem (a projection onto the constraint set) at each step, feasibility is maintained. The method is most effective if this sub-problem is easy to solve, such as in the case of a linearly defined constraint set. The algorithm is outlined next.

Conceptual Algorithm: [Gradient Projection] Consider the constrained minimization problem

\[(P) \quad \min\{f(x) : x \in S\},\]

where \(f\) is a \(C^1\) function and \(S\) a regular set. Perform the following:

**I. Initialize:** Set the iteration counter \(k\) to 0 and select an initial point \(x_0\).

**II. Search Direction:**

Set \(d_k := -\nabla f(x_k)\).

(The true search direction, asymptotically, is the projection of \(d_k\) onto the tangent cone to \(S\) at \(x_k\).)

**III. Step Size:**

Select \(s_k\) by examining \(f(P_S(x_k + s_k d_k))\).

**IV. Update:**

Set \(x_{k+1} := P_S(x_k + s_k d_k)\), then increase \(k\) by one.

**V. Repeat:** From II until the stopping qualification is satisfied.

Approaches to selecting step sizes and convergence for the Gradient Projection method are well researched. For a general overview see [Ber95] or [NW99]. In this work however, we are only concerned with conditions which ensure the gradient projection method will identify the active manifold of the constraint set in a finite number of iterations. As such, we apply the assumption that the method converges and consider what additional assumptions ensure finite identification of the active manifold. Finite identification results for the Gradient Projection method are also well established, though only in the convex setting. This is discussed in Subsection 1.2.2 of this thesis. Theorem 5.8 subsumes many previous results on constraint identification, and broadens them to a nonconvex setting.
Theorem 5.8 (Gradient Projection Identifies Active Manifolds) Suppose $f$ is a $C^1$ function and $S$ is regular set. Suppose the Gradient Projection method is used to create iterates \( \{x_k\} \) which converge to $\bar{x}$. If the step size $s_k$ satisfies $\liminf_k s_k > 0$, then $|\nabla_S f(x_k)| \to 0$.

In this case, if $S$ is partly smooth at $\bar{x}$ relative to a manifold $M$ and $-\nabla f(\bar{x}) \in \text{rint} N_S(\bar{x})$ then $x_k \in M$ for all large $k$.

Proof: Applying the formula $x - P_S(x) \in N_S(P_S(x))$ at the iteration point $x_{k+1}$ immediately yields

\[
(x_k - s_k \nabla f(x_k)) - x_{k+1} \in N_S(x_{k+1}).
\]

Since the normal cone is a cone, this tells us

\[
\frac{1}{s_k}(x_k - x_{k+1}) - \nabla f(x_k) \in N_S(x_{k+1}).
\]

Proposition 5.5 therefore yields

\[
|\nabla_S f(x_{k+1})| \leq \frac{1}{s_k}|x_k - x_{k+1} - \nabla f(x_k) + \nabla f(x_{k+1})| \\
\leq \frac{1}{s_k}|x_k - x_{k+1}| + |\nabla f(x_{k+1}) - \nabla f(x_k)|.
\]

As $x_k$ converges, $f \in C^1$, and $s_k$ is bounded below, the right hand side converges to 0. Thus $|\nabla_S f(x_k)| \to 0$. Applying Theorem 5.7 completes the proof.

Theorem 5.8 makes the strong assumption that the step size in the Gradient Projection method to be bounded below. This condition is also assumed by Bertsekas [Ber76, eq (3)], Dunn [Dun87, eq (1d), (note $\delta$ is fixed], Calamai and Moré [CM87, eq (2.2)], and Wright [Wri93, eq (12, iv)], though they assume the condition in the construction of the algorithm and not in the statements of their results.

We end our discussion on the Gradient Projection method with a brief note on the Subgradient Projection method.

Note 5.9 [Subgradient Projection Method] When $f$ is not $C^1$ a similar method called the "Subgradient Projection" method can be used to solve the problem. Essentially, the only change is to replace the gradient with the projection of zero onto the subdifferential (that is $-\nabla f$ becomes $-P_{\partial f}(0)$) [Flåd92]. This mimics steepest descent for a nondifferentiable function in a constrained minimization setting.

In general, convergence analysis for the Gradient Projection and Subgradient Projection methods are similar. However, Theorem 5.7 of this thesis cannot be applied to get finite identification results for the Subgradient Projection method, as it requires the function $f$ to be differentiable. It seems likely that a more direct approach via Theorem 3.16 of this thesis could be used to show finite constraint identification for the Subgradient Projection method. We defer this to future work.
5.2.2 Newton-like Methods

Newton's method for minimizing a function works by attempting to find the zeros of the function's derivative. This is done by the classical Newton zero finding method. As this is a local method, it is important that a good initial point is employed.

Since Newton's method makes use of second order information, it is often faster than first order methods (such as Gradient Projection). However, this second order information can be difficult to obtain, so the gain in speed may be lost in computational difficulty. To counter this, Newton-like methods replace the Hessian information required by an approximation. Newton-like methods are extensively studied in optimization. A good overview can be found in [NW99].

Because of the approximate Hessian used, results regarding finite identification of active constraints seem to be more difficult to obtain for Newton-like methods. These results are discussed in Subsection 1.2.2 of this thesis. Newton-like methods are outlined next.

Conceptual Algorithm: [Newton-like] Consider the constrained minimization problem

\[(P) \quad \min\{f(x) : x \in S\},\]

where \(f\) is a \(C^1\) function and \(S\) is a regular set. Perform the following:

I. Initialize: Set the iteration counter \(k\) to 0 and select

an initial point \(x_0\),

a positive semi-definite matrix \(H_0\).

II. Search Direction:

Find \(\tilde{x}_k \in \{x \in \arg\min \{\nabla f(x_k)(x - x_k) + \frac{1}{2}\langle x - x_k, H_k(x - x_k) \rangle : x \in S\},\]

set \(d_k = \tilde{x}_k - x_k\).

III. Step size:

Select \(s_k\) by examining \(f(x_k + s_kd_k)\).

IV. Update:

Set \(x_{k+1} = x_k + s_kd_k\),

select the next positive semi-definite matrix \(H_{k+1}\), then

increase \(k\) by one.

V. Repeat: From II until the stopping qualification is satisfied.
As in the Gradient Projection method, approaches to selecting step sizes and general convergence results are beyond the scope of this thesis. Good overviews can be found in [Ber95] and [NW99]. As before, we assume the method converges and consider what additional assumptions ensure the method identifies active manifolds in a finite number of iterations.

The Newton-like method also leaves the choice of the estimate matrices \((H_k)\) out of its description. How to choose these matrices is the focus of much research in Newton-like methods, and beyond the scope of this thesis. However, it is worth mentioning some of the classical approaches to selecting these matrices. If \(H_k = 0\) the “Conditional Gradient” method results, if \(H_k = \nabla^2 f(x_k)\) “Newton’s” method results, while when \(H_k\) approximates \(\nabla^2 f(x_k)\) and a secant equation holds “Quasi-Newton” methods result.

The next theorem makes use of the projected gradient to show when Newton-like methods identify the active manifold in a finite number of iterations.

**Theorem 5.10 (Newton-like Methods Identify Active Manifolds)** Suppose \(f\) is a \(C^1\) function and the set \(S\) is regular. Suppose a Newton-like method is used to create a sequence of iterates \(x_k\) which converge to \(\bar{x}\). If eventually the step size \(s_k\) is always 1, and the matrices \(H_k\) are bounded then \(|\nabla Sf(x_k)| \to 0\).

In this case, if \(S\) is partly smooth at \(\bar{x}\) relative to a manifold \(M\) and \(-\nabla f(\bar{x}) \in \text{rint} \, N_S(\bar{x})\) then \(x_k \in M\) for all large \(k\).

**Proof:** Without loss of generality we will assume the step size is one for all \(k\).

Define the function

\[
q_k(x) := \nabla f(x_k)(x - x_k) + \frac{1}{2}(x - x_k, H_k(x - x_k)).
\]

Note that the algorithm begins by finding \(\tilde{x}_k\) a minimizer of \(q_k\) over \(S\). Thus \(-\nabla q_k(\tilde{x}_k) \in N_S(\tilde{x}_k)\).

By Proposition 5.5 this yields

\[
|\nabla Sf(\tilde{x}_k)| \leq |\nabla f(\tilde{x}_k) - \nabla q_k(\tilde{x}_k)| = |\nabla f(\tilde{x}_k) - \nabla f(x_k) - H_k(\tilde{x}_k - x_k)| \leq |\nabla f(\tilde{x}_k) - \nabla f(x_k)| + ||H_k|| |\tilde{x}_k - x_k|.
\]

Since the step size is always one, \(x_{k+1} = \tilde{x}_k\), thus

\[
|\nabla Sf(x_{k+1})| \leq |\nabla f(x_{k+1}) - \nabla f(x_k)| + ||H_k|| |x_{k+1} - x_k|.
\]

(5.1)

As \(H_k\) is bounded, \(f \in C^1\), and \(x_k\) converges we must have the right hand side converge to zero. Thus \(|\nabla Sf(x_{k+1})| \to 0\). Theorem 5.7 completes the proof.
CHAPTER 5. ALGORITHMIC RESULTS

Theorem 5.10 uses a much bolder condition to ensure finite identification of active manifolds than Theorem 5.8 does. Specifically, in Theorem 5.8 we demand only that the step size is bounded below, while for Newton-like methods we demand the step size is eventually always one. This condition was also used by Burke and Moré in [BMM, Thm 4.1], though they do not explicitly mention it (instead they simply remove $\alpha_k$ from the statement of the theorem). Our second constraint on the algorithm (that the approximate Hessians be bounded) is also used by Burke and Moré to achieve their finite convergence results. Conditions on the algorithm or problem that ensure these requirements are fulfilled are beyond the scope of this thesis.

5.3 Unconstrained Minimization

The previous section focused on minimizing differentiable functions over prox-regular partly smooth constraint sets, and the finite identification of the active manifold which results. In this section we turn our attention to minimizing prox-regular partly smooth functions with no constraint set. We make use of Theorem 3.21 to show that the Proximal Point method identifies the active manifold of the function being minimized.

5.3.1 Proximal Point Methods

To understand the Proximal Point algorithm we must first discuss the proximal envelope. The definition of the proximal envelope is usually attributed to Moreau, and therefore often referred to as the Moreau envelope. We state its definition below.

