Continuously Differentiable Selections and Parametrizations of Multifunctions in One Dimension.

Craig Knuckles
Louisiana State University and Agricultural & Mechanical College

Follow this and additional works at: https://repository.lsu.edu/gradschool_disstheses

Recommended Citation
https://repository.lsu.edu/gradschool_disstheses/5962

This Dissertation is brought to you for free and open access by the Graduate School at LSU Scholarly Repository. It has been accepted for inclusion in LSU Historical Dissertations and Theses by an authorized administrator of LSU Scholarly Repository. For more information, please contact gradetd@lsu.edu.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700  800/521-0600
CONTINUOUSLY DIFFERENTIABLE SELECTIONS
AND
PARAMETRIZATIONS OF MULTIFUNCTIONS
IN
ONE DIMENSION

A Dissertation
Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in
The Department of Mathematics

by
Craig Knuckles
B.S., Tennessee Technological University, 1989
M.S., Tennessee Technological University, 1991
May 1995
ACKNOWLEDGEMENTS

The author would like to thank his advisor, Dr. Peter Wolenski, for working patiently with him and providing many valuable insights into Mathematics. He especially thanks Dr. Wolenski for helping him with the preparation of this dissertation. The author would also like to thank Dr. John Hildebrant, Dr. Guillermo Ferreyra, Dr. Gerald Knapp, Dr. Frank Neubrander, and Dr. Robert Perlis for serving on his committee.

In a much different respect, the author offers profound thanks to his parents Mr. and Mrs. Otis Knuckles. As a youth, they provided him the proper environment in which to instill the fortitude necessary for success in a long collegiate career. Their financial support made surviving a ten year collegiate career possible. They also continually provided him with moral support and encouragement. The author feels that it would be impossible to have better parents. He would like to thank his wife Hope Knuckles for being continually supportive throughout his graduate studies.

The author would also like to thank several long term friends Bryan Moore, Steve Grodrian, Danny Jones, and Luther Johnson for their encouragement along the way. Finally, the author would like to thank the band Black Sabbath for providing most excellent contemplative music during his periods of reclusive mathematical cogitation.
# TABLE OF CONTENTS

**ACKNOWLEDGEMENTS** .............................................................................................. ii  
**INDEX OF DEFINITIONS** ............................................................................................ iv  
**ABSTRACT** ...................................................................................................................... v  

**CHAPTER**  
1  **INTRODUCTION** ............................................................................................... 1  
2  **DEFINITIONS AND PRELIMINARY DISCUSSION** .................................. 9  
3  **STATEMENTS OF THE MAIN RESULTS AND SOME EXAMPLES** ......................... 26  
4  **PROOF OF THEOREM 3.1** ........................................................................... 32  
   Proof of Theorem 3.1 .................................................................................. 38  
5  **PROOF OF THEOREM 3.2** ........................................................................... 48  
   Proof of Theorem 3.2 (sufficiency) ............................................................ 48  
   Proof of Theorem 3.2 (necessity) ............................................................... 62  

**BIBLIOGRAPHY** ................................................................................................. 63  
**VITA** ...................................................................................................................... 64
INDEX OF DEFINITIONS

<table>
<thead>
<tr>
<th></th>
<th>Definition</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Multifunction (Closed, Compact, Lipschitz, etc.)</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>Selection (Measurable, Continuous, (C^1), etc.)</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>Hamiltonian</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>Parametrization (Measurable, Continuous, (C^1), etc.)</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>Lipschitz Function</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>Directional Derivative</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>Generalized Directional Derivative</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>Clarke Generalized Gradient</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>Lower-(C^k) Function</td>
<td>11</td>
</tr>
<tr>
<td>10</td>
<td>(Strictly) Submonotone Multifunction</td>
<td>13</td>
</tr>
<tr>
<td>11</td>
<td>Subdifferentially Regular Function</td>
<td>14</td>
</tr>
<tr>
<td>12</td>
<td>Upper Semi-continuous Function</td>
<td>14</td>
</tr>
<tr>
<td>13</td>
<td>Directional Convergence of Sequence</td>
<td>15</td>
</tr>
<tr>
<td>14</td>
<td>Semismooth Function</td>
<td>15</td>
</tr>
<tr>
<td>15</td>
<td>Support Function of Convex Set</td>
<td>15</td>
</tr>
<tr>
<td>16</td>
<td>Left and Right Derivatives</td>
<td>32</td>
</tr>
</tbody>
</table>
ABSTRACT

Sufficient conditions are given for a multifunction (set-valued function) to admit a continuously differentiable selection in one dimension. These conditions are given in terms of Clarke generalized gradients of the Hamiltonian associated with the multifunction. Also, the multifunctions in one dimension that can be parametrized with continuously differentiable functions are completely characterized. The characterization is again in terms of Clarke generalized gradients of the Hamiltonian associated with the multifunction.
A multifunction (set-valued mapping) $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a function mapping a point of $\mathbb{R}^n$ to a subset of $\mathbb{R}^n$. The double arrows indicate that $F$ takes on as values, subsets of $\mathbb{R}^n$ rather than only single points as is the case with ordinary functions. Multifunctions are playing an increasingly important role in mathematics, and particularly in applications to optimal control theory. This is due to the fact that optimal control problems can be reformulated into a so-called differential inclusion problem, which we next briefly describe. Loosely speaking, a differential inclusion is a generalization of an ordinary differential equation in which the right-hand side of the equation is replaced by a multifunction. Consider the ordinary differential equation

\begin{align*}
\text{(ODE)} & \quad x'(s) = \phi(s, x(s)) \quad \text{for almost all } s \in [0, T], \\
& \quad x(0) = x_0,
\end{align*}

where $\phi : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a function usually assumed to have some continuity properties in $x$ and at least measurability in $s$. Here $x'(s)$ denotes the derivative of $x(\cdot)$ at $s$. A solution of the ordinary differential equation is an absolutely continuous function $x(\cdot)$ satisfying the differential equation and the initial condition $x(0) = x_0$.

A differential inclusion has the form

\begin{align*}
\text{(DI)} & \quad x'(s) \in F(x(s)) \quad \text{for almost all } s \in [0, T], \\
& \quad x(0) = x_0,
\end{align*}

where $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a multifunction.
A solution $x(\cdot)$ of (DI) is an absolutely continuous function whose derivative $x'(s)$ lands in the set $F(x(s))$ for almost all $s$, and which satisfies the initial condition $x(0) = x_0$. We briefly demonstrate the relationship between optimal control systems and differential inclusions.

Consider a control system, which we formulate as

(CS)  \[ x'(s) = \phi(x(s), u(s)) \text{ for almost every } s \in [0, T], \]
\[ u(s) \in U \text{ for almost every } s \in [0, T], \]
\[ x(0) = x_0. \]

Here the control function $u(\cdot) : [0, T] \to \mathbb{R}^m$ is taken to be measurable with almost all of its values in a predetermined control set $U \subseteq \mathbb{R}^m$, $\phi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is a given function (at least continuous), and $x(\cdot) : [0, T] \to \mathbb{R}^n$ is the absolutely continuous function satisfying the resulting ordinary differential equation (which is of the form (ODE) once $u(\cdot)$ has been fixed).

An equivalent and sometimes preferable manner to approach the control system is by studying the related differential inclusion system, in which we define

\[ F(x) = \{ \phi(x, u) : u \in U \}. \]  \hspace{1cm} (1.1)

Note that the control variables are then explicitly suppressed in the differential inclusion formulation; they do, however, determine the structure of the multifunction $F$. Under very mild and reasonable hypotheses, the (DI) and (CS) formulations are equivalent in the sense that both models generate the same trajectories $x(\cdot)$. If the pair $(x(\cdot), u(\cdot))$ satisfies (CS), then it is trivial to see that $x(\cdot)$ solves (DI). The converse is not so obvious, but is nevertheless true. Namely, if $x(\cdot)$ solves (DI), then there exists a measurable function $u(\cdot)$ so that $(x(\cdot), u(\cdot))$ satisfies (CS). This result follows by applying a so-called measurable selection theorem, which
in this context is usually referred to as Filippov’s Lemma. What this amounts to is the following. Let \( x(\cdot) \) be a solution of (DI) with \( F \) as in (1.1). If we define \( V(s) = \{ u \in U : x'(s) = \phi(x, u) \text{ for a.e. } s \in [0, T] \} \), then \( V(\cdot) \) is another multifunction, and the sought-after control function \( u(\cdot) \) is a measurable function such that \( u(s) \in V(s) \) for a.e. \( s \in [0, T] \). If such a \( u(\cdot) \) exists, then the pair \((x(\cdot), u(\cdot))\) satisfies (CS). The function \( u(\cdot) \) is an example of a measurable selection of the multifunction \( V(\cdot) \), and Filippov’s Lemma amounts to providing hypotheses for when such selections exist.

To see why (DI) may be preferable to (CS), let us consider a Mayer-type optimal control problem. This has the form

\[
\text{minimize } l(x(T))
\]

over absolutely continuous functions \( x(\cdot) \) satisfying (CS), where \( l : \mathbb{R}^n \to \mathbb{R} \) is a predetermined function. As demonstrated above, under suitable hypotheses the (CS) and (DI) systems have the same trajectories, and since \( l \) is being minimized only over the endpoints of these trajectories, the explicit nature of the control functions play no essential role. In fact, in a theoretical sense they can complicate the analysis because it is difficult in general to immediately determine if there is a pair \((x(\cdot), u(\cdot))\) which provides the minimum. On the other hand, using (DI) as dynamics, and making the assumptions that \( F \) is compact and convex, one can deduce the existence of optimal solutions. Thus, it can be shown that solutions to the reformulation exist if \( F \) is defined as in (1.1) and \( F(x) \) is replaced by the closed convex hull of \( F(x) \) for each \( x \) (here the nature of the control set \( U \) and the continuity of \( \phi \) usually imply that \( F(x) \) is also bounded for each \( x \)). Let us then note that the compactness and convexity of the values of \( F \), especially as they arise in the context of control theory, is quite a natural assumption. But they are also natural for our purposes, as we shall see.
The study of multifunctions in their own right (not necessarily arising from control theory) has since become extensive. In fact, much mathematical literature has been devoted to multifunctions in general. A large portion of this attention has been devoted to finding selections and parametrizations of multifunctions under various hypotheses (selections and parametrizations are defined below). The advent of much of this attention was highlighted by E. Michael's continuous selection theorem \cite{10} in 1956. As optimal control theory burgeoned in the 1960's, the study of multifunctions in general became more popular. But we now turn our attention away from control theory and focus on the general notion of a multifunction.

Let \( F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n \) be a multifunction. The multifunction \( F \) is nonempty, closed, compact, or convex if at each \( x \in \mathbb{R}^n \), the set \( F(x) \) is nonempty, closed, compact, or convex, respectively. The multifunction \( F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n \) is Lipschitz on \( S \subseteq \mathbb{R}^n \) if there exists a positive scalar \( K \) such that for all \( x_1, x_2 \in S \) and all \( \xi_1 \in F(x_1) \), there exists \( \xi_2 \in F(x_2) \) such that

\[
|\xi_1 - \xi_2| \leq K|x_1 - x_2|.
\]

The multifunction \( F \) is locally Lipschitz at \( x \) if there is a neighborhood of \( x \) on which \( F \) is Lipschitz. We mention parenthetically that Lipschitzian properties and other properties regarding continuity of multifunctions can also be formulated using the Hausdorff metric, which is a metric on the compact subsets of \( \mathbb{R}^n \) (see, for example, \cite{1}). A multifunction \( F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n \) is locally bounded if for every \( x_0 \in \mathbb{R}^n \), there exist \( \epsilon > 0 \) and \( M > 0 \) such that \( |x - x_0| < \epsilon \) implies that \( |y| \leq M \) for all \( y \in F(x) \).

A thorough exposition of multifunctions can be found in \cite{1}.

A function \( f : \mathbb{R}^n \to \mathbb{R}^n \) is a selection of the multifunction \( F \) if \( f(x) \in F(x) \) for each \( x \). The function \( f \) is respectively a measurable selection, continuous selection, \( C^1 \) selection, etc., if the function \( f \) is measurable, continuous, \( C^1 \), etc. Let \( F : \mathbb{R} \rightrightarrows \mathbb{R} \) be a Lipschitz multifunction, which throughout will be assumed to have compact
convex values. Our first main result, Theorem 3.1, provides sufficient conditions for $F$ to admit a $C^1$ selection.

There is an extensive literature on selection theorems for multifunctions, but to our knowledge, there are no results giving conditions for a smooth (i.e. $C^1$) selection. On the other hand, the existing results can be obtained in quite general spaces. For example, Michael’s Continuous Selection Theorem [10] is valid for $F$ being defined on a complete metric space and mapping into the subsets of a Banach space. Note that $F$ is required to be compact and convex in Michael’s Continuous Selection Theorem. Similar general hypotheses are sufficient for many of the measurable-type selection theorems, as illustrated by the book by Castaing and Valadier [3], the survey by Wagner [18], and the many references contained therein. A highly readable treatment of the finite-dimensional case is given by Rockafellar [13]. There is also a Caratheodory-type selection result due to Lojasiewicz [8]. A Caratheodory function is a function $f : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ such that $f(\cdot, x)$ is measurable for each $x \in \mathbb{R}^n$ and $f(t, \cdot)$ is continuous for each $t \in \mathbb{R}$. The book of Aubin and Frankowska [1, Chapter 9] provides an excellent reference to these results, as well as to the parametrization results alluded to below.

A common method used to obtain the selections mentioned above is to obtain a sequence of approximate selections and then take a limit to obtain a selection with the desired properties. Mere generalizations of this approach do not seem to be adequate in obtaining a smooth selection of a multifunction, for the $C^1$ conclusion is difficult to obtain by a limiting argument. Our proof is based upon a discretization procedure, followed by an elaborate patching together of the pieces. (This helps to explain why the results are framed only in dimension one, where such discretizing and pasting is more readily realizable). Moreover, it is not a priori clear exactly which multifunctions should have $C^1$ selections, since there is no obvious set-valued analogue of classical differentiability; obviously, if the values of the multifunction
$F$ should "collapse" to a singleton on some neighborhood $I$, then $F$, viewed as an ordinary function, must be $C^1$ in the classical sense on $I$. The term "singleton" refers to a set consisting of a single point.

The Hamiltonian $H$ of a multifunction $F$ is given by

$$H(x, p) = \sup\{\langle v, p \rangle : v \in F(x)\}.$$ 

Our first main result, Theorem 3.1, asserts that if the Clarke gradient (defined in Chapter 2) of the Hamiltonian $H$ associated with $F$ is submonotone in the sense of Spingarn [17] (also defined in Chapter 2) as a function of $x$ for each $p = \pm 1$, then $F$ admits a $C^1$ selection.

We next discuss parametrizations of multifunctions, which are merely representations of $F$ as a compactly parametrized family of selections. We can write this as

$$F(x) = \{f(x, u) : u \in U\},$$

where $U$ is a compact topological space. A measurable parametrization of the multifunction $F$ is a representation of $F$ as in (1.2), where $f$ is measurable in $x$. A continuous parametrization of $F$ is a representation as in (1.2) where $f$ is jointly continuous in $(x, u)$. A Lipschitz parametrization of $F$ is a representation as in (1.2) where $f$ is Lipschitz in $x$ with Lipschitz rank $K$ independent of $u$ and moreover, $f$ is jointly continuous in $(x, u)$. A $C^1$ parametrization of $F$ is a representation as in (1.2) where $f$ is jointly continuous in $(x, u)$, $f$ is continuously differentiable in $x$ for each $u \in U$ (with derivative denoted by $f_x(\cdot, u)$), and this derivative $f_x(\cdot, \cdot)$ is also jointly continuous in $(x, u)$. Note that in all the parametrization types, $U$ is assumed to be compact. Our second main result, Theorem 3.2, characterizes those $F : \mathbb{R} \rightrightarrows \mathbb{R}$ that admit a $C^1$ parametrization. In this case, $U$ is a compact subset of an Euclidean space $\mathbb{R}^n$. 
For a very simple example of a $C^1$ parametrization, consider the following. Let $F(x)$ be defined for each $x \in [0,1]$ by $F(x) = \{y : 0 \leq y \leq 1\}$. For $x, u \in [0,1]$, let $f(x, u) = u$. Then for fixed $u$, $f(\cdot, u)$ is a $C^1$ selection of $F$. Moreover, the data $(f, U)$ provides a $C^1$ parametrization of $F$.

Again, an extensive literature exists for results (often in abstract spaces) characterizing those multifunctions $F$ that admit a measurable parametrization (Castaing [2]), a continuous parametrization (Ekeland and Valadier [6]), or a Caratheodory-type parametrization (Lojasiewicz [9], Ornelas [12]). As might be expected, multifunctions that admit a $C^1$ parametrization must have some special characteristics that correspond to $C^1$ behavior if the multifunction should collapse to a singleton. We illuminate some simple necessary conditions (these are proven in more detail in Chapter 5) that, as we shall see in Theorem 3.2, are in fact sufficient. Although the characterization in Theorem 3.2 is only in one dimension, the following necessary conditions are valid in higher dimensions as well.

