Essentially Smooth Lipschitz Functions

Jonathan M. Borwein

CECM, Department of Mathematics, Simon Fraser University,
Burnaby, British Columbia, Canada V5A 1S6

and

Warren B. Moors

Department of Mathematics, University of Auckland,
Private Bag 92019, Auckland, New Zealand

Received January 24, 1995; revised November 4, 1996; accepted January 20, 1997

In this paper we address some of the most fundamental questions regarding the
differentiability structure of locally Lipschitz functions defined on separable Banach
spaces. For example, we examine the relationship between integrability, D-represent-
tability, and strict differentiability. In addition to this, we show that on any
separable Banach space there is a significant family of locally Lipschitz functions
that are integrable, D-representable and possess desirable differentiability properties.
We also present some striking applications of our results to distance functions.
© 1997 Academic Press

1. INTRODUCTION

The first goal of this paper is to show that there is a significant class of
locally Lipschitz functions which possesses the property that each of its
members, $f$, satisfies the following three conditions:

($P_1$) $f$ is D-representable, that is, $f$ is Gateaux differentiable on some
dense subset $D$ of its domain and the Clarke subdifferential mapping,
$x \mapsto \partial f(x)$, is generated by the derivatives chosen from any dense subset
of $D$;

($P_2$) $f$ is integrable, that is, we may determine the function $f$, up to
an additive constant, from its Clarke subdifferential mapping, $x \mapsto \partial f(x)$,
(provided of course, that the domain of $f$ is connected);

($P_3$) $f$ possesses differentiability properties similar to those enjoyed
by continuous convex functions.
In addition to the fore-mentioned properties, the class of functions that we exhibit, also possesses very strong closure properties. For example, it is closed under addition, subtraction, multiplication, and division (when this is defined), as well as, the lattice operations. Yet a further advantage with this class of functions is that it enables a unified presentation of many previously known results, which up-till now, appeared unconnected.

The second goal of the paper is to examine the relationship between $D$-representability, integrability and almost everywhere strict differentiability. Of course, this second goal is closely related to our first goal.

We begin by recalling some preliminary definitions regarding the Clarke subdifferential mapping. A real-valued function $f$ defined on a non-empty open subset $A$ of a Banach space $X$, is locally Lipschitz on $A$, if for each $x_0 \in A$ there exists a $K > 0$ and a $\delta > 0$ such that

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

for all $x, y \in B(x_0, \delta) \cap A$.

For functions in this class, it is often instructive to consider the following three (right-hand) directional derivatives.

1. The upper Dini derivative at $x \in A$, in the direction $y$, is given by

$$f^+(x; y) = \limsup_{\lambda \to 0^+} \frac{f(x + \lambda y) - f(x)}{\lambda}.$$

2. The lower Dini derivative at $x \in A$, in the direction $y$, is given by

$$f^-(x; y) = \liminf_{\lambda \to 0^+} \frac{f(x + \lambda y) - f(x)}{\lambda}.$$

3. The Clarke generalized directional derivative at $x \in A$, in the direction $y$, is given by

$$f^0(x; y) = \limsup_{\lambda \to 0^+} \frac{f(z + \lambda y) - f(z)}{\lambda}.$$

It is immediate from these three definitions that for each $x \in A$ and each $y \in X$,

$$f^-(x; y) \leq f^+(x; y) \leq f^0(x; y).$$

Associated with the Clarke generalized directional derivative is the Clarke subdifferential mapping, which is defined by

$$\partial f(x) = \{ x^* \in X^* : x^*(y) \leq f^0(x; y) \text{ for each } y \in X \}.$$
The Clarke subdifferential mapping, $x \mapsto \partial f(x)$, has played a prominent role in the recent development of non-smooth analysis. Two reasons for this success are:

$(\ast_1)$ for each $x \in A$, $\partial f(x)$ is non-empty, convex and weak* compact, and for each weak* open subset $W$ of $X^*$, $\{x \in A : \partial f(x) \subseteq W\}$ is an open subset of $A$;

$(\ast_2)$ at each point $x \in A$, where the Gateaux derivative of $f$ exists, $\nabla f(x) \in \partial f(x)$.

Let us now examine some notions of differentiability that are associated with Lipschitz functions. We say that a function $f$ is differentiable at $x$, in the direction $y$ if,

$$f'(x; y) \equiv \lim_{\lambda \to 0} \frac{f(x + \lambda y) - f(x)}{\lambda}$$

exists, and we say that $f$ is Gateaux differentiable at $x$ if

$$\nabla f(x)(y) \equiv \lim_{\lambda \to 0} \frac{f(x + \lambda y) - f(x)}{\lambda}$$

exists for each $y \in X$ and $\nabla f(x)$ is a continuous linear functional on $X$. Note that a locally Lipschitz function $f$ may be differentiable at a point $x$, in every direction $y \in X$, while still not being Gateaux differentiable at that point. This is because, although the mapping, $y \mapsto f'(x; y)$, is necessarily continuous when $f$ is Lipschitz, it may fail to be linear. If $f$ is Gateaux differentiable at $x$ and

$$\lim_{\lambda \to 0} \frac{f(x + \lambda y) - f(x)}{\lambda}$$

is uniform over $y \in S(X)$; the unit sphere in $X$

then $f$ is said to be Fréchet differentiable at $x$ (and $\nabla f(x)$ is called the Fréchet derivative of $f$ at $x$).

In finite dimensions, the notions of Gateaux and Fréchet differentiability coincide for locally Lipschitz functions, ([17], p. 30). However, outside of finite dimensions these notions are distinct, [7].

Unfortunately, (in infinite dimensional spaces) the notion of Gateaux differentiability does not easily yield “generic” differentiability results and so we will need to consider two slightly stronger notions of differentiability.
A locally Lipschitz function \( f \) is said to be strictly differentiable (strictly Fréchet differentiable) at \( x \), if for each \( \varepsilon > 0 \) there exists a \( \delta > 0 \) such that
\[
\left| \frac{f(z + \lambda y) - f(z)}{\lambda} - \nabla f(x)(y) \right| < \varepsilon
\]
whenever \( 0 < \lambda < \delta \) and \( \|z - y\| < \delta \) (uniformly over \( y \in S(X) \)).

For continuous convex functions, Gateaux differentiability coincides with strict differentiability, as does, Fréchet differentiability with strict Fréchet differentiability. In general however, these concepts are distinct.

**Example 1.1.** Let
\[
f(x) = \begin{cases} 
  x^2 \sin(1/x) & \text{if } x \neq 0 \\
  0 & \text{if } x = 0.
\end{cases}
\]

Then \( f \) is differentiable everywhere on \( \mathbb{R} \), but \( f \) is not strictly differentiable at \( x = 0 \). In fact \( f'(0) = 0 \) while \( \partial f(0) = [-1, 1] \).

Next, we recall the connection between strict differentiability and single-valuedness of the Clarke subdifferential mapping.

**Proposition 1.1** [3, Proposition 3.1]. Let \( f \) be a real-valued locally Lipschitz function defined on a non-empty open subset \( A \) of a Banach space \( X \). Then:

(a) \( \partial f(x) \) is a singleton if, and only if, \( f \) is strictly differentiable at \( x \);

(b) \( \partial f(x) \) is a singleton and has the property that for each \( \varepsilon > 0 \) there exists a \( \delta > 0 \) such that \( \partial f(B(x, \delta)) = \partial f(x) + \varepsilon B(X) \) if, and only if, \( f \) is strictly Fréchet differentiable at \( x \).

For further information regarding the Clarke subdifferential mapping see, [17]. Apart from the fore-mentioned notions of differentiability, the other key concept contained in this paper is that of a minimal cusco.

A set-valued mapping \( \Phi \) from a topological space \( A \) into subsets of a topological (linear topological) space \( X \) is an usco (cusco) on \( A \) if:

(i) for each \( t \in A \), \( \Phi(t) \) is non-empty (convex) and compact;

(ii) for each open subset \( W \subseteq X \), \( \{ t \in A : \Phi(t) \subseteq W \} \) is open in \( A \).

It follows from (\( \ast \)), that the Clarke subdifferential mapping of any real-valued locally Lipschitz function defined on an open subset is a weak* cusco. Amongst the class of usco (cusco) mappings special attention is given to the so-called minimal uscos (minimal cuscos). An usco (cusco) mapping \( \Phi \) from a topological space \( A \) into subsets of a topological (linear
topological) space $X$ is called a minimal usco (minimal cusco) if its graph does not strictly contain the graph of any other usco (cusco) defined on $A$. It is immediate from this definition that all single-valued uscos (cuscos) are minimal, however, there are many important examples of minimal uscos (cuscos) that are not everywhere single-valued. We begin our study of minimal cuscos (minimal uscos) by recalling some of their basic properties.

**Proposition 1.2** [16, p. 649]. Let $\Phi$ be an usco (cusco) mapping from a topological space $A$ into subsets of a topological (linear topological) space $X$. Then there exists a minimal usco (minimal cusco) $\Psi$ defined on $A$ such that $\Psi(t) \subseteq \Phi(t)$ for each $t \in A$.

Let $\Omega$ be a set-valued mapping from a non-empty set $A$ into a non-empty set $X$. Then by the graph of $\Omega$ we mean $\text{Gr}(\Omega) = \{(t, x) \in A \times X : x \in \Omega(t)\}$ and by the (effective) domain of $\Omega$ we mean $\text{Dom}(\Omega) = \{t \in A : \Omega(t) \neq \emptyset\}$. When the domain of $\Omega$ is dense in $A$ we say that $\Omega$ is densely defined.

It is worthwhile observing that for an usco mapping $\Phi$ from a topological space $A$ into subsets of Hausdorff topological space $X$, the graph of $\Phi$ is a closed subset of $A \times X$ (when $A \times X$ is endowed with the product topology). It is also interesting to see that to some extent the converse of this observation is true.

**Proposition 1.3** [16, p. 651]. Let $\Phi$ be an usco mapping from a topological space $A$ into subsets of a topological space $X$ and let $\Omega$ be a set-valued mapping from $A$ into non-empty subsets of $X$. If $\text{Gr}(\Omega)$ is a closed subset of $A \times X$ and $\text{Gr}(\Omega) \subseteq \text{Gr}(\Phi)$, then $\Omega$ is an usco mapping on $A$.

The next proposition gives further information concerning the construction of usco (cusco) mappings.

**Proposition 1.4.** Let $\Omega$ be a densely defined set-valued mapping from a topological space $A$ into subsets of a Hausdorff topological (separated locally convex topological) space $X$. If the graph of $\Omega$ is contained in the graph of an usco (cusco) mapping $\Phi$, then there exists a unique smallest usco (cusco) containing $\Omega$, denoted $\text{USC}(\Omega)(\text{CSC}(\Omega))$, given by,

$$\text{USC}(\Omega)(x) = \bigcap \{V : V \text{ is an open neighbourhood of } x\},$$

$$\text{CSC}(\Omega)(x) = \bigcap \{V : V \text{ is an open neighbourhood of } x\}. $$

**Proof.** We shall only prove that $\text{CSC}(\Omega)$ is the smallest cusco containing $\Omega$, as the proof that $\text{USC}(\Omega)$ is the smallest usco containing $\Omega$, is identical to this.
We begin with the following three observations:

(i) for each \( t \in \text{Dom}(\Omega) \), \( \Omega(t) \subseteq \text{CSC}(\Omega)(t) \);

(ii) for any set-valued mapping \( \Psi \), \( \text{CSC}(\Psi) \) possesses a closed graph;

(iii) if \( \Psi \) is a cusco then \( \Psi = \text{CSC}(\Psi) \).

We now show that \( \text{CSC}(\Omega) \) is a cusco mapping on \( A \). From (iii) and the definition of \( \text{CSC}(\Omega) \) it follows that

\[
\text{Gr}(\text{CSC}(\Omega)) \subseteq \text{Gr}(\text{CSC}(\Phi)) = \text{Gr}(\Phi).
\]

Furthermore, by (ii), we have that the graph of \( \text{CSC}(\Omega) \) is closed, so by Proposition 1.3, it is sufficient to show \( \text{Dom}(\text{CSC}(\Omega)) = A \).

Suppose, for the purpose of obtaining a contradiction, that there exists an element \( t_0 \notin \text{Dom}(\text{CSC}(\Omega)) \). For each \( x \in \Phi(t_0) \) choose open sets \( U_x \subseteq A \) and \( V_x \subseteq X \) such that \( (t_0, x) \in U_x \times V_x \) and \( (U_x \times V_x) \cap \text{Gr}(\text{CSC}(\Omega)) = \emptyset \). Since \( \Phi(t_0) \) is compact and \( \Phi(t_0) \subseteq \bigcup \{ V_x : x \in \Phi(t_0) \} \) there exists a finite subcover \( \{ V_{x_j} : 1 \leq j \leq n \} \) of \( \{ V_x : x \in \Phi(t_0) \} \) such that

\[
\Phi(t_0) \subseteq \bigcup \{ V_{x_j} : 1 \leq j \leq n \}.
\]

Let \( U_1 \equiv \bigcap \{ U_{x_j} : 1 \leq j \leq n \} \), and observe that for each \( t \in U_1 \),

\[
\text{CSC}(\Omega)(t) \cap \bigcup \{ V_{x_j} : 1 \leq j \leq n \} = \emptyset.
\]

On the other hand, \( \Phi \) is a cusco, so there exists an open neighbourhood \( U_2 \) of \( t_0 \) such that \( \Omega(U_2) \subseteq \Phi(U_2) \subseteq \bigcup \{ V_{x_j} : 1 \leq j \leq n \} \). Let \( U \equiv U_1 \cap U_2 \neq \emptyset \). Now, by (i) we have that for each \( t \in \text{Dom}(\Omega) \cap U \neq \emptyset \),

\[
\text{CSC}(\Omega)(t) \cap \bigcup \{ V_{x_j} : 1 \leq j \leq n \} \neq \emptyset.
\]

But this contradicts the fact that \( \emptyset \neq U \cap \text{Dom}(\Omega) \subseteq U_1 \).

Hence \( \text{Dom}(\text{CSC}(\Omega)) = A \), which shows that \( \text{CSC}(\Omega) \) is a cusco on \( A \).

To see that \( \text{CSC}(\Omega) \) is the smallest cusco containing \( \Omega \) it suffices to observe that for any cusco \( \Psi \) containing \( \Omega \),

\[
\text{Gr}(\text{CSC}(\Omega)) \subseteq \text{Gr}(\text{CSC}(\Psi)) = \text{Gr}(\Psi).
\]

Note. In the above proposition, the set-valued mapping, \( \text{CSC}(\Omega) \), is called the \textit{cusco generated by} \( \Omega \) and \( \text{USC}(\Omega) \) is called the \textit{usco generated by} \( \Omega \).
Remark 1.1. It is possible to strengthen the previous proposition so as to only require that for each point \( x \in A \) there is an open neighbourhood \( U_x \) of \( x \) and a usco (cusco) \( \Phi_x \) defined on \( U_x \) such that \( \Omega(y) \subseteq \Phi_x(y) \) for each \( y \in U_x \cap \text{Dom}(\Omega) \). In this way, use see that the graph of any densely defined locally bounded mapping into the dual of a Banach space is contained in the graph of a weak* cusco.

Now that we have established some of the elementary properties and definitions concerning locally Lipschitz functions and minimal cuscos, we may discuss more precisely the connection between locally Lipschitz functions that possess the properties \((P_1)-(P_3)\), listed at the start of this paper, and minimal cuscos. At the heart of this relationship, is the fact that a locally Lipschitz function \( f \) defined on a non-empty open subset \( A \) of a smooth Banach space (or more generally, a class(\( S \) space, see [41])) is \( D \)-representable if, and only if, its Clarke subdifferential mapping, \( x \mapsto \partial f(x) \), is a minimal weak* cusco on \( A \) (see, Corollary 2.2). However, to fully understand this statement we must first make precise what we mean by \( D \)-representable. Let \( f \) be a real-valued locally Lipschitz defined on a non-empty open subset \( A \) of a normed linear space \( X \). Then we say that \( f \) is \( D \)-representable on \( A \) if:

(a) \( D \equiv \{ x \in A : \forall f(x) \text{ exists} \} \) is dense in \( A \) and

(b) for each dense subset \( D^* \) of \( D \) we have that \( \partial f = \text{CSC}(\Omega_{D^*}) \), where \( \Omega_{D^*} : D^* \rightarrow 2^{X^*} \) is defined by \( \Omega_{D^*}(x) \equiv \{ \forall f(x) \} \).

Note that in particular, when \( f \) is \( D \)-representable, \( \partial f = \text{CSC}(\Omega_D) \).

So we see then, that by a desire to consider locally Lipschitz functions that are \( D \)-representable we are inextricably lead to consider locally Lipschitz functions whose Clarke subdifferential mappings are minimal (with respect to the family of weak* cusco mapping). There is however, one significant difference between these definitions, namely, the notion of minimality extends beyond the class of functions that are densely Gateaux differentiable. In addition to the notion of \( D \)-representability we need to also make precise what we mean by “integrable.” Let \( f \) be a real-valued locally Lipschitz function defined on a non-empty open subset \( A \) of a Banach space \( X \). Then we say that \( f \) is integrable on \( A \) if, \( \partial (f - g) \equiv \{ 0 \} \) for each real-valued locally Lipschitz function \( g \) defined on \( A \) with \( g(x) \leq f(x) \) for each \( x \in A \). It follows from this, that if \( A \) is connected and \( f \) is integrable on \( A \), then \( f - g \equiv \text{constant} \) on \( A \), whenever \( \partial g(x) \leq \partial f(x) \) for each \( x \in A \). In particular, this means that \( f \) is determined, up to an additive constant, by its Clarke subdifferential mapping.