**Definition 5.11 (Proximal Point Mapping)** For a function $f$ and a parameter value $R > 0$ the proximal envelope and its related proximal point mapping are defined by

$$e_R(x) := \inf_y \{f(y) + \frac{R}{2} |x - y|^2\}$$

$$P_R(x) := \arg\min_y \{f(y) + \frac{R}{2} |x - y|^2\},$$

respectively.

Closely related to the proximal envelope and proximal point map is the threshold of prox-boundedness, which describes for which $R$ the proximal envelope is well defined.

**Definition 5.12 (Prox-bounded)** A function $f$ is prox-bounded if there exists some point $x$ and scalar $R > 0$ such that

$$e_R(x) > -\infty.$$  

The smallest such $R$ is the threshold of prox-boundedness for $f$. 
The use of the proximal point map in optimization was first discussed by Martinet in [Mar70, Sec 4]. Martinet expanded his work in [Mar72] which focused almost entirely on the proximal point algorithm. The algorithm has gained much interest since; good overviews can be found in [Lem89], [HUL93b] and [Ber95]. We outline the complete algorithm next.

Conceptual Algorithm: [Proximal Point] Consider the unconstrained minimization problem

\[(P) \quad \min_x \{f(x)\}.\]

Perform the following:

I. **Initialize:** Set the iteration counter \(k\) to 0 and select

an initial point \(x_0\),

and a penalty parameter \(R > 0\).

II. **Update:**

Set \(x_{k+1} \in P_R(x_k) := \arg\min_y \{f(y) + \frac{R}{2} |y - x_k|^2\}\), then

increase \(k\) by one.

III. **Repeat:** From II until the stopping qualification is satisfied.

The proximal point algorithm forces the search direction and step size stages to be combined into the update stage. The algorithm could be rewritten to include search direction and step size stages by setting \(d_k = P_R(x_k) - x_k\) and \(s_k = 1\), then updating \(x_{k+1} = x_k + s_k d_k\).

The first finite identification results for the proximal point algorithm appeared shortly after its introduction. In [Roc76, Thm 3], Rockafellar showed that if zero was in the interior of the subdifferential for a convex function then the solution was finitely identified. He furthered his work by showing finite identification results for the Proximal Point algorithm when applied to polyhedral functions [Roc76, Prop 8].

More recent results on the finite identification of active manifolds via the proximal point algorithm appear in Mifflin and Sagastizábal's work with fast tracks [MS02a]. Here they show that the Proximal Point algorithm will finitely identify the fast track on which the optimal point lies.

We seek to generalize these results to a prox-regular partly smooth setting. To do this we will require a result of Poliquin and Rockafellar on the proximal mapping of a prox-regular function.

**Lemma 5.13** [PR96a, Thm 2.3] *Suppose the function \(f\) is prox-regular at \(\bar{x}\) for the subgradient \(\bar{v} = 0 \in \partial f(\bar{x})\), and prox-bounded. If \(R > 0\) is sufficiently large then,*
(i) the proximal envelope $e_R$ is $C^1$ with $e_R(\bar{x}) = f(\bar{x})$ and,

(ii) the proximal point mapping $P_R$ is single valued, continuous, and satisfies $P_R(\bar{x}) = \{\bar{x}\}$.

It is worth noting that 'R sufficiently large' means $R$ must be greater than both the constant of prox-regularity and the threshold of prox-boundedness. For a convex function this reduces to $R > 0$.

By applying Lemma 5.13 to Theorem 3.21 we obtain a result on the finite identification of the active manifold via the Proximal Point algorithm. This result provides an alternate approach to the results of Rockafellar, and Mifflin and Sagastizábal referenced above.

Theorem 5.14 (Proximal Points Identify Active Manifolds) Suppose the function $f$ is partly smooth at the point $\bar{x}$ relative to the manifold $\mathcal{M}$. Suppose also that $f$ is prox-bounded, and prox-regular at $\bar{x}$ with $0 \in \text{rint} \, \partial f(\bar{x})$. If $R$ is sufficiently large, then the proximal point mapping of $f$ satisfies $P_R(x) \in \mathcal{M}$ for all points $x$ near $\bar{x}$.

Therefore, if the sequence of points $\{x_k\}$ is generated via the proximal point algorithm with $R$ sufficiently large, and $x_k \to \bar{x}$ then $x_k \in \mathcal{M}$ for all $k$ sufficiently large.

Proof: We begin by selecting $R$ large enough that Lemma 5.13 may be applied. Suppose the theorem is false, so there must exist a sequence of points $x_k$ converging to $\bar{x}$, such that the related proximal points $y_k = P_R(x_k)$ are never on the active manifold $\mathcal{M}$. (Lemma 5.13 allows us to assume these points are unique.)

By Lemma 5.13 (ii) we know $y_k \to \bar{x}$. This shows that $\frac{R}{2} |y_k - x_k|^2 \to 0$. Therefore $e_R(x_k) = f(y_k) + \frac{R}{2} |y_k - x_k|^2$ combined with Lemma 5.13 (i) shows that $f(y_k) \to f(\bar{x})$.

Next notice, as $y_k \in \text{argmin} \{f(y) + \frac{R}{2} |y - x_k|^2\}$, we must have $0 \in \partial (f(\cdot) + \frac{R}{2} |\cdot - x_k|^2)(y_k)$ for each $k$. This simplifies to

$$0 \in \partial f(y_k) + R(y_k - x_k),$$

showing that $\text{dist}(0, \partial f(y_k)) \leq R|y_k - x_k|$. The righthand side converges to 0, so Theorem 3.21 shows that $y_k \in \mathcal{M}$ for all $k$ large.

As in Lemma 5.13 for a convex function, 'R sufficiently large' reduces to $R > 0$. With this in mind, by applying Theorem 5.14 to Example 2.5, Rockafellar's original result on the Proximal Point method for strict critical points of convex functions ([Roc76, Thm 3]) is recreated (provided some iterate is sufficiently close to the solution). As polyhedral functions are partly smooth, Theorem 5.14 also yields Rockafellar's results for the polyhedral case ([Roc76, Prop 9]), (again provided some iterate is sufficiently close to the active manifold). Lastly, Section 2.5 shows that Theorem 5.14 subsumes Mifflin and Sagastizábal's work on fast tracks.
CHAPTER 5. ALGORITHMIC RESULTS

Notice that the nondegeneracy condition appears in Theorem 5.14, though this time in the functional form. The next example shows that this condition cannot be removed.

Example 5.15 (Necessity of $0 \in \text{rint} \partial f(\bar{x})$) Consider the function $f(x) = \max\{0, x\}$ at the point $\bar{x} = \{0\}$ and the manifold $\mathcal{M} = \{0\}$. It is easy to verify $f$, $\bar{x}$, and $\mathcal{M}$ satisfy all conditions of Theorem 5.14 except $0 \in \text{rint} \partial f(\bar{x})$. Examining the proximal points for $f$ at any point $x < 0$ we see that

$$P_R(x) = \arg\min_y \{ f(y) + \frac{R}{2} |x - y|^2 \} = \{x\}.$$

Thus the proximal points do not identify the manifold $\mathcal{M} = \{0\}$. 

The above example may seem discouraging, but it is not difficult to compensate for. The next theorem shows that a slight tilt to the problem corrects for such an occurrence.

Corollary 5.16 (Tilted Proximal Points) Suppose the function $f$ is partly smooth at the point $\bar{x}$ relative to the manifold $\mathcal{M}$. Suppose also that $f$ is prox-bounded, and prox-regular at $\bar{x}$. Select any subgradient $w \in \text{rint} \partial f(\bar{x})$. Define $h(x) = f(x) - (w, x)$. Then for $R$ sufficiently large and any point $x$ near $\bar{x}$ the proximal point of $h$ at $x$ satisfies $P_R(x) = \arg\min_y \{ h(y) + \frac{R}{2} |x - y|^2 \} \in \mathcal{M}$.

Proof: Apply Theorem 5.14 to $h$ noting that shifting by a linear function does not effect prox-regularity [RW98, Ex 13.35], prox-boundedness (see definition), or partial smoothness (Corollary 3.3).

Note 5.17 (The Proximal Bundle Method) In many ways the Proximal Point method is a conceptual algorithm rather than a practical one. The problem is that calculating the proximal point mapping can be equally difficult as solving the original problem. It is natural to consider methods of simplifying the proximal point mapping. This is typically done by finding the proximal points of an approximation of the original function. By updating and refining the approximation we can ensure convergence. This leads to the "Proximal Bundle" method. Good overviews to this method can be found in [HUL93b, XV.3] and [Kiw89].

Much of the theory for Proximal Bundles appears the same as for proximal points. However, to the author's knowledge, no results on the finite identification of active manifolds have been published for the Proximal Bundle method. The difficulty in finding such a result can be seen by considering the functions $f_k(x) := \max\{x, -x, \frac{1}{k}\} \ (k = 1, 2, ...) \text{ as approximations to the convex function } f(x) := |x|$. Notice $f$ is partly smooth at 0 relative to the manifold $\{0\}$, and $0 \in \text{int} \partial f(0)$. We employ the
**CHAPTER 5. ALGORITHMIC RESULTS**

Proximal Bundle method by setting the initial iterate $x_1 = 1$, fixing $R = 1$, and then finding the subsequent iterates by calculating the proximal point for $f_{k+1}$. This yields

$$x_{k+1} = P_{I_{k+1}} \left( \frac{1}{k} \right) = \arg\min_y \left\{ f_{k+1}(y) + \frac{1}{2} \left\| y - \frac{1}{k} \right\|^2 \right\} = \frac{1}{k+1}$$

($k = 1, 2, \ldots$), a sequence which never identifies the active manifold of partial smoothness. Thus, any finite identification results on the Proximal Bundle method would require a discussion on 'good' methods of selecting the approximation functions $f_k$.

### 5.4 Random Normal Minimization

Until now this thesis focused mainly on partly smooth functions. Adding prox-regularity allowed us to create the active manifold identification results of Section 3.4. The previous two sections of this chapter then showed that the conditions for the results of Section 3.4 were accessible in practice. This concludes our examination of partly smooth functions in nonsmooth optimization.