Suppose $F : \mathbb{R}^n \to \mathbb{R}^n$ admits a $C^1$ parametrization as in (2.1), and let $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be its associated Hamiltonian. In this case, we have that $H$ is given by

$$H(x, p) = \sup \{ \langle f(x, u), p \rangle : u \in U \}$$

We are assuming that $f(\cdot, u)$ is differentiable for each $u \in U$ with the $x$-derivative continuous jointly in $(x, u)$. It is immediate that $H(\cdot, \cdot)$ is lower-$C^1$ in $(x, p)$ (hence strictly submonotone, see Theorem 2.6) in the sense of Spingarn [17] and Rockafellar [15] jointly as a function from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}$. (The definition of lower-$C^1$ is given below in Chapter 2). This condition imposed on $H$ is not in itself sufficient to ensure a parametrization, as a simple example illustrates (see Chapter 3). For another necessary condition of the given parametrization, notice (see (2.4) below)
that the partial Clarke gradient of $H$ with respect to $x$ can be calculated by

$$\partial_x H(x,p) = \text{co} \{ \nabla_x f(x,u) : u \in U \text{ satisfies } H(x,p) = f(x,u) \}, \quad (1.3)$$

where "co" denotes the convex hull. Then it follows from (1.3) that

$$H(x,p) = -H(x,-p) \implies \partial_x H(x,p) = -\partial_x H(x,-p). \quad (1.4)$$

Theorem 3.2 states that, in dimension $n = 1$ the necessary conditions of $\partial_x H(\cdot,p)$ being strictly submonotone and satisfying (1.4) are also sufficient for obtaining a $C^1$ parametrization.

It is instructive to compare our result with those in Spingarn [17], in which it is shown that lower-$C^1$ is equivalent to a strict submonotonicity condition. One of the difficulties encountered in parametrizing a multifunction, and which is not met in representing a lower-$C^1$ function, is to handle the multisided nature of multifunctions, and in particular, the problems this imposes when the multifunction collapses to a singleton. Condition (1.4) is the missing ingredient that allows one to construct a lower-$C^1$ representation of $H(x,p)$ that does not interfere with a corresponding lower-$C^1$ representation of $H(x,-p)$. We mention, however, that we are heavily in debt to some of the constructions in [17] that are made to produce lower-$C^1$ representations of functions.

The organization of the dissertation is as follows. Definitions and some preliminaries are given in Chapter 2. The main results and examples exemplifying the strength of the hypotheses are stated in Chapter 3. The proof of Theorems 3.1 and 3.2 are given Chapters 4 and 5, respectively.
CHAPTER 2
DEFINITIONS AND PRELIMINARY DISCUSSION

In this chapter, we define some tools of nonsmooth analysis, including Clarke generalized gradients, which are useful when analyzing functions in some particular function classes, which we shall subsequently define. We also review some previous work in nonsmooth analysis and prove a new result. We choose to work in \( n \)-dimensional space since the new result mentioned above is proven in that context. However, almost all of our subsequent applications are only in dimension 1.

Let \( f : \mathbb{R}^n \to \mathbb{R} \). The function \( f \) is locally Lipschitz on the set \( S \subseteq \mathbb{R}^n \) if for each \( x_0 \in S \), there exists a neighborhood \( U \) of \( x_0 \) and a positive scalar \( K \) such that
\[
|f(x) - f(y)| \leq K|x - y| \quad \text{for all } x, y \in U.
\]

If there is a global Lipschitz constant, \( K \), for \( f \) on \( S \), then \( f \) is simply called Lipschitz on \( S \). If \( f \) is locally Lipschitz on a compact set \( S \), then it is easily shown that \( f \) is actually Lipschitz on \( S \).

By Rademacher's Theorem, Lipschitz functions are differentiable almost everywhere (a.e.) with respect to Lebesgue measure. Not only may a Lipschitz function \( f \) fail to be differentiable at a point \( x \in \mathbb{R}^n \), but \( f \) may fail to have directional derivatives at \( x \). The directional derivative (when it exists) of \( f \) at \( x \) in the direction \( v \in \mathbb{R}^n \) is the quantity
\[
f'(x; v) = \lim_{t \downarrow 0} \frac{f(x + tv) - f(x)}{t}.
\]

Given a locally Lipschitz function \( f : \mathbb{R}^n \to \mathbb{R} \), the generalized directional derivative
(in the sense of Clarke [5]) at \( x \) in the direction \( v \in \mathbb{R}^n \) is defined by

\[
f^0(x; v) = \limsup_{y \to x} \frac{f(y + tv) - f(y)}{t}.
\]

For \( f : \mathbb{R}^n \to \mathbb{R} \) Lipschitz, \( v \mapsto f^0(x; v) \) exists finitely at each \( x \in \mathbb{R}^n \), is positively homogeneous, and is subadditive (see [5]). If \( f \) is \( C^1 \) (or even strictly differentiable), then \( f'(x; v) = f^0(x; v) \) for all \( x, v \in \mathbb{R}^n \).

Given a locally Lipschitz function \( f : \mathbb{R}^n \to \mathbb{R} \), the Clarke gradient of \( f \) at \( x \) is defined by

\[
\partial f(x) = \{ \xi \in \mathbb{R}^n : f^0(x; v) \geq \langle v, \xi \rangle \text{ for all } v \in \mathbb{R}^n \}.
\]

Here \( \langle \cdot, \cdot \rangle \) is the usual inner product on \( \mathbb{R}^n \). The function \( f^0(x; \cdot) \) has sufficient properties to ensure that \( \partial f(x) \) is a nonempty, convex, and compact subset of \( \mathbb{R}^n \) (see [5]). The Clarke gradient multifunction, \( \partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n \), of a Lipschitz function \( f \) is thus nonempty, closed, compact, and convex. Of course, if \( f \) is \( C^1 \) (or strictly differentiable), then \( \partial f(x) \) reduces to the singleton \( \{ \nabla f(x) \} = \partial f(x) \). Also as Clarke has shown, they are the usual subgradients of convex analysis when \( f \) is concave or convex.

For example, consider the function \( f(x) = |x| \). The multifunction \( \partial f(x) \) can be calculated to be

\[
\partial f(x) = \begin{cases} 
-1 & \text{if } x \in (-\infty, 0), \\
\{y : -1 \leq y \leq 1\} & \text{if } x = 0, \\
1 & \text{if } x \in (0, \infty). 
\end{cases}
\]

Let \( x \in \mathbb{R}^n \) and \( U \) be a neighborhood of \( x \). Let \( \Omega \) be the set of points of \( U \) at which the locally Lipschitz function \( f \) fails to be differentiable (recall that \( \Omega \) has Lebesgue measure 0 by Rademacher's Theorem). For \( x \in U \), a useful characterization
of \( \partial f(x) \) is the following (see [5]):

\[
\partial f(x) = \mathrm{co}\{ \lim_{i \to \infty} \nabla f(x_i) : x_i \to x, x_i \notin \Omega, \text{ and } x_i \notin S \} \tag{2.1}
\]

for any set \( S \) of zero measure. So \( \partial f(x) \) is the convex hull of limits of sequences \( \{\nabla f(x_i)\} \) where \( x_i \) avoids the set \( \Omega \cup S \).

There is a thoroughly developed calculus for Clarke gradients (see [5]), among which is a mean value theorem for Clarke gradients due to Lebourg [7]. The theorem holds in more general spaces, but we state it for \( \mathbb{R}^n \).

**Theorem 2.2.** Let \( x, y \in \mathbb{R}^n \) and suppose that \( f \) is Lipschitz on an open set containing the line segment \([x, y]\) (i.e. the convex hull of the two point set \( \{x, y\}\)). Then there exists a point \( v \in (x, y) \) such that

\[
f(y) - f(x) \in \langle \partial f(v), y - x \rangle.
\]

A function \( f : \mathbb{R}^n \to \mathbb{R} \) is called **lower-\( C^k \)** provided that for each \( x_0 \in \mathbb{R}^n \), there are a compact topological space \( S \), a neighborhood \( U \) of \( x_0 \), and a function \( g : U \times S \to \mathbb{R} \) so that \( f \) has the representation

\[
f(x) = \sup_{s \in S} g(x, s) \quad \text{for each } x \in U \tag{2.3}
\]

where \( g \) and all the partial derivatives of \( g \) up through order \( k \) with respect to the \( x \) variable are jointly continuous on \( U \times S \). For example, lower-\( C^1 \) functions have the representation (2.3) where both \( g(\cdot, \cdot) \) and \( g_x(\cdot, \cdot) \) are jointly continuous. Here \( g_x = \nabla_x g \), the gradient of \( g \) with respect to the \( x \) variable. In words, lower-\( C^k \) functions are obtained as the supremum of a compactly indexed family of \( C^k \)
functions which behave “smoothly” in the parameter variable. It can easily be shown that lower-$C^k$ functions are locally Lipschitz. A theorem of Clarke [4] allows one to obtain $\partial f(x)$ for a lower-$C^k$ function in the following manner.

$$\partial f(x) = \text{co}\{\nabla_x g(x, s) : \text{for all } s \in S \text{ such that } g(x, s) = f(x)\}. \quad (2.4)$$

Clarke did not study lower-$C^k$ functions explicitly. The theorem alluded to above deals with the general case of “max” functions represented as in (2.3) but with $g$ not necessarily differentiable with respect to $x$.

We next survey some properties of lower-$C^k$ functions. We include a result by Rockafellar which shows that the class of lower-$C^2$ functions coincides with the class of lower-$C^k$ functions for each $k > 2$ ($k \in \mathbb{N}$). However, we will focus primarily on lower-$C^1$ functions.

Henceforth $F : \mathbb{R}^n \to \mathbb{R}$ denotes a closed, convex, and locally bounded multifunction. The multifunction $F$ is strictly hypomonotone at $x_0$ if

$$\liminf_{\substack{x_1 \neq x_2 \\
\quad x_1 \to x_0, i=1,2 \\
\quad y_i \in F(x_i), i=1,2}} \frac{(x_1 - x_2, y_1 - y_2)}{|x_1 - x_2|^2} > -\infty.$$  

That all classes of lower-$C^k$ functions coincide for $k \geq 2$ is demonstrated by the following result of Rockafellar (see [15]).

**Theorem 2.5.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz. Then the following are equivalent:

(1) $f$ is lower-$C^2$ in a neighborhood of $x_0$.

(2) $\partial f$ is strictly hypomonotone at each $x_0$.

(3) For each $x_0 \in \mathbb{R}^n$, there is a neighborhood $U$ of $x_0$ and a representation of $f$ as in (2.3) with $S$ a compact topological space, $g(x, s)$ quadratic in the $x$ variable and continuous in the $s$ variable. \qed
Any representation of type 2.5(3) above is a special case of the kind of representation in the definition of $f$ being lower-$C^2$ (in fact lower-$C^\infty$). Hence the following (see [15]):

**Corollary 2.6.** If a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is lower-$C^2$, it is actually lower-$C^\infty$. Thus, for $2 \leq k \leq \infty$, the classes of lower-$C^k$ functions all coincide. □

Examples show however, that the class of lower-$C^1$ functions differs from the class of lower-$C^k$ functions when $k \geq 2$ (see [15]). We now turn our attention to the class of lower-$C^1$ functions. The multifunction $F$ is submonotone at $x_0 \in \mathbb{R}^n$ provided

$$\liminf_{z \rightarrow x_0, \, z \neq x_0 \atop y \in F(z), \, y \neq y_0} \frac{(y - y_0, x - x_0)}{|x - x_0|} \geq 0.$$  

The multifunction $F$ is strictly submonotone at $x_0 \in \mathbb{R}^n$ if

$$\liminf_{z \rightarrow x_0, \, z \neq x_0 \atop y_i \in F(x_i), \, i=1,2} \frac{(x_1 - x_2, y_1 - y_2)}{|x_1 - x_2|} \geq 0.$$  

Strict submonotonicity implies submonotonicity but not the converse (see [17] for an example). Submonotonicity is somewhat of a pointwise condition whereas strict submonotonicity is a local condition. Spingarn's main result in [17] is the following:

**Theorem 2.7.** Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be locally Lipschitz. Then $f$ is lower-$C^1$ if and only if $\partial f$ is strictly submonotone. □

We do not provide a proof for Theorem 2.7 as such. However, the construction Spingarn used to prove Theorem 2.7 will help us in attaining the parametrization in Theorem 3.2. Also, certain properties of submonotone multifunctions are featured in the proof of our selection theorem, Theorem 3.1. So we now illuminate the mechanics used by Spingarn in obtaining Theorem 2.7.
A locally Lipschitz function \( f : \mathbb{R}^n \to \mathbb{R} \) is **subdifferentially regular** (called regular in [5]) at \( x \in \mathbb{R}^n \) if \( f'(x; v) \) exists and \( f'(x; v) = f^o(x; v) \) for all \( v \in \mathbb{R}^n \). Subdifferential regularity allows "lower corners" but no "upper" ones. For a simple example in one dimension, \( |x| \) is subdifferentially regular at \( x = 0 \), but \(-|x|\) is not. Note that any convex or \( C^1 \) function is subdifferentially regular. Many properties of subdifferentially regular functions are collected in [5]. A useful characterization of subdifferentially regular functions in terms of Clarke gradients which is easily deduced is the following: The function \( f : \mathbb{R}^n \to \mathbb{R} \) is subdifferentially regular at \( x \in \mathbb{R}^n \) if and only if

\[
f'(x; v) = \sup_{\xi \in \partial f(x)} \langle \xi, v \rangle \quad \text{for each } v \in \mathbb{R}^n. \tag{2.8}
\]

We state the special case of a theorem due to Clarke [4].

**Theorem 2.9.** If \( f \) is lower-\( C^k \) \((k \geq 1)\), then \( f \) is not only locally Lipschitz but also subdifferentially regular.  

So, in particular, lower-\( C^1 \) functions are subdifferentially regular.

A function \( f : \mathbb{R}^n \to \mathbb{R} \) is **upper semi-continuous** at \( x \in \mathbb{R}^n \) if

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

The function \( f \) is simply u.s.c., if \( f \) is u.s.c. at each \( x \in \mathbb{R}^n \). The following theorem is due to Rockafellar [15].

**Theorem 2.10.** Let \( f : \mathbb{R}^n \to \mathbb{R} \). Then \( f \) is locally Lipschitz and subdifferentially regular if and only if \( f'(x; v) \) exists finitely for all \( x, v \in \mathbb{R}^n \) and \( x \mapsto f'(x; v) \) is upper semi-continuous.
A sequence \( \{x_n\} \) of points of \( \mathbb{R}^n \) is said to converge to \( x \) in the direction \( v \in \mathbb{R}^n \), written \( x_n \overset{v}{\to} x \), provided \( x_n \to x \),

\[
\frac{x_n - x}{|x_n - x|} \to \frac{v}{|v|},
\]
as \( n \to \infty \), and \( x_n \neq x \) for all large \( n \). A locally Lipschitz function \( f : \mathbb{R}^n \to \mathbb{R} \) is semismooth at \( x \), (as defined by Mifflin [11]), provided that for all \( v \in \mathbb{R}^n \), if \( x_n \overset{v}{\to} x \) and \( y_n \in \partial f(x_n) \), then \( (v, y_n) \to f'(x; v) \). In the case of differentiable functions, semismoothness just implies that the derivative is continuous at \( x \). For nonsmooth functions, semismoothness implies that if “corners” bunch up, then they must “flatten” out in doing so. The following theorem is due to Spingarn [17].

**Theorem 2.11.** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be locally Lipschitz. Then \( \partial f \) is submonotone at \( x \) if and only if \( f \) is semismooth and subdifferentially regular at \( x \). \( \square \)

Theorems 2.10 and 2.11 will be of considerable help in proving our results. However, the main machinery we need in proving Theorem 3.2 comes from the strict submonotonicity condition. Although Theorem 2.7 helped motivate us to undertake the parametrization problem, its statement will not be of use in our endeavors. Rather we will use the Spingarn construction, in its proof, which is born from the strict submonotonicity condition. Much of the remainder of this section, modulo some minor alterations, is due to Spingarn [17]. The closed unit ball in \( \mathbb{R}^n \) is denoted by \( B^n = \{x \in \mathbb{R}^n : |x| \leq 1\} \). If \( K \subset \mathbb{R}^n \) is compact and convex, then \( \Psi^*_K \) is the support function of \( K \) defined by

\[
\Psi^*_K(u) = \sup_{x \in K} \langle u, x \rangle.
\]

Note that if \( K = \partial f(x) \), then \( \Psi^*_K(u) = f^\circ(x; u) \).
Lemma 2.12. Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz and $x, y \in \mathbb{R}^n$. For every $\epsilon > 0$, there are neighborhoods $U$ of $x$ and $V$ of $y$ such that if $x' \in U$ and $y' \in V$, then

$$|\Psi^*_f(x')(y) - \Psi^*_f(x')(y')| \leq \epsilon.$$

**Proof.**

Let $k$ be a Lipschitz constant for $f$ on a neighborhood $U$ of $x$. Then $\partial f(x') \subset kB^n$ for all $x' \in U$. It follows that $k$ is a global Lipschitz constant for $\Psi^*_f(x')(\cdot)$. Take $V$ to be the open ball of radius $\frac{\epsilon}{k}$ centered at $y$.

Lemma 2.13. Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz. Then

$$\liminf_{x' \to x \atop t \to 0} \frac{f(x' + ty) - f(x')}{t} - \Psi^*_f(x') (y) \geq 0 \quad \text{for all } y \in \mathbb{R}^n \tag{2.14}$$

if and only if, for any compact $K \subset \mathbb{R}^n$ and any $\epsilon > 0$, there is a neighborhood $U$ of $x$ and $\lambda > 0$ such that

$$\frac{f(x' + ty') - f(x')}{t} - \Psi^*_f(x') (y') \geq -\epsilon \tag{2.15}$$

whenever $x' \in U$, $y' \in K$ and $0 < t < \lambda$.