Life would be simple if all \( D \)-representable functions automatically satisfied the conditions \((P_2)\) and \((P_3)\) given earlier, however, there are numerous examples (even on \( R \)) of Lipschitz functions that are \( D \)-representable, but
which fail to satisfy either \((P_2)\) or \((P_3)\). Furthermore, the class of \(D\)-representable functions is neither closed under addition, multiplication nor either of the lattice operations (see Example 8.1). Therefore, in order to achieve our goal, we are forced to consider a proper subclass of the \(D\)-representable functions.

In this paper, we propose that the appropriate functions to consider (on a separable Banach space) are those functions which are strictly differentiable almost everywhere, that is, strictly differentiable everywhere except on a Haar-null set. The fact that this is a reasonable class of functions to consider, derives from the following facts: (a) on the real line, the locally Lipschitz functions which satisfy \((P_1)\), \((P_2)\) and \((P_3)\) (actually, on the real line, any locally Lipschitz function that satisfies \((P_3)\) automatically satisfies \((P_1)\) and \((P_2)\)) are exactly those functions which are strictly differentiable almost everywhere on their domain, and (b) on separable Banach spaces, the continuous convex functions are strictly differentiable almost everywhere.

We begin Section 2 by recalling some necessary topological prerequisites that are required to show that a densely Gateaux differentiable Lipschitz function is \(D\)-representable if, and only if, its Clarke subdifferential mapping is a minimal weak* cusco. Then, in Section 3, we characterize when the Clarke subdifferential mapping is minimal, in terms of a “quasi continuity” property possessed by the upper Dini derivative mapping, \(x \mapsto f^+(x; y)\), (for each \(y \in S(X)\)). We then use this characterization to show that the distance function \(d_c\) generated by a set \(C\) possesses a minimal subdifferential mapping on \(X\) if, and only if, \(d_c\) possesses a minimal subdifferential mapping on \(X\backslash C\).

We begin Section 4 by showing that on any separable Banach space, the functions that are strictly differentiable almost everywhere satisfy the properties \((P_1)\), \((P_2)\), and \((P_3)\) given at the start of this paper. Moreover, we show that all the pseudo-regular and semi-smooth functions belong to this class (plus many others).

In Section 5 we show how the results of Section 4 maybe applied to perturbation functions. Section 6 concerns distance functions; more specifically, in this section we examine when the Clarke subdifferential mapping of a distance function is a minimal weak* cusco. In doing this, we are able to derive a “Proximal Normal Formula,” which holds for all non-empty closed subsets of any reflexive Banach space which possesses a smooth Kadec–Klee norm. Moreover, we show that if such a proximal normal formula holds for all subsets, then the space is necessarily reflexive and the norm is necessarily a smooth Kadec-Klee norm. In Section 7 we re-examine integrability and \(D\)-representability. In particular, we show that \(D\)-representability does not imply integrability and that integrability does not imply \(D\)-representability. In fact, we show that integrability does not even imply dense strict differentiability. We also show that integrability is not a
hereditary property, that is, it is possible for a Lipschitz function \( f \) to be integrable on a non-empty open set \( A \) while its restriction \( f|_U \) to a non-empty open subset \( U \) of \( A \) may not be integrable on \( U \). Finally, in Section 8, we give some examples which highlight some of the behaviour (both good and bad) possessed by functions whose Clarke subdifferential mappings are minimal.

Since this is not the first article written on the topic of Lipschitz functions with minimal subdifferential mappings, we shall take this opportunity to review some of the known results in this area. For example, it is known that:

(a) on any Banach space, each member of the vector space generated by the pseudo-regular functions possesses a minimal subdifferential mapping;
(b) minimality of the Clarke subdifferential mapping is not preserved under addition; (c) on an Asplund space, those Lipschitz functions which possess a minimal subdifferential mapping are strictly Fréchet differentiable on a dense and \( G_\delta \) subset of their domain, while those on a class(\( S \)) Banach space are strictly differentiable on a dense and \( G_\delta \) subset of their domain; (d) minimality of the Clarke subdifferential mapping is separably determined (see [3, 36] for the details).

2. SOME TOPOLOGICAL PREREQUISITES

Throughout the remainder of this paper we shall be interested in the topological behaviour of minimal cuscos and to a lesser extent minimal uscos. So we take this opportunity to “gather-up” some pertinent facts concerning minimal uscos and minimal cuscos. Perhaps the most important amongst these is the following characterization.

**Theorem 2.1** [26, Lemma 2.5]. A cusco (usco) \( \Phi \) from a topological space \( A \) into subsets of a separated locally convex topological space (Hausdorff topological space) \( X \) is a minimal cusco (minimal usco) on \( X \) if, and only if, given any open subset \( U \) of \( A \) and closed and convex subset (closed subset) \( K \) of \( X \), with \( \Phi(U) \not\subseteq K \), there exists a non-empty open subset \( V \) of \( U \) such that \( \Phi(V) \cap K = \emptyset \).

We shall see next that the minimality of a cusco (usco) mapping is preserved under composition with a continuous linear (continuous) function.

**Theorem 2.2.** Let \( \Phi \) be a minimal cusco (minimal usco) from a topological space \( A \) into subsets of a separated locally convex topological space (Hausdorff topological space) \( X \) and let \( f \) be a continuous linear mapping (continuous mapping) from \( X \) into a separated locally convex topological space (Hausdorff topological space) \( Y \). Then the mapping, \( x \rightarrow f(\Phi(x)) \), is a minimal cusco (minimal usco) on \( A \).
Proof. Clearly, \( f \cdot \Phi \) is a cusco (an usco) on \( A \), so it remains to show that it is a minimal cusco (minimal usco) on \( A \). Consider a closed and convex subset (closed subset) \( K \) of \( Y \) and an open set \( U \) in \( A \) such that \((f \cdot \Phi)(U) \not\subseteq K\). Since \( f \) is continuous and linear (continuous) on \( X \), \( f^{-1}(K) \) is a closed and convex subset (closed subset) of \( X \). Since \( \Phi \) is a minimal cusco (minimal usco) and \( \Phi(U) \not\subseteq f^{-1}(K) \) there exists a non-empty open set \( V \subseteq U \) such that \( \Phi(V) \cap f^{-1}(K) = \emptyset \). Hence, \((f \cdot \Phi)(V) \cap K = \emptyset\). 

The following proposition shows that in general there is a close connection between minimal uscos and minimal cuscos.

**Proposition 2.1 [29].** Suppose \( \Psi \) is a minimal usco and \( \Phi \) is a cusco, which both map from a topological space \( A \) into subsets of a separated locally convex topological space \( X \). If \( \Psi(t) \subseteq \Phi(t) \) for each \( t \in A \), then the set-valued mapping \( \Psi': A \rightarrow 2^X \) defined by \( \Psi'(t) = \overline{\triangledown \Psi(t)} \) is a minimal cusco on \( A \), and \( \Psi(t) \subseteq \Phi(t) \) for all \( t \in A \).

**Proof.** Let us show first that \( \Psi' \) is a cusco on \( A \). It is easy to see that for each \( t \in A \), \( \Psi'(t) \) is non-empty, convex and compact. Let \( W \) be a non-empty open subset of \( X \) and consider the set \( U = \{ t \in A : \Psi'(t) \subseteq W \} \). We may, without loss of generality, assume that \( U \neq \emptyset \). So let \( t_0 \in U \). Since \( X \) is a separated locally convex topological space and \( \Psi'(t_0) \) is compact, there exists a convex open neighbourhood \( N \) of 0 in \( X \) such that \( \Psi(t_0) \subseteq \Psi'(t_0) + N \subseteq \Psi(t_0) + N \subseteq W \). Now, \( \Psi \) is an usco on \( A \) so there exists an open neighbourhood \( V \) of \( t_0 \) such that \( \Psi(V) \subseteq \Psi(t_0) + N \). On the other hand, \( \Psi'(t_0) + N \) is closed and convex and so \( \Psi'(t) = \overline{\triangledown \Psi(t)} \subseteq \Psi'(t_0) + N \subseteq W \) for each \( t \in V \). Therefore \( t_0 \in V \subseteq U \); which shows that \( \Psi' \) is a cusco on \( A \). To see that \( \Psi' \) is a minimal cusco, we merely need to appeal to Theorem 2.1.

**Remark 2.1.** In the above proof, the only place where we used the fact that \( \Psi(t) \subseteq \Phi(t) \) for each \( t \), was where we deduced the compactness of \( \overline{\triangledown \Psi(t)} \), and so this condition is not needed when \( X \) is quasi-complete.

**Theorem 2.3.** Consider a minimal cusco (minimal usco) \( \Phi \) from a topological space \( A \) into subsets of a separated locally convex topological (Hausdorff topological) space \( X \).

(i) Given a continuous real-valued function \( g \) defined on \( A \), the set-valued mapping \( g \cdot \Phi \) is a minimal cusco (minimal usco) on \( A \).

(ii) Given a continuous mapping \( T \) from \( A \) into \( X \), the set-valued mapping \( T + \Phi \) is a minimal cusco (minimal usco) on \( A \).

**Proof.** (i) In the case when \( \Phi \) is a minimal usco, \( g \cdot \Phi \) is the composition of the continuous mapping \( F \) from \( R \times X \) into \( X \) defined by
Let $\Phi$ be a minimal cusco, consider the following. Let $\Psi$ be a minimal usco whose graph is contained in Graph($\Phi$). By above, $g \cdot \Psi$ is a minimal usco on $A$ and $g(t) \cdot \Psi(t) \subseteq g(t) \cdot \Phi(t)$ for all $t \in A$. Now, by Proposition 2.1, $g\circ \Psi(t) = \Phi(t)$ for all $t \in A$. Therefore, $g(t) \cdot \Phi(t) = g(t) \cdot \Phi(t) = g\circ \Phi(t)$ for all $t \in A$. So by again appealing to Proposition 2.1 we have that $t \rightarrow g(t) \cdot \Psi(t)$ is a minimal cusco and so $g \cdot \Phi$ is a minimal cusco.

(ii) The mapping $T + \Phi$ is the composition of the continuous linear mapping $S : X \times X \rightarrow X$ defined by $S(x, y) = x + y$ with the minimal cusco (minimal usco) mapping $t \rightarrow (T(t), \Phi(t))$ from $A$ into $X \times X$, and so $T + \Phi$ is a minimal cusco (minimal usco) by Theorem 2.2.

Recently, the notion of minimality, for a set-valued mapping, has been successfully extended outside the class of cusco (usco) mappings, (see, for example, [25, 27, 36, 31]). The key to these extensions is Theorem 2.1. A set-valued mapping $\Phi$ from a topological space $A$ into non-empty subsets of a linear topological space $X$ is hyperplane minimal if for any open half-space $W$ in $X$ and open set $U$ in $A$ with $\Phi(U) \cap W \neq \emptyset$ there exists a non-empty open subset $V \subseteq U$ such that $\Phi(V) \subseteq W$. Similarly, we say that a set-valued mapping $\Phi$ from a topological space $A$ into non-empty subsets of a topological space $X$ is minimal if for any open set $W$ in $X$ and open set $U$ in $A$ with $\Phi(U) \cap W \neq \emptyset$ there exists a non-empty open subset $V \subseteq U$ such that $\Phi(V) \subseteq W$. It follows then, from Theorem 2.1, that a cusco (usco) mapping from a topological space $A$ into subsets of a separated locally convex topological (Hausdorff topological) space $X$ is a minimal cusco (usco) on $A$ if, and only if, it is hyperplane minimal (minimal) on $A$.

Another important notion in the analysis of set-valued mappings, and minimal mappings in particular, is that of a selection. Let $\Phi$ be a set-valued mapping from a non-empty set $A$ into a non-empty set $X$. Then a function $f : A \rightarrow X$ is called a selection of $\Phi$ if $f(t) \in \Phi(t)$ for each $t \in A$.

**Corollary 2.1.** Let $\Omega$ be a densely defined set-valued mapping from a topological space $A$ into subsets of a separated locally convex topological (Hausdorff topological) space $X$. If the graph of $\Omega$ is contained in Graph of a cusco (usco) $\Phi$, then CSC($\Omega$) (USC($\Omega$)) is a minimal cusco (usco) if, and only if, $\Omega$ is hyperplane minimal (minimal).

**Proof.** The proof is a straight-forward application of Theorem 2.1.

Next, we give several “useful” characterizations of minimality.

**Theorem 2.4.** For a cusco mapping $\Phi$, from a topological space $A$ into subsets of a separated locally convex topological space $X$, the following conditions are equivalent:
(i) \( \Phi \) is a minimal cusco on \( A \);

(ii) there exists a densely defined, hyperplane minimal selection \( \sigma \) of \( \Phi \) such that \( \text{CSC}(\sigma) = \Phi \);

(iii) for any densely defined selection \( \sigma \) of \( \Phi \), \( \text{CSC}(\sigma) = \Phi \);

(iv) there exists a densely defined selection \( \sigma \) of \( \Phi \) such that \( \Phi = \text{CSC}(\sigma|_{\text{Dom}(\sigma)}) \) for each dense subset \( D \) of \( \text{Dom}(\sigma) \).

Proof. Corollary 2.1 gives us that (i) \( \iff \) (ii) and clearly (i) \( \implies \) (iii) and (iii) \( \implies \) (iv). So it remains to show that (iv) \( \implies \) (i). We proceed via the characterization given in Theorem 2.1. To this end, let \( U \) be a non-empty open subset of \( A \) and suppose that \( \Phi(U) \not\subseteq K \), where \( K = \{ x \in X : f(x) \leq x \} \), \( x \in R \) and \( f \in X^* \). Choose \( x_0 \in \Phi(U) \setminus K \) such that \( f(x_0) > x + \varepsilon \), for some \( \varepsilon > 0 \) and set \( D' = \{ t \in \text{Dom}(\sigma) \cap U : f(\sigma(t)) \leq x + \varepsilon \} \). Clearly \( D' \) is not dense in \( U \), because if \( D' \) were dense in \( U \) then by hypothesis \( \Phi = \text{CSC}(\sigma|_{\text{Dom}(\sigma)}) \) where \( \text{Dom}(\sigma) \cap U \) and this would imply that \( \text{sup} \{ f(x) : x \in \Phi(U) \} \leq x + \varepsilon \), which is clearly not true. Therefore, there exists a non-empty open subset \( V \) of \( U \) such that \( V \cap D' = \emptyset \). Now consider \( \text{CSC}(\sigma|_{\text{Dom}(\sigma)}) \). Again by hypothesis, \( \text{CSC}(\sigma|_{\text{Dom}(\sigma)}) = \Phi \), but for each \( t \in V \cap \text{Dom}(\sigma) \), \( f(\sigma(t)) > x + \varepsilon \), therefore,

\[
\Phi(V) \cap K = \text{CSC}(\sigma|_{\text{Dom}(\sigma)})(V) \cap K = \emptyset.
\]

This theorem has some important consequences for differentiability theory.

**Corollary 2.2.** Let \( f \) be a densely Gateaux differentiable real-valued locally Lipschitz function defined on a non-empty open subset \( A \) of a Banach space \( X \). Then \( f \) possesses a minimal Clarke subdifferential mapping if, and only if, \( f \) is \( D \)-representable.

Proof. This result follows from parts (i) and (iv) in Theorem 2.4.

Sometimes it is convenient to express \( D \)-representability in terms of sequences. So our next task is to show that on any class(\( S \)) Banach space, \( D \)-representability maybe characterized in terms of sequential limits of Gateaux derivatives. But first, let us recall that a Banach space \( X \) is said to be of class(\( S \)) if every minimal weak* cusco from a Baire space into subsets of \( X^* \) is single-valued at the points of a dense and \( G^\infty \) of its domain. It is well-known that if a Banach space \( X \) is of class(\( S \)) then every continuous convex function defined on a non-empty open convex subset of \( X \) is Gateaux differentiable on a dense and \( G^\infty \) subset of its domain. In fact, this was the original motivation for this class of spaces. In the other direction, it is still an open question as to whether a Banach space \( X \), which has
the property that, every continuous convex function defined on a non-empty open convex subset of \( X \) is Gateaux differentiable at the points of a dense and \( G_\delta \) subset of its domain (that is, a weak Asplund space), is necessarily of class(\( S \)) (see [20] or [44] for further information on class(\( S \)) spaces).

**Lemma 2.1** [3, Lemma 1.4 part (b)]. Let \( X \) be a Banach space whose dual ball is weak* sequentially compact (that is, every sequence in \( B(X^*) \) possesses a weak* convergent subsequence) and let \( \{ A_n : n \in \mathbb{N} \} \) be a decreasing sequence of bounded non-empty subsets of \( X^* \). Then

\[
\bigcap \{ \overline{a_n}^* : n \in \mathbb{N} \} = \overline{\bigcap \{ a \in X^* : a \text{ weak*} \}^*}^* \cap \{ a_n : n \in \mathbb{N} \}
\]

It is well known that class(\( S \)) Banach spaces possess weak* sequentially compact dual balls (see [33] or [24, p. 203]. We may now give a sequential characterization of D-representability.