The next two sections of this thesis turn their attention to nonsmooth optimization over oracle-based sets. An oracle-based set is a set which is defined by a binary function to whose structure the user does not have access. The user may ask the function whether a point is in the set, but may not examine or manipulate the function mathematically. Thus, the user can accurately determine what points are feasible, but determining the exact normal cone to the set, for example, is impossible. Since the normal cone is a basic tool of nonsmooth optimization our task will be to explore methods of generating approximate normal cones to oracle-based sets.

In a practical sense oracle-based sets are of interest for several reasons. First they eliminate the need for the optimizer to know advanced mathematics (such as how to calculate the normal cone to a set). They also arise in natural circumstances. For example, companies may not wish to divulge data to the optimizer, or feasibility may be determined experimentally. In the second example, checking whether a point is feasible or infeasible could be a lengthy operation. This suggests that methods of creating approximate normal cones should ideally require few set evaluations.

From a practical point of view the generated normal cone should also be "robust". That is, it should pick up normal directions from neighbouring points. When optimizing over a set with interior, a given iteration, for any practical algorithm maintaining feasibility, will almost always be in the interior of the set. Thus the correct normal cone would simply be the zero vector. This however does not tell us much about promising search directions. A robust estimate for the normal cone could improve an algorithm's convergence.

One recent work by Burke, Lewis and Overton suggests an approach for dealing with oracles. In [BLO02], Burke, Lewis and Overton discuss how to approximate the subdifferential map for a function using randomly selected gradients. In applied optimization most functions are differentiable.
almost everywhere. In such circumstances the subdifferential map can often be created as the convex closure of all limiting gradients [BLO02, eq (1.1)]. By randomly selecting nearby points, evaluating the gradients of the function at those points, and then taking the convex hull, Burke, Lewis and Overton create a reasonable approximation of the subdifferential of the function.

We attempt to emulate the gradient sampling algorithm of Burke, Lewis and Overton by creating a randomly generated approximate normal cone. By randomly sampling nearby points an estimate for how the constraint set locally appears can be created. Projecting the points which lie outside the set onto the estimate of the set creates a collection of "approximate normal vectors" (recall the formula \( x - P_S(x) \in N_S(P_S(x)) \)). To add robustness, only normal vectors of a minimum length are included in the approximate normal cone. We outline the proposed algorithm next.

**Conceptual Algorithm: [Random Normal Generation (RNG)]** Given an oracle-based set \( S \), a feasible point \( \bar{x} \), a search radius \( \varepsilon > 0 \), a minimum length parameter \( \mu \in [0, 1) \) and an integer \( n \), perform the following:

I. **Selection**: Pick \( n \) random points within \( \varepsilon \) of \( \bar{x} \), \( \{\bar{x}_i\}_{i=1}^n \).

II. **Organization**: Using the oracle, organize these points into two sets:

\[
P_{in} := \{\bar{x}_i : \bar{x}_i \in S\} \quad \text{and} \quad P_{out} := \{\bar{x}_i : \bar{x}_i \notin S\}.
\]

III. **Convexification**: Take the convex hull of \( P_{in} \cup \bar{x} \) to create an estimate of \( S \):

\[
\tilde{S} := \text{conv} (P_{in} \cup \bar{x}).
\]

IV. **Projection**: Project each point in \( P_{out} \) onto \( \tilde{S} \) to create the approximate normal cone:

\[
\tilde{N} := \mathbb{R}_+ (\text{conv} \{\bar{x}_i - P_{\tilde{S}}(\bar{x}_i) : \bar{x}_i \in P_{out}, |\bar{x}_i - P_{\tilde{S}}(\bar{x}_i)| > \mu \varepsilon\}).
\]

In Section 5.5 we perform some numerical testing on this algorithm, while the remainder of this section examines the question of describing normal cones without knowledge of how the underlying set is constructed. To do this we consider the following two cones.

**Definition 5.18 (Near Normals and Tangents)** Let \( C \) be a closed convex set containing the point \( \bar{x} \). For parameters \( \varepsilon > 0 \) and \( \mu \in [0, 1) \), we define the near normal cone and near tangent cone at \( \bar{x} \) as

\[
N_{C \varepsilon \mu}(\bar{x}) := \{\lambda (x - P_C(x)) : x \in B_{\varepsilon}(\bar{x}) \text{ with } |x - P_C(x)| \geq \mu \varepsilon, \lambda \geq 0\} \cup \{0\}, \quad (5.4)
\]

and

\[
T_{C \varepsilon \mu}(\bar{x}) := \{d : \langle d, x - P_C(x) \rangle \leq 0 \text{ for all } x \in B_{\varepsilon}(\bar{x}) \text{ with } |x - P_C(x)| \geq \mu \varepsilon\}. \quad (5.5)
\]

When \( \bar{x} \), \( C \) and \( \mu \) are apparent we shall simplify notation to \( N_\varepsilon \) and \( T_\varepsilon \).
CHAPTER 5. ALGORITHMIC RESULTS

Note 5.19 When defining the near normal cone the inclusion of the element 0 is necessary to ensure this set is nonempty. (Consider for example the near normal cone of $\mathbb{R}^n$).

To compare Definition 5.18 with RNG, consider a set $S$ containing a point $\bar{x}$, a search radius $\varepsilon > 0$, and a minimum length parameter $\mu \in [0, 1)$. With these parameters, the near normal cone to the convex set

$$S_\varepsilon := \text{cl conv} \{B_\varepsilon(\bar{x}) \cap S\}$$

(5.6)

is the set which would result if our sampling in step I of RNG filled out the entire ball $B_\varepsilon(\bar{x})$. The near tangent cone is then the polar to the near normal cone. Of course for practical purposes this is impossible, but if RNG has any hope of success we must hope the near normal and near tangent cones converge to the correct normal and tangent cones as $\varepsilon \searrow 0$. The goal for the remainder of this section is to determine when this occurs.

We begin with a useful lemma.

Lemma 5.20 (Creating Normals) Let the set $S$ be regular at the point 0. Select a normal vector $\bar{n} \in N_S(0)$ and define, for each scalar $\varepsilon > 0$, the set $S_\varepsilon := \text{cl conv} \{B_\varepsilon(0) \cap S\}$ and the vector

$$\bar{n}_\varepsilon := \varepsilon \bar{n} - P_{S_\varepsilon}(\varepsilon \bar{n}).$$

Then

$$\lim_{\varepsilon \searrow 0} \bar{n}_\varepsilon = \bar{n}.$$

Proof: As $S$ is regular at 0, $T_S(0)$ and $N_S(0)$ are mutually polar [RW98, Cor 6.30]. Therefore $\bar{n} \in N_S(0)$ implies that

$$P_{B_1(0) \cap T_S(0)}(\bar{n}) = 0.$$ 

The regularity of $S$ at 0 also yields

$$\lim_{\varepsilon \searrow 0} (B_1(0) \cap \varepsilon^{-1}S) = B_1(0) \cap T_S(0),$$

[RW98, Prop 6.2 & Prop 4.30 (b)]. Lastly [RW98, Prop 4.9] and the knowledge $\varepsilon^{-1}P_{S_\varepsilon}(\varepsilon \bar{n}) = P_{B_1(0) \cap \varepsilon^{-1}S}(\bar{n})$ shows that

$$\lim_{\varepsilon \searrow 0} \varepsilon^{-1}P_{S_\varepsilon}(\varepsilon \bar{n}) = \lim_{\varepsilon \searrow 0} P_{B_1(0) \cap \varepsilon^{-1}S}(\bar{n}) = P_{B_1(0) \cap T_S(0)}(\bar{n}) = 0.$$ 

Hence $\lim_{\varepsilon \searrow 0} \frac{\bar{n}_\varepsilon}{\varepsilon} = \bar{n}$ as required.

Our next theorem shows that the approximate tangent directions created by RNG will likely maintain asymptotic feasibility.
Chapter 5. Algorithmic Results

Theorem 5.21 (Valid Search Directions) Let the set $S$ be regular at the point $\bar{x}$ and fix a parameter $\mu \in [0,1)$. Define the set $S_\varepsilon := \text{cl conv}\{B_\varepsilon(\bar{x}) \cap S\}$. Then the near normal and near tangent cones to $S_\varepsilon$ satisfy

$$\liminf_{\varepsilon \to 0} N_{S_\varepsilon}^{\varepsilon,\mu}(\bar{x}) \supseteq N_S(\bar{x})$$

(5.7)

and

$$\limsup_{\varepsilon \to 0} T_{S_\varepsilon}^{\varepsilon,\mu}(\bar{x}) \subseteq T_S(\bar{x})$$

(5.8)

Proof: We will use the simplified notation $N_\varepsilon$ and $T_\varepsilon$ to represent $N_{S_\varepsilon}^{\varepsilon,\mu}(\bar{x})$ and $T_{S_\varepsilon}^{\varepsilon,\mu}(\bar{x})$ respectively. Furthermore, by shifting the set $S$, we may assume the point $\bar{x} = 0$.

Consider a normal vector $\bar{n} \in N_S(0)$, a sequence of parameters $\varepsilon \searrow 0$. To show equation (5.8) we must show that any convergent sequence of near tangents $t_\varepsilon \to \bar{t}$, $t_\varepsilon \in T_\varepsilon$ satisfies $(\bar{t}, \bar{n}) \leq 0$. If $\bar{n} = 0$ we are done, therefore we only consider the case where $|\bar{n}| > 0$. Thus the vector

$$n_\varepsilon := \frac{\varepsilon}{|\bar{n}|} \bar{n} - P_{S_\varepsilon}\left(\frac{\varepsilon}{|\bar{n}|}\right)$$

is well defined. By Lemma 5.20 we know $\varepsilon^{-1} n_\varepsilon \to \bar{n}/|\bar{n}|$, which shows that eventually $|n_\varepsilon| \geq \mu \varepsilon$. Combining this with $\varepsilon^{-1} n_\varepsilon \in B_\varepsilon(0)$ and $t_\varepsilon \in T_\varepsilon$, shows that

$$\left\langle t_\varepsilon, \frac{\varepsilon}{|\bar{n}|} \bar{n} - P_{S_\varepsilon}\left(\frac{\varepsilon}{|\bar{n}|}\right) \right\rangle \leq 0.$$

Taking limits yields $(\bar{t}, \bar{n}/|\bar{n}|) \leq 0$, as required.