**Proof.**

Assume (2.14) holds. Fix a compact $K \subset \mathbb{R}^n$ and any $\epsilon > 0$. Since $f$ is locally Lipschitz, (2.14) implies

$$\liminf_{x' \to x \atop y' \to y \atop t \to 0} \frac{f(x' + ty') - f(x')}{t} - \Psi^*_f(x') (y) \geq 0 \quad \text{for all } y \in \mathbb{R}^n.$$
This and Lemma 2.12 imply that for each \( y \in K \) we may find neighborhoods \( U_y \) of \( x \), \( V_y \) of \( y \), and a scalar \( \lambda_y > 0 \) such that \( \Psi^*_{\partial f(x')} (y) - \Psi^*_{\partial f(x')} (y') \geq \frac{-\varepsilon}{2} \) and

\[
\frac{f(x' + ty') - f(x')}{t} - \Psi^*_{\partial f(x')} (y) \geq \frac{-\varepsilon}{2}
\]

whenever \( x' \in U_y \), \( y' \in V_y \), and \( 0 < t < \lambda_y \). Pick a finite subcover \( V_{y_1}, \ldots, V_{y_m} \) of \( K \), and let \( U = V_{y_1} \cap \cdots \cap V_{y_m} \). Let \( \lambda = \min\{\lambda_{y_1}, \ldots, \lambda_{y_m}\} \). For any \( x' \in U \), \( y' \in K \), and \( t \in (0, \lambda) \), let \( i \) be such that \( y' \in V_{y_i} \). Then we get

\[
\frac{f(x' + ty') - f(x')}{t} - \Psi^*_{\partial f(x')} (y') \\
= \left( \frac{f(x' + ty') - f(x')}{t} - \Psi^*_{\partial f(x')} (y_i) \right) \\
+ \left( \Psi^*_{\partial f(x')} (y_i) - \Psi^*_{\partial f(x')} (y') \right) \\
\geq -\frac{\varepsilon}{2} - \frac{\varepsilon}{2} = -\varepsilon,
\]

as desired. The opposite direction of the lemma is obvious. \( \Box \)

**Proposition 2.16.** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be locally Lipschitz. Then \( \partial f \) is strictly submonotone at \( x \) if and only if (2.14) holds.

**Proof.**

\((\Rightarrow)\) If \( y = 0 \) the assertion is trivial. Assume without loss of generality that \( |y| = 1 \). Fix \( \varepsilon > 0 \). Since \( \partial f \) is strictly submonotone at \( x \), there is \( r > 0 \) such that

\[
\frac{(x_1 - x_2, y_1 - y_2)}{|x_1 - x_2|} \geq -\varepsilon
\]

whenever \( |x_i - x| \leq 2r \), \( y_i \in \partial f(x_i) \), for \( i = 1,2 \), and \( x_1 \neq x_2 \). Let \( x' \) and \( t \) be chosen so that \( |x' - x| < r \) and \( t < r \). We will complete the proof by showing that

\[
\frac{f(x' + ty) - f(x')}{t} - \Psi^*_{\partial f(x')} (y) \geq -\varepsilon.
\]
By the mean value theorem for Clarke gradients (Theorem 2.2), we may find \( s \in (0, t) \) and \( y_2 \in \partial f(x' + sy) \) such that \( f(x' + ty) - f(x') = t(y, y_2) \). Letting \( x_1 = x' \) and \( x_2 = x' + sy \), we have

\[
\frac{f(x' + ty) - f(x')}{t} - \Psi^*_{\partial f(x')}(y) = \langle y, y_2 - y_1 \rangle \\
= \frac{(x_2 - x_1, y_2 - y_1)}{|x_2 - x_1|} \geq -\varepsilon.
\]

\((\Leftarrow)\) Next, suppose (2.14) holds, and let \( \varepsilon > 0 \) be given. By Lemma 2.13, there is a neighborhood \( U \) of \( x \) and \( \lambda > 0 \) such that

\[
\frac{f(x' + tu) - f(x')}{t} - \Psi^*_{\partial f(x')}(u) \geq -\frac{\varepsilon}{2}
\]

whenever \( x' \in U \), \( |u| \leq 1 \), and \( 0 < t < \lambda \). We may also assume that \( U \) is small enough so that \( |z - z'| < \lambda \) for all \( z, z' \in U \). Fix \( x_1 \in U \), \( y_i \in \partial f(x_i) \) for \( i = 1, 2 \), with \( x_1 \neq x_2 \). Let \( t = |x_2 - x_1| \) and \( u = \frac{(x_2 - x_1)}{t} \). Then

\[
\frac{(x_1 - x_2, y_1 - y_2)}{|x_1 - x_2|} = -\langle u, y_1 \rangle - \langle -u, y_2 \rangle
\]

\[
\geq -\Psi^*_{\partial f(x_1)}(u) - \Psi^*_{\partial f(x_2)}(-u)
\]

\[
= \frac{f(x_1 + tu) - f(x_1)}{t} - \Psi^*_{\partial f(x_1)}(u)
\]

\[
+ \frac{f(x_2 - tu) - f(x_2)}{t} - \Psi^*_{\partial f(x_2)}(-u)
\]

\[
\geq -\frac{\varepsilon}{2} - \frac{\varepsilon}{2} = -\varepsilon,
\]

which shows that \( \partial f \) is strictly submonotone at \( x \). \( \square \)

**Lemma 2.17.** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be locally Lipschitz. Let \( C \) and \( K \) be compact sets in \( \mathbb{R}^n \), and suppose that \( \partial f \) is strictly submonotone on \( C \). Then

\[
\liminf_{x \in C, y \in K, t \downarrow 0} \frac{f(x + ty) - f(x)}{t} - \Psi^*_{\partial f(x)}(y) \geq 0.
\]
Proof.

Let $\epsilon > 0$ be given. By Proposition 2.16 and Lemma 2.13, for each $x \in C$, there is a $\lambda_x > 0$ such that

$$\frac{f(x' + ty) - f(x')}{t} - \Psi_{\partial f(x')}^*(y) \geq -\epsilon$$

whenever $|x' - x| < \lambda_x$, $y \in K$, and $0 < t < \lambda_x$. Let $x_1, ..., x_r \in C$ be such that for every $x \in C$ we have $|x - x_i| < \lambda_{x_i}$ for some $i$. Let $\lambda = \min\{\lambda_{x_1}, ..., \lambda_{x_r}\}$. Then, for any $x \in C$ and $y \in K$,

$$\frac{f(x + ty) - f(x)}{t} - \Psi_{\partial f(x)}^*(y) \geq -\epsilon$$

whenever $0 < t < \lambda$. \hfill \Box

Lemma 2.18. Let $\phi : [0, 1] \rightarrow \mathbb{R}$. Suppose that $\phi$ is bounded above on $[0, 1]$ and $\lim_{t \to 0} \phi(t) = 0$. Then there is a $C^1$ function $\alpha(t)$ defined on $[0, 1]$ such that

1. $\alpha(0) = \alpha'(0) = 0$
2. $\alpha(t) > t\phi(t)$ for all $t \in (0, 1]$.

Proof.

Let $M$ be an upper bound for $\phi$ on $[0, 1]$. Extend $\phi$ to $[0, 2]$ by setting $\phi(t) = M$ for each $t \in (1, 2]$. For $k = 0, 1, 2, ...$ let $a_k = \frac{1}{2k}$. Let $\beta$ be the infimum of all affine functions $l : \mathbb{R} \to \mathbb{R}$ which satisfy $l(a_k) \geq \phi(t)$ for all $t \in (0, 2a_k]$ and $k = 0, 1, 2, ...$ Then the following properties are easily checked:

- $\beta$ is continuous, concave, and nondecreasing on $[0, 1]$,
- $\beta(0) = 0$,
- $\beta(t) \geq \phi(t)$ for all $t \in (0, 1]$,
- $\beta$ is affine on $[a_{k+1}, a_k]$, $k = 0, 1, 2, ...$
Also, $\beta'_+, \text{ the right derivative of } \beta$, has these properties:

$\beta'_+$ is finite, nonnegative, and nonincreasing on $(0, 1)$,
$\beta'_+$ is constant on $[a_{k+1}, a_k)$, $k = 0, 1, 2, ...$
$\beta'_+$ is integrable on $[0, 1]$

The last assertion is proven as follows. Whenever $0 < u < v < 1$,

$$\beta(v) - \beta(u) = \int_u^v \beta'_+(s) \, ds$$

(see [14, Chap.24.2.1]). Since $\beta'_+ \geq 0$ and $\beta$ is continuous,

$$\int_0^1 \beta'_+(s) \, ds = \lim_{u \to 0^+} \int_u^v \beta'_+(s) \, ds = \beta(1) - \beta(0) < \infty.$$ 

So $\beta$ is integrable. Note that $\beta(t) = \int_0^t \beta'_+(s) \, ds$ for all $t \in [0, 1]$ since $\beta(0) = 0$.

For each $k = 0, 1, 2, ...$, pick $c_k$ such that $\frac{1}{2}(a_k + a_{k+1}) < c_k < a_k$, and $(a_k - c_k)(\beta'_+(a_{k+1}) - \beta'_+(a_k)) < a_{k+1}$. Define $\mu : (0, 1) \to \mathbb{R}$ to be a function that agrees with $1 + \beta'_+$ on the intervals $[a_{k+1}, c_k]$ and on $[a_1, a_0)$, and is affine on the intervals $[c_k, a_k]$. Then $\mu$ is continuous, nonnegative, and nonincreasing on $(0, 1)$. Moreover,

$$\int_{a_{k+1}}^t \mu(s) - \beta'_+(s) \, ds \geq 0 \quad \text{for all } k = 0, 1, 2, ... \text{ and } t \in [a_{k+1}, a_k].$$

Since $0 \leq \mu \leq \beta'_+ + 1$ and $\beta'_+$ is integrable, it follows that $\mu$ is integrable. Then for all $t \in [0, 1]$,

$$\int_0^t \mu(s) \, ds \geq \int_0^t \beta'_+(s) \, ds = \beta(t).$$

Define $\alpha(t) = t \int_0^t \mu(s) \, ds$ for all $t \in [0, 1]$. Clearly, the following properties hold:

$\alpha$ is $C^1$ on $(0, 1)$,
$\alpha(0) = 0$, and
$\alpha(t) \geq t\phi(t)$ for all $t \in (0, 1]$. 

We now show that $\alpha$ is $C^1$ at 0. We have

$$\alpha'(0) = \lim_{t \to 0} \frac{\alpha(t)}{t} = \lim_{t \to 0} \int_0^t \mu(s) \, ds = 0.$$  

Also, for $t > 0$,

$$\alpha'(t) = \int_0^t \mu(s) \, ds + t \mu(t) = \int_0^t (\mu(s) + \mu(t)) \, ds$$

$$\leq 2 \int_0^t \mu(s) \, ds \quad \text{(since $\mu$ is nondecreasing)}.$$  

So $\lim_{t \to 0} \alpha'(t) = 0$.

Thus, $\alpha$ is the desired function except that $\alpha$ satisfies $\alpha(t) \geq t \phi(t)$ rather than the strict inequality of 2.18(2). To remedy this, define $\hat{\alpha}(t) = t^2 + \alpha(t)$. Clearly then $\hat{\alpha}$ is $C^1$ on $[0,1]$ and satisfies 2.18(1). Moreover, $\hat{\alpha}(t) = t^2 + \alpha(t) > \alpha(t) \geq t \phi(t)$ for all $t \in (0,1]$ satisfying 2.18(2). \qed

Consider a multifunction $F : \mathbb{R}^n \to \mathbb{R}^n$ and its associated Hamiltonian $H(x,p) = \sup_{y \in F(x)} \langle y, p \rangle$. If $F$ can be parametrized with $C^1$ functions, then $H$ is lower-$C^1$ as a function from $\mathbb{R}^{2n}$ to $\mathbb{R}$ (see Proof of Theorem 3.2 (necessity)). To characterize those $F$ which admit a $C^1$ parametrization, the multifunction $(x,p) \mapsto \partial H(x,p)$ being strictly submonotone is a necessary condition, and thus is a natural assumption to make. However, our results are framed in one dimension, and so the values $p = \pm 1$ in the Hamiltonian are sufficient to use in describing $F$. In higher dimensions, the $p$ variable in the Hamiltonian plays a more significant role in describing $F$ since the set $\{p \in \mathbb{R}^n : |p| = 1\}$ is no longer topologically discrete. So, in one dimension, the hypothesis of $\partial H$ being strictly submonotone jointly in $(x,p)$ is replaced by each of the multifunctions $\partial H(x,1)$ and $\partial H(x,-1)$ being strictly submonotone. To be able to extend our results to higher dimensions, we suspect that $\partial H$ being strictly submonotone jointly in $(x,p)$ would play a significant role. We next elaborate further on this point.
Roughly speaking, Spingarn obtained the lower-$C^1$ representation of a function $f$ whose Clarke gradient is strictly submonotone as follows. For $\xi_0 \in \partial f(x_0)$, the function $g(x) = f(x_0) + \langle x - x_0, \xi_0 \rangle$ is the tangent hyperplane of $f$ at $x_0$ with normal $\xi_0$. Clearly, $g(x_0) = f(x_0)$, but there is no guarantee that $g(x) \leq f(x)$ for any $x \neq x_0$ (unless $f$ is convex). To obtain the lower-$C^1$ representation of $f$, Spingarn contrived the existence of a $C^1$ function $\alpha$ such that $\tilde{g}(x) = f(x_0) + \langle x - x_0, \xi_0 \rangle - \alpha(|x - x_0|)$ stays below $f(x)$ for all $x$. In words, the function $\alpha$ curves the hyperplane $g$ enough to keep it below $f$. Moreover, the function $\alpha$ is fixed for all $x_0$ in a compact set and all $\xi_0 \in \partial f(x_0)$. For a suitable choice of $\phi$, Lemma 2.18 gives the proper $\alpha$.

As Spingarn’s construction does not deal with the nuance of having the extra $p$ variable, which must be considered separately from the $x$ variable, we alter his construction at this point by adding an extra parameter variable $p$, which is natural viewed from the Hamiltonian context. Although, as noted above, the $p$ variable in the Hamiltonian does not play a significant role in one dimension, we offer the following result in higher dimensions because it could be valuable in further efforts to parametrize multifunctions with $C^1$ functions in higher dimensions.

To parametrize the multifunction $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ with $C^1$ functions one would perhaps start by obtaining a lower-$C^1$ representation in the $x$ variable of $H(x,p)$ for each $|p| = 1$. To use Spingarn’s construction as discussed above for a fixed $p_0$, one merely needs the function $\alpha$ to curve the hyperplane $g(x) = H(x_0, p_0) + \langle x - x_0, \xi_0 \rangle$ where $\xi_0 \in \partial_x H(x_0, p)$. However, for $p \neq p_0$, the same alpha will in general not give the lower-$C^1$ representation of $H(\cdot, p)$. Our result below gives the lower-$C^1$ representation in the $x$ variable of $H(x, p)$, via the function $\alpha$ alluded to above, where the choice of $\alpha$ is independent of the choice of $p$ from a compact set.

In the following $\partial_x H(x, p)$ denotes the the Clarke gradient of the function $H(\cdot, p)$ with respect to the $x$ variable. The multifunction $\partial H(x, p)$ denotes the the Clarke gradient of $H$ with respect to the $(x, p)$ variables (recall that we are assuming $F$ to
be Lipschitz, from which it follows that $H$ is a locally Lipschitz function). For each $(x, p)$, $\partial H(x, p)$ is then a subset of $\mathbb{R}^{2n}$. Define the projection $\pi_x \partial H(x, p)$ as the set

$$\{ \xi \in \mathbb{R}^n : (\xi, \mu) \in \partial H(x, p) \text{ for some } \mu \in \mathbb{R}^n \}.$$

Clarke showed in [5, Proposition 2.1.16] that

$$\partial_x H(x, p) \subseteq \pi_x \partial H(x, p). \quad (2.19)$$

In the following, let $B^n$ denote the closed unit ball in $\mathbb{R}^n$ and $S^{n-1}$ the unit sphere in $\mathbb{R}^n$.

**Proposition 2.20.** Suppose that $H : B^n \times \mathbb{R}^n \to \mathbb{R}$ is such that $\partial H(\cdot, \cdot)$ is strictly submonotone. Then, there is a $C^1$ function $\alpha : [0, 1] \to \mathbb{R}$ such that $\alpha(0) = \alpha'(0) = 0$ and

$$H(x + ty, p) \geq H(x, p) + t \langle \xi, y \rangle - \alpha(t)$$

whenever $0 < t \leq 1$, $y, p \in S^{n-1}$, $x \in B^n$, and $\xi \in \partial_x H(x, p)$.