**Theorem 2.5.** Let \( f \) be a real-valued locally Lipschitz function defined on a non-empty open subset \( A \) of a class(\( S \)) Banach space \( X \). Let \( D \equiv \{ x : \exists x_n \to x : f(x_n) \exists \} \). Then, \( x \to \partial f(x) \), is a minimal weak* cusco on \( A \) if, and only if, for each dense subset \( D^* \) of \( D \)

\[
\partial f(x) = \overline{\partial f(x)_n : x_n \to x} \}
\]

We complete this section by using the results obtained thus far, to determine some properties of locally Lipschitz functions whose Clarke subdifferential mappings are minimal. To do this, we need to recall that a real-valued function \( f \) defined on a non-empty open subset \( A \) of a normed linear space \( X \) is strictly differentiable on \( A \) if, and only if, \( \forall x \in A \) and the mapping, \( x \to \partial f(x) \), is continuous on \( A \), with respect to the weak* topology on \( X^* \) (see [17, p. 32]).

**Theorem 2.6.** Let \( f \) and \( g \) be real-valued locally Lipschitz functions defined on a non-empty open subset \( A \) of a Banach space \( X \). If, \( x \to \partial f(x) \), is a minimal weak* cusco on \( A \) and \( g \) is strictly differentiable on \( A \) then:

(i) \( x \to \partial (f + g)(x) \) is minimal on \( A \) and \( \partial (f + g) = \partial f + \partial g \);

(ii) \( x \to \partial (f \cdot g)(x) \) is minimal on \( A \) and \( \partial (f \cdot g) = f \cdot \partial g + \partial f \cdot g \);

(iii) If \( h : R \to R \) is a strictly differentiable locally Lipschitz function defined on \( R \) then \( x \to \partial (h \cdot f)(x) \) is minimal on \( A \) and \( \partial (h \cdot f) = (\partial h \cdot f) \cdot \partial f \).

**Proof.** (i) By Proposition 2.3.3 in [17], we have that \( \partial (f + g)(x) \subseteq \partial f(x) + \partial g(x) \) for each \( x \in A \). Moreover, since \( g \) is strictly differentiable on \( A \),
\(g(x) = \{Vg(x)\}\) for each \(x \in A\) and so the mapping, \(x \to Vg(x)\), from \(A\) into \((X^*, \text{weak}^*)\) is continuous. Hence, from Theorem 2.3 part (ii), the mapping, \(x \to \partial (f + g)(x)\), is a minimal weak* cusco. On the other hand, the mapping \(x \to \partial (f + g)(x)\) is a weak* cusco on \(A\) and \(\partial (f + g)(x) \subseteq \partial f(x) + \{Vg(x)\}\) for each \(x \in A\). Therefore, \(\partial (f + g)(x) = \partial f(x) + \{Vg(x)\}\) for each \(x \in A\) and \(x \to \partial (f + g)(x)\) is a minimal weak* cusco on \(A\).

(ii) By Proposition 2.3.13 in [17] we have that \(\partial (f \cdot g)(x) \subseteq f(x) \partial g(x) + g(x) \partial f(x)\) for each \(x \in A\). As in part (i), the mapping \(x \to Vg(x)\) is continuous on \(A\) and so the mapping, \(x \to f(x) \partial g(x)\), is continuous on \(A\). Further to this, we have from Theorem 2.3 part (i), that the mapping \(x \to g(x) \partial f(x)\) is a minimal weak* cusco on \(A\). Therefore, we may deduce from Theorem 2.3 part (ii) that the mapping, \(x \to f(x) \partial g(x) + g(x) \partial f(x)\), is a minimal weak* cusco. However, as \(\partial (f \cdot g)(x) \subseteq f(x) \partial g(x) + g(x) \partial f(x)\) for each \(x \in A\), it follows that \(\partial (f \cdot g)(x) = f(x) \partial g(x) + g(x) \partial f(x)\) for all \(x \in A\) and it also follows that, \(x \to \partial (f \cdot g)(x)\), is a minimal weak* cusco on \(A\).

(iii) Theorem 2.3.9 part (ii) of [17] says that \(\partial (h \cdot f)(x) = Vh(f(x)) \partial f(x)\) for each \(x \in A\). Now the mapping, \(x \to Vh(f(x))\), is continuous on \(A\), therefore with the aid of Theorem 2.3 part (i), we see that, \(x \to Vh(f(x)) \partial f(x)\), is a minimal weak* cusco on \(A\). From this we may deduce that \(\partial (h \cdot f)(x) = Vh(f(x)) \partial f(x)\) for each \(x \in A\) and so also deduce that the subdifferential mapping, \(x \to \partial (h \cdot f)(x)\), is a minimal weak* cusco on \(A\).

Note that equality in (i) and (ii) is usually deduced from regularity. Therefore, the new information contained in (i) and (ii) is that the composite function is minimal. In order to establish some further information about minimal subdifferential mappings, we will need to examine more closely the differential structure of the underlying functions.

3. A CHARACTERIZATION OF MINIMAL SUBDIFFERENTIAL MAPPINGS

We begin this section by characterizing minimality of the Clarke subdifferential mapping in terms of a continuity property possessed by the upper Dini directional derivative. We will then use this characterization in conjunction with the results from Section 2 to establish some further properties enjoyed by those locally Lipschitz functions whose subdifferential mappings are minimal.

**Theorem 3.1 (Lebesgue Mean-value Theorem).** Let \(f\) be a real-valued locally Lipschitz function defined on a non-empty open subset of the real line,
which contains the non-degenerate interval \([a, b]\). Then there exists a Borel subset \(M\) of \([a, b]\), with positive measure, such that for each \(t \in M\), \(f'(t)\) exists and

\[
f'(t) \geq \frac{f(b) - f(a)}{b - a}.
\]

Using this theorem we may obtain a well-known characterization of the Clarke generalized directional derivative.

**Proposition 3.1.** Let \(f\) be a real-valued locally Lipschitz function defined on a non-empty open subset \(A\) of a Banach space \(X\). Then for each \(x \in A\) and each \(y \in X\)

\[
f^0(x; y) = \limsup_{z \to x} f^+(z; y) = \limsup_{z \to x} f^-(z; y).
\]

In order to expedite the rest of this section we will introduce the following definition. Let \(A\) be a non-empty Borel subset of a Banach space \(X\). Then a Borel subset \(S\) of \(A\) is 1-D almost everywhere in \(A\), in the direction \(y\), if for each \(x \in A\)

\[
\lambda(\{t \in \mathbb{R} : x + ty \in A \text{ and } x + ty \notin S\}) = 0
\]

(here and later \(\lambda\) will denote the Lebesgue measure on \(\mathbb{R}\)).

For us, the most important example of a 1-D almost everywhere set is the following.

**Proposition 3.2.** Let \(f\) be a locally Lipschitz function defined on a non-empty open subset \(A\) of a Banach space \(X\). Then for each \(y \in S(X)\), \(D_y \equiv \{x \in A : f'(x; y) \text{ exists}\}\) is 1-D almost everywhere in \(A\), in the direction \(y\).

**Theorem 3.2** [36, Theorems 2.14 and 2.16]. Let \(f\) be a locally Lipschitz function defined on a non-empty open subset \(A\) of a Banach space \(X\). Then, \(x \to \partial f(x)\), is a minimal weak* cusc on \(A\) if, and only if, for each \(y \in S(X)\), one of the following conditions holds.

(i) The mapping \(T_y : A \to 2^A\) defined by \(T_y(x) = y(\partial f(x))\) is a minimal cusc.

(ii) The function \(D_y : A \to R\) defined by \(D_y(x) = f^+(x; y)\) is hyperplane minimal on \(A\).

(iii) The restriction of \(D_y\) to a Borel subset \(P_y\), which is 1-D almost everywhere in \(A\), in the direction \(y\), is hyperplane minimal on \(P_y\).
By breaking-down the notion of hyperplane minimality, into its two constituent parts, we are able to refine Theorem 3.2. Let \( f \) be a real-valued function defined on a topological space \( A \). Then \( f \) is quasi lower semi-continuous (quasi upper semi-continuous) on \( A \) if for each \( t_0 \in A, \varepsilon > 0 \) and open neighbourhood \( U \) of \( t_0 \) there exists a non-empty open subset \( V \) of \( U \) such that \( \inf \{ f(t) : t \in V \} > f(t_0) - \varepsilon \) (\( \sup \{ f(t) : t \in V \} < f(t_0) + \varepsilon \)) [30].

From these definitions, it follows that \( f \) is hyperplane minimal on \( A \) if, and only if, it is both quasi upper and quasi lower semi-continuous on \( A \). Let us also make the following observations; (i) \( f \) is quasi lower semi-continuous on \( A \) if, and only if, \( -f \) is quasi upper semi-continuous on \( A \); (ii) if \( D \) is a dense subset of \( A \) and \( f \) is quasi lower semi-continuous on \( A \) (quasi upper semi-continuous on \( A \)) then the restriction of \( f \) to \( D \) is quasi lower semi-continuous on \( D \) (quasi upper semi-continuous on \( D \)).

**Theorem 3.3.** Let \( f \) be a real-valued locally Lipschitz function defined on a non-empty open subset \( A \) of a Banach space \( X \). Then, \( x \rightarrow \partial f(x) \), is a minimal weak* cusco on \( A \) if, and only if, for each \( y \in S(X) \), there exists a Borel subset \( P_y \) of \( A \), which is 1-D almost everywhere in \( A \), in the direction \( y \), such that the function \( D_y : P_y \rightarrow R \) defined by \( D_y(x) \equiv f^+(x; y) \) is quasi lower semi-continuous (quasi upper semi-continuous) on \( P_y \).

**Proof.** Suppose that the mapping, \( x \rightarrow \partial f(x) \), is a minimal weak* cusco on \( A \). Fix \( y \in S(X) \) and set \( P_y \equiv A \). By Theorem 3.2 part (ii) we have that there exists a Borel subset \( P_y \) of \( A \) which is 1-D almost everywhere in \( A \), in the direction \( y \), such that the mapping \( D_y : P_y \rightarrow R \) defined by \( D_y(x) \equiv f^+(x; y) \) is quasi lower semi-continuous (quasi upper semi-continuous) on \( P_y \). Conversely, suppose that for each \( y \in S(X) \) there exists a subset \( P_y \) of \( A \) which is 1-D almost everywhere in \( A \), in the direction \( y \), such that the mapping \( x \rightarrow f^+(x; y) \), is hyperplane minimal on \( A \). Fix \( y \in S(X) \), we will show that there exists a Borel set \( R_y \), let \( S_y \equiv \{ t \in A : f(t; y) \text{ exists} \} \), and define \( R_y \equiv P_y \cap S_y \cap P_{-y} \). Since \( P_y \), \( S_y \), and \( P_{-y} \) are 1-D almost everywhere in \( A \), in the direction \( y \), so is \( R_y \). Now, \( R_y \subseteq S_y \), therefore \( f^+(x; y) = f^-(x; y) = f^+(x; y) = f^-(x; y) \) and so the mapping \( D_y \), restricted to \( R_y \), is both quasi upper and quasi lower semi-continuous on \( R_y \) (that is, \( D_y \) is hyperplane minimal on \( R_y \)), which completes the proof (via Theorem 3.2 part (iii)).

**Theorem 3.4.** Let \( f \) be a real-valued locally Lipschitz function defined on a non-empty open subset \( A \) of a Banach space \( X \). Let \( M \equiv \{ x \in A : f(x) = \inf \{ f(A) \} \} \). Then, \( x \rightarrow \partial f(x) \), is a minimal weak* cusco on \( A \) if, and only if, \( x \rightarrow \partial f(x) \), is a minimal weak* cusco on \( A \setminus M \).
ESSENTIALLY SMOOTH FUNCTIONS

Proof. It follows directly from Theorem 2.1 that if, \( x \rightarrow \partial f(x) \), is a minimal weak* cusco on \( A \) then, \( x \rightarrow \partial f(x) \), is a minimal weak* cusco on \( A \setminus M \). So now we consider the converse. We proceed via the characterization given in Theorem 3.3. To this end, fix \( y \in S(X) \) and let \( P_y \equiv \{ x \in A : f'(x; y) \text{ exists} \} \). By Proposition 3.2, \( P_y \) is 1-D almost everywhere in \( A \), in the direction \( y \). We will show that the mapping \( D_y : P_y \rightarrow R \) defined by

\[ D_y(x) \equiv f'(x; y) = f^+(x; y) \]

is quasi lower semi-continuous on \( P_y \). We may of course, assume that without loss of generality, \( M \neq \emptyset \). Consider a point \( x_0 \in P_y \). Clearly, if \( x_0 \in (\text{int } M \cup A \setminus M) \cap P_y \), then \( D_y \) is quasi lower semi-continuous at \( x_0 \) (see, Theorem 3.2 part (ii)). So we consider the case when \( x_0 \) is on the boundary of \( M \). Let \( U \) be a convex open neighbourhood of \( x_0 \) contained in \( A \), and let \( \varepsilon > 0 \). We may assume, by possibly making \( U \) smaller, that \( f \) is Lipschitz on \( U \) with Lipschitz constant \( K \). Choose \( 0 < t_0 < 1 \) such that \( x_0 + t_0 y \in U \), and choose \( 0 < r < \varepsilon t_0 / K \) such that \( B(x_0 + t_0 y, r) \subseteq U \). Now since \( x_0 \in M \cap P_y \), \( D_y(x_0) = 0 \). Next, we show that there exists a non-empty open subset \( V \subseteq U \) such that \( D_y(z) > -\varepsilon \) for each \( z \in V \cap P_y \). Clearly, if \( B(x_0 + t_0 y, r) \cap \text{int } M \neq \emptyset \) then we are done (choose \( V \equiv B(x_0 + t_0 y, r) \cap \text{int } M \)). In the other case, choose \( x_0 + y' \in B(x_0 + t_0 y, r) \setminus M \). Let \( s \equiv \max \{ t \in [0, 1] : x_0 + t y' \in M \} \). Then,

\[
\frac{f(x_0 + y') - f(x_0 + sy')}{1 - s} > 0.
\]

Hence, by the Lebesgue mean-value theorem there exists a number \( s_0 \in (s, 1) \) such that \( f'(x_0 + s_0 y', y') > 0 \). Moreover, since \( x_0 > s \), \( x_0 + s_0 y' \notin M \). Therefore, by the minimality of, \( x \rightarrow \partial f(x) \), on \( A \setminus M \), there exists a non-empty open subset \( V \subseteq U \setminus M \) such that \( f^+(z; y') > 0 \) for each \( z \in V \), and by positive homogeneity, \( f^+(z; t_0^{-1} y') > 0 \) for each \( z \in V \). However, by our choice of \( y' \), \( \| t_0 y - y' \| < r < \varepsilon t_0 / K \) and so, \( D_y(z) = f^+(z; y) = f^+(z; t_0^{-1} y') + (f^+(z; y) - f^+(z; t_0^{-1} y')) \geq f^+(z; t_0^{-1} y') - \varepsilon > -\varepsilon \) for each \( z \in V \cap P_y \). This ends the proof. 

COROLLARY 3.1. Let \( f \) and \( g \) be real-valued locally Lipschitz functions defined on a non-empty open subset \( A \) of a Banach space \( X \). If, \( x \rightarrow \partial f(x) \), is a minimal weak* cusco on \( A \) and \( g \) is strictly differentiable on \( A \) then:

(i) \( f^+ \) and \( f^- \) possess minimal subdifferential mappings, (here \( f^+(x) \equiv \max \{ f(x), 0 \} \) and \( f^-(x) \equiv \min \{ f(x), 0 \} \));

(ii) \( |f| \) possesses a minimal subdifferential mapping;

(iii) \( x \rightarrow \max \{ f(x), g(x) \} \) and \( x \rightarrow \min \{ f(x), g(x) \} \) possess minimal subdifferentials.
Proof. (i) The proof that \( f^+ \) and \( f^- \) possess minimal subdifferential mappings follows directly from Theorem 3.4. (ii) Similarly, the proof that \( |f| \) possesses a minimal subdifferential mapping also follows directly from Theorem 3.4. (iii) Observe that \( \max \{ f(x), g(x) \} = (f-g)^+(x) + g(x) \) and that \( \min \{ f(x), g(x) \} = (f-g)^-(x) + g(x) \). Now by Theorem 2.6 part (i) \( f-g \) possesses a minimal subdifferential mapping and so, by part (i) above, \( (f-g)^+ \) and \( (f-g)^- \) both possess minimal subdifferential mappings. The proof is completed by again appealing to Theorem 2.6 part (i).

By far and away the most important application of Theorem 3.4 is to distance functions. Let \( C \) be a non-empty closed subset of a Banach space \((X, \| \cdot \|)\). The distance function associated with the set \( C \) and the norm \( \| \cdot \| \), (denoted by \( d_C \)), is defined by, \( d_C(x) = \inf \{ \|x-c\| : c \in C \} \). We may now obtain a notable fact concerning the minimality of the Clarke subdifferential mapping of a distance function.

**Theorem 3.5.** Let \( C \) be a non-empty closed subset of a Banach space \( X \). Then \( d_C \) possesses a minimal subdifferential mapping on \( X \) if, and only if, \( x \to \partial d_C(x) \) is a minimal weak* cusco on \( X^* \).

4. ESSENTIALLY SMOOTH LIPSCHITZ FUNCTIONS

In this section of the paper we will define a class of locally Lipschitz functions whose subdifferential mappings are both minimal and integrable. This class of functions contains all the sub-regular and all the semi-smooth functions considered in [18, 35]. In this way, we are able to generalize, in a unified manner, the various results contained in [3, 7, 18, 19, 28, 36, 39, 42, 43] (at least in the case of Lipschitz functions).