Examining the above we note that we have also shown that for any $\bar{n} \in N_S(0) \setminus \{0\}$ and sequence of parameters $\varepsilon \searrow 0$, the existence of a sequence of near normals $\varepsilon^{-1} n_\varepsilon \in N_\varepsilon$ converging to $\bar{n}/|\bar{n}|$. As $0 \in N_\varepsilon$ for all $\varepsilon$ equation (5.7) holds.

Theorem 5.21 requires very few conditions for its proof. Indeed, the only condition imposed is that the set be regular at the point in question. This is somewhat surprising as we are attempting to create normal cones via the projection mapping. As Lemma 1.14 shows there is an intimate connection between projection mappings, normal cones, and prox-regularity. So why is Theorem 5.21 so simple?

To begin, we are not projecting onto the set $S$, but onto a convex approximation $S_\varepsilon$. Thus prox-regularity is not needed to ensure the projection is well defined. Furthermore, Theorem 5.21 only shows an outer approximation for the normal cone. That is, if the near normal cone converges then the limit is at least as big as the normal cone. The next example shows that this limit can in fact be larger than the correct normal cone. The tangent directions created will still be feasible directions, but they will not include all tangent directions. This can be corrected by forcing the parameter $\mu$ to be strictly positive.
Example 5.22 (Necessity of \( \mu > 0 \)) Consider the set \( S := \{(x, \pm x^2) : x \in \mathbb{R}\} \) at the point \((0,0)\). This set is regular at \((0,0)\) with cone \( N_S(0,0) = \{(0,y) : y \in \mathbb{R}\} \).

For fixed \( \varepsilon > 0 \) and \( \mu = 0 \) consider the set \( S_{\varepsilon} := \text{cl conv} \{S \cap B_{\varepsilon}(0,0)\} \) and its near normal cone \( N_{S_{\varepsilon}}^{e,\mu}(0,0) \). It is easy to see that the near normal cone is the coordinate axes
\[
N_{S_{\varepsilon}}^{e,\mu}(0,0) = \{(x,y) : xy = 0\}.
\]

Figure 5.1: Effect of \( \mu \) on Random Normal Generation

This does not converge to \( N_S(0,0) \) as \( \varepsilon \searrow 0 \). However if \( \mu > 0 \), once \( \varepsilon \) is sufficiently small, the near normal cone becomes \( \{(0,y) : y \in \mathbb{R}\} \), as desired.

The above example suggests that if we want the near normal cone to converge to the correct object, we might need to select the parameter \( \mu \) to be strictly positive. In Theorem 5.23 we see this is the only addition required to ensure the near normal cone converges to the normal cone.

Theorem 5.23 (Normal Convergence) Let the set \( S \) be regular at the point \( \bar{z} \) and fix the parameter \( \mu \in (0,1) \). Define the set \( S_{\varepsilon} := \text{cl conv} \{B_{\varepsilon}(\bar{z}) \cap S\} \). Then the near normal cone to \( S_{\varepsilon} \) converges to the normal cone of \( S\):
\[
\lim_{\varepsilon \searrow 0} N_{S_{\varepsilon}}^{e,\mu}(\bar{z}) = N_S(\bar{z}).
\]

Proof: As in the proof of Theorem 5.21 we will use the simplified notation \( N_\varepsilon \) to represent \( N_{S_{\varepsilon}}^{e,\mu}(\bar{z}) \) and assume the point \( \bar{z} = 0 \).

By Theorem 5.21 we know \( \lim \inf_{\varepsilon \searrow 0} N_\varepsilon \supseteq N_S(0) \), thus we need only show that \( \lim \sup_{\varepsilon \searrow 0} N_\varepsilon \subseteq N_S(0) \). To that end, let the near normal vectors \( n_\varepsilon \in N_\varepsilon \) converge to \( \bar{n} \). If \( |\bar{n}| = 0 \) we are done, otherwise we have \( n_\varepsilon /|n_\varepsilon| \) are well defined (eventually) with \( n_\varepsilon /|n_\varepsilon| \in N_\varepsilon \) and \( n_\varepsilon /|n_\varepsilon| \rightarrow \bar{n}/|\bar{n}| \).
Therefore, for each $\varepsilon$, there exists $y_\varepsilon \in B_\varepsilon(0)$ with $|y_\varepsilon - P_{S_\varepsilon}(y_\varepsilon)| \geq \mu \varepsilon$ and

$$\frac{n_\varepsilon}{|n_\varepsilon|} = \frac{y_\varepsilon - P_{S_\varepsilon}(y_\varepsilon)}{|y_\varepsilon - P_{S_\varepsilon}(y_\varepsilon)|}.$$ 

Notice both $|y_\varepsilon|$ and $|P_{S_\varepsilon}(y_\varepsilon)|$ are bounded above by $\varepsilon$, therefore, dropping to a subsequence as necessary, we may assume $\varepsilon^{-1}y_\varepsilon \rightarrow \bar{y}$ and $\varepsilon^{-1}P_{S_\varepsilon}(y_\varepsilon) \rightarrow \bar{P}$ for some points $|\bar{y}| \leq 1$, and $|\bar{P}| \leq 1$.

We next claim $\bar{P}$ is actually equal to $P_{B_1(0) \cap T_\varepsilon(0)}(\bar{y})$. Indeed $\varepsilon^{-1}P_{S_\varepsilon}(y_\varepsilon) = P_{B_1(0) \cap T_\varepsilon(0)}(\varepsilon^{-1}y_\varepsilon)$, so [RW98, Prop 4.9] shows that

$$\varepsilon^{-1}(P_{S_\varepsilon}(y_\varepsilon) - y_\varepsilon) = P_{B_1(0) \cap T_\varepsilon(0)}(0) \rightarrow P_{B_1(0) \cap T_\varepsilon(0)}(0) = P_{B_1(0) \cap T_\varepsilon(0)}(\bar{y}) - \bar{y}.$$ 

As $y_\varepsilon \rightarrow \bar{y}$ we have $\varepsilon^{-1}P_{S_\varepsilon}(y_\varepsilon) \rightarrow P_{B_1(0) \cap T_\varepsilon(0)}(\bar{y})$ as desired.

Next, notice $\bar{y} \neq \bar{P}$ as

$$|\varepsilon^{-1}y_\varepsilon - \varepsilon^{-1}P_{S_\varepsilon}(y_\varepsilon)| = \frac{1}{\varepsilon}|y_\varepsilon - P_{S_\varepsilon}(y_\varepsilon)| \geq \mu.$$  

Since $\bar{y} \neq \bar{P}$ and both are in $B_1(0)$, the Cauchy Schwartz equation yields $\langle \bar{y}, \bar{P} \rangle < 1$. Since $\bar{P} = P_{B_1(0) \cap T_\varepsilon(0)}(\bar{y})$ we have

$$0 \leq \langle \bar{P}, \bar{y} - \bar{P} \rangle = \langle \bar{y}, \bar{P} \rangle - |\bar{P}|^2 < 1 - |\bar{P}|^2,$$

so $|\bar{P}| < 1$. Therefore $\bar{P}$ is actually the projection of $\bar{y}$ onto the tangent cone $T_\varepsilon(0)$, so $\bar{y} - \bar{P} \in N_S(0)$.

Lastly,

$$\frac{n_\varepsilon}{|n_\varepsilon|} = \frac{y_\varepsilon - P_{S_\varepsilon}(y_\varepsilon)}{|y_\varepsilon - P_{S_\varepsilon}(y_\varepsilon)|} \rightarrow \frac{\bar{y} - \bar{P}}{|\bar{y} - \bar{P}|} \quad \text{and} \quad \frac{n_\varepsilon}{|n_\varepsilon|} \rightarrow \frac{\bar{n}}{|\bar{n}|},$$

implies that

$$\frac{\bar{n}}{|\bar{n}|} = \frac{\bar{y} - \bar{P}}{|\bar{y} - \bar{P}|} \in N_S(0),$$

and the proof is complete.

Since the normal and tangent cones are polar at points where a set is regular, one might expect convergence of the near normal cone to be sufficient to ensure the near tangent cone also converges. Surprisingly, this is not the case. The next example shows that at least one more condition is required to ensure the near tangent cone converges to the tangent cone.

**Example 5.24 (Necessity of Interior Tangent Vectors)** Consider the set $S := \{(x, x^2) : x \in \mathbb{R}\}$ at the point $(0, 0)$. The tangent cone to $S$ at $(0, 0)$ is $T_S(0, 0) = \{(x, 0) : x \in \mathbb{R}\}$, which does not have interior.

For fixed $\mu \in (0, 1)$ and $\varepsilon > 0$ consider the set $S_\varepsilon := \text{cl conv} \{B_\varepsilon(0, 0) \cap S\}$ and its near tangent cone $T_{S_\varepsilon}^{\text{near}}(0, 0)$. It is easy to see that the near tangent cone is always the singleton $\{(0, 0)\}$.

This does not converge to $T_S(0, 0)$ as $\varepsilon \searrow 0$. 
Example 5.24 shows that the parameter \( p \) being strictly positive, along with the set being regular is insufficient to ensure the near tangent cone converges to the tangent cone. The problem may lie in fact the tangent cone to the set does not have interior. We end this section by showing that the additional assumption that the tangent cone has interior is enough to ensure convergence of the near tangent cone to the tangent cone.

**Theorem 5.25 (Tangent Convergence)** Suppose \( S \) is regular at the point \( \bar{x} \) and the tangent cone \( T_S(\bar{x}) \) has interior. Fix the parameter \( p \in (0,1) \). For any parameter \( \varepsilon > 0 \), define the set \( S_\varepsilon := \text{cl conv} \{ B_\varepsilon(\bar{x}) \cap S \} \). Then the near tangent cone to \( S_\varepsilon \) at \( \bar{x} \) converges to the tangent cone to \( S \) at \( \bar{x} \):

\[
\lim_{\varepsilon \to 0} T_{S_\varepsilon}^{\varepsilon}(\bar{x}) = T_S(\bar{x}).
\]

Moreover in this case the near normal cone satisfies the stronger convergence

\[
\lim_{\varepsilon \to 0} \text{conv} \, N_{S_\varepsilon}^{\varepsilon}(\bar{x}) = N_S(\bar{x}).
\]

**Proof:** As above we will use the simplified notation \( N_\varepsilon \) and \( T_\varepsilon \) and assume the point \( \bar{x} = 0 \).

The regularity of \( S \) and Theorem 5.21 show that \( \lim \sup_{\varepsilon \to 0} T_\varepsilon \subseteq T_S(0) \), so we need only show that \( \lim \inf_{\varepsilon \to 0} T_\varepsilon \supseteq T_S(0) \).