**Proof.**

Let the compact set $K \subset \mathbb{R}^{2n}$ be defined by $K = \{(y, 0) : y \in \mathbb{R}^n \text{ and } |y| = 1\}$. Also note that $C = B^n \times S^{n-1}$ is compact in $\mathbb{R}^{2n}$. For each $v \in \partial H(x, p)$ we write $v = (\xi, \mu)$ where $\xi, \mu \in \mathbb{R}^n$ (here $\xi \in \pi_1 \partial H(x, p)$). Employing Lemma 2.17 to obtain
the first line, and using 2.19 to obtain line 3 below

\[
\lim_{t \downarrow 0} \sup_{(x,p) \in C} \left\{ \sup_{y \in \partial H(x,p)} \langle \xi, y \rangle - \frac{H((x,p) + ty') - H(x,p)}{t} \right\} \leq 0
\]

\[
\iff \lim_{t \downarrow 0} \sup_{(x,p) \in C} \left\{ \sup_{y \in S^{n-1}} \langle \xi, y \rangle - \frac{H(x + ty, p) - H(x,p)}{t} \right\} \leq 0
\]

\[
\iff \lim_{t \downarrow 0} \sup_{(x,p) \in C} \left\{ \sup_{y \in S^{n-1}} \langle \xi, y \rangle - \frac{H(x + ty, p) - H(x,p)}{t} \right\} \leq 0
\]

for each \( \xi \in \partial_x H(x,p) \). Now define

\[
\phi(t) = \sup_{x \in B^n} \max_{y,p \in S^{n-1}} \left\{ \langle \xi, y \rangle - \frac{H(x + ty, p) - H(x,p)}{t} \right\}. \]

Then from the above calculation,

\[
\lim_{t \downarrow 0} \phi(t) = 0.
\]

So we may apply Lemma 2.18 obtaining a \( C^1 \) function \( \alpha : [0,1] \rightarrow \mathbb{R} \) such that \( \alpha(0) = \alpha'(0) = 0 \) and \( \alpha(t) > t \phi(t) \) for all \( t > 0 \). Then, for each \( 0 < t \leq 1, x \in B^n, y,p \in S^{n-1}, \) and each \( \xi \in \partial_x H(x,p) \) we have

\[
\alpha(t) > t \langle \xi, y \rangle - H(x + ty, p) + H(x,p)
\]

from which the conclusion follows. \( \Box \)
Remark.

(1) In Proposition 2.40, \( x \in B^n \) and \( p \in S^{n-1} \) could be replaced by \( x \) and \( p \) belonging to any compact sets.

(2) Proposition 2.40 holds for any \( H \) (\( H \) not necessarily the Hamiltonian) if \( (x, p) \mapsto H(x, p) \) is lower \( C^1 \).
Let $I$ be the closed interval $[0,1]$. Throughout the rest of this dissertation, we will consider a Lipschitz multifunction $F : I \rightrightarrows \mathbb{R}$ with nonempty, compact, and convex values. Specifically, then, the multifunction assigns to each $x \in I$ a nonempty (but possibly degenerate) closed bounded interval. Thus the values of the multifunction $F$ can be described by writing $F(x) = [h(x), H(x)]$ for each $x \in I$, where $h$ and $H$ are real-valued functions satisfying $h(x) < H(x)$ for $x \in I$. The Lipschitz assumption on the multifunction $F$ translates into nothing more than that each of the functions $h$ and $H$ are Lipschitz continuous in the ordinary sense of functions.

One of the many simplifications in working in one dimension is that only the two functions $h$ and $H$ are required to know everything about $F$. In higher dimensions, where convex sets can be quite complicated, one needs to consider the support functions in all directions to recapture a convex set. A convenient way to express this is in terms of the Hamiltonian function $H$. Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a given multifunction, and recall that the Hamiltonian $H : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ associated to $F$ is defined as

$$H(x,p) = \sup_{y \in F(x)} \langle y, p \rangle.$$

Then if $F$ has nonempty, compact, and convex values, the relationship

$$F(x) = \{ y \in \mathbb{R}^n : \langle y, p \rangle \leq H(x, p) \text{ for all } p \in \mathbb{R}^n \}$$

holds. Since we shall be working exclusively in dimension $n = 1$, only the values $p = \pm 1$ in the Hamiltonian are needed to describe the multifunction; one has
$H(x) := H(x, 1)$ describing the "top" and $h(x) := -H(x, -1)$ describing the "bottom." Thus in the statements of our results, we shall use only the functions $h$ and $H$. It seems reasonable to conjecture that our main theorems will remain true in higher dimensional spaces if the assumptions stated below for only the two functions $−h$ and $H$ are rather assumed to hold for all of the functions $x \to H(x, p)$, where $p$ is a unit vector in $\mathbb{R}^n$.

We now are ready to state the main results of this dissertation. The first is Theorem 3.1, which gives a sufficient condition for $F$ to admit a $C^1$ selection. The second is Theorem 3.2, which gives necessary and sufficient conditions for $F$ to admit a $C^1$ parametrization.

**Theorem 3.1.** If both of the subgradient multifunctions $\partial H(\cdot)$ and $\partial (−h)(\cdot)$ are submonotone, then $F$ admits a $C^1$ selection.

**Theorem 3.2.** The multifunction $F$ admits a $C^1$ parametrization if and only if both of the following hold:

1. both of the subgradient multifunctions $\partial H(\cdot)$ and $\partial (−h)(\cdot)$ are strictly sub-monotone,
2. if $x \in I$ satisfies $h(x) = H(x)$, then $\partial H(x) = \partial h(x)$

The proofs of Theorem 3.1 and 3.2 are given in Chapters 4 and 5, respectively. The proof of Theorem 3.2 depends quite heavily upon particular selections that will be realized from Theorem 3.1.

We next discuss under what kind of assumptions multifunctions might admit a $C^1$ selection. Suppose for example that $F$ is Lipschitz, and each of its values is convex with nonempty interior. For such an $F$, there is a "tube" of values $(f(x) - \epsilon, f(x) + \epsilon) \subseteq F(x)$ where $f(x)$ is continuous. Such an $F$ is uninteresting in this context because for every such tube, there is a $C^1$ function $g(\cdot)$ with $g(x) \in (f(x) - \epsilon, f(x) + \epsilon)$ for all $x \in I$, and hence the multifunction admits a $C^1$ selection.
trivially. It is much more intriguing to allow multifunctions to take on singleton values. A natural starting place is to re-examine Michael's Continuous Selection Theorem ([10]), which says that lower-semicontinuous multifunctions (with closed and convex values) admit continuous selections. Recall that a multifunction $F$ is called lower-semicontinuous at $x$ if for any $y \in F(x)$ and for any sequence $x_n$ converging to $x$, there exists a sequence of elements $y_n \in F(x_n)$ which converge to $y$. The following simple example shows that this assumption is not sufficient to insure a $C^1$ selection, even though almost all of the values of $F$ have an interior.

**Example.** Consider the following multifunction defined on $[-1,1]$. Let

$$F(x) = \begin{cases} 
|x|, & \text{if } x = \pm \frac{1}{n} \text{ for } n \in \mathbb{N}, \\
[0,1], & \text{if } x \neq \pm \frac{1}{n} \text{ for } n \in \mathbb{N}.
\end{cases}$$

It is easy to see that $F$ is lower semicontinuous. Moreover, any selection $f$ of $F$ must satisfy $f(\pm \frac{1}{n}) = \frac{1}{|n|}$ for each $n \in \mathbb{N}$, and thus clearly cannot be differentiable at 0. \qed

We now further examine the strength of the hypotheses in Theorem 3.1. First recall (Theorem 2.11) that the submonotonicity condition on the subgradient multifunction $\partial H(\cdot)$ is equivalent to $H(\cdot)$ being semismooth and subdifferentially regular, and a similar statement holds with regard to $-h$. The first example provides a multifunction $F$ for which $H(\cdot)$ and $-h(\cdot)$ are semismooth, but $F$ does not admit a $C^1$ selection.

**Example.** Consider the following multifunction defined on $[-1,1]$. For $x \in [-1,1]$ let

$$F(x) = \{ y : |x| \leq y \leq 2|x| \}.$$ 

Then $F$ does not admit a $C^1$ selection, due to the nature of the corner at 0. Indeed,
if \( f(x) \in F(x) \) for all \( x \in [-1,1] \), then
\[
\limsup_{x \to 0^-} \frac{f(x) - f(0)}{x} \leq -1 < 1 \leq \liminf_{x \to 0^+} \frac{f(x) - f(0)}{x},
\]
and thus cannot be differentiable at 0. The only hypothesis not satisfied in Theorem 3.1 is that \(-h(x) = -|x|\) is not subdifferentially regular at \( x = 0 \). □

Next we show the existence of a multifunction \( F \) for which \( H(\cdot) \) and \(-h(\cdot)\) are both subdifferentially regular, but still \( F \) does not admit a \( C^1 \) selection due to the lack of semismoothness.

**Example.** In the following, the multifunction \( F \) will be defined on the interval \([-1,1]\). We first let
\[
F(x) := [x,-x] \quad \text{whenever} \quad x \in [-1,0].
\]
Whenever \( x \in (0,1] \), the value \( F(x) \) will be a singleton (say, \( F(x) = \{ f(x) \} \)), but is specifically somewhat difficult to describe (we do so below). The properties we seek for \( f \) is that it be \( C^1 \) on \((0,1]\), with its derivative \( f'(x) \) taking on both the values of 0 and 1 for arbitrarily small \( x > 0 \), but in such a manner that the limit
\[
\lim_{x \to 0^+} \frac{f(x)}{x}
\]
exists and equals one. Suppose we have such an \( f \), then it is clear that the multifunction \( F \) can have no \( C^1 \) selection, because any selection will have an oscillating derivative as \( x \to 0^+ \). But it can be verified that both \( H \) and \(-h\) are subdifferentially regular (when \( x \neq 0 \), this is trivial; the case of \( x = 0 \) is somewhat less obvious, but still easy to verify in lieu of the representation of Clarke gradients in (2.1)). The hypotheses of Theorem 3.1 not satisfied here is that \( H \) and \(-h\) are not semismooth.
We now describe a function $f$ as indicated above. First, we define $f(x)$ for some values of $x$. Let

$$f(x) = \begin{cases} \frac{3}{16} & \text{for } x \in \left[\frac{1}{4}, 1\right] \\ x - x^2 & \text{for } x = \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \ldots \end{cases}$$

We next describe the derivative $f'(x)$ of $f$ on certain intervals. The function $f$ itself is then found by integrating. The slightly complicating aspect of this procedure is that it must be ensured that these integrations agree with the above values already assigned to $f$. For $n = 5, 7, 9, \ldots$, $f'(\cdot)$ is continuous and decreases from 1 to 0 on the interval $\left[\frac{1}{n+1}, \frac{1}{n}\right]$. For $n = 5, 7, 9, \ldots$, $f'(x) = 1$ whenever $x \in \left[\frac{n^2+n^2+4n+2}{n^2(n+1)^2}, \frac{1}{n}\right]$, and $f'$ increases linearly from 0 to 1 on the interval $\left[\frac{1}{n+1}, \frac{n^2+n^2+4n+2}{n^2(n+1)^2}\right]$. Note that $\frac{1}{n+1} < \frac{n^2+n^2+4n+2}{n^2(n+1)^2} < \frac{1}{n}$, so the latter is well defined.

To reiterate, we have demonstrated with the previous two examples that neither of the conditions, subdifferential regularity nor semismoothness, alone are sufficient to obtain a $C^1$ selection.

We now consider some of the difficulties in obtaining a parametrization. A reasonable starting place is the hypothesis of Theorem 3.1, since in essence, a parametrization is a bunch of selections. Let us again recall Spingarn’s result [17], which says that the Lipschitz functions $g$ (defined on $\mathbb{R}^n$) that admit a representation of the form

$$g(x) = \sup \{ f(x, u) : u \in U \}, \tag{3.3}$$

where $U$ is a compact subset of a metric space, and $f$ and $f_x$ are both continuous jointly in $(x, u)$, are precisely those functions $g$ for which the subgradient multifunction $\partial g$ is strictly submonotone. If we obtain a $C^1$ parametrization for $F$, then as both $H$ and $-h$ will subsequently have the representation (3.3), we must at
the very least have the submonotonicity conditions in Theorem 3.1 strengthened to strict submonotonicity. So the hypothesis 3.2(1) is natural.

We have already indicated in the Introduction that hypothesis 3.2(2) is necessary for \( F \) to be \( C^1 \) parametrized, but we further illuminate its nature in the following example.

**Example.** Consider the multifunction \( F \) defined by

\[
F(x) = \{ y : 0 \leq y \leq |x| \}.
\]

Suppose \( F \) admits a \( C^1 \) parametrization

\[
F(x) = \{ f(x, u) : u \in U \}.
\]

Then in particular there are \( C^1 \) selections \( f(\cdot, u_n) \) of \( F \) such that \( f(\frac{1}{n}, u_n) = \frac{1}{n} \) for each \( n \in \mathbb{N} \), where \( u_n \in U \) is a sequence of parameters. In this case, \( f_x(\frac{1}{n}, u_n) = 1 \) for each \( n \in \mathbb{N} \), where \( f_x \) denotes the derivative of \( f \) in the \( x \)-variable. Since the parameter space is compact, there exists a parameter \( u \) such that some subsequence of \( \{ u_n \} \) (which we do not relabel) converges to \( u \). Since \( f_x \) is jointly continuous in \( (x, u) \), we must have

\[
1 = \lim_{n \to \infty} f_x(\frac{1}{n}, u_n) = f_x(0, u).
\]

But then, it must be the case that \( f(x, u) \) exits \( F(x) \) for all small \( x < 0 \), and thus can not be a selection. \( \square \)
Theorem 3.1 will be proven later in this chapter after some preliminaries have been established. For simplicity of notation, we set

\[ H(x) = H(x, 1) \quad \text{and} \quad h(x) = -H(x, -1). \]

For a function \( f : \mathbb{R} \to \mathbb{R} \) we also set

\[ f'_+(x) = \lim_{y \to x^+} \frac{f(y) - f(x)}{y - x} \quad \text{and} \quad f'_-(x) = \lim_{y \to x^-} \frac{f(x) - f(y)}{x - y}. \]

Of course \( f'_+ \) is the right derivative and \( f'_- \) is the left derivative. Then \( f'_+(\cdot) = f'(\cdot) \) and \( f'_-(\cdot) = -f'(\cdot; -1) \).

Of course, it is a simple matter to find a \( C^1 \) selection when \( F \) has an interior. That is, when

\[ h(x) = \inf\{y : y \in F(x)\} < \sup\{y : y \in F(x)\} = H(x). \]

Indeed, if \( F \) has an interior on some closed interval \( J \subseteq I \), then this is trivial since \( C^1 \) functions are dense in continuous functions. However, when \( F \) collapses to a single value, (that is, when \( h(x) = H(x) \)), a selection must be chosen carefully so that it may pass through such a point while remaining smooth. If \( F \) remains a singleton on an entire closed interval \( J \subseteq I \), it is clear that \( F \) must be \( C^1 \) on \( J \) in the sense of ordinary functions, since any selection must take on the singleton value. The following lemma deals with this case. Note that, in the following, the notation
$F'$ is meaningless since $F$ is a multifunction. However, we will abuse notation when $F$ behaves like an ordinary function.

Lemma 4.1. Suppose $F(x) = \{H(x)\} = \{h(x)\}$ is a singleton for each $x$ in a open interval $J \subseteq I$. If each $\partial H(x)$ and $\partial h(x)$ is submonotone, then $F = \{H(x)\} = \{h(x)\}$ is $C^1$ on $J$ in the sense of ordinary functions. Moreover, $F = \{H(x)\} = \{h(x)\}$ is $C^1$ on $J$ closure.

Proof.

Suppose both $\partial H(x)$ and $\partial h(x)$ are submonotone. Then, recalling Theorem 2.11, each of $H(x)$ and $h(x)$ are semismooth and subdifferentially regular. In particular by Theorem 2.10, the one-sided derivatives of each of $H(x)$ and $h(x)$ exist and are upper semi-continuous (u.s.c.). Consequently, $h'_+(\cdot)$ is lower semi-continuous (l.s.c.). In this case, $H'_+(x) = h'_+(x)$ is both u.s.c. and l.s.c. at each $x \in J$. Thus, $F'_+ = \{H'_+\} = \{h'_+\}$ is continuous on $J$ being both u.s.c. and l.s.c. there. Since $H(x) = h(x)$ is semismooth, $F'_+(\cdot)$ is continuous on $J$ closure. □

We now need to distinguish among points according to the behavior of $F$. To accomplish this, we define the sets.

$$V = \{x \in I : h(x) < H(x)\},$$

$$A = \{x \in I : h(x) = H(x)\}.$$

Now, since each of $H(x)$ and $h(x)$ is continuous, each $x \in V$ lies in an open interval contained in $V$. So $V$ is open and, therefore, $V$ is the union of disjoint open intervals

$$V = \bigcup_{i=1}^{\infty} (c_i, d_i). \quad (4.2)$$

We first will construct a $C^1$ selection of $F$ defined on $V$ with properties that will enable us to extend this selection to the closure of $V$ without losing any smoothness.
We begin by developing some tools. The following lemma will enable us to "splice" two $C^1$ functions together without losing any differentiability.

**Lemma 4.3.** Let $[x_1, x_2] \subset \mathbb{R}$ be given and $f, g \in C^1[x_1, x_2]$. Then there exists a function $v \in C^1[x_1, x_2]$ such that

1. $v(x_1) = f(x_1), v'(x_1) = f'(x_1), v(x_2) = g(x_2),$ and $v'(x_2) = g'(x_2)$.
2. $v(x)$ lies between or equal to $f(x)$ and $g(x)$ for each $x \in [x_1, x_2]$.
3. Let $M$ and $N$ be such that $M \leq f'(x), g'(x) \leq N$ for each $x \in [x_1, x_2]$ and $S = \sup_{x \in [x_1, x_2]} |f(x) - g(x)|$. Then $M - \frac{2S}{x_2 - x_1} \leq v'(x) \leq N + \frac{2S}{x_2 - x_1}$ for each $x \in [x_1, x_2]$.