We will call a Borel subset \( N \) of a separable Banach space \( X \), a Haar-null set if there exists a (not necessarily unique) Borel probability measure \( p \) on \( X \), such that \( p(x+N) = 0 \) for each \( x \in X \). (In such a case, we shall call the measure \( p \) a test-measure for \( N \)). More generally, we say that a subset \( N \subseteq X \) is a Haar-null set if it is contained in a Borel Haar-null set.

The Haar-null sets are closed under translation and countable unions, [14]. It follows therefore, that if \( N \) is a Haar-null set then \( X\setminus N \) is dense in \( X \). In finite dimensions, the Haar-nulls sets coincide with the Lebesgue null sets. Further, we shall say that a property \( P \) holds almost everywhere in \( A \) if \( \{ x \in A : P(x) \) is not true \} is a Haar-null set. Using this terminology J. P. R. Christensen has shown (see, [15]) that each real-valued locally Lipschitz function defined on a non-empty open subset of a separable Banach space, is Gateaux differentiable almost everywhere (in its domain). In fact, the following even stronger result is known.

---

Codes: 3211 Signs: 2616 . Length: 45 pic 0 pts, 190 mm
**Theorem 4.1** [46, Proposition 2.2]. Let $f$ be a real-valued locally Lipschitz function defined on a non-empty open subset $A$ of a separable Banach space $X$ and let $D \equiv \{ x \in A : \nabla f(x) \text{ exists} \}$. Then for each Haar-null set $N \subseteq X$ and each $x \in X$ we have that

$$
\partial f(x) = \overline{conv}^* \{ x^* \in X^* : x^* = \text{weak}^* \lim_{x_n \to x} \nabla f(x_n), \text{ and } x_n \in D \setminus N \}.
$$

**Corollary 4.1.** Let $f$ be a real-valued locally Lipschitz function defined on a non-empty open subset $A$ of a separable Banach space $X$ and let $D \equiv \{ x \in A : \nabla f(x) \text{ exists} \}$. Then, $x \to \partial f(x)$, is a minimal weak* cusco on $A$ if, and only if, the mapping, $x \to \nabla f(x)$, (defined almost everywhere on $D$) is weak* hyperplane minimal almost everywhere on its domain.

**Proof.** Let $N$ be any Haar-null subset of $X$ such that, $x \to \nabla f(x)$, is defined, and weak* hyperplane minimal on $D \setminus N$. Then, by Theorem 4.1, we have that $\partial f = \text{CSC}(\nabla f)$. The result now follows from Theorem 2.4 part (ii).

The significance of the previous result is that it entitles us to neglect certain “small” subsets when determining the global minimality of the Clarke subdifferential mapping. Next, we shall consider an important sub-class of the $D$-representable locally Lipschitz functions. Let $A$ be a non empty open subset of a separable Banach space $X$. Following ([3], p. 68) we will say that a real-valued locally Lipschitz function $f$ defined on $A$ is essentially smooth or smooth almost everywhere on $A$, if $f$ is strictly differentiable everywhere on $A$ except possibly on a Haar-null set. We will denote by $S_e(A)$ the family of all real-valued essentially smooth locally Lipschitz functions defined on $A$. Let us also note that this class of functions has also been considered in [42], at least in the case when $X$ is finite dimensional. Our first two tasks are to show that, each member of $S_e(A)$ possesses a minimal subdifferential mapping and to show that $S_e(A)$ contains a significant class of functions. We begin with the following characterization.

**Theorem 4.2.** Let $f$ be a real-valued locally Lipschitz function defined on a non-empty open subset $A$ of a separable Banach space $X$ and let $D \equiv \{ x \in A : \nabla f(x) \text{ exists} \}$. Then $f \in S_e(A)$ if, and only if, the mapping, $x \to \nabla f(x)$, (defined on $D$) is norm to weak* continuous almost everywhere in $D$.

**Proof.** This follows from Theorem 4.1 and Proposition 1.1 part (a).
Proof. This result follows from Theorem 4.2 and Corollary 4.1.

Let $f$ be a real-valued locally Lipschitz function defined on a non-empty open subset $A$ of a Banach space $X$. Then $f$ is upper hemi-smooth (lower hemi-smooth) at a point $x \in A$, in the direction $y$ if,

$$f^+(x; y) \geq \limsup_{t \to 0^+} f^0(x + ty; y) \quad (f^-(x; y) \leq \liminf_{t \to 0^+} f^0(x + ty; -y))$$

Remark 4.1. If we define

$$T_y: A \to R \text{ by, } T_y(x) = \limsup_{t \to 0^+} f^0(x + ty; y)$$

and

$$S_y: A \to R \text{ by, } S_y(x) = \liminf_{t \to 0^+} -f^0(x + ty; -y)$$

then it is easy to check that both $T_y$ and $S_y$ are Borel measurable on $A$. Hence, the set of points in $A$ where $f$ is upper (lower) hemi-smooth in the direction $y$, is always a Borel subset of $A$. Indeed, to see that $T_y$ is Borel measurable, it suffices to observe that

$$T_y(x) = \lim_{n \to \infty} g_n(x) \text{ where, } g_n(x) = \sup \{ f^0(x + ty; y) : 0 < t \leq 1/n \}$$

and

$$g_n(x) = \lim_{m \to \infty} f^m_n(x) \text{ where, } f^m_n(x) = \max \{ f^0(x + ty; y) : 1/m \leq t \leq 1/n \}$$

(for each $m > n$) is upper semi-continuous on $A$. A similar argument shows that $S_y$ is also Borel measurable.

If $X$ is a separable Banach space then we say that $f$ is essentially upper hemi-smooth (essentially lower hemi-smooth) on $A$, if for each $y \in S(X)$ the set of all points in $A$ where $f$ is not upper hemi-smooth (lower hemi-smooth) is a Haar-null set. We shall also say that $f$ is pseudo-regular at $x$ in the direction $y$ if, $f^0(x; y) = f^+(x; y)$ and we shall say that $f$ is pseudo-regular at $x$, if it is pseudo-regular at $x$, in every direction $y$.

**Lemma 4.1.** Let $f$ be a real-valued locally Lipschitz function defined on a non-empty open subset $A$ of a Banach space $X$. Then for each $y \in S(X)$ the (Borel) set

$$F_y = \{ x \in A : f^0(x; y) > T_y(x) \} \quad (E_y = \{ x \in A : -f^0(x; -y) < S_y \})$$
has the property that for each \(a \in A\), \(F_y(a) \equiv \{ r \in R : a + ry \in F_y \} \) (\(E_y(a) \equiv \{ r \in R : a + ry \in E_y \} \)) is at most countable.

**Proof.** Fix \(y \in S(X)\) and \(a \in A\). We will show that \(F_y(a)\) is at most countable (the proof that \(E_y(a)\) is countable is identical to this). Note that without loss of generality we may assume that \(F_y(a)\) is non-empty. So in this case, we define \(s: F_y(a) \to \mathbb{Q}^2\) by, \(s(t) \equiv (r_1, r_2)\) where \(r_1 \in (T_y(a + ty), f^0(a + ty; y)) \cap \mathbb{Q}\) and \(r_2 \in (t, \infty) \cap \mathbb{Q}\) is chosen so that

\[
\sup \{ f^0(a + ty; y) : t < r < r_2 \} < r_1
\]

It is easy to see that \(s\) is 1-to-1 and so, \(F_y\) must be at most countable (here, \(\mathbb{Q}\) denotes the rational numbers).

**Remark 4.2.** If \(X\) is a separable Banach space, then for each \(y \in S(X)\), \(\{ x \in A : f\) is upper hemi-smooth, but not pseudo-regular, in the direction \(y\}\) is contained in \(F\), and hence is a Haar-null set (in this case, we may take the normalised Lebesgue measure, supported on \(sp[y]\), as a test-measure for the Borel set \(F\)).

**Proposition 4.1.** Let \(f\) be a real-valued locally Lipschitz function defined on a non-empty open subset \(A\) of a separable Banach space \(X\) and let \(\{ y_n : n \in \mathbb{N} \} \) be a dense subset of \(S(X)\). If for each \(n \in \mathbb{N}\), \(f\) is almost everywhere pseudo-regular in the direction \(y_n\), then \(f \in S_x(A)\).

**Proof.** For each \(n \in \mathbb{N}\), let \(P_n\) be the set of all points in \(A\) where \(f\) is pseudo-regular in the direction \(y_n\). Let \(D \equiv \{ x \in A : \nabla f(x) \text{ exists} \}\). By Theorem 7.5 in [15], \(A \setminus D\) is a Haar-null set. Now, let \(S \equiv \bigcap \{ P_n : n \in \mathbb{N} \} \cap D\). We claim that \(f\) is strictly differentiable at each point of \(S\). To see this, consider \(x_0 \in S\). Then \(f^0(x_0; y_n) = \nabla f(x_0)(y_n)\) for each \(n \in \mathbb{N}\). However, since both mappings, \(y \to f^0(x_0; y)\), and \(y \to \nabla f(x_0)(y)\), are continuous on \(X\) we must have that \(f^0(x_0; y) = \nabla f(x_0)(y)\) for each \(y \in S(X)\). This shows that \(f\) is strictly differentiable at \(x_0\).

We may now establish a fundamental (and initially surprising) fact.

**Corollary 4.3.** If \(f\) is a real-valued essentially upper hemi-smooth (essentially lower hemi-smooth) locally Lipschitz function defined on a non-empty open subset \(A\) of a separable Banach space \(X\), then \(f \in S_x(A)\).

**Proof.** The proof follows from Lemma 4.1 (and the subsequent remark) and Proposition 4.1.

Next, we show that each member of \(S_x(A)\) is integrable.
PROPOSITION 4.2 [3, Proposition 4.4]. Let $A$ be a non-empty open subset of a separable Banach space $X$. Then each member of $S_e(A)$ is integrable.

Proof. Suppose that $f \in S_e(A)$ and $g$ is a real-valued locally Lipschitz function defined on $A$ such that $\partial g(x) \subseteq \partial f(x)$ for all $x \in A$. Let $h = f - g$, then $\nabla h(x) = 0$ almost everywhere in $A$, since $\nabla f(x) = \nabla g(x)$ at each point of $A$ where $\partial f(x)$ is a singleton. The result now follows directly from Theorem 4.1.

Let us now establish some stability properties for $S_e(A)$. A first but naive guess might be that if $f_1, f_2, \ldots, f_n \in S_e(A)$ and $g \in S_e(R^n)$, then $g \circ f \in S_e(A)$, where $f = (f_1, f_2, \ldots, f_n)$. However, the following example shows that in general this is not true (when $n \geq 2$).

EXAMPLE 4.1. Let $X$ be a separable Banach space let $C$ be a Cantor subset of $R$ with positive Lebesgue measure. Define the functions $f_1, \ldots, f_n$ on $X$ by, $f_j \in X^* \setminus \{0\}$ and $f_j \equiv 0$ for each $1 \leq j < n$. Further, we define $d_C : R \rightarrow R$ by $d_C(t) \equiv \text{dist}(t, C)$ and $g : R^n \rightarrow R$ by $g(x_1, x_2, \ldots, x_n) \equiv \text{dist}((x_1, x_2, \ldots, x_n), \{0\} \times \{0\} \times \cdots \times C)$ (here the distance is with respect to the Euclidean norm on $R^n$). Clearly, each $f_j$ is strictly differentiable on $X$. Moreover, by Theorem 6.2 we have that $g \in S_e(R^n)$.

We claim that $g \circ f \not\in S_e(X)$, where $f \equiv (f_1, f_2, \ldots, f_n)$. To see this, observe that $g \circ f(x) \equiv d_C(f_j(x))$. Now, it is standard that $d_C$ is not strictly differentiable at any point of $C$. Hence, it follows that $g \circ f$ is not strictly differentiable at any point of $f_n^{-1}(C)$ which is not a Haar-null set (see the remark just after Theorem 6 in [14]). Therefore, $g \circ f \not\in S_e(X)$.

Despite this example $S_e(A)$ does possess very strong closure properties. In the next part of this paper we will need to consider vector-valued functions. Let $A \subseteq R$ and $x : A \rightarrow R^n$ be defined by

\[
x(t) \equiv (x_1(t), x_2(t), \ldots, x_n(t)) \quad \text{where} \quad x_j : A \rightarrow R.
\]

Then we say that the vector-valued function $x$ is essentially smooth on $A$ if $x_j \in S_e(A)$ for each $1 \leq j \leq n$, and in this case we write: $x \in S_e(A; R^n)$. Further to this, we will say that a real-valued locally Lipschitz function $f$ defined on a non-empty open subset $U$ of $R^n$ is arc-wise essentially smooth on $U$, if for each locally Lipschitz function $x \in S_e((0, 1); R^n)$

\[
\lambda(\{ t \in (0, 1) : f_i'(x(t)) \neq f_i'(x(t)) \}) = 0
\]

where $x'(t) \equiv (x_1'(t), x_2'(t), \ldots, x_n'(t))$.

We shall denote by $A_e(U)$, the family of all arc-wise essentially smooth functions on $U$. 
Remark 4.3. It is easily seen that the definition of arc-wise essential smoothness is unaffected by replacing the open set \((0, 1)\) (given in the definition) by any other non-empty open subset of \(R\).

**Theorem 4.3.** Let \(A\) be a non-empty open subset of a separable Banach space \(X\). If \(f_1, f_2, \ldots, f_n \in S(A)\) and \(U\) is a non-empty open subset of \(R^n\) that contains \(f(A)\), where \(f \equiv (f_1, f_2, \ldots, f_n)\), then for each \(g \in A(U)\), \(g \circ f \in S(A)\).

**Proof.** It suffices (see Theorem 4.5) to show that for each \(y \in S(X)\), \((g \circ f)^0(x; y) = -(g \circ f)^0(x; -y)\) almost everywhere in \(A\). Fix \(y \in S(X)\). Let \(D\) be the \((G_\delta)\) set of all points in \(A\) where \(f_j^0(x; y) = -f_j^0(x; -y)\) for each \(1 \leq j \leq n\) and let \(P_n = \{ x \in A : (g \circ f)^0(x; y) = -(g \circ f)^0(x; -y)\}\). Clearly \(P_n\) is a Borel set, in fact \(P_n\) is a \(G_\delta\) set. Let \(H\) be any closed hyperplane in \(X\) such that \(y \notin H\). Now consider the isomorphism \(T : H \times R \to X\) defined by \(T(h, t) = h + ty\). Let

\[
H_0 = \{ h \in H : \lambda(\{ t \in R : T(h, t) \in A \setminus D \}) = 0 \}.
\]

By the remark just after Theorem 6 in [14] we see that \(H \setminus H_0\) is a Haar-null set in \(H\), since \(A \setminus D\) is a Haar-null set in \(X\). To show that \(A \setminus P_n\) is a Haar-null set in \(X\) it suffices (also because of the remark made after Theorem 6 in [14]) to show that for each \(h \in H_0\), \(\lambda(\{ t \in R : T(h, t) \in A \setminus P_n \}) = 0\).

To this end, consider \(h_0 \in H_0\). Let \(A_{h_0} = \{ t \in R : T(h_0, t) \in A \}\). If \(A_{h_0} = \emptyset\) then we are done, so let us suppose that \(A_{h_0} \neq \emptyset\). Define \(z : A_{h_0} \to U\) by, \(z(t) = f(h_0 + ty)\). Since \(h_0 \in H_0, z \in S(A_h; R^n)\). Let

\[
D_j \equiv \{ t \in A_{h_0} : f_j^0(h_0 + ty; y) = -f_j^0(h_0 + ty; -y) \text{ for all } 1 \leq j \leq n \}
\]

and

\[
D_x \equiv \{ t \in A_{h_0} : g^0(z(t); z'(t)) = -g^0(z(t); -z'(t)) \}.
\]

Now define \(D_0 \equiv D_j \cap D_x\). Clearly, \(\lambda(A_{h_0} \setminus D_0) = 0\). We claim that

\[
(g \circ f)^0(h_0 + ty; y) = -(g \circ f)^0(h_0 + ty; -y)
\]

at each point \(t \in D_0\). To see this, consider an arbitrary point \(t_0 \in D_0\). Set \(x_0 = h_0 + t_0y\), then \(g^0(f(x_0); f'(x_0; y)) = -g^0(f(x_0); -f'(x_0; y))\) as \(t_0 \in D_x\) and \(f_j^0(x_0; y) = -f_j^0(x_0; -y)\) for each \(1 \leq j \leq n\) as \(t_0 \in D_j\). It is now standard that

\[
(g \circ f)^0(x_0; y) = -(g \circ f)^0(x_0; -y).
\]

This completes the proof. 

\[\square\]
The fact that Theorem 4.4 provides us with some strong closure properties for $S_*(A)$ derives from the following proposition.

**Proposition 4.3** [9]. Let $U$ be a non-empty open subset of $\mathbb{R}^n$. Then:

(a) $A_*(U) \subseteq S_*(U)$;

(b) $A_*(U)$ is closed under composition, that is, if $f_1, f_2, \ldots, f_n \in A_*(U)$ and $g \equiv (f_1, f_2, \ldots, f_n)$ then $g \cdot f \in A_*(U)$ where $f \equiv (f_1, f_2, \ldots, f_n)$;

(c) $A_*(U)$ contains all the upper semi-smooth (lower semi-smooth) locally Lipschitz functions defined on $U$.

Recall that a real-valued locally Lipschitz function $f$ defined on a non-empty open subset $A$ of a Banach space $X$ is called upper semi-smooth (lower semi-smooth), if for each $x \in A$ and $y \in S(X)$

$$f^+(x; y) \geq \limsup_{t \to 0^+} f^+(x + ty; y) \quad (f^-(x; y) \leq \liminf_{t \to 0^+} f^-(x + ty; y))$$

Note that if $f$ is upper semi-smooth (lower semi-smooth) on $A$, then it is upper hemi-smooth (lower hemi-smooth) on $A$.