Select any \( \bar{\ell} \in \text{int} \, T_S(0) \). We claim for any \( \varepsilon \) sufficiently small that \( \bar{\ell} \in T_\varepsilon \). Suppose this is false, then there exists a sequence \( \varepsilon_k \downarrow 0 \) \( (k = 1, 2, \ldots) \) such that \( \bar{\ell} \notin T_{\varepsilon_k} \) for all \( k \). For this to be true, for each \( k \) there must exist some \( n_k \in N_{\varepsilon_k} \) such that \( |n_k| = 1 \) and \( \langle \bar{\ell}, n_k \rangle > 0 \). Taking a subsequence as necessary, we may assume \( n_k \to \bar{n} \) for some vector with \( |\bar{n}| = 1 \) and \( \langle \bar{\ell}, \bar{n} \rangle \geq 0 \). By Theorem 5.23, \( \bar{n} \in N_S(0) \). As \( \bar{\ell} \in \text{int} \, T_S(0) \), this implies that \( \langle \bar{\ell}, \bar{n} \rangle < 0 \), a contradiction. Therefore, for any \( \varepsilon \) sufficiently small, \( \bar{\ell} \in T_\varepsilon \) as claimed.

Thus we have

\[
\lim \inf_{\varepsilon \to 0} T_\varepsilon \supseteq \text{int} \, T_S(0).
\]
As set limits are always closed ([RW98, Prop 4.4]), this yields \( \lim_{\epsilon \to 0} T_\epsilon \supseteq \text{cl int } T_S(0) = T_S(0) \).

Equation (5.12) follows from Theorem 5.23 and [RW98, Prop 4.30].

\[ \text{Note 5.26} \text{ If the regular tangent cone to a set has interior (at a point) the set is called epi-Lipschitz (at the point). Loosely speaking, epi-Lipschitz sets look like the epi-graphs of Lipschitz functions: such sets have been studied since the late 1970's, and many equivalent definitions have been developed [RW98, p. 418 & Ex 9.42]. Using this language, Theorem 5.25 requires the set } S \text{ be regular and epi-Lipschitz at the point } \mathbf{z}. \]

5.5 Numerical Results

In this section we present some numerical results for the Random Normal Generation (RNG) algorithm presented in Section 5.4. We begin by mentioning that the algorithm has been coded in MATLAB®, and all code is currently available at [Har]¹.

Our examination of the algorithm will take several forms. To begin, we study the effect of search radius on the accuracy of the approximate normal cone. We then examine the effect of dimension and the number of randomly selected points. We end by testing the algorithm in some minimization problems. To do this we create a second algorithm which makes use of RNG to create descent directions which are likely to maintain feasibility.

To gain a baseline for comparison, we consider an alternate approach to creating normal cone approximations. Specifically, we consider the same algorithm with the convexification step (III) removed. That is, instead of projecting each point in \( P_{\text{out}} \) onto \( \bar{S} \) (we use notation laid out in Section 5.4 page 94), simply project each point in \( P_{\text{out}} \) onto \( P_{\text{in}} \). Code for this alternate approach is also available at [Har].

5.5.1 Effect of Search Radius on Normal Convergence

In this subsection we examine the effect of the search radius on the accuracy of the approximate normal cone. We consider the unit ball, \( \{ x : |x| \leq 1 \} \), and the complement ball, \( \{ x : |x| \geq 1 \} \), in 2 and 5 dimensions. In both cases we attempt to find the normal cone at \( (1,0,...,0) \), by selecting 100 random points. To see the effect of the search radius we perform the experiment for three values of \( \epsilon \): 1.25, 0.5, and 0.05. We do the experiment both with and without the convexification step.

¹The author feels it is only fair to mention the web is a highly unstable archive. (I.e. "currently" refers to the spring of 2004.)
In order to get a qualitative measure of how accurate the approximate normal cone is, we compare it to the normal cone via the following formula:

\[
\max \left\{ \max \{ \arcsin(|n - P_{N_S}(\bar{x})|) : n \in \tilde{N}, |n| = 1 \} \right\},
\]

where \( N_S(\bar{x}) \) is the correct normal cone, and \( \tilde{N} \) is the approximate normal cone produced via RNG. Formula (5.13) measures the maximum angle between correct and approximate normal cones.

One hundred trials of each test were performed. The following figure displays the results of approximating the normal cone to the unit ball. In each bar graph, the interval [0,2] is divided into 20 equally sized subintervals. The height of each bar represents the number of trials whose error lands within the sub-interval. We also provide a table with the mean and standard deviation for the errors produced.

Figure 5.3: Effect of \( \varepsilon \) on RNG for the Unit Ball

Examining Figure 5.3 and Table 5.1 reveal several results.

First, in all cases, taking the convex hull creates a better estimate for the normal cone than skipping the convexification step. This suggests that RNG with the convexification step is making better use of the data given by the oracle. In cases where checking feasibility is difficult, this may be important.

The data also shows that when the search radius is large, the normal cone produced is less accurate. This is expected as the algorithm picks up normal vectors for points within the search radius. If the radius is large, the algorithm should create normal vectors to points farther from the center point. The boundary of the unit ball curves, and as a result, the normal vectors "fan out". This causes a larger approximate normal cone. As the search radius decreases, one expects the approximate normal
CHAPTER 5. ALGORITHMIC RESULTS

<table>
<thead>
<tr>
<th>Problem</th>
<th>Dimension &amp; Approach</th>
<th>Search Radius</th>
<th>mean error</th>
<th>standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit Ball</td>
<td>2, hull</td>
<td>1.25</td>
<td>1.1897</td>
<td>0.2161</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.50</td>
<td>0.7635</td>
<td>0.2890</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>0.4144</td>
<td>0.2432</td>
</tr>
<tr>
<td>( \bar{x} = [1, 0, ...0] )</td>
<td>2, no hull</td>
<td>1.25</td>
<td>1.3420</td>
<td>0.2043</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.50</td>
<td>1.0472</td>
<td>0.2281</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>0.8404</td>
<td>0.2220</td>
</tr>
<tr>
<td></td>
<td>5, hull</td>
<td>1.25</td>
<td>1.5458</td>
<td>0.0525</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.50</td>
<td>1.3607</td>
<td>0.1274</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>1.1620</td>
<td>0.1021</td>
</tr>
<tr>
<td></td>
<td>5, no hull</td>
<td>1.25</td>
<td>1.5708</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.50</td>
<td>1.5311</td>
<td>0.0620</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>1.4315</td>
<td>0.0879</td>
</tr>
</tbody>
</table>

Table 5.1: Mean and Standard Deviations for RNG on the Unit Ball

cone to become more accurate.

The last thing the data shows is that in higher dimensions it appears much harder to create an accurate approximation for the normal cone. We will discuss this further in the next subsection.

As mentioned, the same test was also performed on the complement of the unit ball. The interest in this set is in examining how RNG might behave in a nonconvex setting. Like the unit ball, 100 trials were performed, both with and without the convexification step.

The next figure displays the results of approximating the normal cone to the complement of the unit ball. Again, we follow it by a table presenting the mean and standard deviation for the errors produced.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Dimension &amp; Approach</th>
<th>Search Radius</th>
<th>mean error</th>
<th>standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complement Ball</td>
<td>2, hull</td>
<td>1.25</td>
<td>0.0850</td>
<td>0.0724</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.50</td>
<td>0.2189</td>
<td>0.2138</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>0.3827</td>
<td>0.2774</td>
</tr>
<tr>
<td>( \bar{x} = [1, 0, ...0] )</td>
<td>2, no hull</td>
<td>1.25</td>
<td>1.3334</td>
<td>0.2155</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.50</td>
<td>0.9393</td>
<td>0.2506</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>0.8386</td>
<td>0.2387</td>
</tr>
<tr>
<td></td>
<td>5, hull</td>
<td>1.25</td>
<td>0.7629</td>
<td>0.1732</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.50</td>
<td>0.9988</td>
<td>0.1274</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>1.1602</td>
<td>0.1215</td>
</tr>
<tr>
<td></td>
<td>5, no hull</td>
<td>1.25</td>
<td>1.5044</td>
<td>0.1100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.50</td>
<td>1.4444</td>
<td>0.0993</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>1.4301</td>
<td>0.0772</td>
</tr>
</tbody>
</table>

Table 5.2: Mean and Standard Deviations for RNG on the Complement Ball
At first glance the results for the complement ball may appear flawed. Although the errors for the approximate normal cone created without the convexification step appear the same as in the previous test; the errors from testing with the convexification step are actually smaller when the search radius is large. A moment's thought may explain this phenomenon. The convexification step causes the approximation of the set to become polyhedral; including a large face covering many points outside of the set. The normal vectors created will only include those produced from points on the outside of this face. These will be projected onto this face, causing (by good fortune) a high accuracy in the approximate normal cone. When the convexification step is skipped, the approximate normal cone does not have this advantage and is considerably less accurate.

When the search radius decreases the two problems (the unit ball and complement ball) should become similar. The results when $\epsilon = 0.05$ support this idea.
Like the results for the unit ball, the data suggests that as dimension increases the normal cone becomes more difficult to approximate. We examine this next.

5.5.2 Effects of Dimension and Number of Points

Our next two tests examine the effect of dimension and the number of points tested on the accuracy of the approximate normal cone. As before accuracy is measured by formula (5.13).

The two sets we consider are the positive half plane, \( \{x : x_1 \geq 0\} \), and the positive orthant, \( \{x : x_i \geq 0 \text{ for all } i\} \). In each case we approximate the normal cone at the origin. To see the effect of dimension we consider each set in 2, 5, and 11 dimensions. We further our experiment by testing each dimension with 10, 100, and 1000 randomly selected points. One hundred trials of each test were performed.