**Proof.**

Let $\delta = \frac{x_2 - x_1}{2}$. Define, for $x \in [x_1, x_2]$,

$$a'(x) = \begin{cases} \frac{x_1 - x}{\delta x}, & x \in [x_1, \frac{x_1 + x_2}{2}] \\ \frac{x - x_2}{\delta x}, & x \in [\frac{x_1 + x_2}{2}, x_2] \end{cases}$$

$$b'(x) = -a'(x).$$

For $x \in [x_1, x_2]$, let

$$a(x) = 1 + \int_{x_1}^{x} a'(t) \, dt \quad \text{and} \quad b(x) = \int_{x_1}^{x} b'(t) \, dt.$$

The following properties can be easily verified.

1. $a(x), b(x) \in [0, 1]$ and $a(x) + b(x) = 1$ for each $x \in [x_1, x_2]$.
2. $a'(x_1) = a'(x_2) = b'(x_1) = b'(x_2) = 0$, $a(x_2) = b(x_1) = 0$.
3. $|a'(x)| = |b'(x)| \leq \frac{2}{x_2 - x_1}$ for each $x \in [x_1, x_2]$. 

Now, we set \( v(x) = a(x)f(x) + b(x)g(x) \) for \( x \in [x_1, x_2] \). Then (4.3)(1) and (4.3)(2) follow easily from (4.4) and (4.5). We also see that

\[
v'(x) = a'(x)f(x) + a(x)f'(x) + b'(x)g(x) + b(x)g'(x) \\
\leq b'(x)(g(x) - f(x)) + a(x)N + b(x)N \quad \text{since} \quad a'(x) = -b'(x) \\
\leq \frac{1}{\delta} |f(x) - g(x)| + N(a(x) + b(x)) \quad \text{since} \quad b' \geq 0 \\
\leq N + \frac{S}{\delta}
\]

where \( S = \sup_{x \in [x_1, x_2]} |f(x) - g(x)| \). A similar argument yields the lower bound thereby establishing (4.3)(3).

**Remark 4.7.**

1. Consider the functions \( a(\cdot) \) and \( b(\cdot) \) above as functions of the endpoints of the interval \([x_1, x_2]\) as well as of \( x \) (i.e. \( a(x, x_1, x_2) = 1 + \int_{x_1}^{x} a'(t) \, dt \) for \( x \in [x_1, x_2] \) and \( b(\cdot, \cdot, \cdot) \) similarly). Then each of \( a(\cdot, \cdot, \cdot), \ a_x(\cdot, \cdot, \cdot), \ b(\cdot, \cdot, \cdot) \), and \( b_x(\cdot, \cdot, \cdot) \) is jointly continuous from \( \mathbb{R}^3 \to \mathbb{R} \). So the functions \( v(\cdot, \cdot, \cdot) \) and \( v_x(\cdot, \cdot, \cdot) \) above are also jointly continuous for \( x \in [x_1, x_2] \). Note that \( v(x, x_1, x_2) \) may blow up to infinity if \( x_1 \to x_2 \).

2. The estimate (4.3)(3) suffices nicely for our purposes in Chapter 4. However, Chapter 5 requires a somewhat more pointwise version. Suppose \( f(x) \geq g(x) \) for all \( x \in [x_1, x_2] \). Then the estimate can be stated in a more pointwise fashion by replacing \( S \) with \( (f(x) - g(x)) \).

**Lemma 4.8.** Suppose each of \( H(\cdot) \) and \( -h(\cdot) \) is subdifferentially regular. If \( H(x) = h(x) \), then

\[
\partial h(x) \cap \partial H(x) \neq \emptyset.
\]
Proof.

Suppose \( H(x) = h(x) \). We have by regularity that

\[
\partial H(x) = [H'_-(x), H'_+(x)] \quad \text{and} \\
\partial h(x) = [h'_+(x), h'_-(x)].
\]

Also, since \( h(x) \leq H(x) \) with equality at \( x \), we have

\[
h'_+(x) \leq H'_+(x) \quad \text{and} \\
H'_-(x) \leq h'_-(x).
\]

If \( h'_+(x) \in \partial H(x) \), we are done. Otherwise \( h'_+(x) < H'_-(x) \). Then either \( h'_-(x) \in \partial H(x) \) and we are done or \( h'_-(x) > H'_+(x) \) in which case \( \partial H(x) \subset \partial h(x) \).

When we construct a selection of \( F \) on \( V \), we must use the fact that the directional derivatives of subdifferentially regular functions are upper semi-continuous (u.s.c.). The following proposition will enable us to exploit this.

**Proposition 4.9.** Suppose \( f : I \to \mathbb{R} \) is u.s.c. Then there is a monotone decreasing sequence of continuous functions \( C_n : I \to \mathbb{R} \) such that

1. \( C_n \) converges to \( f \) pointwise on \( I \).
2. \( \inf_{y \in I} f(y) \leq C_n(x) \leq \sup_{y \in I} f(y) \) for each \( x \in I \) and \( n \in \mathbb{N} \).

Proof.

See [16, page 50] for an outline of the proof.

We now apply this to certain continuous selections of \( F \). For each continuous function \( \beta : I \to I \) such that \( \beta'_+(x) \) exists for each \( x \in I \), define

\[
K_\beta(x) = \beta(x)H(x) + (1 - \beta(x))h(x) \quad \text{for each} \ x \in I.
\]
The function $K_\beta$ stays between $h(x)$ and $H(x)$ since $\beta(\cdot)$ has values in $I$. Moreover, $(K_\beta)'_+(x)$ exists for each $x \in I$ since each of $h'_+(x)$, $H'_+(x)$, and $\beta'_+(x)$ does.

**Proposition 4.10.** Suppose each of $H(\cdot)$ and $-h(\cdot)$ is subdifferentially regular. Let $\beta : I \to I$ be a continuous function such that $\beta'_+(x)$ exists for each $x \in I$. Then there exists a sequence of continuous functions $C_n : I \to \mathbb{R}$ such that

1. $\lim_{n \to \infty} C_n(x) = (K_\beta)'_+(x)$ for each $x \in I$;
2. $M^\pi \leq C_n(x) \leq N^\pi$ for each $x \in I$ and $n \in \mathbb{N}$ where

$$N^\pi = \beta'_+(x)H(x) + \beta(x) \sup_{y \in I} H'_+(y)$$
$$- \beta'_+(x)h(x) + (1 - \beta(x)) \sup_{y \in I} h'_+(y)$$

and

$$M^\pi = \beta'_+(x)H(x) + \beta(x) \inf_{y \in I} H'_+(y)$$
$$- \beta'_+(x)h(x) + (1 - \beta(x)) \inf_{y \in I} h'_+(y).$$

**Proof.**

Recall that Rockafellar has shown that, in this case, each of $H'_+(\cdot)$ and $-h'_+(\cdot)$ is u.s.c. (see Theorem 2.10). So we may apply Proposition 4.9 obtaining sequences of continuous functions, $\Phi_n, \Psi_n : I \to \mathbb{R}$, such that

$$\lim_{n \to \infty} \Phi_n(x) = H'_+(x)$$
$$\lim_{n \to \infty} \Psi_n(x) = -h'_+(x)$$

for each $x \in I$.

For each $n \in \mathbb{N}$, let

$$C_n(x) = \beta'_+(x)H(x) + \beta(x)\Phi_n(x) - \beta'_+(x)h(x) - (1 - \beta(x))\Psi_n(x).$$

Then the sequence $\{C_n\}$ satisfies 4.10(1).
By Proposition 4.9(2) we have
\[ \Phi_n(x) \leq \sup_{y \in I} H'_+(y) \quad \text{and} \quad -\Psi_n(x) \leq \sup_{y \in I} h'_+(y). \]

Then, for each \( x \in I \) and \( n \in \mathbb{N} \),
\[
C_n(x) = \beta'_+(x)H(x) + \beta(x)\Phi_n(x) - \beta'_+(x)h(x) - (1 - \beta(x))\Psi_n(x) \\
\leq \beta'_+(x)H(x) + \beta(x)\sup_{y \in I} H'_+(y) \\
- \beta'_+(x)h(x) + (1 - \beta(x))\sup_{y \in I} h'_+(y)
\]
since \( 0 \leq \beta(x) \leq 1 \). A similar argument holds for the lower estimate thereby yielding 4.10(2).

We are now ready to prove our selection theorem.

Proof of Theorem 3.1.

Lemma 4.8 guarantees that, for each \( x \in A \), we may choose a fixed \( \xi_x \in \partial h(x) \cap \partial H(x) \). For each \( x \in I \), fix such a \( \xi \); call it \( \xi_x \). Our selection will have the property that \( f'(x) = \xi_x \) for each \( x \in A \). We fix this in advance because of situations similar to the following. Suppose \((c_i, d_i) \) and \((c_j, d_j)\) are distinct intervals in \( V \) with \( d_i = c_j \) (recall 4.2). Of course, here \( d_i = c_j \in A \). Suppose we have \( C^1 \) selections \( S_i \) and \( S_j \) defined on \([c_i, d_i]\) and \([c_j, d_j]\) respectively. If it was predetermined that \( S_i'(d_i) = S'_j(c_j) \), then we actually have a \( C^1 \) selection on the entire interval \([c_i, d_j]\). We also note that, if \( \{x_n\} \) is a sequence of points of \( A \) such that \( x_n \to x \) as \( n \to \infty \), then the choice of \( \xi_x \) is unique. To see this, assume without loss of generality that \( x_n < x \) for each \( n \). Since \( H \) and \( h \) are each semismooth, \( \xi_{x_n} \to H'(x) \) and \( \xi_{x_n} \to h'(x) \) as \( n \to \infty \). Therefore, there is only one element in \( \partial h(x) \cap \partial H(x) \), namely \( \xi_x \).
Fix an interval \((c, d) = (c_i, d_i)\) of \(V\). We will first construct a selection on \([c, d]\) with derivatives equal to \(\xi_c\) and \(\xi_d\) at the points \(c\) and \(d\) respectively. In this situation

\[
\begin{align*}
    h'_+(c) &\leq \xi_c \leq H'_+(c) \\
    H'_-(d) &\leq \xi_d \leq h'_-(d).
\end{align*}
\]

Consequently, the quantities

\[
\beta(c) = \frac{\xi_c - h'_+(c)}{H'_+(c) - h'_+(c)} \quad \text{and} \quad \beta(d) = \frac{h'_-(d) - \xi_d}{h'_-(d) - H'_-(d)}
\]

are in \([0,1]\). We claim that there exists a \(C^1\) function \(\beta : [c, d] \to I\) agreeing with the above quantities \(\beta(c)\) and \(\beta(d)\) at the points \(c\) and \(d\) respectively, and such that \(0 < \beta(x) < 1\) for all \(x \in (c, d)\). Moreover, we require that

\[
|\beta'_+(x)| \leq \frac{1}{d - c} \quad \text{for each } x \in [c, d].
\]

Indeed this is possible by taking \(\beta\) to be linear \((\beta(x) = \frac{1}{d - c} \{(x - c)\beta(d) + (d - x)\beta(c)\})\) unless \(\beta(c) = \beta(d) = 0\) or \(\beta(c) = \beta(d) = 1\). If \(\beta(c) = \beta(d) = 0\), set

\[
\beta(x) = \begin{cases} 
    \frac{x - c}{d - c}, & x \in [c, \frac{c + d}{2}] \\
    \frac{d - x}{d - c}, & x \in [\frac{c + d}{2}, d]
\end{cases}
\]

and similarly if \(\beta(c) = \beta(d) = 1\).

The basic strategy now is to construct the selection on \([c, d]\) so that it behaves in a similar fashion to the function \(K_\beta(\cdot)\). Choose a strictly increasing sequence \(\sigma = y_0, y_1, y_1', y_2, y_2', \ldots, y_i, y_i', \ldots\) of points of \((c, d)\) such that \(\sigma \to d\) as \(n \to \infty\). For each \(i \in \mathbb{N}\), denote \(I_i = [y_{i-1}, y_i']\). Now, since \((c, d) \subseteq V\) and \(\beta(x) \in (0, 1)\) for
all \( x \in (c, d) \), there exists an \( \epsilon_i > 0 \) such that \( K_\beta(x) \pm \epsilon_i \in F(x) \) for each \( x \in I_i \).

Furthermore, these may be chosen so that

\[
\epsilon_i \leq \frac{(y_i^1 - y_i)(d - c)}{2i}
\]

and \( \epsilon_{i+1} \leq \epsilon_i \). Note that \( \lim_{i \to \infty} \epsilon_i = 0 \) since \( d \in A \).

Now, we have a sequence of continuous functions, \( \{C_n^i\}_{n=1}^\infty \), that converge pointwise on \( I_i \) to \( (K_\beta)'_+ \) as prescribed by Proposition 4.10. By (4.10)(2), we see that \( \{C_n^i\}_{n=1}^\infty \) is uniformly bounded in \( L^1[I_i] \). So, by Lesbegue's Dominated Convergence Theorem,

\[
\lim_{n \to \infty} \int_{I_i} C_n^i(t) \, dt = \int_{I_i} (K_\beta)'_+(t) \, dt
\]

for each \( i \in \mathbb{N} \). Note that the derivative \( K_\beta' \) exists almost everywhere since \( K_\beta \) is Lipschitz so we can recover \( K_\beta \) from \( (K_\beta)'_+ \). In particular, the above gives us that, for each \( x \in I_i \), we have

\[
K_\beta(x) = K_\beta(y_{i-1}) + \int_{y_{i-1}}^x (K_\beta)'_+(t) \, dt = K_\beta(y_{i-1}) + \lim_{n \to \infty} \int_{y_{i-1}}^x C_n^i(t) \, dt. \quad (4.12)
\]

Thus, if we set

\[
f_n^i(x) = K_\beta(y_{i-1}) + \int_{y_{i-1}}^x C_n^i(t) \, dt,
\]

then \( \{f_n^i\}_{n=1}^\infty \) is a sequence of \( C^1 \) functions converging pointwise on \( I_i \) to the continuous function \( K_\beta \). Furthermore, from (4.10)(2), it is clear that \( \{f_n^i\}_{n=1}^\infty \) is equicontinuous for each \( i \in \mathbb{N} \). Consequently, the convergence is uniform. So, for each \( i \in \mathbb{N} \), there exists \( Q_i \in \mathbb{N} \) large enough so that

\[
|f_{Q_i}^i(x) - K_\beta(x)| < \epsilon_i \quad (4.13)
\]

for each \( x \in I_i \).
To simplify notation, we set \( f^i(x) = f^i_{Q_i}(x) \) on each interval \( I_i \). Note that, since the estimate (4.10)(2) is independent of \( n \), we have

\[
M_{I_i}^x \leq (f^i)'(x) \leq N_{I_i}^x \quad \text{for each } x \in I_i. \tag{4.14}
\]

The domains of the functions \( f^i \) and \( f^{i+1} \) overlap only on the interval \([y_i, y'_i]\). For each \( i \in \mathbb{N} \), we invoke Lemma 4.3 on \([y_i, y'_i]\) obtaining a \( C^1 \) function \( v^i \) matching up with \( f^i \) and \((f^i)'\) at \( y_i \) and with \( f^{i+1} \) and \((f^{i+1})'\) at \( y'_i \) as prescribed by (4.3)(1). Define a function \( f \) on \([y_0, d)\) by

\[
f(x) = \begin{cases} 
  f^i(x), & x \in [y_{i-1}, y_i] \quad \text{(here } y_0 = y'_0) \\
  v^i(x), & x \in [y_i, y'_i].
\end{cases}
\]

Then \( f \) is \( C^1 \) on \([y_0, d)\) and, from (4.3)(2) and (4.13), it follows that \( f(x) \in F(x) \) for each \( x \in [y_0, d) \).

The selection \( f \) must now be extended to the point \( d \). Since \( F \) is a singleton at \( d \), it is clear that \( \lim_{x \to d} f(x) = F(d) \). Thus, \( f \) can be extended continuously to \( d \). We now must show that \( f'(x) \to \xi_d \) as \( x \uparrow d \). We now consider the behavior of \( f' \) near \( d \).

Let \( \{x_n\} \subset [y_0, d) \) be an arbitrary sequence converging to \( d \). Now, for each \( n \in \mathbb{N} \), \( x_n \in I_{i(n)} \) for some \( i(n) \) (i.e. for some \( i \) depending on \( n \)). Clearly then; \( i(n) \to \infty \) as \( n \to \infty \). Two possibilities arise here; \( x_n \in (y_{i(n)-1}, y_{i(n)}) \) or \( x_n \in [y_{i(n)-1}, y'_{i(n)-1}] \cup [y_{i(n)}, y'_{i(n)}] \). In each case we need a suitable estimate on \( f'(x_n) \).