**Corollary 4.4**. Let $A$ be a non-empty open subset of a separable Banach space $X$, then $S_*(A)$ is closed under addition, subtraction, multiplication and division (when this is defined), as well as, the lattice operations.

**Proof.** In each case $g$ is upper semi-smooth on $\mathbb{R}^2$.

Although in general, $S_*(A)$ is not closed under composition, we have from the next theorem, that if $f \in S_*(A)$ and $g \in S_*(\mathbb{R})$, then $g \cdot f \in S_*(A)$.

**Theorem 4.4**. If $U$ is a non-empty open subset of $\mathbb{R}$ then $A_*(U) = S_*(U)$.

**Proof.** We see from the above proposition that $A_*(U) \subseteq S_*(U)$ and so we need only show the reverse inclusion. To this end let $f \in S_*(U)$ and let $x \in S_*(\{(0, 1); U\})$. Now define, $C = \{t \in U : f'\!(t; 1) = -f'\!(t; -1)\}$ and $D = \{t \in (0, 1) : x(t) \in C \text{ or } x'(t) = 0\}$. We need to show that $\lambda(D) = 1$, since $f'\!(x(t); x'(t)) = -f'\!(x(t); -x'(t))$ at almost all points of $D$. However, this follows from the fact that if $E \subseteq (0, 1)$ (and $x$ is differentiable at each point of $E$) and the Lebesgue outer-measure of $x(E)$ is zero, then $x'(t) = 0$ for almost all $t \in E$ (see, Lemma 6.92 in [45]).
Remark 4.4. It follows from Theorem 4.4 that the (distance) function \( g \)
defined in Example 4.1 lies in \( S_e(R^n) \setminus A_e(R^n) \) \((n \geq 2)\). However, by translation and dilation one can show that for every non-empty open subset \( U \) of \( R^n \) \((n \geq 2)\), \( A_e(U) \) is a proper subset of \( S_e(U) \).

Our investigation of the properties of \( S_e(A) \) is ended with the following theorem, which provides a condition sufficient to ensure membership in \( S_e(A) \).

**Theorem 4.5.** Let \( f \) be a real-valued locally Lipschitz function defined on a non-empty open subset \( A \) of a separable Banach space \( X \). Let \( B \) be a subset of \( X \) such that \( \overline{B} = X \). If for each \( b \in B \), \( f^b(x; b) = -f^b(x; -b) \) almost everywhere in \( A \), then \( f \in S_e(A) \).

**Proof.** The proof of this follows directly from the fact that if (for some point \( x_0 \in A \)) \( f^0(x_0; y) = -f^0(x_0; -y) \) for each \( y \in Y \subseteq X \), then
\[
f^0(x_0; y) = -f^0(x_0; -y) \quad \text{for each} \quad y \in \overline{Y}.
\]

We finish this section of the paper with some general comments.

Our first comment concerns our choice of null set. Indeed, we note here that our choice of using Haar-null sets (as defined by J. P. R. Christensen) as our “null” sets was reasonably arbitrary (except for the fact that the larger the class of null sets, the larger \( S_e(A) \) becomes: Recall that in [38] it is shown that the Haar-null sets contain all the Gaussian null sets, which in turn contain all the Aronszajn null sets; see also [2] for further information). In fact, the only properties that we required of our \( \sigma \)-ideal of null sets were:

(i) no open set is a null set;
(ii) the formula in Theorem 4.1 holds;
(iii) a Borel subset \( A \subseteq H \times R \) is a null set if, and only if,
\[
\lambda\left\{ t \in R : (h, t) \in A \right\} = 0 \quad \text{for almost all} \ h \in H.
\]

Our other comment pertains to some recent extensions of Haar-null sets to spaces which are not necessarily Polish (We say a that Borel subset \( N \) of a Banach space \( X \) is a Haar-null set if there exists a Radon probability measure \( p \) on \( X \) such that \( p(x + N) = 0 \) for all \( x \in X \).) In this way, we can define the essentially smooth functions on any Banach space \( X \), in the following manner. We will say that a real-valued locally Lipschitz function \( f \) defined on a non-empty open subset \( A \) of \( X \) is essentially smooth on \( A \) if for each \( y \in S_e(X) \), \( \{ x \in A : f^0(x; y) \neq -f^0(x; -y) \} \) is a Haar-null set. Using this definition the authors in [10, 11] have extended some of the results in this paper to arbitrary Banach spaces.
5. PERTURBATION FUNCTIONS

In this section of the paper, we apply the results of Section 4 to perturbation functions. Let $A$ be a non-empty open subset of a Banach space $X$ and let $T$ be a topological space. We say that a real-valued function $f: A \times T \to R$ is locally Lipschitz on $A$, uniformly in $T$ if for each $x_0 \in A$ there exists an $K > 0$ and $\delta > 0$ such that

$$|g(x, t) - g(y, t)| \leq K|x - y|$$

for all $x, y \in B(x_0, \delta)$ and $t \in T$.

Further, we say that an extended real-valued function $f$ defined on $A$ is a sup-marginal function if $f(x) = \sup\{g(x, t); t \in T\}$ for some function $g: A \times T \to R$. If more stringently, we have that $f(x) = \max\{g(x, t); t \in T\}$ and $g$ is locally Lipschitz on $A$, uniformly in $T$, then $f$ is real-valued and locally Lipschitz on $A$. A set-valued mapping $M$ from a topological space $A$ into non-empty subsets of a topological space $T$ will be said to be semi-continuous on $A$ if, for each $x \in A$ and each net $(x_n)_{n \in I}$ in $A$, converging to $x$, there exists a point $y \in M(x)$ and elements $y_n \in M(x_n)$ such that $y$ is an accumulation point of the set $\{y_n; z \in I\}$, that is, $y \in \overline{\{y_n; z \in I\}}$. (Note that this definition is less arduous than that given in [18, 19].) The following theorem unifies Theorems 6.1 and 6.2 in [18] and Proposition 2.6 in [19].

**Theorem 5.1.** Let $A$ be a non-empty open subset of a separable Banach space $X$ and let $T$ be a Hausdorff topological space. Let $g: A \times T \to R$ be locally Lipschitz on $A$, uniformly in $T$ and let $f: A \to R$ be defined by $f(x) = \max\{g(x, t); t \in T\}$. Furthermore, suppose that (i) the set-valued mapping $M: A \to 2^T$, defined by $M(x) = \{t \in T; f(x) = g(x, t)\}$, is semi-continuous on $A$ and that (ii) for each $x \in A$ and each $y \in B \cup -B$, the mapping $(r, y', t) \in R^+ \times X \times T \to g^+(x + ry', t; y)$ is upper semi-continuous (as a real-valued function) at each point of $[0] \times \{y\} \times M(x)$ (here $B$ is any subset of $X$ such that $\overline{\{B = X\}$, then $f \in S(A)$).

**Proof.** To show that $f \in S(A)$, it suffices by Remark 4.2, to show that $f$ is upper semi-smooth in the direction $y$ on $A$, for each $y \in B \cup -B$. Let $x$ be a fixed element of $A$ and $y$ be a fixed element of $B \cup -B$. We will show that for any sequence of positive real numbers $\{s_n; n \in N\}$ converging to 0 and any sequence $\{y_n; n \in N\}$ of elements of $X$ converging to $y$, we have that $\liminf\{f^+(x + s_n y_n; y); n \to \infty\} \leq f^+(x; y)$. Indeed, by a standard subsequence argument this will show that $f$ is upper semi-smooth at $x$, in the direction $y$. So let $\{s_n; n \in N\}$ be a sequence of positive real numbers converging to 0 and let $\{y_n; n \in N\}$ be a sequence of elements of...
X converging to \( y \). For each \( n \in \mathbb{N} \), we may choose \( 0 < \lambda_n < s_n \) such that

\[
\frac{f^+(x + s_n y_n; y) - f(x + s_n y_n; y)}{\lambda_n} < \frac{f(x + s_n y_n + \lambda_n y) - f(x + s_n y_n)}{\lambda_n} + 1/n.
\]

Since \( M \) is semi-continuous on \( A \) and

\[
\lim_{n \to \infty} (x + s_n y_n + \lambda_n y) = x,
\]

there exists a point \( t \in M(x) \) and a sequence \( \{t_n; n \in \mathbb{N}\} \) in \( T \) such that \( t_n \in M(x + s_n y_n + \lambda_n y) \) for each \( n \in \mathbb{N} \) and \( n \in \{t_n; n \in \mathbb{N}\} \setminus \{t\} \). Now, for each \( n \in \mathbb{N} \), we have that

\[
\frac{f(x + s_n y_n + \lambda_n y) - f(x + s_n y_n)}{\lambda_n} \leq \frac{g(x + s_n y_n + \lambda_n y, t_n) - g(x + s_n y_n, t_n)}{\lambda_n}.
\]

Furthermore, by the Lebesgue mean-value theorem we have that for each \( n \in \mathbb{N} \) there exists a real number \( s_n' \) such that \( 0 < s_n' < \lambda_n \) and

\[
\frac{g(x + s_n y_n + \lambda_n y, t_n) - g(x + s_n y_n, t_n)}{\lambda_n} \leq \frac{g^+(x + s_n y_n + s_n' y, t_n; y) + 1/n}{1/n}.
\]

Therefore, for each \( n \in \mathbb{N} \),

\[
f^+(x + s_n y_n; y) \leq g^+(x + s_n y_n + s_n' y, t_n; y) + 2/n.
\]

Now, let \( s_n'' = (s_n + s_n') \) and \( y_n'' = (s_n y_n + s_n' y)/s_n'' \). Then clearly,

\[
\lim_{n \to \infty} y_n'' = y \quad \text{and} \quad \lim_{n \to \infty} s_n'' = 0.
\]

Hence,

\[
\liminf_{n \to \infty} f^+(x + s_n y_n; y) \leq \liminf_{n \to \infty} g^+(x + s_n y_n + s_n'' y; t_n; y)
\]

\[
= \lim_{n \to \infty} g^+(x + s_n'' y_n, t_n; y)
\]

\[
\leq g^+(x, t; y) \quad \text{(since } t \in M(x) \text{).}
\]

In particular, condition (i) holds if \( T \) is compact and the function \( t \to g(x, t) \) is upper-semi-continuous on \( T \) (or more generally, if \( M \) is an usco mapping on \( A \)); (ii) is fulfilled if the mapping, \( (x, t) \to g^+(x, t; y) \), is upper semi-continuous on \( A \times T \), for each \( y \in X \).
6. DISTANCE FUNCTIONS

Let us first examine distance functions defined on finite-dimensional Banach spaces. For the most part, we will only consider distance functions that are defined by smooth norms. The reason for this is revealed in the next theorem.

**Theorem 6.1.** Let $(X, \cdot \cdot)$ be a Banach space. If each distance function on $X$ possesses a minimal subdifferential mapping, then the norm $\cdot \cdot$ on $X$ is smooth.

**Proof.** Suppose that the norm $\cdot \cdot$ is not smooth at a point $x_0 \in S(X)$ (Note that there is no loss of generality in assuming that $x_0 \in S(X)$). Then there exist two distinct linear functionals $x_1^* \notin dom \phi(x_0)$ and $x_2^* \notin dom \phi(x_0)$ such that $x_1^*(x_0) = x_2^*(x_0) = 1$. Let $x_0 = x_1^* + x_2^*$. Let $K_1 = \ker(x_1^*)$, $K_2 = \ker(x_2^*)$ and $K_3 = \ker(x_1^* + x_2^*)$. Clearly, $K_1 \cap K_2 \subseteq K_3$. Choose $z \in K_3 \setminus (K_1 \cap K_2)$ such that $x_1^*(z) = 1$ and $x_2^*(z) = -1$. Let us recall that on p. 216, Example 6 part (e) of [45] (see also, Example 8.2), an example is given of an everywhere differentiable Lipschitz function $f: R \rightarrow R$ which is strictly increasing on $R$ and for which the set $\{x \in R: f'(x) = 0\}$ is dense in $R$. Moreover, this function $f$ is a strict contraction on $R$, that is, $|f(x) - f(y)| < |x - y|$ whenever $x \neq y$. Let us note that each element $x \in X$ can be uniquely expressed as $x = x_1^* + x_2^* x_0$, where $x_1^* \in K_1 \cap K_2$ and $x_2^* \in K_2$. Furthermore, $\mu_x = x_1^*(x)$ and $\lambda_x = 1/2(x_1^*(x) - x_2^*(x))$, and so, both mappings, $x \mapsto \mu_x$, and, $x \mapsto \lambda_x$, are continuous and open on $X$. Let $C = \{x \in X: \mu_x \geq f(\lambda_x)\}$ (it is instructive to think of $C$ as the epigraph of the real-valued function $f: K_1 \rightarrow R$, defined by $f_x(k + \lambda x) \equiv f(\lambda)$). Clearly, $C$ is a proper, non-empty closed subset of $X$. We will show that $x \mapsto \partial d_C(x)$ is not a minimal weak* cusco on $X \setminus C$. We claim that $\sigma: X' \setminus C \rightarrow \sigma(\lambda_x)$, defined by $\sigma(x) \equiv x + (f(\lambda_x) - \mu_x) x_0$ is a selection of the metric projection on $X \setminus C$. (Note that if this is the case, then $d_C(x) = f(\lambda_x) - \mu_x$.) To prove this, consider a point $x \in X \setminus C$. We will show first that $\sigma(x) \in C$. To see this, consider the following:

$\mu_{\sigma(x)} = \mu_{x} + (f(\lambda_x) - \mu_x) = f(\lambda_{\sigma(x)}) \quad \text{since} \quad \lambda_x = \hat{\lambda}_{\sigma(x)}. \quad (*)$

Therefore, $f(\lambda_{\sigma(x)}) \leq \mu_{\sigma(x)}$ and so $\sigma(x) \in C$. Next, we show that $d_C(x) = f(\lambda_x) - \mu_x$, which will complete the proof of the claim. Let

$T_x = \{ y \in X: x_1^*(\sigma(x)) \leq x_1^*(y) \text{ or } x_2^*(\sigma(x)) \leq x_2^*(y) \}$

$= \{ y \in X: \mu_{\sigma(x)} + \lambda_{\sigma(x)} \leq \mu_y + \lambda_y \text{ or } \mu_{\sigma(x)} - \lambda_{\sigma(x)} \leq \mu_y - \lambda_y \}.$
Indeed, we need only do some arithmetic. Suppose that

\[ f(\lambda_{m(x)}) - f(\lambda_x) \leq \lambda_{m(x)} - \lambda_x \]

or

\[ f(\lambda_{m(x)}) - f(\lambda_x) \leq \lambda_y - \lambda_{m(x)} \]

**Case (i).** \( f(\lambda_{m(x)}) - \lambda_{m(x)} \leq f(\lambda_y) - \lambda_y \). By \((\ast)\) \( f(\lambda_{m(x)}) = \mu_{m(x)} \) and since

\[ y \in C, \ f(\lambda_y) = \mu_y. \]

Therefore, \( \mu_{m(x)} - \lambda_{m(x)} \leq \mu_y - \lambda_y \).

**Case (ii).** \( f(\lambda_{m(x)}) + \lambda_{m(x)} \leq f(\lambda_y) + \lambda_y \). As before, we have that \( f(\lambda_{m(x)}) = \mu_{m(x)} \) and \( f(\lambda_y) = \mu_y \). Therefore, \( \mu_{m(x)} + \lambda_{m(x)} \leq \mu_y + \lambda_y \). Hence \( y \in T_x \) and so

\[ C \subseteq T_x. \]

Now, it is easy to see that \( \{ y \in X : \| x - y \| < f(\lambda_x) - \mu_x \} \subseteq X \setminus T_x \subseteq X \setminus C. \)

Indeed, we need only do some arithmetic. Suppose that \( \| x - y \| < f(\lambda_x) - \mu_x \), then

\[
x^*_1(y) &= x^*_1(x) + x^*_1(y-x) \\
&< x^*_1(x) + (f(\lambda_x) - \mu_x) \quad \text{since} \quad \| x^*_1 \| = 1 \\
&= x^*_1(\sigma(x) - (f(\lambda_x) - \mu_x) x_0) + (f(\lambda_x) - \mu_x) \\
&= x^*_1(\sigma(x)) \\
x^*_2(y) &= x^*_2(x) + x^*_2(y-x) \\
&< x^*_2(x) + (f(\lambda_x) - \mu_x) \quad \text{since} \quad \| x^*_2 \| = 1 \\
&= x^*_2(\sigma(x) - (f(\lambda_x) - \mu_x) x_0) + (f(\lambda_x) - \mu_x) \\
&= x^*_2(\sigma(x)).
\]

Therefore, \( d_C(x) \geq f(\lambda_x) - \mu_x \), but \( \sigma(x) \in C \), and so \( d_C(x) = f(\lambda_x) - \mu_x \).

Hence, \( \text{Vd}_d(x) = f'(x^*_1(x)) \cdot x^*_1 - x^*_1 \) on \( X \setminus C \), where \( x^*_1 = 1/2(x^*_1 - x^*_2) \).

