A chain rule for essentially strictly differentiable Lipschitz functions

Jonathan M. Borwein*
CECM, Department of Mathematics and Statistics,
Simon Fraser University,
Burnaby, BC V5A 1S6, Canada jborwein@cecm.sfu.ca

Warren B. Moors†
Department of Mathematics,
The University of Auckland,
Private Bag 92019, Auckland, New Zealand
moors@math.auckland.ac.nz

January 16, 1996

Abstract
In this paper we introduce a new class of real-valued locally Lipschitz functions, (that are similar in nature and definition to Valadier’s same functions) which we call arc-wise essentially smooth, and we show that if \( g : \mathbb{R}^n \rightarrow \mathbb{R} \) is arc-wise essentially smooth on \( \mathbb{R}^n \) and each function \( f_j : \mathbb{R}^m \rightarrow \mathbb{R}, 1 \leq j \leq n \) is strictly differentiable almost everywhere in \( \mathbb{R}^m \), then \( g \circ f \) is strictly differentiable almost everywhere in \( \mathbb{R}^m \), where \( f \equiv (f_1, f_2, \ldots, f_n) \). We also show that all the semi-smooth and pseudo-regular functions are arc-wise essentially smooth. Thus, we provide a large and robust lattice algebra of Lipschitz functions whose generalized derivatives are well-behaved.

Keywords: Lipschitz functions, Chain rule, Haar-null sets, Differentiability, Essentially strictly differentiable AMS (1991) subject classification: Primary: 49J520; 46N10 Secondary: 58C20.

*Research supported by NSERC and the Shrum Endowment at Simon Fraser University.
†Research supported by a New Zealand Science and Technology Post Doctoral Fellowship
1 Introduction.

In this paper we show that those Lipschitz functions which are strictly differentiable almost everywhere, possess extremely strong closure properties. Essentially strictly differentiable functions were studied in detail in [3]. In particular, it was shown therein that all such functions possess very well-behaved Clarke generalized gradients.

We begin by recalling some preliminary definitions regarding the Clarke subdifferential mapping, [6]. 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

$$|f(x) - f(y)| \leq K||x - y|| \quad \text{for all} \quad x, y \in B(x_0, \delta)$$

For functions in this class, it is often instructive to consider the following three directional derivatives:

(i) The upper Dini derivative at $x \in A$, in the direction $y$, is given by,

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

(ii) The lower Dini derivative at $x \in A$, in the direction $y$, is given by,

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

(iii) The Clarke generalised 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 $y \in X$,

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

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

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

exists
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 $X$. In this paper we will also be interested in two slightly stronger notions of differentiability. A locally Lipschitz function $f$ is said to be strictly differentiable at $x$, in the direction $y$, if

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

exists and we say that $f$ is strictly differentiable at $x$, if $f$ is strictly differentiable at $x$, in every direction $y \in X$. Let us recall that a function $f$ is strictly differentiable at $x$, in the direction $y$ if, and only if,

$$
f^1(x; y) = f'(x; y) = -f^1(x; -y)
$$

In addition to the various notions of differentiability we will need the notion of a ‘null’ set. We say that a subset $N$ of a separable Banach space $X$ is universally measurable if it belongs to the $m$-completion of the Borel subsets, $\mathcal{B}(X)$, for each finite Borel measure $m$ on $\mathcal{B}(X)$. A universally measurable subset $N$ of $X$ is called a Haar-null set if there exists a probability measure $P$ on $\mathcal{B}(X)$, (which extends canonically to the universally measurable sets on $X$) such that $P(N + x) = 0$ for all $x \in X$. The Haar-null sets are closed under translation and countable unions, [5]. It follows therefore, that if $N$ is a Haar-null set then $X \setminus N$ is dense in $X$. In finite dimensions, the Haar-null sets coincide with the universally measurable Lebesgue null sets. We will say that a property $P$ holds almost everywhere if $\{x \in X : P(x) \text{ is not true}\}$ is contained in a Haar-null set. A further important class of sets that we will need to consider is the following. Let $A$ be a non-empty open subset of a Banach space $X$. Then a subset $S$ of $A$ is $1 - D$ almost everywhere in $A$, in the direction $y$, if for each $x \in A$

$$
\mu(\{t \in R : x + ty \in A \text{ and } x + ty \notin S\}) = 0
$$

(Here $\mu$ represents the Lebesgue measure on $R$.)