If \( x_n \in (y_{i(n)-1}, y_{i(n)}) \), then \( f(x_n) = f^{i(n)}(x_n) \) in which case

\[
f'(x_n) = (f^{i(n)})'(x_n).
\]

Combining this with (4.14) we have

\[
M_{I_{i(n)}}^{x_n} \leq f'(x_n) \leq N_{I_{i(n)}}^{x_n}.
\]
Now suppose $x_n \in [y_{i(n)}-1, y_{i(n)}'] \cup [y_{i(n)}, y'_{i(n)}]$. Suppose that $x_n \in [y_{i(n)}, y'_{i(n)}]$ since the other case is similar. Then $f(x_n) = v^{i(n)}(x_n)$ for some $i(n)$. Recall that $v^{i(n)}(x_n)$ is constructed from $f^{i(n)}$ and $f^{i(n)+1}$ on $[y_{i(n)}, y'_{i(n)}]$. So, from (4.3)(3) and (4.14),

$$f'(x_n) \leq \max_{j=i(n),i(n)+1} \left\{ N^x_{f_i} \right\} + \frac{2(2\epsilon_i(n))}{y_{i(n)} - y_i(n)}$$

$$\leq \max_{j=i(n),i(n)+1} \left\{ N^x_{f_i} \right\} + \frac{2(d - c)}{t_i(n)}$$

which, together with a similar lower estimate, gives

$$\max_{j=i(n),i(n)+1} \left\{ M^x_{f_i} \right\} - \frac{2(d - c)}{t_i(n)} \leq f'(x_n) \leq \max_{j=i(n),i(n)+1} \left\{ N^x_{f_i} \right\} + \frac{2(d - c)}{t_i(n)}. \tag{4.15}$$

Thus, in both cases, we have a similar bound for $f'(x_n)$. The estimate 4.15 holds for both cases.

In order to evaluate

$$\lim_{n \to \infty} \max_{j=i(n),i(n)+1} \left\{ N^x_{f_i} \right\},$$

it suffices to evaluate

$$\lim_{n \to \infty} N^x_{f_i(n)}.$$ 

Now

$$\lim_{n \to \infty} N^x_{f_i(n)} = \lim_{n \to \infty} \left( \beta'_+(x_n) H(x) + \beta(x_n) \sup_{y \in f_i(n)} H'_+(y) \right. \left. - \beta'_+(x_n) h(x) + (1 - \beta(x_n)) \sup_{y \in f_i(n)} h'_+(y) \right).$$

Since each of $H(\cdot)$ and $h(\cdot)$ is semismooth and all of the other functions above are continuous, all limits in question exist thereby enabling us to distribute the limit.
Also the intervals \( I_{i(n)} \) are collapsing to \( d \), so we can continue the above obtaining

\[
\lim_{n \to \infty} N_{I_{i(n)}}^x = \beta_+(d)[H(d) - h(d)] + \beta(d) \lim_{n \to \infty} \sup_{y \in I_{i(n)}} H'_+(y) + (1 - \beta(d)) \lim_{n \to \infty} \sup_{y \in I_{i(n)}} h'_+(y)
\]

\[
= \beta(d) H'_-(d) + (1 - \beta(d)) h'_-(d)
\]

\[
= \left( \frac{h'_-(d) - \xi_d}{h'_-(d) - H'_-(d)} \right) H'_-(d) + \left( 1 - \frac{h'_-(d) - \xi_d}{h'_-(d) - H'_-(d)} \right) h'_-(d)
\]

\[
= \frac{H'_-(d) h'_-(d) - H'_-(d) \xi_d + h'_-(d) \xi_d - H'_-(d) h'_-(d)}{h'_-(d) - H'_-(d)}
\]

\[
= \xi_d \left( \frac{h'_-(d) - H'_-(d)}{h'_-(d) - H'_-(d)} \right)
\]

\[
= \xi_d.
\]

It follows similarly that

\[
\lim_{n \to \infty} M_{i(n)}^x = \xi_d.
\]

Also note that

\[
\lim_{n \to \infty} \frac{2(d - c)}{i(n)} = 0
\]

since \( i(n) \to \infty \) as \( n \to \infty \). So, letting \( n \to \infty \) in (4.15), we obtain \( \lim_{n \to \infty} f'(x_n) = \xi_d \) as desired.

We can employ a similar argument to obtain a selection on \([c, y_0]\) with derivative at \( c \) equal to \( \xi_c \). We now have the desired selection on \([c, d]\). In this way, we have a selection on \( f \) defined on the set

\[
W = \bigcup_{i=1}^{\infty} [c_i, d_i]
\]

where the intervals \([c_i, d_i]\) are the closed counterparts of the ones in (4.2). Furthermore, if \([c_i, d_i]\) and \([c_j, d_j]\) are distinct intervals of \( W \) such that \( d_i = c_j \), then \( f \) is \( C^1 \) on \([c_i, d_j]\) since we judiciously chose \( f' \) at the endpoints of the intervals comprising
W. So, we may extend \( f \) to \( I \), and define our final selection:

\[
S(x) = \begin{cases} 
  f(x), & x \in W \\
  F(x), & x \in A.
\end{cases}
\]

The only points left in question are the points of \( A \), some of which may already be taken care of if they are isolated as in the above case. If a point of \( A \) is not isolated, then either it is contained in a open interval \( J \) of \( A \) or it is in the closure of \( W \). Suppose \( x \in W \setminus W \) (here \( W \) is the closure of \( W \)). Let \( \{w_n\} \) be a sequence of points of \( W \) such that \( w_n \rightarrow x \) as \( n \rightarrow \infty \). Then \( w_n \in [c_i(n), d_i(n)] \) where \( \{(c_i(n), d_i(n))\}_{n=1}^{\infty} \) is a sequence of intervals of \( W \). Necessarily,

\[
\lim_{n \rightarrow \infty} c_i(n) = \lim_{n \rightarrow \infty} d_i(n) = x.
\]

As shown in the first paragraph of this proof, if the intervals \( [c_i(n), d_i(n)] \) bunch up to \( x \) from below, then \( H'_-(x) = h'_-(x) \) making \( \xi_x \) uniquely determined. If a sequence of intervals of \( W \) also bunch up to \( x \) from above, then \( H'_+(x) = h'_+(x) \) making \( F \) differentiable in the sense of ordinary functions at \( x \).

If each \( w_n \) is an endpoint of one of the intervals \( [c_i(n), d_i(n)] \). Then \( w_n \in A \) so that \( f'(w_n) = \xi_{w_n} \) for some \( \xi_{w_n} \in \partial_x h(w_n) \cap \partial_x H(w_n) \). Then, by semismoothness, \( \xi_{w_n} \) converges to \( \xi_x \). Suppose each \( w_n \) is in the interior of an interval comprising \( W \). That is, \( w_n \in (c_i(n), d_i(n)) \). Recall the construction of \( f \) above. To each interval \( (c_i(n), d_i(n)) \), corresponds a function \( \beta_{i(n)} \), where \( f \) models the function \( K_{\beta_{i(n)}} \) on \( (c_i(n), d_i(n)) \). Furthermore, from (4.15), the estimate

\[
M^{w_n}_{(c_i(n), d_i(n))} - 2(d_i(n) - c_i(n)) \leq f'(w_n) \leq N^{w_n}_{(c_i(n), d_i(n))} + 2(d_i(n) - c_i(n)) \quad (4.16)
\]
holds. Consider the upper estimate

\[ f'(w_n) \leq (\beta_i(n))' + (w_n)H(w_n) + \beta_i(n)(w_n) \left( \sup_{y \in (c_i(n), d_i(n))} \{ H'_+(y) \} \right) \]

\[ - (\beta_i(n))' + (w_n)h(w_n) + [1 - \beta_i(n)](w_n) \left( \sup_{y \in (c_i(n), d_i(n))} \{ h'_+(y) \} \right) \]

\[ + 2(d_i(n) - c_i(n)) \]

\[ = (\beta_i(n))' + (w_n)(H(w_n) - h(w_n)) \]

\[ + \beta_i(n)(w_n) \left( \sup_{y \in (c_i(n), d_i(n))} \{ H'_+(y) \} \right) - \left( \sup_{y \in (c_i(n), d_i(n))} \{ h'_+(y) \} \right) \]

\[ + \sup_{y \in (c_i(n), d_i(n))} \{ h'_+(y) \} + 2(d_i(n) - c_i(n)). \]

The terms \( \sup_{y \in (c_i(n), d_i(n))} h'_+(y) \) and \( 2(d_i(n) - c_i(n)) \) above limit at \( \xi_x \) and 0 respectively as \( n \to \infty \). In order to evaluate \( f'(w_n) \) as \( n \to \infty \), we must establish that the first two terms above have limits as \( n \to \infty \). Since \( |(\beta_i(n))' + (w_n)| \leq \frac{1}{d_i(n) - c_i(n)} \)

we calculate

\[ (\beta_i(n))' + (w_n)(H(w_n) - h(w_n)) \leq \frac{H(w_n) - h(w_n)}{d_i(n) - c_i(n)} \]

\[ \leq \frac{H(w_n) - h(w_n)}{d_i(n) - w_n} \]

\[ = \frac{H(w_n) - H(d_i(n))}{d_i(n) - w_n} + \frac{H(d_i(n)) - h(w_n)}{d_i(n) - w_n}. \]

Taking limits we obtain

\[ \lim_{n \to \infty} (\beta_i(n))' + (w_n)(H(w_n) - h(w_n)) \leq -H'_-(x) + h'_-(x) \]

\[ = 0. \]

A similar argument yields

\[ \lim_{n \to \infty} (\beta_i(n))' + (w_n)(H(w_n) - h(w_n)) \geq 0. \]
We now turn our attention to the second term. Since $0 \leq \beta_{i(n)}(w_n) \leq 1$ and

$$\lim_{n \to \infty} \left( \sup_{y \in (c_i(n), d_i(n))} \{H'_+(y)\} - \sup_{y \in (c_i(n), d_i(n))} \{h'_+(y)\} \right) = 0.$$

Taking limits yields

$$\lim_{n \to \infty} \beta_{i(n)}(w_n) \left( \sup_{y \in (c_i(n), d_i(n))} \{H'_+(y)\} - \sup_{y \in (c_i(n), d_i(n))} \{h'_+(y)\} \right) = 0.$$

So we can evaluate the upper estimate of $f'(w_n)$ as $n \to \infty$ by distributing the limit obtaining

$$\lim_{n \to \infty} f'(w_n) \leq \xi_x.$$

A similar argument yields the lower estimate

$$\xi_x \leq \lim_{n \to \infty} f'(w_n).$$

Thus, we conclude that

$$\lim_{n \to \infty} f'(x_n) = \xi_x.$$

The function $S(\cdot)$ is then a $C^1$ selection of $F$ on $W$. Lemma 4.1 guarantees that $s(\cdot)$ is a $C^1$ selection of $F$ on the entire interval $I$.

\begin{remark}

The above construction works under slightly different hypotheses. The submonotone condition is equivalent to the two conditions; semismoothness and subdifferential regularity. The regularity of $H(\cdot)$ and $-\partial h(\cdot)$ guaranteed the existence of some $\xi \in \partial H(x) \cap \partial h(x)$ whenever $H(x) = h(x)$ and then enabled us to suitably approximate the functions $H'_+(\cdot)$ and $h'_+(\cdot)$. Semismoothness then enabled convergence of the estimates (4.16) and (4.17). That was the basic strategy.

Suppose we assume that both of $\partial H(\cdot)$ and $\partial h(\cdot)$ are submonotone (in Theorem 3.1, $-\partial h(\cdot)$ was assumed to be submonotone). Furthermore assume that there
\end{remark}
exists some $\xi \in \partial H(x) \cap \partial h(x)$ whenever $H(x) = h(x)$. Under these hypotheses, the approximations of $H'_+ (\cdot)$ and $h'_+ (\cdot)$ are still realized. Moreover, semismoothness is still preserved. Thus, the estimates (4.16) and (4.17) are still obtained and their convergence guaranteed. So, under these hypotheses, a $C^1$ selection is still obtained.

$\Box$
CHAPTER 5

PROOF OF THEOREM 3.2

Proof of Theorem 3.2 (sufficiency).

Assume that $F$ satisfies the sufficient conditions of Theorem 3.2. That is, we will assume that each of $H(x)$ and $-h(x)$ is strictly submonotone and

$$\partial h(x) = \partial H(x) \text{ whenever } h(x) = H(x). \quad (5.1)$$

The next few pages will be concerned primarily with $H(x)$, the "top" of $F$. Under our assumption, $H(x)$ is lower-$C^1$ and thus is the supremum of a family of $C^1$ functions indexed by some compact parameter space. If these functions all were to stay within $F$, we would be well on our way in obtaining a parametrization of $F$. Spingarn's construction of this family of functions does not guarantee this, however, so more delicacy is required.

Recall the sets $A$ and $V$ of Chapter 4. In light of (5.1), the set $A$ consists of at most two types of points: those at which $\partial H(x) = \partial h(x)$ is a singleton (i.e. both of $H(x)$ and $h(x)$ are differentiable with equal derivatives) and those at which $\partial H(x)$ is not a singleton (i.e. $\partial h(x) = \partial H(x) = [H'_-(x), H'_+(x)]$ with $H'_-(x) < H'_+(x)$). To distinguish between these in the future, we define the following subsets of $A$.

$$B = \{ x \in A : \partial H(x) = [H'_-(x), H'_+(x)] \text{ with } H'_-(x) < H'_+(x) \}$$

$$D = \{ x \in A : \partial H(x) = \{H'_-(x)\} = \{H'_+(x)\} \}$$

Lemma 5.2. The points of $B$ are isolated from each other. Moreover, any cluster point of $A$ is contained in $D$. 

48
Proof.

The points of $B$ being isolated from each other follows from the continuity of each of $h(x)$ and $H(x)$. Suppose $\{x_n\}$ is a sequence of points of $A$ such that $x_n \to x$ and $x_n < x$ for each $n \in \mathbb{N}$. Let $\xi_n \in \partial h(x_n) = h(x_n)$. By semismoothness, $\xi_n \to H'_-(x)$ and $\xi_n \to h'_-(x)$ as $n \to \infty$ and, thus $H'_-(x) = h'_-(x)$. Similarly, one can see that $H'_+(x) = h'_+(x)$. Now, since $\partial h(x) = [h'_{+}(x), h'_-(x)]$ and $\partial H(x) = [H'_-(x), H'_+(x)]$, (5.1) implies that $H'_-(x) = H'_+(x)$ and $h'_-(x) = h'_+(x)$. Consequently, $x \in D$. □

We now define a subset $P$ of $\mathbb{R}^2$ for which each point $p \in P$ will correspond to a $C^1$ selection, $x \mapsto f(x,p)$, of $F$ with the desired continuity properties in the parameter variable. Moreover, we will have

$$\sup_{p \in P} f(x,p) = H(x) \quad \text{for each } x \in I.$$ 

Let

$$P = \{(z, \xi) : z \in V \text{ and } \xi \in \partial H(z)\} \cup \{(z, \xi) : z \in A \text{ and } \xi = H'_\pm(z)\}.$$ 

Throughout the rest of this paper, when a pair $(z, \xi)$ is mentioned, it will be implicit that $\xi \in \partial H(z)$. We will need the following notation in order to readily describe certain subsets of $P$. For $X \subset I$ define

$$P_X = \{(z, \xi) : z \in X \text{ and } (z, \xi) \in P\}.$$ 

Each $z \in D$ gives rise to a unique point $(z, \xi) \in P$; however, each $z \in B$, gives rise to exactly two points of $P$ which, to simplify notation, will be labeled $(z, \xi^-)$ and $(z, \xi^+)$. Of course then $\xi^- = H'_-(x)$ and $\xi^+ = H'_+(x)$. In many cases, we
will discuss pairs \((z, \xi^\pm)\) with \(z \in A\) being fully aware that, if \(x \in D\), then the superscripts of \(\xi^\pm\) are superfluous.

So, considering \(P\) topologically as a subspace of \(\mathbb{R}^2\), there is a separation of \(P\) at each point \(z \in B\) consisting of the sets \((z, \xi^-) \cup P_{[0,z)} \) and \((z, \xi^+) \cup P_{(z,1]}\). Moreover, if \(z_1 < z_2\) are consecutive points of \(B\), then the set \(\{(z_1, \xi^+), (z_2, \xi^-)\} \cup P_{(z_1,z_2)}\) is a (connected) component of \(P\). However, all of the components of \(P\) may not be of this form since the set \(B\) need not be closed, even though the points of \(B\) are isolated from each other. With this in mind, given \(z \in I\), we label the component of \(P\) containing \(z\) by

\[
C(z) = \{(z_1, \xi^+), (z_2, \xi^-)\} \cup P_{(z_1,z_2)}
\]

where

\[
z_1 = \sup \{x : x \in B \text{ and } x < z\} \quad \text{and} \quad z_2 = \inf \{x : x \in B \text{ and } z < x\}.
\]

For \(z \in I\), since the points \(z_1\) and \(z_2\) are expedient in describing \(C(z)\), we will frequently have need to use the notation

\[
C(z) = \overline{z_1, z_2}.
\]

Note that if the supremum (or infimum) for \(z_1\) (or \(z_2\)) is not attained for some \(x \in B\), then considering Lemma 5.2, the superscript of \(\xi^+(\text{or} \xi^-)\) is superfluous. We also note that it is also possible that \(z_1 = z_2\), in which case, \(z \in D\) and \(C(z) = \overline{z, z} = (z, \xi)\).