We begin by examining the results for approximating the normal cone to the half plane. The next table contains the mean and standard deviations produced.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Dimension &amp; Approach</th>
<th>Number of points</th>
<th>mean error</th>
<th>standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Half Plane</td>
<td>2, hull</td>
<td>10</td>
<td>0.7100</td>
<td>0.3685</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.4465</td>
<td>0.2789</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>0.2013</td>
<td>0.1268</td>
</tr>
<tr>
<td>( \bar{x} = [0, 0, \ldots 0] )</td>
<td>2, no hull</td>
<td>10</td>
<td>0.9876</td>
<td>0.3372</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.8558</td>
<td>0.2406</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>0.5075</td>
<td>0.0891</td>
</tr>
<tr>
<td></td>
<td>5, hull</td>
<td>10</td>
<td>1.2662</td>
<td>0.1807</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>1.1606</td>
<td>0.1097</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>1.0340</td>
<td>0.1081</td>
</tr>
<tr>
<td></td>
<td>5, no hull</td>
<td>10</td>
<td>1.3991</td>
<td>0.1375</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>1.4322</td>
<td>0.0869</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>1.4974</td>
<td>0.0400</td>
</tr>
<tr>
<td></td>
<td>11, hull</td>
<td>10</td>
<td>1.4012</td>
<td>0.1184</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>1.3494</td>
<td>0.0497</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>1.3736</td>
<td>0.0329</td>
</tr>
<tr>
<td></td>
<td>11, no hull</td>
<td>10</td>
<td>1.4902</td>
<td>0.0879</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>1.5303</td>
<td>0.0287</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>1.5446</td>
<td>0.0151</td>
</tr>
</tbody>
</table>

Table 5.3: Mean and Standard Deviations for RNG on the Half Plane

The results for this series of tests show several interesting trends. First, and least surprising, is that as the dimension of the problem increases, the accuracy of the approximate normal cone decreases. If this trend did not appear we would be extremely puzzled. Second, when convexification is applied, increasing the number of points increases the accuracy of the normal cone produced. This too was expected, and reassuring to see.
CHAPTER 5. ALGORITHMIC RESULTS

A third trend is somewhat surprising. When no convexification is applied, increasing the number of points seems to decreases the accuracy of the approximate normal cone. A possible reason for this demonstrates the effect of the convexification step. When convexification is applied, each point outside the set is projected onto a polyhedral set. This means only some normal directions are possible. Only directions which are normal to the surface of the polyhedral set and have points outside the set lying in that direction can occur. The result is that many points outside the set end up creating the same direction vector. If the convexification step is not applied, this is not true. Since points outside the set are projected only to the nearest point inside the set, each vector created is unique. More direction vectors increase the probability that one of them is in error. Since, formula (5.13) makes no allowance for how many vectors are in error, but only determines the worst one, the more points selected the less likely the results from the nonconvex approach will score well.

The same series of tests was performed on the positive orthant. The interest in this lies in the fact the normal cone is no longer a single ray. Indeed the normal cone to the positive orthant at the origin is the negative orthant. We would like to see if this larger normal cone causes more difficulty in approximating. As before the tests were run 100 times each. The results appear in the next table.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Dimension &amp; Approach</th>
<th>Number of points</th>
<th>mean error</th>
<th>standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Orthant</td>
<td>2, hull</td>
<td>10</td>
<td>0.8133</td>
<td>0.3671</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.4208</td>
<td>0.2635</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>0.2059</td>
<td>0.1177</td>
</tr>
<tr>
<td>( \bar{x} = [0, 0, \ldots, 0] )</td>
<td>2, no hull</td>
<td>10</td>
<td>0.9529</td>
<td>0.3644</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.8350</td>
<td>0.2296</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>0.4787</td>
<td>0.0880</td>
</tr>
<tr>
<td>5, hull</td>
<td>10</td>
<td>1.2852</td>
<td>0.1960</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.1166</td>
<td>0.1613</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.9888</td>
<td>0.1056</td>
<td></td>
</tr>
<tr>
<td>5, no hull</td>
<td>10</td>
<td>1.3081</td>
<td>0.1797</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.3916</td>
<td>0.1231</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1.3713</td>
<td>0.0879</td>
<td></td>
</tr>
<tr>
<td>11, hull</td>
<td>10</td>
<td>1.2922</td>
<td>0.1213</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.4217</td>
<td>0.0861</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1.4056</td>
<td>0.1486</td>
<td></td>
</tr>
<tr>
<td>11, no hull</td>
<td>10</td>
<td>1.2814</td>
<td>0.1175</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.4219</td>
<td>0.0809</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1.5046</td>
<td>0.0450</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.4: Mean and Standard Deviations for RNG on the Positive Orthant

The results appear very similar to the results for the half plane. This suggests that the size of the normal cone plays very little role in the difficulty of approximating it. This is reassuring from an algorithmic point of view.
5.5.3 Testing Problem Solving

Our final series of tests examines how RNG might be applied to optimization. To do this we create a second algorithm, Random Normal Minimization, which makes use of RNG to determine descent directions which are likely to maintain feasibility. We describe the Random Normal Minimization algorithm next.

**Conceptual Algorithm: [Random Normal Minimization (RNM)]** Consider the constrained minimization problem

\[
(P) \quad \min_{x} \{\langle c, x \rangle : x \in S\},
\]

where \(c\) is a fixed vector in \(\mathbb{R}^n\), and \(S\) is an oracle-based set. Perform the following:

**I. Initialize:** Set the iteration counter \(k\) to 0 and select

- an initial point \(x_0 \in S\),
- an initial search radius \(\varepsilon_0\),
- an initial search number \(n_0\) and
- an inner threshold parameter \(\mu \in (0, 1)\).

**II. Search Direction:** Use the Approximate Normal Creation algorithm on \(S, x_k, \varepsilon, \mu,\) and \(n_k\) to create an approximate normal cone \(\tilde{N}\), then find descent direction by minimizing

\[
\min_{d} \{\langle c, d \rangle : |d| \leq 1, \langle d, n \rangle \geq 0 \text{ for all } n \in \tilde{N}\}.
\]

**III. Step Size:**

Find the smallest integer \(m \geq 0\) such that \(x_k + \frac{1}{2}^m d_k\) is feasible, then set \(s_k = \frac{1}{2}^m\).

**IV. Update:**

Set \(x_{k+1} = x_k + s_k d_k\),

select \(\varepsilon_{k+1}\) and \(n_{k+1}\) for next iteration, then

increase \(k\) by one.

**V. Repeat:** From II until the stopping criterion is satisfied.

The algorithm, as described, leaves several decisions to the user: how to select \(\varepsilon_{k+1}\) and \(n_{k+1}\) for the next iteration, and what stopping criterion to use. Good choices for these options are open to further research. We describe the methods used in our code next.

After each iteration, our code compares the step length taken with the current search radius, \(\varepsilon_k\). If the step length is sufficiently large (0.1\(\varepsilon_k\)) the iteration is considered successful. In this case the
search radius remains the same for the next iteration while the number of points checked is returned to its initial value (defined by the user, or code defaults). If the iteration is unsuccessful the search radius is reduced by 25%, and the number of points checked is increased by 33%. The search radius is never increased.

To see how useful RNG might be in practice, we examine how the RNM algorithm performs on two series of problems.

The first problem we solve is

$$\min\{(5,1,1,...,1), \mathbf{x} : |x_i| \leq 1 \text{ for } i = 1,2,\ldots,n\}.$$  

To add challenge we set the initial point to be $x_0 := (1,1,...1)$. The difficulty in solving this problem appears after several iterations. The first iteration should move in the descent direction $(-5,-1,-1,...,-1)$ toward the far side of the unit box. Eventually iterates will become close enough to the boundary that future descent in this direction is unproductive. At this point the algorithm must use the approximate normal cone to search for directions more parallel to the boundary.

We perform this test in 2, 5, and 11 dimensions. Since, in each dimension the optimal value is easy to calculate, our stopping criterion will be the function value becomes “sufficiently close” to the correct solution. Sufficiently close varies by dimension as follows:

$$\min\{(5,1), \mathbf{x} : |x_i| \leq 1 \text{ for } i = 1,2\} \text{ requires } (5,1), x_k < -5.6;$$

$$\min\{(5,1,1,1,1), \mathbf{x} : |x_i| \leq 1 \text{ for } i = 1,2,\ldots,5\} \text{ requires } (5,1,1,1,1), x_k < -8;$$

$$\min\{(5,1,1,1,1,1,1,1,1,1,1), \mathbf{x} : |x_i| \leq 1 \text{ for } i = 1,2,\ldots,11\} \text{ requires } (5,1,\ldots,1), x_k < -9.$$  

As one can see, our stopping criterion becomes weaker as the dimension of the problem increases. This is to reduce the time required to complete the test.

A test is considered successful (pass) if it nears the minimum value within 50 iterations, otherwise it is consider unsuccessful (fail). Each test was run 100 times. We compared the mean and standard deviation of the number of set evaluations below.

In all dimensions tested, RNM performed better when the convex hull operation was applied. Not only was its success rate considerably better, but the number of set evaluations was (on average) noticeable lower than when no convexification took place. However, in either case, the number of set evaluations is too large for practical use.

The second problem we consider is

$$\min\{(1,0,...0), \mathbf{x} : 1 \leq |x| \leq 2\}.$$  

The algorithm is begun at the initial point $x_0 := (1,0,0...0)$. The interest in this problem is the initial point is a critical point. The direction of steepest descent, $(-1,0,0,...0)$, immediately leaves
CHAPTER 5. ALGORITHMIC RESULTS

Problem & Dimension & Approach & Pass & Set & Fail & Set
\[ c = [5, 1, 1, ... 1] \]
<table>
<thead>
<tr>
<th>Problem</th>
<th>Dimension &amp; Approach</th>
<th>Pass</th>
<th>Set Evaluations</th>
<th>Fail</th>
<th>Set Evaluations</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ D = {x :</td>
<td>x_i</td>
<td>\leq 1} ]</td>
<td>2, hull</td>
<td>100</td>
<td>mean: 229.33 std: 125.36</td>
</tr>
<tr>
<td></td>
<td>2, no hull</td>
<td>99</td>
<td>mean: 762.71 std: 751.36</td>
<td>1</td>
<td>mean: 5424 std: 0</td>
</tr>
<tr>
<td></td>
<td>5, hull</td>
<td>82</td>
<td>mean: 739.95 std: 312.15</td>
<td>18</td>
<td>mean: 1634.4 std: 232.08</td>
</tr>
<tr>
<td></td>
<td>5, no hull</td>
<td>36</td>
<td>mean: 1009.5 std: 528.34</td>
<td>64</td>
<td>mean: 2573.8 std: 754.87</td>
</tr>
<tr>
<td></td>
<td>11, hull</td>
<td>84</td>
<td>mean: 1117.3 std: 412.67</td>
<td>16</td>
<td>mean: 2161.6 std: 241.19</td>
</tr>
<tr>
<td></td>
<td>11, no hull</td>
<td>59</td>
<td>mean: 1366.1 std: 407.98</td>
<td>41</td>
<td>mean: 2277.0 std: 296.26</td>
</tr>
</tbody>
</table>

Table 5.5: Mean and Standard Deviations for RNM on the Unit Square

the constraint set. Despite this RNM tends to discover descent directions. This is likely because errors in the approximate normal cone yield "false" feasible descent directions. As the code attempts a unit step size first, the fact that the descent direction leaves and reenters the set is undetected by the algorithm.