Note that, in the above definition, it is implicit that for each $x \in A$,

$$
\{t \in R : x + ty \in A \text{ and } x + ty \notin S\}
$$

is Lebesgue measurable. For us, the most important example of a $1 - D$ almost everywhere set is the following.
Proposition 1.1 ([7], Remark 2.4) 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 unit sphere),

$$D_y \equiv \{ x \in A : f'(x; y) \text{ exists} \}$$

is $1 - D$ almost everywhere in $A$, in the direction $y$.

Not surprisingly, Haar-null sets are related to $1 - D$ almost everywhere subsets.

Proposition 1.2 ([3], Proposition 5.1) Let $S$ be a universally measurable subset of a non-empty open subset $A$ of a separable Banach space $X$. If for some $y \in S(X)$, $S$ is $1 - D$ almost everywhere in $A$, in the direction $y$, then $A \setminus S$ is a Haar-null set.

In [3] the present authors said that a real-valued locally Lipschitz function $f$, defined on a non-empty open subset $A$, of a separable Banach space $X$, is essentially strictly differentiable on $A$, if $\{ x \in A : f$ is not strictly differentiable at $x \}$ is a Haar-null set, and they denoted by $\mathcal{S}_{e}(A)$, the family of all real-valued essentially strictly differentiable locally Lipschitz functions defined on $A$. We may also call such functions essentially smooth since strict differentiability is an appropriate localization of continuous differentiability.

On a few occasions we will need to consider vector-valued functions. Let $A$ be a non-empty open subset of a Banach space $X$ and let $f : A \to \mathbb{R}^n$ be defined by

$$f(x) \equiv (f_1(x), f_2(x), \ldots, f_n(x)) \quad \text{where } f_j : A \to \mathbb{R}.$$ 

Then we say that the vector-valued function $f$ is essentially strictly differentiable on $A$ if $f_j \in \mathcal{S}_{e}(A)$ for each $1 \leq j \leq n$, and in this case we write: $f \in \mathcal{S}_{e}(A; \mathbb{R}^n)$. In addition, we will say that $f$ is strictly differentiable at $x \in A$, in the direction $y$ if,

$$f^0_j(x; y) = -f^0_j(x; -y) \quad \text{for each } 1 \leq j \leq n$$

In this case we write: $f^0(x; y) = -f^0(x; -y)$.

Let us now return to the purpose of this paper, which as mentioned earlier, is to show that the family of functions $\mathcal{S}_{e}(A)$, possesses very striking closure properties. A first and optimistic guess might be, that if $f_1, f_2, \cdots, f_n \in \mathcal{S}_{e}(A)$ and $g \in \mathcal{S}_{e}(\mathbb{R}^n)$, then $g \circ f \in \mathcal{S}_{e}(A)$, where $f \equiv (f_1, f_2, \ldots, f_n)$. However, the following example shows that in general this is not true.
Example 1.1 Let $A$ denote the open interval $(0,1)$ in $R$ and let $C$ denote a Cantor subset of $A$ with $\mu(C) > 0$. Define the real-valued functions $f_1, f_2$ and $d_C$ on $A$ by: $f_1 \equiv 0$, $f_2(t) \equiv t$ and $d_C(t) \equiv \text{dist}(t, C)$. Further, define $g : R^2 \to R$ by $g(x, y) \equiv \text{dist}(x, y); \{0\} \times C)$. Clearly, $f_1$ and $f_2$ are strictly differentiable almost everywhere, in fact $f_1$ and $f_2$ are strictly differentiable everywhere. Moreover, by Theorem 8.5 in [3] we have that $g \in \mathcal{S}_c(R^2)$. We claim that $g \circ f \notin \mathcal{S}_c(A)$, where $f \equiv (f_1, f_2)$.

To see this, observe that $g \circ f(x) = d_C(x)$ for each $x \in A$. Now, it is standard that $d_C$ is not strictly differentiable at any point $x \in C$. Hence, it follows that $g \circ f$ is not strictly differentiable at any point of $C$, and so $g \circ f \notin \mathcal{S}_c(A)$.

Remark 1.1 By exploiting a little more about Haar-null sets, [3], we could have shown the following.

1. On any non-empty open subset $A$, of any separable Banach space $X$, there exist $f_1, f_2 \in \mathcal{S}_c(A)$ such that $g \circ f \notin \mathcal{S}_c(A)$, where $f \equiv (f_1, f_2)$.

2. It is true however, that for any non-empty open subset $A$ of any separable Banach space $X$, if $f \in \mathcal{S}_c(A)$ and $g \in \mathcal{S}_c(R)$ then $f \circ g \in \mathcal{S}_c(A)$, (see [3], Theorem 5.17).