Recall Proposition 2.20. There is a \(C^1\) function \(\alpha\) such that \(\alpha(0) = \alpha'(0) = 0\) and

\[
H(z + ty) > H(z) + \xi ty - \alpha(t) \tag{5.5}
\]
for each $0 < t \leq 1$, $z \in I$, $\xi \in \partial H(z)$, and $y = \pm 1$. For each $(z, \xi) \in P$ define a function $g(\cdot, z, \xi) : \mathbb{R} \to \mathbb{R}$ by

$$g(x, z, \xi) = H(z) + \xi(x - z) - \alpha(|x - z|).$$

It is clear from (5.5) that $g(x, z, \xi) \leq H(x)$ for each $x$, with equality only at $z = x$. Also, $g(\cdot, z, \xi)$ is $C^1$ for each $(z, \xi) \in P$, and has derivative $g_x(x, z, \xi) = \xi - \alpha'(|x - z|)$. Moreover, $g(\cdot, \cdot, \cdot)$ and $g_x(\cdot, \cdot, \cdot)$ are continuous as functions from $I \times P$ to $\mathbb{R}$.

Step 1. For each $z \in I$, we construct a special $C^1$ selection $S_z(\cdot)$ of $F$ on $I$ such that if $z', z'' \in C(z) = \overline{z_1, z_2}$, then $S_z(x) = S_{z'}(x)$ for each $x \in I$. In words, we construct a selection unique to each component of $P$. Let $S(x)$ be a fixed $C^1$ selection of $F$ on $I$ ($S$ will remain fixed henceforth). The selection $S_z(\cdot)$ will have the following properties:

1. $S'_z(c) = H'_+(c)$ and $S'_z(d) = H'_-(d)$, for each interval $(c, d)$ of $V$ with $P(c, d) \subset C(z)$ (the intervals $(c, d)$ refer to the disjoint intervals of $V$ in the representation 4.2).

2. $S_z(x) > g(x, c, \xi^+)$ and $S_z(x) > g(x, d, \xi^-)$ for each $x \in (c, d)$ for each interval $(c, d)$ of $V$ with $P(c, d) \subset C(z)$.

3. for each $z \in I$, $S_z(x)$ coincides with the fixed selection $S(x)$ for each $x \notin (x_1, x_2)$ where $x_1 = \sup\{y : y \in A$ and $y < z_1\}$ and $x_2 = \inf\{y : y \in A$ and $y > z_2\}$

4. Let $\{x_n\}$ and $\{z_n\}$ be sequences of points of $I$. If $x_n \to x \in I$ and $z_n \to z \in D$, then $S'_{z_n}(x_n) \to S'_z(x)$. Moreover, if $x = z$, then $S'_n(x_n) \to H'(x)$.

Let $(c, d)$ be an interval of $V$ (remember that intervals $(c, d)$ refer to the disjoint intervals of $V$ in the representation 4.2). We first construct a special $C^1$ selection $S_{[c, d]}(\cdot)$ of $F$ on $[c, d]$. For $x \in [c, d]$ define

$$\hat{h}(x) = \max\{S(x), g(x, c, \xi^+), g(x, d, \xi^-)\}.$$
Then \( \hat{h}(x) < H(x) \) for each \( x \in (c, d) \), \( \hat{h}(c) = H(c) \), and \( \hat{h}(d) = H(d) \). Moreover, \( \hat{h}(\cdot) \) is lower-\( C^1 \) on \([c, d]\) since it is the max of \( C^1 \) functions; in particular, \( \partial \hat{h}(\cdot) \) is submonotone. Also, for \( x \in [c, d] \) sufficiently near \( c \), \( \hat{h}(x) = \max\{S(x), g(x, c, \xi^+)\} \) since \( g(x, d, \xi^-) < H(x) \) for all \( x \neq d \). Thus \( \hat{h}'_+(c) \) exists and, furthermore, \( \hat{h}'_+(c) = H'_+(c) \), since \( g'(c, c, \xi^+) = \xi^+ = H'_+(c) \). Similarly, \( \hat{h}'_-(d) = H'_-(d) \). Keeping in mind Remark 4.17, there is a \( C^1 \) selection \( S_{[c,d]}(\cdot) \) of the multifunction \( \mathcal{F}(x) = \{y : \hat{h}(x) \leq y \leq F(x)\} \) defined on \([c, d]\). Necessarily, \( S'_{[c,d]}(c) = H'_+(c) \) and \( S'_{[c,d]}(d) = H'_-(d) \). It follows from the definition of \( \hat{h} \) that, for each \( x \in (c, d) \),

\[
S_{[c,d]}(x) > g(x, c, \xi^+) \quad \text{and} \quad S'_{[c,d]}(x) > g(x, d, \xi^-).
\]

Moreover, as in (4.16), we have the estimate

\[
M_{[c,d]}^x - 2(d - c) \leq S'_{[c,d]}(x) \leq N_{[c,d]}^x - 2(d - c) \quad (5.6)
\]

where, in this case,

\[
N_{[c,d]}^x = \beta^+(x)(H(x) - \hat{h}(x)) + \beta(x) \sup_{y \in [c,d]} \{H'_+(y)\} + (1 - \beta(x)) \sup_{y \in [c,d]} \{\hat{h}'_+(y)\}
\]

and similarly

\[
M_{[c,d]}^x = \beta^+(x)(H(x) - \hat{h}(x)) + \beta(x) \inf_{y \in [c,d]} \{H'_+(y)\} + (1 - \beta(x)) \inf_{y \in [c,d]} \{\hat{h}'_+(y)\}.
\]

Let \( z \in I \). We now define a \( C^1 \) selection \( \mathcal{S}_z(\cdot) \) of \( F \) on the interval \([z_1, z_2]\) corresponding to the component \( C(z) = z_{1,2} \). It is clear that there are no points of \( B \) strictly between \( z_1 \) and \( z_2 \) (otherwise \( C(z) \) would not be connected). However, points of \( D \) may abound there. Thus, if \((c, d)\) is an interval of \( V \) with \( z_1 < c < d < z_2 \), then \( S'_{[c,d]}(c) = H'(c) \) and \( S'_{[c,d]}(d) = H'(d) \). Consequently, if \((c_i, d_i)\) and \((c_j, d_j)\) are such intervals with \( d_i = c_j \), then the selections \( S_{[c_i,d_i]}(\cdot) \)
and $S_{[c_i, d_j]}(\cdot)$ combine to form a $C^1$ selection of $F$ on the combined interval $(c_i, d_j)$.

Moreover, the estimates (5.6) ensure that, if given a sequence of intervals of $V$ (still between $z_1$ and $z_2$), say $\{(c_i, d_i)\}$, the selections $S_{[c_i, d_i]}(\cdot)$ combine to form a $C^1$ selection of $F$ on the set $\bigcup_{i=1}^{\infty} [c_i, d_i]$. So define

$$\hat{S}_z(x) = \begin{cases} 
S_{[c, d]}(x), & \text{if } x \in [c, d] \subset [z_1, z_2] \text{ where } (c, d) \text{ is an interval of } V, \\
F(x), & \text{if } x \in [z_1, z_2] \setminus V.
\end{cases}$$

$\hat{S}_z(\cdot)$ is then a $C^1$ selection of $F$ defined on $[z_1, z_2]$ since $x \in A \cap (z_1, z_2)$ implies that $x \in D$. Note that $\hat{S}_z(z_1) = H'_+(z_1)$ and $\hat{S}_z(z_2) = H'_-(z_2)$.

The selection $\hat{S}_z(\cdot)$ must now be extended to the entire interval $I$. There are two possibilities for the nature of the points $z_1$ and $z_2$; they may be in $B$ or $D$. If they are in $D$, then $\hat{S}_z'(z_1) = H'(z_1) = S'(z_1)$ and $\hat{S}_z'(z_2) = H'(z_2) = S'(z_2)$. In this case, $\hat{S}_z(\cdot)$ is easily extended to $I$ using $S(\cdot)$. However, $z_1$ and $z_2$ may not be in $D$.

Let $x_1 = \sup\{y : y \in A \text{ and } y < z_1\}$ and $x_2 = \inf\{y : y \in A \text{ and } y > z_2\}$. Then $z_1, z_2 \in B$ and $x_1 \leq z_1 < z_2 \leq x_2$. Choose $C^1$ selections $T_1(\cdot)$ and $T_2(\cdot)$ of $F$ on the intervals $(x_1, z_1)$ and $(z_2, x_2)$ respectively such that $T_1'(x_1) = S'(x_1)$, $T_2'(x_2) = S'(x_2)$, $T_1'(z_1) = H'_+(z_1)$, and $T_2'(z_2) = H'_-(z_2)$. Define

$$S_z(x) = \begin{cases} 
\hat{S}_z(x), & \text{if } x \in [z_1, z_2], \\
T_1(x), & \text{if } x \in [x_1, z_1], \\
T_2(x), & \text{if } x \in [z_2, x_2], \\
S(x), & \text{if } x \in [0, x_1] \cup [x_2, 1].
\end{cases}$$

Then $S_z(\cdot)$ is a $C^1$ selection of $F$ on $I$. Note that if $z_1 \in D$ or $z_2 \in D$, then $x_1 = z_1$ or $x_2 = z_2$ respectively. In that case, the selection $T_1(\cdot)$ or $T_2(\cdot)$ is trivial and, as noted before, $\hat{S}_z(\cdot)$ is extended by using only $S(\cdot)$. If $z_1 \in B$ or $z_2 \in B$, then $x_1 < z_1$ or $x_2 > z_2$ respectively and the selection $T_1(\cdot)$ or $T_2(\cdot)$ is definitely not trivial.
Remark 5.7. Given $z \in I$, the selection $S_z(\cdot)$ is a bit of patchwork. However, we have estimates on $S'_z(x)$ for each $x \in I$. If $S_z(x) = \hat{S}_z(x)$, the estimate (5.6) holds. If $S_z(x) = S(x)$ or $S_z(x) = T_i(x)$ for $i = 1, 2$, we have the estimate (4.16). □

Step I(1) and Step I(2) are inherited from the fact that $s_z(x) = S_{[c,d]}(x)$ on each interval $(c, d)$ of $V$ with $P_{(c,d)} \subset C(z)$. Step I(3) follows from the definition of $s_z(\cdot)$. We must verify Step I(4). If there exists $N \in \mathbb{N}$ such that $C(z_n) = C(z_m)$ for all $m, n \geq N$, then the selections $S_{z_n}(\cdot)$ all coincide for $n \geq N$. In this case $C(z) = C(z_n)$ for all $n \geq N$. Thus, \[ \lim_{n \to \infty} S'_{z_n}(x_n) = \lim_{n \to \infty} S'_{z_n}(x_n) = S'_z(x) \]. So assume that $C(z_n) \neq C(z_m)$ if $m \neq n$. Here the selections $S_{z_n}(\cdot)$ vary as $n \to \infty$.

Denote $C(z_n) = \left[z_1^n, z_2^n\right]$. Then $z_1^n \to z$ and $z_2^n \to z$ as $n \to \infty$. Also, $x_1^n \to z$ and $x_2^n \to z$ ($x_1$ and $x_2$ are as in Step I(3)). If $x \neq z$, then for large $n$, Step I(3) implies that $S_{z_n}(x_n) = S(x_n)$. Thus \[ \lim_{n \to \infty} S'_{z_n}(x_n) = \lim_{n \to \infty} S'(x_n) = S'(x) = S'_z(x) \]. If $x = z$, a calculation like the one following (4.16) enables the estimates on $S'_{z_n}(x_n)$ discussed in Remark (5.7) to each converge to $H'(x)$. It follows that $S'_{z_n}(x_n) \to H'(x) = S'_z(x)$ verifying Step I(4). □

Step II. Let $(c, d)$ be an interval of $V$. For each $z \in (c, d)$, we splice the function $g(\cdot, z, \xi)$ onto the selection $S_z(\cdot)$ creating a function $f(\cdot, z, \xi)$ with the following properties:

1. $f(z, z, \xi) = H(z)$ for each $(z, \xi) \in P_{(c,d)}$.
2. $f(\cdot, z, \xi)$ is a $C^1$ selection of $F$ on $I$ for each $(z, \xi) \in P_{(c,d)}$.
3. $f(\cdot, \cdot, \cdot)$ and $f_x(\cdot, \cdot, \cdot)$ are jointly continuous on $I \times P_{(c,d)}$.

It is clear that $g(z, z, \xi) = H(z)$ for each $z \in I$, but there is no guarantee that $g(x, z, \xi) \in F(x)$ for any $x$ other than $z$. If $z \in V$, however, we are guaranteed that $g(x, z, \xi) \in F(x)$ for $x$ sufficiently near $z$. In particular,

$$g(x, z, \xi) > S_z(x)$$
for \( x \) near \( z \). Now fix an interval \((c, d)\) of \( V \) and let \( z_0 \in (c, d) \). The set \( P_{(c, d)} \) is then entirely contained \( C(z_0) = \bar{z_1}, \bar{z_2} \). Since the inequality in (5.5) is strict except for \( t = 0 \), to each point \((z, \xi) \in P_{(c, d)}\), there correspond two points \( x^-(z, \xi) \) and \( x^+(z, \xi) \) in \((c, d)\) such that \( x^-(z, \xi) < z < x^+(z, \xi), g(x^-(z, \xi), z, \xi) = S_{z_0}(x^-(z, \xi)) \), \( g(x^+(z, \xi), z, \xi) = S_{z_0}(x^+(z, \xi)) \), and \( g(x, z, \xi) > S_{z_0}(x) \) for all \( x < (x^-(z, \xi), x^+(z, \xi)) \). In other words, \( x^-(z, \xi) \) is the first point to the left of \( z \) where \( g(\cdot, z, \xi) \) is equal to the selection \( S_{z_0}(\cdot) \) and \( x^+(z, \xi) \) is the first point to the right of \( z \) where \( g(\cdot, z, \xi) \) is equal to the selection \( S_{z_0}(\cdot) \). Moreover, The points \( x^\pm(z, \xi) \) vary continuously in \((z, \xi)\).

We will now use Lemma 4.3 to splice each function \( g(\cdot, z, \xi) \) onto the selection \( S_{z_0}(\cdot) \) on the interval \([x^-(z, \xi), x^+(z, \xi)]\) thereby keeping it in \( F \) and allowing it to smoothly exit \((c, d)\). Of course, the function \( v \in C^1[x^-(z, \xi), x^+(z, \xi)] \) created by Lemma 4.3 will vary with the points \( x^\pm(z, \xi) \) so, for each \((z, \xi) \in P_{(c, d)}\), we will be creating a \( C^1 \) function \( v(\cdot, z, \xi) \). Remark 4.7 indicates that \( v(\cdot, \cdot, \cdot) \) and \( v_\xi(\cdot, \cdot, \cdot) \) are jointly continuous on \((c, d) \times P_{(c, d)}\). Moreover, we will still require that \( v(z, z, \xi) = H(z) \). To accomplish this, we alter \( g(\cdot, z, \xi) \) separately on each interval \([x^-(z, \xi), z]\) and \([z, x^+(z, \xi)]\).

For each \((z, \xi) \in P_{(c, d)}\), and each \( x \in I \) define

\[
f(x, z, \xi) = \begin{cases} 
v^-(x, z, \xi), & x \in [x^-(z, \xi), z] \\
v^+(x, z, \xi), & x \in [z, x^+(z, \xi)] \\
S_{z_0}(x), & x \notin [x^-(z, \xi), x^+(z, \xi)]
\end{cases}
\]

where \( v^+(x, z, \xi) \) is the function created on the interval \([z, x^+(z, \xi)]\) by Lemma 4.3 agreeing with \( g(\cdot, z, \xi) \) and \( g_\xi(\cdot, z, \xi) \) at the point \( z \) and agreeing with \( S_{z_0}(\cdot) \) and its derivative at the point \( x^+(z, \xi) \). The function \( v^-(x, z, \xi) \) is its counterpart on the interval \([x^-(z, \xi), z]\). So, for fixed \((z, \xi) \in P_{(c, d)}\), the function \( f(\cdot, z, \xi) \) is \( C^1 \) on \( I \).
and \( f(x, z, \xi) \in F(x) \) for each \( x \in I \). More importantly,

\[
f(x, x, \xi) = H(x).
\]  

(5.6)

Also, \( f(\cdot, \cdot, \cdot) \) and \( f_x(\cdot, \cdot, \cdot) \) are continuous on \( I \times P_{(c,d)} \). Step II(1), Step II(2), and Step II(3) are satisfied.

In this way, for each \( x \in I \) and \( (z, \xi) \in P_V \), the function \( f(x, z, \xi) \) is defined. For each \( (z, \xi) \in P_A \) and \( x \in I \), define

\[
f(x, z, \xi) = S_z(x).
\]

The function \( f(\cdot, \cdot, \cdot) \) is now defined on all of \( I \times P \). However, many issues of its continuity remain to be verified.

Step III. We verify that \( f(\cdot, \cdot, \cdot) \) and \( f_x(\cdot, \cdot, \cdot) \) are jointly continuous on \( I \times P_{(c,d)} \) for each interval \( (c, d) \) of \( V \).

Let \((c, d)\) be an interval of \( V \) and \( z_0 \in (c, d) \). Then \( P_{(c,d)} \subseteq C(z_0) \) and we have the selection \( S_{z_0}(\cdot) \) corresponding to \( C(z_0) \). For each \((z, \xi) \in P_{(c,d)}\) we have the associated points \( \xi^\pm(z, \xi) \) as before. We first verify that \( f(\cdot, \cdot, \cdot) \) and \( f_x(\cdot, \cdot, \cdot) \) are jointly continuous on \( I \times P_{(c,d)} \).