Now, if \( x \to \text{Vd}_d(x) \) were a minimal weak* cusco on \( X \setminus C \) then \( x \to \text{Vd}_d(x) \) would be hyperplane minimal on \( X \setminus C \), but then \( x \to f'(x^*_1(x)) \cdot x^*_1 \) would be hyperplane minimal on \( X \setminus C \). However, since \( x^*_1 \) is both continuous and open on \( X \), this would imply that \( t \to f'(t) \) is hyperplane minimal on some non-empty open subset of \( R \) (this follows from the general fact if \( \Phi : T \) is hyperplane minimal and \( T \) is both continuous and open, then \( \Phi \) is hyperplane minimal) but we know this is not true (by Example 8.2). Therefore we may conclude that \( x \to \text{Vd}_d(x) \) is not a minimal weak* cusco on \( X \). □

**Remark 6.1.** It is interesting to observe the following facts about the set

\[ C \]

constructed in Theorem 6.1:

(a) \( d_C \) is Gateaux differentiable on \( X \setminus C \);

(b) \( \sigma(x) = x + (f(\lambda_x) - \mu_x) x_0 \) is Lipschitz-continuous on \( X \setminus C \), and this means that \( C \) is almost convex, (see [47] or [24], p. 240);
(c) $\partial d_C(x) = \partial f(x_0(x) \cdot x^*_y - x^*_y$ on $X \setminus C$. In particular, $d_C$ is not integrable on $X \setminus C$. For example, let $d_*(x) \equiv h(x_0(x)) - x^*_y(x)$, where $h: R \to R$ is chosen so that $h - f$ is not a constant function on $x_0^*(X \setminus C)$ and $\partial h(t) \subseteq \partial f(t)$ for each $t \in R$. Then $\partial d_*(x) \subseteq \partial d_C(x)$ for each $x \in X \setminus C$, but $d_* - d_C$ is not a constant function on $X \setminus C$ (note that $h \equiv 0$ will do the job).

So we see then, that even in $R^2$ there are distance functions whose Clarke subdifferential mappings are not minimal, (of course there are no such examples on $R$). However, the situation is dramatically better for smooth norms. A normed linear space $X$ is said to have a uniformly Gateaux differentiable norm if for each $y \in X$, and each $\varepsilon > 0$, there exists a $\delta(\varepsilon, y) > 0$ such that for every $x \in X, \|x\| = 1$, there is a continuous linear functional $f_x$ on $X$ and

$$\frac{\|x + ty\| - \|x\| - f_x(y)}{t} < \varepsilon$$

for all $0 < t < \delta(\varepsilon, y)$.

Every Hilbert space and $L^p$ space ($1 < p < \infty$) has a uniformly Gateaux differentiable norm. Furthermore, any separable Banach space can be equivalently renormed to have a uniformly Gateaux differentiable norm [48] as can any super-reflexive Banach space.

**Proposition 6.1** [6, Theorem 8]. *If the norm $\|\cdot\|$ on a Banach space $X$ is uniformly Gateaux differentiable, then for each non-empty closed subset $C$ of $X$, $-d_C$ is regular (and hence pseudo-regular) on $X \setminus C$.*

**Corollary 6.1** [3, Theorem 5.2]. *Let $\|\cdot\|$ be a uniformly Gateaux differentiable norm on a Banach space $X$. Then for each non-empty closed subset $C$ of $X$, $d_C$ is $D$-representable on $X$.*

In finite dimensions all smooth norms are uniformly Gateaux differentiable. Therefore we may deduce the next result.

**Proposition 6.2.** *The norm $\|\cdot\|$ on a finite dimensional Banach space $X$ is smooth if, and only if, each distance function defined on $X$ possesses a minimal Clarke subdifferential mapping.*

For a smooth finite dimensional Banach space $X$ we can characterize those subsets $C$ of $X$ such that $d_C \in S^1(X)$. Indeed, since no point of $\partial C$, (the boundary of $C$) can be a point of strict differentiability (recall that in a finite dimensional Banach space the notions of strict Fréchet differentiability and strict Gateaux differentiability coincide) we immediately have a necessary condition for $d_C \in S^1(X)$, namely, $\partial C$ must be a Lebesgue-null
set. However, we have from ([6], Theorem 8) that \( -d_C \) is regular on \( X \backslash C \cup \text{int} C \). Therefore, if \( \partial C \) is a Lebesgue-null set then \( d_C \) is strictly differentiable almost everywhere in \( X \), since any locally Lipschitz function which is both Gateaux differentiable and pseudo-regular at a given point is necessarily strictly differentiable at that point. Hence we may deduce the following.

**Theorem 6.2.** Let \( \| \cdot \| \) be a smooth norm on a finite dimensional Banach space \( X \). Then for each non-empty closed subset \( C \) of \( X \), we have that \( d_C \in S_e(X) \) if, and only if, \( \partial C \) is a Lebesgue-null set.

It is natural to ask whether the characterization given in Theorem 6.2 still holds for an arbitrary separable Banach space. Unfortunately the answer to this is "no." However, we do have the following corollary.

**Corollary 6.2.** Let \( \| \cdot \| \) be a uniformly Gateaux differentiable norm on a separable Banach space \( X \). Then for each non-empty closed subset \( C \) of \( X \), \( d_C \in S_e(X) \), whenever \( \partial C \) is a Haar-null set.

Next, we show that the converse of this result does not hold.

**Example 6.1.** In Example 6.2 part (b) of [3], the author gives an example of a closed and convex subset of \( c_0(N) \), such that \( C \) is not a Haar-null set. However, as \( d_C \) is convex on \( X \) (and hence pseudo-regular on \( X \)), we must have that \( d_C \in S_e(X) \). Furthermore, from [34] we know that such sets exist in any separable non-reflexive space. Note also, that such sets necessarily have empty interior.

We say that a norm \( \| \cdot \| \) on a Banach space \( X \) is a Kadec-Klee norm if the relative norm and relative weak topologies agree sequentially on the unit sphere, \( S(X) \) (that is, if a sequence \( \{ x_n : n \in \mathbb{N} \} \subseteq S(X) \) converges to an element \( x \in S(X) \) in the weak topology, then it converges to \( x \) in the norm topology). Using this definition we can prove another important result regarding the minimality of the subdifferential mappings of a distance function.

**Theorem 6.3.** Let \( \| \cdot \| \) be a smooth Kadec-Klee norm on an Asplund space \( X \). Let \( C \) be a non-empty closed subset of \( X \) such that \( C \cap B(0, r) \) is relatively weakly compact for each \( r > 0 \), (that is, \( C \cap B(0, r)^{\text{weak}} \) is weak compact for each \( r > 0 \)). Then \( d_C \) is D-representable on \( X \). In particular, \( \partial d_C \) is generated by the strict Fréchet derivatives, and the set of points in \( X \backslash C \) which admit a closest point in \( C \) contains a dense and \( G_\delta \) subset of \( X \backslash C \).

**Proof.** By [40, Theorem 2.5] we know that \( \partial d_C = \text{CSC}(\Omega_D) \) where \( D = \{ x \in X : d_C \text{ is Fréchet differentiable at } x \} \) and \( \Omega_D : D \rightarrow 2^{X^*} \) is defined.
by, \( \Omega_D(x) \equiv \{ \text{Vd}_C(x) \} \). Hence, to show that \( \partial d_C \) is a minimal weak* cusco on \( X \), we need only show by Corollary 2.1 and Theorem 3.5 that \( \Omega_D \) is hyperplane minimal on \( D \setminus C \). To this end, we consider the following set-valued mapping \( p_C : D \setminus C \to 2^C \) defined by \( p_C(x) \equiv \{ z \in C : \|x - z\| = d_C(x) \} \). We proceed from here in two steps.

(i) Our first step is to show that \( p_C \) is a norm usco mapping on \( D \setminus C \). We recall from Proposition 1.4 in [4] that for each \( x \in D \setminus C \) we have that \( d_C(x) = \lim_{n \to \infty} \text{Vd}_C(x)(x - z_n) \) for any sequence \( \{z_n : n \in N\} \subseteq C \) such that \( \lim_{n \to \infty} \|z_n - x\| = d_C(x) \). Let us show now that for each \( x \in D \setminus C \), \( p_C(x) \) is non-empty. Let \( x_0 \in D \setminus C \) and let \( \{z_n : n \in N\} \) be any sequence in \( C \) such that \( \lim_{n \to \infty} \|z_n - x_0\| = d_C(x_0) \). Since the sequence \( \{z_n : n \in N\} \) is bounded there exists a point \( z \in X \) and a subsequence \( \{z_{n_k} : k \in N\} \) of \( \{z_n : n \in N\} \) such that weak-\( \lim_{k \to \infty} z_{n_k} = z \). Since any norm on \( X \) is lower semi-continuous, with respect to the weak topology on \( X \), we have that, \( \|z - x_0\| \leq \liminf_{k \to \infty} \|z_{n_k} - x_0\| = d_C(x_0) \). However, by above we have that

\[
\|z - x_0\| \geq \|\text{Vd}_C(x_0)(z - x_0)\| = \lim_{k \to \infty} \|\text{Vd}_C(x_0)(z_{n_k} - x_0)\| = d_C(x_0)
\]

(note that since \( d_C \) is Lipschitz-1, \( \|\text{Vd}_C(x_0)\| \leq 1 \)). Hence, \( \|z - x_0\| = d_C(x_0) \). Now, since the norm on \( X \) is Kadec-Klee and \( \lim_{n \to \infty} \|z_n - x_0\| = \|z - x_0\| \) we have that \( \{z_{n_k} : k \in N\} \) converges to \( z \) in the norm topology on \( X \). In particular, this implies that \( z \in C \). (since \( C \) is closed). Therefore, \( z \in p_C(x_0) \) and so \( p_C(x_0) \) is non-empty. Next we show that \( p_C \) is an usco mapping on \( D \setminus C \). To do this, it suffices to show that for any \( x \in D \setminus C \) and any sequences \( \{x_n : n \in N\} \subseteq D \setminus C \) and \( \{z_n : n \in N\} \subseteq C \) such that \( \{x_n : n \in N\} \) converges to \( x \) and \( z_n \in p_C(x_n) \) for each \( n \in N \), \( \{z_n : n \in N\} \) possesses a subsequence which converges to some element \( z \) of \( p_C(x) \) (in the norm topology). So let \( x \in D \setminus C \) and let \( \{x_n : n \in N\} \) be a sequence in \( D \setminus C \) which converges to \( x \). Further, let \( \{z_n : n \in N\} \) be a sequence in \( C \) such that \( z_n \in p_C(x_n) \) for each \( n \in N \). Now,

\[
d_C(x) = \liminf_{n \to \infty} \|z_n - x\| = \limsup_{n \to \infty} \|z_n - x\| \\
\leq \lim_{n \to \infty} \|z_n - x_n\| + \lim_{n \to \infty} \|x_n - x\| \\
= \lim_{n \to \infty} d_C(x_n) = d_C(x).
\]

Therefore, \( \lim_{n \to \infty} \|z_n - x\| = d_C(x) \). Now, by repeating the argument above, we have that there exists a subsequence \( \{z_{n_k} : k \in N\} \) of \( \{z_n : n \in N\} \) which converges to some point \( z \in p_C(z) \) (in the norm topology). This completes part (i) of the proof.
(ii) In this step we show that $\Omega_D$ is hyperplane minimal on $D\backslash C$. Let $x \in D\backslash C$ and let $z$ be any element of $p_C(x)$, then for each $y$

$$\nabla d_C(x)(y) = \lim_{\lambda \to 0} \frac{d_C(x + \lambda y) - d_C(x)}{\lambda}$$

$$\leq \lim_{\lambda \to 0} \frac{\|x + \lambda y - z\| - \|x - z\|}{\lambda}$$

$$= \|x - z\| (y).$$

However, $y \to \|x - z\| (y)$ is linear on $X$, therefore we must have that $\{\nabla d_C(x)\} = \partial \|x - z\|$. Furthermore, since $z$ was an arbitrary element of $p_C(x)$ we must in fact, have that $\partial \|x - p(x)\| = \{\nabla d_C(x)\}$. Therefore, $x \to \partial \|x - p_C(x)\|$ is a single-valued weak* usco on $D\backslash C$ (since it is the composition of two usco mappings) and hence hyperplane minimal on $D\backslash C$. This completes the proof.

Recall, that a set $C$ is densely proximinal if the set $D(C)$ of $X$ for which best approximations exist is dense in $X$, that is, if $x \in D(C)$ then there exists a point $p(x) \in C$ such that $d_C(x) = \|x - p(x)\|$. When $X$ is reflexive and the norm is a Kadec-Klee norm, Lau’s Theorem shows that every closed set is densely proximinal.

**COROLLARY 6.3.** A Banach space $(X, \| \cdot \|)$, is reflexive with a smooth Kadec-Klee norm if, and only if, each non-empty closed subset of $X$ is densely proximinal and the corresponding distance function possesses a minimal subdifferential mapping.

**Proof.** If $X$ is reflexive and $\| \cdot \|$ is a smooth Kadec-Klee norm, then it follows from above, that each non-empty closed subset $C$ of $X$ is densely proximinal and the corresponding distance function $d_C$ possesses a minimal subdifferential mapping. Conversely, if each non-empty closed subset of $X$ is densely proximinal then by [32], $X$ is reflexive and $\| \cdot \|$ is a Kadec-Klee norm. However, by Theorem 6.1, if each distance function possesses a minimal subdifferential mapping, then the norm $\| \cdot \|$ is smooth on $X$.

For a smooth Banach space we define a proximal normal selection $p_D$ on a subset $D$ of $D(C)$, by setting $p_D(x) = f_x - p(x)$ (where $f_x - p(x)$ is any element of $\partial \|x - p(x)\|$ for some nearest point $p(x)$), when $x \in D\backslash C$ and setting $p_D(x) = 0$ when $x \in D \cap C$. This terminology is justified, since in ([3], Lemma 5.3) the author shows that $p_D(x) \in \partial d_C(x)$ for each $x \in D(C)$.

**THEOREM 6.4.** Let $C$ be a non-empty closed subset of a smooth Banach space $X$. Suppose that $C$ is densely proximinal. Then $d_C$ is $D$-representable on $X$ if, and only if, for each proximal normal selection $p_D$, $CSC(p_D) = \partial d_C$. 

---


Codes: 3019 Sigs: 2179. Length: 45 pic 0 pts, 190 mm
Proof. This follows directly from Theorem 2.4 part (iv).

**Corollary 6.4 (Proximal Normal Formula).** Let $C$ be a non-empty closed subset of a reflexive Banach space $X$. If the norm on $X$ is a smooth Kadec-Klee norm, then for each dense subset $D$ of $D(C)$,

$$\partial d_C(x) = \overline{\lim}_{x_n \to x} \rho_D(x_n),$$

and $x_n \in D$.

It follows from Corollary 6.3 and Theorem 6.4 that we cannot weaken the hypothesis in Corollary 6.4 and still have a “Proximal Normal Formula” holding for all non-empty closed subsets of $X$.

### 7. The Relationship Between Integrability, D-representability, and Strict Differentiability

We can see from Example 8.1 part (2) that minimality of the Clarke subdifferential mapping, and so $D$-representability of the underlying function, is not enough to guarantee integrability. So we begin this section by examining the converse question, namely, does integrability imply $D$-representability? The answer to this question is a little more delicate than one might first expect. Indeed, on $R$, integrability does imply $D$-representability (see Corollary 1.3 in [5], or Proposition 8.1 part (b)), in fact on $R$, integrability implies strict differentiability, almost everywhere. However, we will show next that in general, integrability does not imply $D$-representability.

**Example 7.1.** Let $f$ be a real-valued Lipschitz function defined on $R$ such that $\partial f \equiv [0, 1]$. Let $C \equiv \text{epi}(f) = \{(x, y) \in R^2 : f(x) \leq y\}$. Next, consider the distance function $d_C$ defined on $R^2$ by the $l_1$ norm and the set $C$. Then $d_C$ is integrable on $R^2$, but not $D$-representable on $R^2$, in fact $d_C$ is not even densely strictly differentiable on $R^2$.

**Proof.** Suppose that $g$ is a real-valued locally Lipschitz function defined on $R^2$ such that $\partial g \equiv [0, 1]$. Let $C \equiv \text{epi}(f) = \{(x, y) \in R^2 : f(x) \leq y\}$. Now, $\partial d_C(x, y) = \partial f(x, y) = \partial f(x)$ for each $(x, y) \in R^2$. Therefore, $g$ is constant on $\text{int} C$, and so $\partial g(x, y) = \partial f(x, y) = \partial f(x)$ for each $(x, y) \in C$. But $\text{int} C$ is connected, therefore $g$ is constant on $\text{int} C$, and so constant on $C$, that is, $g \equiv c_1$ for some real number $c_1$. Next, we observe that $d_C(x, y) = f(x) - y$ for each $(x, y) \notin C$, (see Theorem 6.1 for a more detailed explanation). Therefore, $\partial d_C(x, y) = \partial f(x) \times \{1\} = \partial f(x) \times \{1\} = [0, 1] \times \{1\}$ on $R^2 \setminus C$. Let $x_0$ be a fixed (but arbitrary) element on $R$. We know, from above, that $g(x_0, f(x_0)) = c_1$. Therefore, by the mean-value theorem (for
differentiable functions) applied to, \( y \to g(x_0, y) \), we have that \( g(x_0, y) = c_1 + (f(x_0) - y) \) for each \( y < f(x_0) \). Hence, \( g(x, y) = (f(x) - y) + c_1 = d_c(x, y) + c_1 \) on \( R^2 \setminus C \). But from above, we have that \( g(x, y) = c_1 = d_c(x, y) + c_1 \) on \( C \). Therefore, \( g = d_c + c_1 \) on \( R^2 \).