Like the first problem, we examine the problem in 2, 5, and 11 dimensions. In each dimension the optimal value is easy to calculate, so we once again base the stopping criterion on the iterates value becoming sufficiently small. The stopping criterion varies by dimension, as follows:

\[
\min \{ ((1, 0), x) : 1 \leq |x| \leq 2 \} \text{ requires } ((1, 0), x_k) < -1.92;
\]

\[
\min \{ ((1, 0, 0, 0, 0), x) : 1 \leq |x| \leq 2 \} \text{ requires } ((1, 0, 0, 0, 0), x_k) < -1.88;
\]

\[
\min \{ ((1, 0, 0, 0, 0, 0, 0, 0, 0, 0), x) : 1 \leq |x| \leq 2 \} \text{ requires } ((1, 0, ..., 0), x_k) < -1.82.
\]

Again as dimension increases the stopping criterion is weakened to decrease computing time.

As before, we will consider a test successful (pass) if it nears the minimum value within 50 iterations, otherwise we will consider it unsuccessful (fail). Each test was run 100 times. Next we present the mean and standard deviations for the number of set evaluations.

Although, not taking the convex hull showed some difficulty in solving the problem in 2 dimensions, in higher dimensions this difficulty appeared to vanish. One possible answer for this relates to the reason the code is successful at all. As mentioned the likely reason the code succeeds in moving away from the initial point is the detection of false feasible descent directions. As Subsection 5.5.2 shows that as dimension increases RNG produces increasingly less accurate approximate normal cones. Therefore the likelihood of a false positives increases.
CHAPTER 5. ALGORITHMIC RESULTS

<table>
<thead>
<tr>
<th>Problem</th>
<th>Dimension &amp; Approach</th>
<th>Pass</th>
<th>Set Evaluations</th>
<th>Fail</th>
<th>Set Evaluations</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c = [1, 0, 0, \ldots 0]$</td>
<td>2, hull</td>
<td>100</td>
<td>mean: 191.47 std: 113.99</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$D = {x : 1 \leq</td>
<td>x</td>
<td>\leq 2}$</td>
<td>2, no hull</td>
<td>66</td>
<td>mean: 266.65 std: 360.97</td>
</tr>
<tr>
<td></td>
<td>5, hull</td>
<td>100</td>
<td>mean: 204.11 std: 90.864</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5, no hull</td>
<td>97</td>
<td>mean: 386.57 std: 403.33</td>
<td>3</td>
<td>mean: 5888.0 std: 37.749</td>
</tr>
<tr>
<td></td>
<td>11, hull</td>
<td>100</td>
<td>mean: 417.91 std: 144.97</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11, no hull</td>
<td>99</td>
<td>mean: 610.62 std: 212.87</td>
<td>1</td>
<td>mean: 6144 std: 0</td>
</tr>
</tbody>
</table>

Table 5.6: Mean and Standard Deviations for RNM on the Donut

This also suggests something about the errors produced in RNG. Despite the fact that the error in 11 dimensions is quite high (see Subsection 5.5.2), descent directions are still found. This suggests the error is a result of a large angle produced by a small number of incorrect vectors. That is the error localized on one "side" of the normal cone. If error occurred on all sides, then RNM would find that there are no feasible descent directions. Since some directions are found, the error must be small on some "sides" of the normal cone, and large on others. So, RNG may be more promising than Subsection 5.5.2 suggests.

Though RNM was successful without taking the convex hull, using the convex hull operation considerably reduced the number of set evaluations. Therefore in circumstances where it is expensive to determine if a point is inside or outside of a set, the convexification step seems favourable.

RNG and RNM are essentially theoretical algorithms. In higher dimensions the number of set evaluations becomes too large for practical use.
Chapter 6

Conclusions and Open Questions

This thesis's primary goal was to examine partial smoothness in nonsmooth optimization. In doing this, it has accomplished two things. First, it has shown that partly smooth functions and sets are a natural idea of smooth substructure in the study of nonsmooth optimization. Second, it has shown that partial smoothness provides a powerful tool for nonsmooth optimization, especially when combined with the notion of prox-regularity.

In Chapter 2 we introduced the definition of $C^p$-partial smoothness, and provided a plethora of examples to demonstrate the concept. Chapter 2 ended by showing that partial smoothness is a nonconvex extension of the convex ideas of $C^p$-identifiable surfaces and fast tracks. Thus we saw that partial smoothness is a natural tool in the study of nonsmooth optimization.

Our study of partial smoothness continued in Chapter 3, where the calculus of partial smoothness was examined. We began by showing that past results of Lewis on $C^2$-partial smoothness could easily be extended to the broader definition used in this thesis. These results were used to reprove the relationship between a partly smooth function and its epigraph; (the relationship was proven directly in Section 2.3). A new calculus rule on infimal convolution was then created and explored.

Chapter 3 continued by examining the effects of prox-regularity on partial smoothness. Previous work of Poliquin and Rockafellar has shown that prox-regularity is a natural extension of convexity. The addition of prox-regularity to partial smoothness culminated in the Smooth Projection Theorem of Section 3.3. Besides extending results on the smoothness of the projection mapping, the Smooth Projection Theorem was shown to have many other uses in examining partly smooth sets and functions. Applying the Smooth Projection Theorem, the active manifold of prox-regular partial smoothness was shown to be unique. The Smooth Projection Theorem was further applied in Section 3.4 where many results on the identification of active manifolds (and therefore active constraint sets) were developed. These results demonstrated that prox-regular partial smoothness is a powerful tool in the study of nonsmooth optimization.
The effect of prox-regularity on partial smoothness was further examined in Section 3.5, where the question of when a critical point of a prox-regular partly smooth function is a local minimum was examined. Results again supported the power of prox-regular partly smooth functions, showing that only points on the active manifold need to be considered when examining optimality.

Chapter 3 concluded with an example showing that these results can fail when prox-regularity is not present.

In Chapter 4 we returned to comparing partial smoothness to previously defined notions of smooth substructure. Primal-dual gradient structures, and \( g \circ F \) decompositions are both broad classes of nonconvex functions with smooth substructure. In Chapter 4 we provided definitions for these notions and compared them to partial smoothness. The results were seen to support the statement that partial smoothness arises naturally in nonsmooth optimization. Functions with primal-dual gradient structures were seen to be partly smooth under some additional structure on their index sets, while \( g \circ F \) decompositions were shown to be the composition of partly smooth functions with smooth functions. Hence, under the transversality condition of Shapiro, \( g \circ F \) decomposable functions are partly smooth. Section 4.2 further showed how the transversality condition of Shapiro was a simplification of the amenability condition of Poliquin and Rockafellar.

Chapter 5 focused on how the results of this thesis might effect future algorithmic research. We began by examining the results on active manifold identification created in Section 3.4 in an applied setting. Using these results we showed that several different algorithms identified the active manifolds (and therefore the active constraints) of a minimization problem in a finite number of iterations. In the case of optimization over a constraint set, Section 5.2 showed that the Gradient Projection method, and Newton-like methods identify the active constraints of the constraint set in a finite number of iterations. For unconstrained optimization, Section 5.3 showed that the Proximal Point method identifies the active constraints of the function being minimized in a finite number of iterations. These results reiterated the power of prox-regular partly smooth functions and sets in nonsmooth optimization.

A secondary goal of this thesis was a new method of generating normal cone information for an oracle-based set. This was accomplished in Chapter 5, where the "Random Normal Generation" algorithm was developed.

The algorithm was outlined in Section 5.4, which also included several results arguing that the algorithm might be successful. These results also showed a new method of describing the normal and tangent cones for regular sets. The algorithm was further examined in Section 5.5, where numerical testing was performed on the algorithm. Though the results showed some promise, in higher dimensions the number of set evaluations became too large for practical use.

This thesis leads to many open questions.

The first argument presented by this thesis proposed, partial smoothness is a natural tool in
CHAPTER 6. CONCLUSIONS AND OPEN QUESTIONS

the study of nonsmooth optimization. This is illustrated both by a plethora of examples, and by comparing partial smoothness to many previously defined notions of smooth substructure. Strong connections are found between partial smoothness, $C^p$-identifiability, fast tracks, and $g \circ F$ decomposable functions; but, the connections between partial smoothness and primal-dual gradient structures is less clear. Theorem 4.8 shows that every partly smooth function has a dual feasible primal-dual gradient structure, but does not retain the active manifold in the creation of this structure. This leads to the question:

\textit{To what degree can a stronger version of Theorem 4.8 be created?}

In a related question, the connection between amenable functions and partial smoothness has not yet been examined. Given the results of Example 2.6 (Finite Max Functions) and Corollary 4.16 (Transversal $g \circ F$ Decompositions), it is clear that some connection exists. How strong this connection is remains unclear. Given the strong connections between strongly amenable functions and lower-$C^2$ functions it may be equally interesting to examine the connections between partial smoothness and lower-$C^2$ functions.

\textit{What connections exist between partly smooth functions, strongly amenable functions, and lower-$C^2$ functions?}

Incidentally, the term Partial Smoothness has been used elsewhere in mathematical literature. In [BTZ99] Borwein, Treiman and Zhu define a Banach space $X$ to be partially smooth relative to a subspace $Y \subseteq X$ if $Y$ has a smooth renorm. Their work continues by exploring when a Banach space being partially smooth implies the entire space had a smooth renorm. The primary focus of the definition of Borwein, Treiman and Zhu is infinite dimensions, but it would be interesting to explore the precise relationship between this and the notion used in this thesis.