Step I(2) implies that

\[
\lim_{(z, \xi) \to (d, \xi^-)} x^\pm(z, \xi) = d.
\]  

(5.8)

As a result, for each \( x \neq d \), \( f(x, z, \xi) = S_{z_0}(x) \) for all \((z, \xi) \in P_{(c,d)}\) with \( z \) sufficiently near \( d \).

In order to show joint continuity, suppose \((x_n, z_n, \xi_n)\) converges to \((x, d, \xi^-)\) where each \((x_n, z_n, \xi_n) \in I \times P_{(c,d)} \). If \( x_n \notin (x^-(z_n, \xi_n), x^+(z_n, \xi_n)) \), then by definition \( f(x_n, z_n, \xi_n) = S_{z_0}(x_n) \) for all \( n \). So

\[
\lim_{n \to \infty} f(x_n, z_n, \xi_n) = \lim_{n \to \infty} S_{z_0}(x_n) = S_{z_0}(x) = f(x, d, \xi^-)
\]
and
\[ \lim_{n \to \infty} f_z(x_n, z_n, \xi_n) = \lim_{n \to \infty} S'_{z_0}(x_n) = S'_{z_0}(x) = f_z(x, d, \xi^-) \]
since \( S_{z_0} (\cdot) \) is \( C^1 \). In this case, we are done. Assume \( x_n \in (x^-(z_n, \xi_n), x^+(z_n, \xi_n)) \) for each \( n \). Then by (5.8), \( x = d \). If \( x_n = z_n \), then \( f(x_n, z_n, \xi_n) = H(x_n) \) and \( f_z(x_n, z_n, \xi_n) = \xi_n \). Clearly then
\[ \lim_{n \to \infty} f(x_n, z_n, \xi_n) = f(d, d, \xi^-). \]
Moreover, by semismoothness and Step I(1),
\[ \lim_{n \to \infty} f_z(x_n, z_n, \xi_n) = H'_-(d) = S'_{z_0}(d) = f_z(d, d, \xi^-). \]
So we are done in this case also.

Further assume that \( x_n \in (z_n, x^+(z_n, \xi_n)) \), the other case being similar. Then
\[ f(x_n, z_n, \xi_n) = v^+_n(x_n, z_n, \xi_n). \]
To simplify notation in the following calculation, set \( x^+(z_n, \xi_n) = x^+_n \) for each \( n \in \mathbb{N} \).

By the Mean Value Theorems for derivatives and Clarke subgradients (Theorem 2.2) respectively, there exist points \( v_n, w_n \in (x_n, x^+_n) \) and \( \gamma_n \in \partial H(w_n) \) such that
\[ S'_{z_0}(v_n) = \frac{S_{z_0}(x^+_n) - S_{z_0}(x_n)}{x^+_n - x_n} \quad \text{and} \quad \gamma_n = \frac{H(x^+_n) - H(x_n)}{x^+_n - x_n}. \quad (5.9) \]
Note that \( v_n \to d \) and \( w_n \to d \). Let
\[ N_n = \sup_{y \in [z_n, x^+_n]} \{ \max \{ g_z(y, z_n, \xi_n), S'_{z_0}(y) \} \}. \]
Now by Step I(1) and the continuity of \( S'_{z_0}(\cdot), S'_{z_0}(d) = H'_-(d) \). Also, \( g_z(x, z_n, \xi_n) = \xi_n - \alpha'(|x - z_n|) \) with \( \alpha'(\cdot) \) continuous and \( \alpha'(0) = 0 \), so that \( g_z(d, d, \xi_n) = H'_-(d) \).
Thus
\[ \lim_{n \to \infty} N_n = H'_-(d). \quad (5.10) \]
Employing the pointwise estimate discussed in Remark 4.7(2) and using (5.9), we calculate

\[
\frac{(v_n^+)_a(x_n, z_n, \xi_n)}{2} \leq \frac{g(x_n, z_n, \xi_n) - S_{z_0}(x_n)}{x_n^+ - z_n} + \frac{N_n}{2}
\]

\[
\leq \frac{H(x_n) - S_{z_0}(x_n)}{x_n^+ - z_n} + \frac{N_n}{2}
\]

\[
= \frac{H(x_n) - H(x_n^+)}{x_n^+ - z_n} + \frac{H(x_n^+) - S_{z_0}(x_n^+)}{x_n^+ - z_n} + \frac{S_{z_0}(x_n^+) - S_{z_0}(x_n)}{x_n^+ - z_n}
\]

\[
+ \frac{N_n}{2}
\]

\[
= \frac{-\gamma_n(x_n^+ - x_n)}{x_n^+ - z_n} + \frac{H(x_n^+) - S_{z_0}(x_n^+)}{x_n^+ - z_n} + \frac{S'_{z_0}(v_n)(x_n^+ - x_n)}{x_n^+ - z_n}
\]

\[
+ \frac{N_n}{2}
\]

\[
= (S'_{z_0}(v_n) - \gamma_n)\frac{(x_n^+ - x_n)}{x_n^+ - z_n} + \frac{H(x_n^+) - S_{z_0}(x_n^+)}{x_n^+ - z_n} + \frac{N_n}{2}.
\]

Now, again by Step I(1) and the continuity of \(S'_{z_0}()\), \(\lim_{v_n \to d} S'_{z_0}(v_n) = S'_{z_0}(d) = H_-'(d)\). Moreover, since \(w_n \to d\) and \(\gamma_n \in \partial H(w_n)\), it follows from semismoothness that \(\lim_{n \to \infty} \gamma_n = H_-'(d)\). Since the terms \(\frac{x_n^+ - x_n}{x_n^+ - z_n}\) are bounded between 0 and 1, we calculate

\[
\lim_{n \to \infty} \left( (S'_{z_0}(v_n) - \gamma_n)\frac{(x_n^+ - x_n)}{x_n^+ - z_n} \right) = 0. \tag{5.11}
\]

We now turn our attention to the second term. Since \(g(x_n^+, z_n, \xi_n) = S_{z_0}(x_n^+)\) it follows that

\[
S_{z_0}(x_n^+) = H(x_n) + \xi_n(x_n^+ - z_n) - \alpha(x_n^+ - z_n).
\]

Subtracting \(H(x_n^+)\) from each side and dividing by \(x_n^+ - z_n\)

\[
\frac{S_{z_0}(x_n^+) - H(x_n^+)}{x_n^+ - z_n} = \xi_n - \frac{\alpha(x_n^+ - z_n)}{x_n^+ - z_n} + \frac{H(z_n) - H(x_n^+)}{x_n^+ - z_n}.
\]

Taking limits above
\[
\lim_{n \to \infty} \frac{S_{x_0}(x_n^+) - H(x_n^+)}{x_n^+ - z_n} = H'_-(d) - H'_-(d) = 0
\]  \hspace{1cm} (5.12)

since \(\alpha'(0) = 0\).

Using (5.10), (5.11), and (5.12) we take limits obtaining

\[
\lim_{n \to \infty} v_n^+(x_n, z_n, \xi_n) \leq \lim_{n \to \infty} N_n = H'_-(d).
\]

This together with a similar lower estimate yields

\[
\lim_{n \to \infty} (v_n^+) = (x_n, z_n, \xi_n) = H'_-(d).
\]

Then, by Step I(1)

\[
\lim_{n \to \infty} f_x(x_n, z_n, \xi_n) = \lim_{n \to \infty} v_n^+(x_n, z_n, \xi_n) = H'_-(d) = S'_{x_0}(d) = f_x(d, d, \xi^-).
\]

We have verified that \(f(\cdot, \cdot, \cdot)\) and \(f_x(\cdot, \cdot, \cdot)\) are jointly continuous on \(I \times P(c, d]\)

If \((x_n, z_n, \xi_n) \to (x, c, \xi^+)\), a similar examining of cases and taking of limits yields

\[
\lim_{n \to \infty} f_x(x_n, z_n, \xi_n) = f_x(x, c, \xi^+).
\]

We have verified Step III. □

**Step IV. We verify that \(f(\cdot, \cdot, \cdot)\) and \(f_x(\cdot, \cdot, \cdot)\) are jointly continuous on all of \(I \times P\).**

Let \((x_0, z_0, \xi_0) \in I \times P\) and \((x_n, z_n, \xi_n) \to (x_0, z_0, \xi_0)\) where \((x_n, z_n, \xi_n) \in I \times P\) for each \(n \in \mathbb{N}\). It is clear that \(f(\cdot, \cdot, \cdot)\) is jointly continuous on \(I \times P\). We must show

\[
\lim_{n \to \infty} f_x(x_n, z_n, \xi_n) = f_x(x_0, z_0, \xi_0).
\]  \hspace{1cm} (5.13)

Suppose \(z_n \in A\) for each \(n \in \mathbb{N}\). Then \(z_0 \in D\) so that \(f(x_0, z_0, \xi_0) = S_{x_0}(x_0)\).

Also, \(f(x_n, z_n, \xi_n) = S_{x_n}(x_n)\) since \(z_n \in A\) for each \(n \in \mathbb{N}\). Thus, using Step I(4),

\[
\lim_{n \to \infty} f_x(x_n, z_n, \xi_n) = \lim_{n \to \infty} S'_{x_n}(x_n) = S'_{x_0}(x_0) = f_x(x_0, z_0, \xi_0).
\]

In this case, (5.13) is verified.
Assume $z_n \in V$ for each $n \in \mathbb{N}$. If $z_0 \in V$, (5.13) follows from Step II. So assume $z_0 \in A$. If for large $n$, $z_n \in (c, d)$ where $(c, d)$ is a fixed interval of $V$, then $z_0 = c$ or $z_0 = d$ and (5.13) follows from Step III. Suppose for large $n$ that $z_n \in (c_i, d_i) \cup (c_j, d_j)$ where $(c_i, d_i)$ and $(c_j, d_j)$ are fixed intervals of $V$ such that $c_i = d_j$. We examine the case where $z_0 = c_i = d_j$ (otherwise we are reduced to Step III again). Denote those $z_n \in (c_i, d_i)$ by $z^i_n$ and those $z_n \in (c_j, d_j)$ by $z^j_n$ (assuming that there are infinitely many points in each of the sequences, $\{z^i_n\}$ and $\{z^j_n\}$, or we are yet again reduced to Step III). If $z_0 \in B$, then $(z^i_n, \xi^i_n) \to (z_0, H'_-(z_0))$ and $(z^j_n, \xi^j_n) \to (z_0, H'_+(z_0))$. However, $H'_-(z_0) < H'_+(z_0)$ contradicting the fact that $(x_n, z_n, \xi_n) \to (x_0, z_0, \xi_0)$. So $z_0 \in D$. Thus the intervals $(c_i, d_i)$ and $(c_j, d_j)$ are in the same component, $C(z_0)$ of $P$. If $x_0 \neq z_0$, then for large enough $n$

\[
f_{x}(x_n, z^i_n, \xi^i_n) = S'_{z_0}(x_n) = f_{x}(x_n, z^j_n, \xi^j_n)
\]

from which (5.13) follows. Otherwise $x_0 = z_0$. Then

\[
\lim_{n \to \infty} f_{x}(x_n, z^i_n, \xi^i_n) = f_{x}(x_0, x_0, H'_-(x_0)) = f_{x}(x_0, x_0, H'_+(x_0)) = f_{x}(x_n, z^i_n, \xi^i_n)
\]

and (5.13) follows.

Finally, assume that for each $n \in \mathbb{N}$, $z_n \in (c_i(n), d_i(n))$ for some interval $(c_i(n), d_i(n))$ of $V$. For convenience of notation, denote $(c_i(n), d_i(n))$ by $(c_n, d_n)$. Here the components $C(z_n)$ may vary with each $n$. Also, $c_n \to z_0$ and $d_n \to z_0$. If $x_0 \neq z_0$, then Step I(3) implies that $f(x_n, z_n, \xi_n) = S(x_n)$ for all $n$ large enough. So (5.13) follows. Suppose $x_0 = z_0$. The nature of $f(x_n, z_n, \xi_n)$ depends upon the position of $x_n$ relative to $z_n$. Recall that the function $g(\cdot, z_n, \xi_n)$ was spliced onto the function $S_{z_n}(\cdot)$ on an interval $[x_n^-(z_n, \xi_n), x_n^+(z_n, \xi_n)]$ to create $f(\cdot, z_n, \xi_n)$ there. Otherwise, $f(\cdot, z_n, \xi_n)$ coincides with $S_{z_n}(\cdot)$. We first suppose that $x_n \notin [x_n^-(z_n, \xi_n), x_n^+(z_n, \xi_n)]$ for each $n \in \mathbb{N}$. Then $f(x_n, z_n, \xi_n) = S_{z_n}(x_n)$.
and (5.13) follows from Step I(4). Otherwise assume without loss of generality that
\[ x_n \in (z_n, x_n^+(z_n, \xi_n)) \]. Here \( f(x_n, z_n, \xi_n) = u_n^+(x_n, z_n, \xi_n) \). To show (5.13), it must
shown that
\[
\lim_{n \to \infty} (u_n^+)_x (x_n, z_n, \xi_n) = H'(x_0).
\] (5.14)
The estimate and calculation required is identical to the one immediately following
(5.10) except that the selections \( S_{x_n}(\cdot) \) may vary in the present case. The convergence of that estimate was established by showing convergence of the terms (5.10),
(5.11), and (5.12). Now, the estimate for (5.10) becomes
\[
N_n = \sup_{y \in [z_n, x_n^+]} \{ \max\{g_x(y, z_n, \xi_n), S_{x_n}'(y)\} \}.
\]
Considering Step I(4), we have the same result as in (5.10). Step I(4) also ensures
the same result as in (5.11). Finally, the right hand side of the estimate for (5.12) is
completely unaffected by any variation of the selection \( S_{x_n}(\cdot) \). So (5.12) also holds
in the present case. Thus, (5.14) is established. □

Step V. We parametrize \( F \) with \( C^1 \) functions.

In similar fashion, we can construct a function \( \hat{f} : I \times Q \to \mathbb{R} \) where \( Q \) is a
compact space and such that \( \hat{f}(\cdot, \cdot, \cdot) \) and \( \hat{f}_x(\cdot, \cdot, \cdot) \) are jointly continuous on \( I \times P \).
Moreover,
\[
h(x) = \inf_{q \in Q} \hat{f}(x, q).
\]
Define the compact set \( S \) by
\[
S = P \times Q \times [0, 1].
\]
Define \( p : I \times S \to \mathbb{R} \) by
\[
p(x, s) = p(x, p, q, l) = lf(x, p) + (1 - l)\hat{f}(x, q).
\]
This is the desired parametrization of \( F \). □
Proof of Theorem 3.2 (necessity).

We offer the first part of this proof in higher dimensions to illustrate that the Hamiltonian being jointly lower-$C^1$ is a necessary condition in higher dimensions. Again, let $B^n$ be the closed unit ball in $\mathbb{R}^n$. Suppose that $F : B^n \to \mathbb{R}^n$ is such that $F(x) = \{ p(x, s) : s \in S \}$ where $p(\cdot, s)$ is $C^1$ for each $s \in S$, $p(\cdot, \cdot)$ and $\nabla_x p(\cdot, \cdot)$ are jointly continuous, and $S$ is compact. Now, for $x \in B^n$ and $p \in \mathbb{R}^n$,

$$H(x, p) = \sup_{y \in F(x)} \langle y, p \rangle$$

$$= \sup_{s \in S} \langle p(x, s), p \rangle.$$

Define $f : \mathbb{R}^{2n} \times S \to \mathbb{R}$ by $f(x, p, s) = \langle p(x, s), p \rangle$. Then $f$ is differentiable with respect to $(x, p)$. Moreover, $f(\cdot, \cdot, \cdot)$ and $\nabla_{x,p}f(\cdot, \cdot, \cdot)$ are continuous.

To show the second necessary condition, we revert to $n = 1$. Again, let $H(x, 1) = H(x)$ and $h(x) = -H(x, -1)$ for each $x \in [0, 1]$. Suppose that $H(x_0) = h(x_0)$. Let

$$S_x = \{ s \in X : p(x, s) = H(x) \}$$

and

$$S'_x = \{ s \in X : p(x, s) = h(x) \}.$$

Clearly $S_{x_0} = S'_{x_0}$. Recall the representation (2.4). Then

$$\partial H(x_0) = \text{co} \{ p_x(x_0, s) : s \in S_{x_0} \}$$

$$= \text{co} \{ p_x(x_0, s) : s \in S'_{x_0} \}$$

$$= \partial h(x_0).$$
BIBLIOGRAPHY


VITA

The author, Craig Knuckles, was born the youngest of two children in Fort Wayne, Indiana on April 19, 1967 to Mr. and Mrs. Otis Knuckles. He attended High School at Heritage High School in Monroeville, Indiana and graduated in 1985. He received the Bachelor of Science degree in Mathematics at Tennessee Technological University in 1989. He subsequently received the Master of Science degree in Mathematics at Tennessee Technological University in 1991. He then received a Board of Regents Fellowship at Louisiana State University where he earned the Doctor of Philosophy degree in Mathematics in 1995. Craig was married to Hope. E. Baskerville in 1992.
DOCTORAL EXAMINATION AND DISSERTATION REPORT

Candidate: Craig Knuckles

Major Field: Mathematics

Title of Dissertation: Continuously Differentiable Selections and Parametrizations of Multifunctions in One Dimension

Approved:

[Signatures]

Major Professor and Chairman

Dean of the Graduate School

EXAMINING COMMITTEE:

[Signatures]

Date of Examination: March 31, 1995