Remark 7.1. It is very important to observe that \( d_c \) is not integrable on \( R^2 \setminus C \). Indeed, let \( f_1(x) \equiv x - f(x) \), then \( \partial f_1 = \partial f \) on \( R \), and so \( \partial g(x, y) = \partial f(x) \times \{ -1 \} \) on \( R^2 \setminus C \), where \( g(x, y) \equiv f_1(x) - y \) on \( R^2 \setminus C \). So we see then that in general, integrability is not hereditary with respect to open subsets. This is a striking contrast with the situation for \( D \)-representability.

The previous example leads us to consider a stronger notion of integrability. We will say that a real-valued locally Lipschitz function \( f \), defined on a non-empty open subset \( A \) of a Banach space \( X \) is hereditarily integrable on \( A \) if, for each non-empty open subset \( U \) of \( A \) the function \( f \mid U \) is integrable on \( U \). It is immediate, that if \( f \) is hereditarily integrable on \( A \) then it is integrable on \( A \), however, the previous example shows, that the converse of this is false, even when \( A \) is connected. We should also note then, that if \( f \in S(f) \) then \( f \) is not only integrable on \( A \), but also hereditarily integrable on \( A \). Let us also observe, that since integrability is not a hereditary property (with respect to open sets) one cannot expect to characterize this property in terms of a local differentiability property, (as was done for \( D \)-representability), but rather, one must expect, such a characterization, to be in terms of some global differentiability property. We give next, a sufficient condition for a Lipschitz function to be integrable. It is not worth, that this condition is, as we mentioned above, expressed in terms of a global property. Recall that a subset \( A \) of a topological space \( X \) is locally connected if for each \( x \in X \) and each open neighbourhood \( U \) of \( x \), there exists an open subset \( V \) of \( U \), which contains \( x \), such that \( V \cap A \) is connected.

**Theorem 7.1.** Let \( f \) be a real-valued locally Lipschitz function defined on a non-empty open subset \( A \) of a Banach space \( X \) and let \( U \) be any open subset of \( A \). Then, (a) if \( f \mid U \) is integrable on \( U \) and the connected components of \( U \) are locally finite, then \( f \) is integrable on \( \text{int}(U \cap A) \), and (b) if \( f \mid U \) is hereditarily integrable on \( U \) and \( U \) is a finite union of locally connected open subsets, then \( f \) is hereditarily integrable on \( \text{int}(U \cap A) \).

**Proof.** (a) Suppose that \( g \) is a real-valued locally Lipschitz function defined on \( B = \text{int}(U \cap A) \) such that \( \partial g(x) \equiv \partial(f \mid B)(x) \) for each \( x \in B \). Let \( \{ U_\gamma : \gamma \in \Gamma \} \) denote the connected components of \( U \). It is easy to see that the family of sets \( \{ U_\gamma : \gamma \in \Gamma \} \) is also locally finite (and \( U = \bigcup \{ U_\gamma : \gamma \in \Gamma \} \)).
Let $x_0$ be an arbitrary point of $B$. We need to show that \( \partial(f - g)(x_0) = \{0\} \). We may choose open neighbourhoods $V$ of $x_0$ (contained in $B$) so that $\{ \gamma \in \Gamma : \bar{U}_\gamma \cap V \neq \emptyset \} = \{ \gamma : 1 \leq j \leq n \}$ and $x_0 \in \bar{U}_\gamma$ for each $1 \leq j \leq n$. Now since $f|_U$ is integrable on $U$, we must have that for each $1 \leq j \leq n$, there exists a $c_j \in R$ such that $g(x) = f(x) + c_j$ for all $x \in U_\gamma$. It follows now from the continuity of $f$ and $g$ that $g(x) = f(x) + c_j$ for each $x \in U_\gamma \cap B$. In particular, we must have that $g(x_0) = f(x_0) + c_j$ for each $1 \leq j \leq n$. Hence, $g(x) = f(x) + c_1$ on $\bar{U} \cap V$ (and so on $B \cap V$). This shows that $\partial(f - g)(x_0) = \{0\}$.

(b) Let $B'$ be an arbitrary non-empty open subset of $B \equiv \text{int } \bar{U} \cap A$ and suppose that $g$ is a real-valued locally Lipschitz function defined on $B'$ such that $\partial g(x) \equiv \partial f|_{B'}(x)$ for each $x \in B'$. Let $x_0$ be an arbitrary point in $B'$. We need to show that $\partial(f - g)(x_0) = \{0\}$. Let $\{U_j : 1 \leq j \leq n\}$ denote the locally connected open subset of $U$ (note that $U = \bigcup \{U_j : 1 \leq j \leq n\}$). We may choose open neighbourhoods $V$ of $x_0$ (contained in $B'$) so that $\{ j : U_j \cap V \neq \emptyset \} = \{ j : x_0 \in U_j \} = A$. Now, for each $j \in A$ there exists an open neighbourhood $V_j$ of $x_0$ (contained in $V$) such that $V_j \cap U_j$ is connected. Therefore, since $f|_U$ is hereditarily integrable on $U$ we must have that for each $j \in A$, there exists a $c_j \in R$ such that $g(x) = f(x) + c_j$ for all $x \in U_j \cap V_j$. It follows (as above) from the continuity of $f$ and $g$ that $f(x) = g(x) + c_j$ for each $x \in U_j \cap V_j \subseteq U_j \cap V_j$. In particular, we have that $g(x_0) = f(x_0) + c_j$ for each $j \in A$. Hence, $f - g \equiv \text{constant on } \cap \{ V_j : j \in A \} \cap \bar{U}$ (and so on $\cap \{ V_j : j \in A \} \cap B'$). This shows that $\partial(f - g)(x_0) = \{0\}$.

These apparently harmless observations provide us with a technique for constructing integrable functions that are not almost everywhere strictly differentiable.

**Corollary 7.1.** Let $\| \cdot \|$ be a uniformly Gateaux differentiable norm on a separable Banach space $X$. Let $C$ be a non-empty closed subset of $X$. Then,

(a) the distance function $d_C$, associated with the set $C$, is integrable if, the connected components of both $\text{int } C$ and $X \setminus C$ are locally finite and (b) the distance function $d_C$, associated with the set $C$, is hereditarily integrable if $\text{int } C$ and $X \setminus C$ are both locally connected subsets of $X$.

**Proof.** This follows from Theorem 7.1 and the fact that $d_C$ is hereditarily integrable on $X \setminus C \cup \text{int } C$.

We may conclude then, that even for distance functions, with respect to uniformly smooth norms, it is possible to be both integrable and $D$-representable, while still not being a member of $S_d(X)$. Indeed, with a little more work, we can show an even stronger result:
Example 7.2. There exists a compact nowhere dense subset $C$ of $\mathbb{R}^2$ such that (i) $d_C$ is $D$-representable; (ii) $d_C$ is hereditarily integrable; (iii) $d_C$ is not strictly differentiable almost everywhere in $\mathbb{R}^2$; that is, $d_C \notin S_a(\mathbb{R}^2)$. (Actually, there are many such examples.)

Proof. Let $C_1$ be a Cantor subset of $[0, 1]$ with $\mu(C_1) > 0$. Let $C = C_1 \times C_1 \subseteq \mathbb{R}^2$. Let $d_C$ be the distance function generated by the set $C$, with the Euclidean norm. Then by Proposition 6.2, $d_C$ is $D$-representable on $\mathbb{R}^2$. To justify that $d_C$ is hereditarily integrable it suffices by Corollary 7.1 part (b) to show that $X \setminus C$ is locally connected. So let $x \in X$ and $U$ be an open neighbourhood of $x$. It is easy to see that the only non-trivial case is when $x \in \partial C$. So let us assume that $x \in \partial C$. We may now choose an $r > 0$ such that $B_{\infty}(x, r) \subseteq U$, where $B_{\infty}(x, r)$ is the $l_\infty$ ball around $x$, of radius $r$. It now only remains to observe that $B_{\infty}(x, r) \cap (X \setminus C)$ is a connected subset (in fact, it is polygonally connected). Now, to see that $d_C \notin S_a(\mathbb{R}^2)$ we need only use the standard fact that $d_C$ cannot be strictly differentiable at any point of $\partial C = C$.

Let $A$ be a non-empty open subset of a Banach space $X$. Let $\mathcal{F}(A)$ denote the family of all real-valued integrable, locally Lipschitz functions defined on $A$ and let $\mathcal{M}(A)$ denote the family of all real-valued locally Lipschitz functions defined on $A$ whose Clarke subdifferential mappings are minimal. It follows then, that $\mathcal{N}(A) \equiv \mathcal{F}(A) \cap \mathcal{M}(A)$ is the largest class of functions that satisfy both the conditions $(P_1)$ and $(P_2)$ given at the start of this paper. So why then, have we not considered the class of functions $\mathcal{N}(A)$? A partial answer to this is revealed in the next example.

Example 7.3. $\mathcal{N}(\mathbb{R}^2)$ is not closed under addition, multiplication nor either of the lattice operations. (Note that we also show that $\mathcal{F}(\mathbb{R}^2)$ is not closed under addition, multiplication, nor either of the lattice operations).

Proof. (a) We show first that $\mathcal{N}(\mathbb{R}^2)$ is not closed under addition. Let $f$ be a non-integrable real-valued Lipschitz-1 function defined on $\mathbb{R}$ such that, $x \mapsto f(x)$, is a minimal cusco. (Note that such functions exist, see Example 8.1.) Let $K_1 = \{(x, y) : f(x) \leq y\}$ and $K_2 = \{(x, y) : y \leq f(x)\}$. Next consider the distance functions $d_{K_1}$ and $d_{K_2}$ defined on $\mathbb{R}^2$ by the $l_1$ norm and the sets $K_1$ and $K_2$, respectively. Then,

\[ d_{K_1}(x, y) = \begin{cases} f(x) - y & \text{if } (x, y) \notin K_1 \\ 0 & \text{if } (x, y) \in K_1 \end{cases} \]

and

\[ d_{K_2}(x, y) = \begin{cases} y - f(x) & \text{if } (x, y) \notin K_2 \\ 0 & \text{if } (x, y) \in K_2 \end{cases} \]
It follows from our earlier work that both \(d_{K_i}\) and \(-d_{K_i}\) are integrable on \(R^2\) and \(D\)-representable on \(R^2\). However, \(d = d_{K_i} + (-d_{K_i})\) is not integrable on \(R^2\). In fact, \(d(x, y) = f(x) - y\) on \(R^2\) and so \(\partial d(x, y) = \partial f(x) \times \{-1\}\) on \(R^2\). Hence for any real-valued function \(g\) defined on \(R\) such that \(\partial g = \partial f\) and \(g - f\) is not a constant function on \(R\), the function \(G(x, y) = g(x) - y\), shares that same Clarke subdifferential mapping as \(d\) (while not differing from \(d\) by a constant). Therefore, \(d\) is not integrable on \(R^2\).

(b) Next, we show that \(\mathcal{A}(R^2)\) is not closed under multiplication. Let \(d_{K_i}^* \equiv d_{K_i} + 1\) and \(d_{K_i}^* \equiv (-d_{K_i}) + 1\). Then \(d^* \equiv d_{K_i}^* \cdot d_{K_i}^*\) is not integrable on \(R^2\). To see this, we compute \(d^*(x, y) = (f(x) - y) + 1\) on \(R^2\). Then as in (a) we see that \(d^*\) is not integrable on \(R^2\).

(c) Finally, we show that \(\mathcal{A}(R^2)\) is not closed under the lattice operations. Let \(C\) be a Cantor subset of \([0, 1]\) with \(\mu(C) > 0\). We define two sets; \(C_1 \equiv \{ (x, y) \in R^2 : x \in C \text{ and } y \leq 0 \}\) and \(C_2 \equiv \{ (x, y) \in R^2 : x \in C \text{ and } y \geq 0 \}\). Now, consider the distance functions \(d_{C_1}\) and \(d_{C_2}\) defined on \(R^2\) by the Euclidean norm and the sets \(C_1\) and \(C_2\) respectively. Then \(d_{C_1}(x, y) = \min\{d_{C_1}(x, y), d_{C_2}(x, y)\}\) is the distance function to the set \(C^* \equiv \{ (x, y) \in R^2 : x \in C \text{ and } y \in R \}\). Moreover, it is easy to see that \(d_{C_1}(x, y) = d(x)\), where \(d: R \to R\) is defined by, \(d(x) = \min\{ |x - c| : c \in C \}\). However, \(d\) is not integrable on \(R\) since \(d\) is not strictly differentiable almost everywhere on \(R\), in particular, \(d\) is not strictly differentiable at any point of \(C\), (see, Proposition 8.1 part (b)). Therefore there exists a Lipschitz function \(g\) such that \(\partial g = \partial d\) and \(g - d\) is not a constant function on \(R\). Let \(G(x, y) = g(x)\). Clearly then, \(\partial G = \partial d_{C_1}\), but \(G - d_{C_2}\) is not a constant function on \(R^2\). To show that \(\mathcal{A}(R^2)\) is not closed under “max” we need only consider \(-d_{C_2}\).

Another reason why we have not considered the class \(\mathcal{A}\) is that thus far, we have not been able to deduce a reasonable characterization for membership in this class of functions.

Despite the previous examples there is an important inter-play between integrability and minimality of the Clarke subdifferential mapping.

**Theorem 7.2 (Identity Theorem).** Suppose that \(f\) and \(g\) are real-valued locally Lipschitz functions defined on a non-empty connected open subset \(A\) of a Banach space \(X\). If \(f \in \mathcal{I}(A)\) and \(g \in \mathcal{M}(A)\), then \(f - g \equiv \text{constant on } A\) if, and only if, \(\{ x \in A : \partial g(x) \cap \partial f(x) \neq \emptyset \}\) is dense in \(A\).

**Proof.** Suppose that \(\{ x \in A : \partial g(x) \cap \partial f(x) \neq \emptyset \}\) is dense in \(A\). Consider the set-valued mapping \(T: A \to 2^{X^*}\) defined by \(T(x) \equiv \partial g(x) \cap \partial f(x)\). Since both \(x \to \partial f(x)\) and \(x \to \partial g(x)\) are upper semi-continuous on \(A\) (and possess compact images), \(T\) possesses non-empty weak* compact, convex images. Moreover, since the graphs of both \(\partial f\) and \(\partial g\) are closed in \(A \times X^*\), with \(X^*\) equipped with the weak* topology, so is the graph of \(T\).
Therefore, by Proposition 1.3, \( T \) is a cusco on \( A \). But, for each \( x \in A \), \( T(x) \subseteq \partial f(x) \) and \( T(x) \subseteq \partial g(x) \). Hence, by the minimality of \( \partial g \) we must have that \( \partial g = T \) (that is, \( \partial g(x) \subseteq \partial f(x) \) for each \( x \in A \)). The result now follows from the fact that \( f \) is integrable. The converse is obvious.

We end this section by giving a comment concerning integrability with respect to the approximate subdifferential mapping. It is possible to construct two Lipschitz functions \( f \) and \( g \) mapping from \( R^2 \) into \( R \) such that \( \partial f = \partial g \) is minimal, while \( \partial f \) and \( \partial g \) differ on a set of positive measure. This cannot happen on the real-line, where \( \partial f \) determines \( \partial g \) [5] (here \( \partial g \) denotes the approximate subgradient of \( f \)). On the other hand we should observe that our conditions for integrability imply integrability with respect to any subdifferential mapping, \( x \mapsto \partial f_u(x) \), which has the property that \( \overline{\text{co}}^{-1} \partial f_u(x) = \partial f(x) \) for each \( x \). Let us also comment that in general \( x \mapsto \partial_g f(x) \) is a weak* usco, however, it is very rarely a minimal usco. Indeed, even the approximate subgradient of the absolute value function fails to be a minimal usco.

8. EXAMPLES AND MISCELLANEOUS RESULTS

Let us begin this section by showing that the family of all \( D \)-representable functions is not closed under addition, multiplication, nor either of the lattice operations.

**Example 8.1.** Let \( C \) be a Cantor subset of \([0, 1]\) (symmetric about \( \frac{1}{2} \)) with \( 1 > \mu(C) > 0 \), and let \( \{(a_n, b_n) : n \in \mathbb{N}\} \) be an enumeration of the disjoint open intervals of \([0, 1]\) \( \setminus C \). Further, for each \( n \in \mathbb{N} \), let \( c_n \equiv (a_n + b_n)/2 \) and \( d_n \equiv (b_n - a_n)/2 \). Now, consider the Lipschitz functions \( f : (0, 1) \to [0, 1] \) and \( g : (0, 1) \to [-1, 1] \) defined by

\[
g(x) = \begin{cases} 0 & \text{if } |x - c_n| \geq d_n \text{ for all } n, \\ 2(x - (c_n - d_n)) & \text{if } x \in (c_n - d_n, c_n - \frac{1}{2} d_n], \\ -(x - c_n) & \text{if } x \in (c_n - \frac{1}{2} d_n, c_n], \\ -2(x - c_n) & \text{if } x \in (c_n, c_n + \frac{1}{2} d_n], \\ x - (c_n + d_n) & \text{if } x \in (c_n + \frac{1}{2} d_n, c_n + d_n) 
\end{cases}
\]

and

\[
f(x) \equiv \int_{[0, x]} \vartheta(t) \, dt
\]
where

$$
\theta(t) = \begin{cases} 
0 & \text{if } t \in [0, 1] \setminus C \\
2 & \text{if } t \in C \cap [0, 1/2] \\
-2 & \text{if } t \in C \cap (1/2, 1].
\end{cases}
$$

(1) $g, f + g$, and $f - g$ possess minimal subdifferential mappings on $(0, 1)$.