The second focus of this thesis was to show the power of prox-regular partly smooth functions in optimization. The examination of prox-regular partly smooth functions led to the Smooth Projection Theorem (Theorem 3.13). This theorem extends many results on the smoothness of the projection mapping for convex sets to a nonconvex setting. This also leads to an open question. Examining Fitzpatrick and Phelps's 1982 work, [FP82], we see that the projection mapping for a convex set is $C^{p-1}$ if and only if the boundary is $C^p$. The Smooth Projection Theorem only provides an extension for one direction of this result. This leads to the open question:

\textit{Can the smoothness of the boundary of a prox-regular set be determined by the smoothness of the projection mapping?}

It would be surprising if the answer to this question turned out to be no.

The power of the Smooth Projection Theorem was largely demonstrated by the many results in active constraint identification. Of course this leads to the question:
CHAPTER 6. CONCLUSIONS AND OPEN QUESTIONS

*How can active manifold identification aid in algorithm design?*

This question has already been approached by Mifflin and Sagastizábal in their research on fast tracks. The VU-Proximal Point algorithm attempts to find the active manifold, and then move along it to speed convergence [MS02b]. Results are preliminary, but hopeful.

Another obvious question arising from the constraint identification results is,

*What other algorithms allow for the finite identification of active constraints?*

One algorithm which is likely to have finite identification properties is the subgradient projection method discussed in [Flâ92]. Flåm’s work shows that the subgradient projection method shares many of the same convergence properties as the gradient projection algorithm. Therefore it is likely that the finite convergence results of Flåm could be extended to this broader setting.

The final focus of this thesis was on the “Random Normal Generation” algorithm, which creates normal cone information for oracle-based sets. Clearly this research is far from complete. For example, this thesis presented results showing that if every point was examined then as the search radius decreases the approximate normal cone converged to the correct normal cone. Clearly examining every point is a computational impossibility, so results still need to be developed to show that as the number of points selected increases, the approximate normal cone approaches the near normal cone (Definition 5.18). Hence the open question,

*Under what conditions must the approximate normal (tangent) cone approach the near normal (tangent) cone?*

In our numerical experimentation on the “Random Normal Generation” algorithm, we briefly examined how the approximate normal cone could be used to minimize a linear function over an oracle-based set. This resulted in the creation of the “Random Normal Minimization” algorithm. Like “Random Normal Generation”, the results for this algorithm are far from complete. We end with the open question:

*What conditions ensure the “Random Normal Minimization” algorithm converges (to a local minimum)?*
CHAPTER 6. CONCLUSIONS AND OPEN QUESTIONS

A Index of Notation

General Notation

\( f, g \) functions (\( \mathbb{R}^n \to \mathbb{R} \) unless otherwise stated)

\( C, S \) a convex set, an arbitrary set

\( x, \bar{x} \) a variable point, a fixed point

\( d, \bar{d} \) a variable direction vector, a fixed direction vector

\( n, \bar{n} \) a variable normal vector, a fixed normal vector

\( t, \bar{t} \) a variable tangent vector, a fixed tangent vector

\( w, \bar{w} \) a variable subgradient, a fixed subgradient

\( c \) a vector (used in minimization problems)

\( H \) a matrix (approximate Hessian, Section 5.2)

\( \lim \inf \) lower or inner limit (Definition 1.3, page 17)

\( \lim \sup \) upper or outer limit (Definition 1.3, page 17)

\( \text{lsc, usc} \) lower semi-continuous, upper semi-continuous (Definition 1.4, page 18)

\( \text{isc, osc} \) inner semi-continuous, outer semi-continuous (Definition 1.4, page 18)

\( \text{min}\{f(x)\} \) minimum of \( f \) over \( \mathbb{R}^n \)

\( \text{min}_S\{f(x)\} \) minimum of \( f \) over \( S \)

\( \text{min}_y\{f(x, y)\} \) minimum of \( f \) with respect to \( y \)

\( \text{min}\{f : A\} \) minimum of \( f \) such that condition \( A \) holds

\( \text{max}, \inf, \sup \) maximum, infimum, supremum (used as in min)

\( \arg\min\{f\}, \arg\max\{f\} \) solution set of \( \text{min}\{f\}, \text{max}\{f\} \) (used as in min)

\( C^p \) the set of all functions with \( p \) continuous derivatives

\( \text{PDG} \) Primal-dual gradient (Definition 4.1, page 68)

\( \text{RNG} \) the Random Normal Generation algorithm (page 94)

\( \text{RNM} \) the Random Normal Minimization algorithm (page 106)
Special Sets

- $\mathbb{R}$: the real numbers
- $\mathbb{R}_+$: the positive real numbers
- $\mathbb{R}$: the extended real numbers ($\mathbb{R} \cup \{\infty\}$)
- $\mathbb{R}^n$: the real vector space of $n$ dimensions
- $\mathbb{R}^+_n$: the positive orthant in $\mathbb{R}^n$
- $B_\varepsilon(\bar{z})$: the open ball of radius $\varepsilon$ about the point $\bar{z}$
- $C^p$: the set of all functions with $p$ continuous derivatives
- $U$: the "smooth" subspace, (Definition 2.22, page 40)
- $V$: the "sharp" subspace, (Definition 2.22, page 40)
- $\Delta$: convex set for PDG structures (Definition 4.1, page 68)
- $K, K_f, K_\phi$: index sets for PDG structures (Definition 4.4, page 70)
- $A(\bar{z})$: set of active constraints at $\bar{z}$ (Example 2.6, page 30)
- $N_S(\bar{z})$: the normal cone to $S$ at $\bar{z}$ (Definition 1.5, page 19)
- $T_S(\bar{z})$: the tangent cone to $S$ at $\bar{z}$ (Definition 1.5, page 19)
- $N^{\varepsilon, \mu}_C(\bar{z}), N_\varepsilon$: the near normal cone to $C$ at $\bar{z}$ relative to $\varepsilon$ and $\mu$ (Definition 5.18, page 94)
- $T^{\varepsilon, \mu}_C(\bar{z}), T_\varepsilon$: the near tangent cone to $C$ at $\bar{z}$ relative to $\varepsilon$ and $\mu$ (Definition 5.18, page 94)
- $S_\varepsilon$: the closed convex hull of $S \cap B_\varepsilon(\bar{z})$ (equation (5.6), page 95)
- $P_{\text{in}}, P_{\text{out}}$: Point organization sets for Random Normal Algorithm (Section 5.4, page 106)
CHAPTER 6. CONCLUSIONS AND OPEN QUESTIONS

Special Functions

\[ |\cdot| \quad \text{the } (l_2) \text{ norm of a vector} \]
\[ \|\cdot\| \quad \text{the } (l_2) \text{ norm of a matrix} \]
\[ \delta_S \quad \text{the indicator function of the set } S \text{ (Equation 1.4, page 16)} \]
\[ \text{dist}(x, S) \quad \text{the distance function (Definition 1.11, page 21)} \]
\[ P_S(x) \quad \text{the projection mapping (Definition 1.11, page 21)} \]
\[ \text{cl } S \quad \text{the closure of } S \]
\[ \text{int } S \quad \text{the interior of } S \]
\[ \text{rint } S \quad \text{the relative interior of } S \]
\[ \text{conv } S \quad \text{the convex hull of } S \]
\[ \text{par } S \quad \text{the subspace parallel to } S \text{ (Equation 1.2, page 16)} \]
\[ S^\circ \quad \text{the polar cone to } S \text{ (Equation 1.9, page 18)} \]
\[ \text{lev } f \quad \text{the level set of } f \text{ (Definition 1.5, page 16)} \]
\[ \text{epi } f \quad \text{the epigraph of } f \text{ (Definition 1.6, page 16)} \]
\[ \text{rng } f \quad \text{the range of } f \text{ (Theorem 1.19, page 24)} \]
\[ \text{nul } f \quad \text{the null space of } f \text{ (Theorem 1.19, page 24)} \]
\[ \text{lin } f \quad \text{the linearity space of } f \text{ (Equation 1.3, page 16)} \]
\[ o(\cdot) \quad \text{the little 'oh' function, (Equation 1.7, page 17)} \]
\[ O(\cdot) \quad \text{the big 'oh' function, (Equation 1.8, page 17)} \]
\[ f \# g \quad \text{the infimal convolution of } f \text{ and } g \text{ (Equation 3.3, page 49)} \]
\[ L_U(u, w) \quad \text{the } U\text{-Lagrangian (Definition 2.24, page 41)} \]
\[ v(u) \quad \text{fast track function (Definition 2.25, page 41)} \]
\[ e_R, P_R \quad \text{proximal envelope, proximal point mapping (Definition 5.11, page 89)} \]
CHAPTER 6. CONCLUSIONS AND OPEN QUESTIONS

Variational Calculus

\(\nabla f(\bar{x})\) the gradient vector of \(f\) at \(\bar{x}\)
\(\nabla_y f(x, \bar{y})\) the gradient vector of \(f(x, \cdot)\) at \(\bar{y}\)
\(df(\bar{x}, d)\) the subderivative of \(f\) at \(\bar{x}\) in the direction \(d\) (Definition 1.7, page 20)
\(\delta f(\bar{x})\) the regular subdifferential of \(f\) at \(\bar{x}\) (Definition 1.8, page 20)
\(\delta_y f(x, \bar{y})\) the regular subdifferential of \(f(x, \cdot)\) at \(\bar{y}\) (Definition 1.8, page 20)
\(\partial f(\bar{x})\) the (limiting) subdifferential of \(f\) at \(\bar{x}\) (Definition 1.8, page 20)
\(\partial_y f(x, \bar{y})\) the (limiting) subdifferential of \(f(x, \cdot)\) at \(\bar{y}\) (Definition 1.8, page 20)
\(\partial f(\bar{x})\) the Clarke subdifferential of \(f\) at \(\bar{x}\) (Definition 1.9, page 21)
\(\partial N_S(\bar{x})\) the regular normal cone to \(S\) at \(\bar{x}\) (Definition 1.5, page 19)
\(\partial T_S(\bar{x})\) the regular tangent cone to \(S\) at \(\bar{x}\) (Definition 1.5, page 19)
\(N_S(\bar{x})\) the (limiting) normal cone to \(S\) at \(\bar{x}\) (Definition 1.5, page 19)
\(T_S(\bar{x})\) the (limiting) tangent cone to \(S\) at \(\bar{x}\) (Definition 1.5, page 19)
Bibliography


W. L. Hare. http://www.cec.m.sfu.ca/~whare/mm.html.


BIBLIOGRAPHY


C. Sagastizábal. Personal correspondence.