2 A chain rule.

In this section of the paper, we show that there is a large subclass of the essentially strictly differentiable functions, that is closed under composition. This class of functions is closely related to Valadier’s saine functions, [9]. Let $A$ be a non-empty open subset of $R^n$. We say that a real-valued locally Lipschitz function $f$ defined on $A$ is arc-wise essentially smooth on $A$ if for each locally Lipschitz function $x \in \mathcal{S}_c((0,1); R^n)$,

$$\mu(\{t \in (0,1) : f^0(t; x^+(t)) \neq f^0(t; -x^+(t))\}) = 0$$

[Valadier requires this to hold for all absolutely continuous arcs.] We shall denote by $\mathcal{A}_c(A)$, the family of all arc-wise essentially smooth functions on $A$.

Remark 2.1 It is easily seen that the definition of $\mathcal{A}_c(A)$ is unaffected by replacing the open set $(0,1)$ by any (other) non-empty open subset of $R$. 
Proposition 2.1  For each non-empty open subset $A$ of $\mathbb{R}^n$, $\mathcal{A}_c(A) \subseteq \mathcal{S}_c(A)$.

Proof. Consider $f \in \mathcal{A}_c(A)$. For each fixed $y \in S(\mathbb{R}^n)$ we will show that the $G_\delta$ set $S_y \equiv \{ x \in A : f^0(x; y) = -f^0(x; -y) \}$ is $1 - D$ almost everywhere in $A$, in the direction $y$. To this end, consider $a \in A$. Let $U \equiv \{ t \in R : a + tx \in A \}$. Clearly $U$ is non-empty and open. Now, consider the mapping $x : U \rightarrow A$ defined by $x(t) \equiv a + ty$. It is obvious that $x \in S_c(U; R^n)$. We may now perform a little set arithmetic.

\[
\{ t \in R : a + ty \in A \text{ and } a + ty \notin S_y \} \\
= \{ t \in U : f^0(a + ty; y) \neq -f^0(a + ty; -y) \} \\
= \{ t \in U : f^0(x(t); x'(t)) \neq -f^0(x(t); -x'(t)) \}
\]

and so,

\[
\mu(\{ t \in R : a + ty \in A \text{ and } a + ty \notin S_y \}) = 0
\]

which shows that $S_y$ is $1 - D$ almost everywhere in $A$, in the direction $y$. Hence, by Proposition 1.2, $A \setminus S_y$ is a Haar-null (Lebesgue-null) set.

Next, let $\{ y_n : n \in N \}$ be a dense subset of $S(\mathbb{R}^n)$, and let $S \equiv \bigcap \{ S_{y_n} : n \in N \}$. It is easy to see that $A \setminus S$ is a Haar-null (Lebesgue-null) set. We shall complete the proof by showing that $f$ is strictly differentiable at each point of $S$. So consider $x_0 \in S$. Since both of the mappings $y \rightarrow f^0(x_0; y)$ and $y \rightarrow -f^0(x_0; -y)$ are continuous on $\mathbb{R}^n$ and $f^0(x_0; y_n) = -f^0(x_0; -y_n)$ for each $n \in N$, we have that $f^0(x_0; y) = -f^0(x_0; -y)$ for each $y \in S(\mathbb{R}^n)$. This shows that $f$ is strictly differentiable at $x_0$.

Proposition 2.2  Let $A$ be a non-empty open subset of $\mathbb{R}$. Then $\mathcal{A}_c(A) = \mathcal{S}_c(A)$.

Proof. We know from Proposition 2.1 that $\mathcal{A}_c(A) \subseteq \mathcal{S}_c(A)$. The proof that they are equal follows (as a special case) from Theorem 5.17 in [3].

Our next task is to show that the class of functions $\mathcal{A}_c(A)$ is reasonably large. Let us begin with the obvious observation that $\mathcal{A}_c(A)$ contains all the $C^1$ functions defined on $A$.

Lemma 2.1  Let $A$ be a non-empty open subset of $\mathbb{R}^n$. Let $f$ be a real-valued locally Lipschitz function defined on $A$. Then $f \in \mathcal{A}_c(A)$, if for each essentially strictly differentiable curve $x : (0, 1) \rightarrow A$,

\[
\mu(\{ t \in (0, 1) : f^0(x(t); x'(t)) = f'(x(t); x'(t)) \}) = 1
\]
A chain rule for Lipschitz functions

Proof. Let $x: (0, 1) \to A$ be essentially strictly differentiable on $(0, 1)$ that is, $x \in S_e((0, 1); R^n)$. We need to show that

$$\mu(\{ t \in (0, 1) : f^0(x(t); x'(t)) = -f^0(x(t); -x'(t)) \}) = 1$$

Consider the mapping $y: (0, 1) \to A$, defined by, $y(t) = x(1 - t)$. Note that almost everywhere $y'(t) = -x'(1 - t)$.