(2) $\partial g = \partial (f + g)$, but $(f + g) - g$ is not a constant function on $(0, 1)$, that is, we cannot determine $g$, up to an additive constant, from its Clarke subdifferential mapping.

(3) If $h = f + g$ and $k = f - g$, then $h + k$ does not possess a minimal subdifferential mapping.

(4) If $M$ and $m$ are defined by $M(x) \equiv \max \{k(x), h(x)\}$ and $m(x) \equiv \min \{k(x), h(x)\}$, then neither $M$ nor $m$ possesses a minimal subdifferential mapping on $(0, 1)$.

(5) If $f : [0, 1] \to [0, 2]$ is defined by $j(x) = f(x) + 1$, then the functions $j + g$ and $j - g$ possess minimal subdifferential mappings on $(0, 1)$, but $(j + g) - (j - g)$ does not.

**Proof.** Let us first observe, that by direct calculation one can show that $g'(x) = 0$ for each $x \in C$. (1) It is easy to see that, $x \mapsto \partial g(x)$, is a minimal cusco on $(0, 1) \setminus C$. Indeed, $g$ is strictly differentiable on $(0, 1) \setminus C$ except for countably many points. It is also easy to see that $\partial g(x) = [-2, 2]$ at each point of $C$. In fact, $\partial g = \operatorname{CSC}((\partial g)_y)$. Therefore, we may conclude, from Corollary 2.1 that $\partial g$ is a minimal cusco on $(0, 1)$. Next, we observe that $(f + g)'(x) = f'(x) + g'(x) \in \partial g(x)$ almost everywhere in $(0, 1)$. Therefore, by Theorem 2.5 in [17], $\partial (f + g) = \partial g(x)$ for each $x \in (0, 1)$. However, since $x \mapsto \partial g(x)$, is a minimal cusco on $(0, 1)$ we must have that $\partial (f + g) = \partial g$. A similar argument shows that $\partial (g - f) = \partial g$ on $(0, 1)$. From this we can deduce that $\partial (f - g)$ is a minimal cusco by observing that $\partial (f - g)(x) = (1 - 1) \cdot \partial (f - g)f(x)$ at each $x \in (0, 1)$. (2) This follows immediately from part (1). (3) From part (1) we see that both $h$ and $k$ possess minimal subdifferential mappings, but $h + k = 2f$, which clearly does not possess a minimal subdifferential mapping, Since in particular, $0 \in \partial f(x)$ for each $x \in (0, 1)$, while $\partial f$ is not identically equal to $\{0\}$. (4) $M(x) \equiv \max \{k(x), h(x)\} = f(x) + |g(x)|$ and $m(x) \equiv \min \{k(x), h(x)\} = f(x) - |g(x)|$. Moreover,
Therefore, $M^+(x) \geq -1$ at each point of $(0,1) \setminus C$. However, there exists a set of positive measure $A \subseteq C \cap (1/2,1)$ such that $M^+(x) = f'(x) + |g'| (x) = f'(x) - 2$ at each point of $A$. Hence, $\partial g$ is not generated by the derivatives chosen from $(0,1) \setminus C$ (which are dense in $[0,1]$), that is, $g$ is not $D$-representable on $(0,1)$, and so $g$ does not possess a minimal subdifferential mapping on $(0,1)$. A similar argument shows that $m$ does not possess a minimal subdifferential mapping. (5) Clearly, both $j + g$ and $j - g$ possess minimal subdifferential mappings. In fact, $\partial (j + g) = \partial (f + g)$ and $\partial (j - g) = \partial (f - g)$. However, $(j + g) - (j - g) = j^2 - g^2$ which does not possess a minimal subdifferential mapping, because $(j^2 - g^2)' (x) = 2(j(x) j'(x) - g(x) g'(x))$ almost everywhere in $(0,1)$ and so $(j^2 - g^2)' (x) \leq 2$ almost everywhere in $(0,1) \setminus C$ (note: $-1 \leq g(x) \leq 1$ on $(0,1)$). However, there exists a set of positive measure $A \subseteq C \cap (0,1/2)$ such that $(j^2 - g^2)' (x) = 2(j(x) j'(x) + g(x) g'(x)) = 2j(x) j'(x) > 4$ at each point of $A$. Hence, as in (4), it follows that $(j + g) - (j - g)$ is not $D$-representable, and so $(j + g) - (j - g)$ does not possess a minimal subdifferential mapping.

Next, we gather-up a few special facts concerning locally Lipschitz functions defined on $R$.

**Proposition 8.1.** Let $I$ be a non-empty open interval of $R$, then:

(a) Each minimal cusco $\Phi: I \to 2^R$ is the Clarke subdifferential mapping of some $D$-representable locally Lipschitz function defined on $I$. (Note: not every cusco from $I$ into $2^R$ is the Clarke subdifferential mapping of some locally Lipschitz function defined on $I$.)

(b) A locally Lipschitz function $f: I \to R$ is integrable on $I$ if, and only if, $f \in S(I)$.

(c) $S(I)$ is closed under addition, subtraction, multiplication and division (when this is defined) as well as, the lattice operations. Moreover, $S(I)$ is closed under composition (when this is defined).

(d) $f \in S(I)$ if, and only if, the mapping $x \mapsto f^+(x;1)$ is Riemann integrable on $I$.

**Proof.** (a) Define $h: I \to R$ by, $h(x) \equiv \max \{ \Phi(x) \}$. Then $h$ is upper semi-continuous on $I$ and hence Borel measurable on $I$. Define $f: I \to R$ by, $f(x) \equiv \int_{(a,x)} h(t) \, dt$ (here $a$ is any element of $I$), then $f'(x) = h(x)$ almost everywhere in $I$, (let us call this set $D$). Then, $\partial f(x) = \text{CSC}(f^+, |D|)(x) \subseteq \Phi(x)$ for each $x \in I$. But since $\Phi$ is a minimal cusco, we may deduce that $\partial f = \Phi$. (b) This is Corollary 1.3 in [5]. (c) This is just Corollary 4.4 and Theorem 4.4. (d) A famous theorem due to Lebesgue says that a real-valued function defined on an interval of $R$ is Riemann integrable if, and
only if, the function is continuous almost everywhere. When we combine this, with the fact that \( f^+ \) is continuous at a point \( x \in I \) if, and only if, \( \partial f \) is single-valued at \( x \), we obtain the desired result.

Ironically, \( D \)-representable Lipschitz functions are also useful in constructing highly pathological Lipschitz functions.

**Theorem 8.1** [12, Theorem 1]. Let \( f_1, f_2, \ldots, f_n \) be real-valued locally Lipschitz functions defined on a non-empty open subset \( A \) of a separable Banach space \( X \). If each function \( f_j \) possesses a minimal Clarke subdifferential mapping on \( A \), then there exists a real-valued locally Lipschitz function \( g \) defined on \( A \) such that \( \partial g(x) = \text{co} \{ \partial f_1(x), \partial f_2(x), \ldots, \partial f_n(x) \} \) for each \( x \in A \).

**Remark 8.1.** It was noted in [12] that the function \( g \), given above, is not integrable, except perhaps when, \( \partial f_1 = \partial f_2 = \cdots = \partial f_n \).

We have seen so far in this paper that those Lipschitz functions which are \( D \)-representable possess very desirable differentiability properties. Hence, the following result due to Preiss [40] is very surprising. In [40] the author show that there is a Lebesgue null set, \( G_0 \) which is also a \( G^\delta \) subset of \( \mathbb{R}^n \), such that

\[
\text{CSC}(\partial f|_{G_0 \cap D_f}) = \partial f = \text{CSC}(\partial f|_{D_f \setminus G_0})
\]

for all locally Lipschitz mappings \( f: \mathbb{R}^n \to \mathbb{R} \). Thus, paradoxically, both sets \( G_0 \) and \( D_f \setminus G_0 \) reconstruct any Lipschitz function, where \( D_f = \{ x \in \mathbb{R}^n: \nabla f(x) \text{ exists} \} \).

Next, we give a slight improvement of Theorem 2.6 and Corollary 3.1.

**Proposition 8.2.** Let \( f \) and \( g \) be real-valued locally Lipschitz functions defined on a non-empty open subset \( A \) of a separable Banach space \( X \). If, \( x \to \partial f(x) \), is a minimal weak* cusco on \( A \) and \( g \in S_\infty(A) \) then:

\[(i) \quad x \mapsto \partial(f + g)(x) \text{ is a minimal weak* cusco on } A;\]
\[(ii) \quad x \mapsto \partial(f \cdot g)(x) \text{ is a minimal weak* cusco on } A;\]
\[(iii) \quad x \mapsto \max\{ f(x), g(x) \} \text{ and } x \mapsto \min\{ f(x), g(x) \} \text{ possess minimal subdifferential mappings on } A.\]

**Proof.** (i) It follows from Theorem 7.5 in [15] that \( \nabla(f + g)(x) = \nabla f(x) + \nabla g(x) \) almost everywhere in \( A \) (call this set \( D \)). Let \( S \) denote the set of all points of strict differentiability of \( g \) in \( A \). It now follows, in a similar manner to Theorem 2.3 part (ii), that \( x \mapsto \nabla(f + g)(x) \) is hyperplane minimal on \( S \cap D \). The result can now be deduced from Corollary 4.1. (ii) Again by Theorem 7.5 in [15] we have that \( \nabla(f \cdot g)(x) = \nabla f(x) g(x) + f(x) \nabla g(x) \) almost everywhere in \( A \) (call this set \( D \)). As before, let \( S \) denote
the set of all points of strict differentiability of $g$ in $A$. Then by Theorem 2.3 part (i) and (ii) we have that, $x \mapsto V(f \cdot g)$, is hyperplane minimal on $S \cap D$. The result may now be obtained from Corollary 4.1. (iii) \[ \max \{ f(x), g(x) \} = (f - g)^+ (x) + g(x) \] and \[ \min \{ f(x), g(x) \} = (f - g)^- (x) + g(x) . \] Now by part (i) $f - g$ possesses a minimal subdifferential mapping. Therefore, by Corollary 3.1 part (i) $(f - g)^+$ and $(f - g)^-$ possess minimal subdifferential mappings. The proof is completed by again appealing to part (i) above.

Remark 8.2. It follows from this, that in Example 8.1 none of $g$, $h$, $k$, $M$ nor $m$ could belong to $S_d((0,1))$. Of course this is easily checked directly.

It is important to realize that everywhere differentiability of a Lipschitz function is not sufficient to imply either $D$-representability (see, example below) or integrability. In fact, in [13] the authors give an example of two distinct differentiable Lipschitz functions which share the same (approximate subgradient) Clarke subgradient at all points.

Example 8.2. On p. 216, Example 6 part (e) of [45] an example is given of an everywhere differentiable Lipschitz function $f: \mathbb{R} \to \mathbb{R}$ which is strictly increasing on $\mathbb{R}$ and for which the set \{ $t \in \mathbb{R}$: $f(t) = 0$ \} is dense in $\mathbb{R}$. We claim that the Clarke subdifferential mapping of $f$ is not a minimal cusco on $\mathbb{R}$.

Proof. To see this, consider the mapping $\sigma: \mathbb{R} \to 2^\mathbb{R}$ defined by $\sigma(t) \equiv \{ 0 \}$. Since the mapping, $t \mapsto \partial f(t)$, is upper semi-continuous on $\mathbb{R}$ we must have that $\sigma(t) \in \partial f(t)$ for each $t \in \mathbb{R}$. However, $\sigma$ is a cusco on $\mathbb{R}$, and so if, $t \mapsto \partial f(t)$, were minimal on $\mathbb{R}$ then $\partial f = \sigma$. But this is not possible, because if $\partial f \equiv \{ 0 \}$, then by the mean-value theorem (for differentiable functions) $f$ would be constant on $\mathbb{R}$, which it is not.

The next result can be considered to be an “abstract invariance” result.

Theorem 8.2 (Abstract invariance). Let $A$ be a topological space and let $T$ be a minimal usco (minimal cusco) from $A$ into subsets of a Hausdorff topological space (separated linear topological space) $X$. Let $T$ be a set-valued mapping from $A$ into (convex) subsets of $X$. If the graph of $T$ is closed in $A \times X$, then the following conditions are equivalent:

(i) \{ $t \in A$: $\Omega(t) \cap T(t) \neq \emptyset$ \} is dense in $A$;
(ii) $\Omega(t) \cap T(t) \neq \emptyset$ for each $t \in A$;
(iii) $\Omega(t) \subseteq T(t)$ for each $t \in A$.
Proof. It follows from a standard compactness argument, that (i) \(\Rightarrow\) (ii). Let \(\Omega^*(t) = \Omega(t) \cap T(t)\) for each \(t \in T\). Then \(\Omega^*\) possesses a closed graph in \(A \times X\). Moreover, \(\Omega^*(t) \subseteq \Omega(t)\) for each \(t \in A\). Therefore, by Proposition 1.3 we have that \(\Omega^*\) is an usco (a cusco) mapping on \(A\). However, \(\Phi\) is a minimal usco (minimal cusco) on \(A\), therefore, \(\Omega(t) \subseteq T(t)\) for each \(t \in A\), which shows that (ii) \(\Rightarrow\) (iii). Finally, it is obvious that (iii) implies (i).

This last result has interesting more concrete applications to fixed point theory and to differential inclusions. Now standard assumptions force \(\Omega\) to be a cusco while \(T\) needs to be a closed tangent cone set-valued mapping. This is often not the case (for example, \(C\) might be the orthant which has problems at the origin) unless \(C\) is a reasonable manifold. However, if \(T \equiv N_C\) is the Clarke normal cone and \(C\) is epi-Lipschitz, as in [17], or convex then \(T\) is closed.) Then (ii) is a weak inwardness condition ensuring the existence of a solution in \(C\) to \(0 \in \Omega(x)\). Also (ii) is a standard hypothesis for weak invariance and ensures under appropriate conditions that the differential inclusion \(x'(t) \in \Omega(x(t)), x(0) \in C\) has a viable solution: remaining (locally) in \(C\) (see [1]). Correspondingly, (iii) is a strong invariance condition used to show that every solution remains in \(C\) [1]. Thus for a minimal cusco and closed tangent cone we need only check (i) to determine (iii). Moreover, weak and strong invariance are then effectively co-determinate.

We end the paper by examining the minimality of vector-valued functions. Let \(f: \mathbb{R}^n \to \mathbb{R}^k\) be a locally Lipschitz function defined on a non-empty open subset \(A \subseteq \mathbb{R}^n\), defined by, \(f(x) = (f_1(x), f_2(x), ..., f_k(x))\). Then the (Clarke) generalized Jacobian of \(f\) (denoted \(\partial f\)) is defined by \(\partial f(x) = \text{CSC}(\text{D}(f))\), where \(\text{D}(f)\) is the classical Jacobian of \(f\) defined on \(x \in A: \text{D}(f)\) exists.

**Proposition 8.3.** Let \(f: \mathbb{R}^n \to \mathbb{R}^k\) be a locally Lipschitz function defined on a non-empty open set \(A \subseteq \mathbb{R}^n\), defined by \(f(x) = (f_1(x), f_2(x), ..., f_k(x))\). If \(f_j \in \text{S}_j(A)\) for each \(j \in \{1, 2, ..., k\}\), then the (Clarke) generalized Jacobian, \(x \to \partial f(x)\), is a minimal cusco on \(A\).

**Proof.** We may, without loss of generality, assume that \(A \equiv B(x, r)\), for some \(x \in X\), and \(r > 0\) (this is because minimality is locally determined, see Theorem 2.1). Let \(S \equiv \{x \in A: \text{each } f_j, 1 \leq j \leq k, \text{is strictly differentiable at } x\}\). By the Theorem in [21] we have that \(\partial f = \text{CSC}(\text{D}(f))\), where \(\text{D}(f)\) is the classical Jacobian of \(f\) restricted to \(S\). It is easy to see that \(x \to \text{D}(f)(x)\) is continuous (and hence hyperplane minimal) on \(S\), and so the result follows from Corollary 2.1.

In the same fashion we may define \(\partial f = \text{CSC}(\text{D}(f))\) whenever \(f: E \to F, E \text{ separable and } F \text{ RNP}\). In this case it is not known whether \(\partial f\)
is insensitive to Haar-null sets. Nonetheless, if $f: E \to F$ is $C^{1,1}$ while $E^*$ is separable one may define a \textit{generalized Hessian} $\tilde{\nabla}^2 f = \text{CSC}(\nabla^2 f)$ and study it accordingly. As with first-order derivatives $\partial f$ is a cusco, and hence maybe manipulated in a similar manner.

**ACKNOWLEDGMENTS**

The authors thank John R. Giles for suggesting the refinement of Theorem 3.2, given in Theorem 3.3. We also thank the anonymous referees for their many valuable suggestions which improved the paper in many places.

**REFERENCES**


42. L. Qi, The maximal normal operator space and integration of subdifferentials of non-
43. R. T. Rockafellar, Favourable classes of Lipschitz-continuous functions in subgradient
optimization, in “Progress in Nondifferentiable Optimization” (E. Nurminski, Ed.),
Systems Analysis, Laxenberg, Austria, 1982.
44. S. D. Sciffer, “Differentiablility Properties of Locally Lipschitz Functions on Banach
Mathematics Series, 1981.
46. L. Thibault, On generalized differentials and subdifferentials of Lipschitz vector-valued
48. V. Zizler, Some notes on various rotundity and smoothness properties of separable