Let $E_1 \equiv \{ t \in (0, 1) : f^0(x(t); x'(t)) = f'(x(t); x'(t)) \}$ and $E_2 \equiv \{ t \in (0, 1) : -f^0(x(t); -x'(t)) = f'(x(t); x'(t)) \}$. Clearly, $f^0(x(t); x'(t)) = -f^0(x(t); -x'(t))$ on $E_1 \cap E_2$, so it remains to show that $\mu(E_2) = 1$.

Now, by the hypothesis, $E_3 \equiv \{ t \in (0, 1) : f'(y(t); y'(t)) = f^0(y(t); y'(t)) \}$ has measure one. On the other hand, $E_2 = h(E_3)$ where $h(t) \equiv 1 - t$. Therefore, $\mu(E_2) = 1$. ⋆

The next two lemmas are quite standard, but we include them for the sake of completeness.

Lemma 2.2 ([3], Lemma 5.13) Let $f_1, f_2, \ldots, f_n$ be real-valued locally Lipschitz functions defined on a non-empty open subset $A$ of a Banach space $X$. Let $g$ be a real-valued locally Lipschitz function defined on an open subset $B$ of $R^n$ which contains $f(A)$; where $f \equiv (f_1, f_2, \ldots, f_n)$.

Suppose that $f$ is differentiable in the direction $y$, at some point $x_0$, and that either, (i) $(g \circ f)'(x_0; y)$ exists or, (ii) $g'(f(x_0); f'(x_0; y))$ exists.

Then

$$(g \circ f)'(x_0; y) = g'(f(x_0); f'(x_0; y)).$$

Remark 2.2 From Lemma 2.2 we see that for each fixed $y \in X \setminus \{0\}$

$$\{ x \in A : (g \circ f)'(x; y) = g'(f(x); f'(x; y)) \}$$

is $1 - D$ almost everywhere in $A$, in the direction $y$.

Lemma 2.3 ([7], Lemma 2.3) Let $f$ be a real-valued locally Lipschitz function defined on a non-empty open subset $A$ of a Banach space $X$. Suppose that for some $y \in X \setminus \{0\}$, a subset $S$ of $A$, is $1 - D$ almost everywhere in $A$, in the direction $y$. Then for each $x \in A$,

$$f^0(x; y) = \limsup_{z \to x} f^+(z; y).$$
If \( f^i(z; y) \) exists for each \( z \in S \), then

\[
-f^0(x; -y) = \liminf_{z \in S} f^+(z; y)
\]

**Remark 2.3** In the case when \( f^0(x; y) = -f^0(x; -y) \)

\[
\lim_{z \in S} f^i(z; y) \text{ exists, and equals } f^i(x; y)
\]

The following lemma encapsulates the heart of our chain rule.

**Lemma 2.4** Let \( A \) be a non-empty open subset of a Banach space \( X \). Let \( f = (f_1, f_2, \ldots, f_n) \) be a locally Lipschitz function from \( A \) into \( \mathbb{R}^n \). Furthermore, let \( g \) be a real-valued locally Lipschitz function defined on a non-empty open subset \( B \) of \( \mathbb{R}^n \) which contains \( f(A) \). Suppose \( f \) is strictly differentiable at some point \( x_0 \in A \), in the direction \( y \) and that either, (i) \( g \) is strictly differentiable at \( f(x_0) \), in the direction \( f'(x_0; y) \) or, (ii) \( f'(x_0; y) = 0 \). Then \( g \circ f \) is strictly differentiable at \( x_0 \), in the direction \( y \).

**Proof.** Let \( D = \{ x \in A : (g \circ f)'(x; y) = g'(f(x); f'(x; y)) \} \). It follows from Remark 2.2 that \( D \) is \( 1 - D \) almost everywhere in \( A \), in the direction \( y \).

(i) Let us first observe that \( x_0 \in D \). From Lemma 2.3 we see that, to show \( (g \circ f)^0(x_0; y) = -(g \circ f)^0(x_0; -y) \) we need only show that

\[
\lim_{z \in D} (g \circ f)^i(z; y) \text{ exists, and equals } (g \circ f)^i(x_0; y)
\]

So suppose \( \epsilon > 0 \). By the continuity of the mapping \( w \to g^+(w; f'(x_0; y)) \) at \( f(x_0) \) there exists an open neighbourhood \( U \) of \( x_0 \) such that

\[
|g^+(f(z); f'(x_0; y)) - g^+(f(x_0); f'(x_0; y))| < \epsilon/2
\]

for each \( z \in U \). Now, as \( g \) is locally Lipschitz there exists an open neighbourhood \( V \) of \( x_0 \) and a constant \( M > 0 \) such that

\[
|g'(f(z); f'(z; y)) - g'(f(z); f'(x_0; y))| \leq M\|f'(z; y) - f'(x_0; y)\|
\]

for each \( z \in D \cap V \). On the other hand, \( x \to f^+(x; y) \) is continuous at \( x_0 \), therefore there exists an open neighbourhood \( W \) of \( x_0 \) such that

\[
\|f'(x_0; y) - f^+(x_0; y)\| < \epsilon/2M
\]
A chain rule for Lipschitz functions

for each \( x \in W \). Hence we have that for each \( x \in D \cap W \cap V \cap U \)

\[
\left| (g \circ f)'(x_0; y) - (g \circ f)'(x; y) \right|
= \left| g'(f(x_0); f'(x_0; y)) - g'(f(x); f'(x; y)) \right|
\leq \left| g'(f(x_0); f'(x_0; y)) - g'(f(x); f'(x_0; y)) \right|
+ \left| g'(f(x); f'(x_0; y)) - g'(f(x); f'(x; y)) \right|
\leq \epsilon/2 + \epsilon/2 = \epsilon
\]

Therefore,

\[
\lim_{z \in D} (g \circ f)'(z; y) = (g \circ f)'(x_0; y)
\]

(ii) From Lemma 2.3, we have that

\[
(g \circ f)^0(x_0; y) = \limsup_{z \in D} (g \circ f)'(z; y)
\]

and

\[
-(g \circ f)^0(x_0; -y) = \liminf_{z \in D} (g \circ f)'(z; y)
\]

Now, \( g \) is locally Lipschitz, so there exists an open neighbourhood \( U \) of \( x_0 \) and a \( M > 0 \) such that

\[
\left| (g \circ f)'(z; y) \right| = \left| g'(f(z); f'(z; y)) \right| \leq M \left\| f'(z; y) \right\|
\]

for each \( z \in U \cap D \).

However, as \( f \) is strictly differentiable at \( x_0 \), in the direction \( y \), and \( f'(x_0; y) = 0 \) we have that

\[
\lim_{z \in D} (g \circ f)'(z; y) = 0
\]

and so by Lemma 2.3 we get that \( (g \circ f)^0(x_0; y) = -(g \circ f)^0(x_0; -y) = 0 \).

We need a few more definitions before we can show that \( A(A) \) contains a significant class of functions, (above and beyond the \( C^1 \) functions). 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 semi-smooth (lower semi-smooth)

at a point \( x \in A \), in the direction \( y \) if,

\[
\begin{align*}
     f^+(x; y) &\geq \limsup_{t \to 0^+, y' \to -y} f^+(x + ty'; y) \\
     f^-(x; y) &\leq \liminf_{t \to 0^+, y' \to -y} f^-(x + ty'; y)
\end{align*}
\]
Moreover, we say that \( f \) is semi-smooth at a point \( x \in A \), in the direction \( y \) if,

\[
\limsup_{t \to 0^+} f^+(x + ty; y) = f^+(x; y) = \liminf_{t \to 0^+} f^-(x + ty; y)
\]

If \( X \) is finite dimensional then we say that \( f \) is arc-wise essentially upper semi-smooth (arc-wise essentially lower semi-smooth) on \( A \), if for each \( x \in \mathcal{S}_c((0, 1); A) \)

\[ \mu(\{t \in (0, 1) : f \text{ is upper (lower) semi-smooth at } x(t) \text{ in the direction } x'(t)\}) = 1 \]

We recall that a real-valued function \( g : (a, b) \to \mathbb{R} \) is approximately continuous at a point \( x \in (a, b) \) if for each \( \epsilon > 0 \),

\[ \lim_{\delta \to 0^+} \frac{\mu(\{t \in [x - \delta, x + \delta] : |g(t) - g(x)| > \epsilon\})}{2\delta} = 0 \]

The following is to be found in ([8], Theorem 35.3) and elsewhere.

**Theorem 2.1** Let \( g \) be a real-valued Lebesgue measurable function defined on \((a, b)\), then \( g \) is approximately continuous almost everywhere in \((a, b)\).

The following technical result allows us to perform only “unilateral” analysis in our main theorem.

**Theorem 2.2** Let \( f \) be a real-valued locally Lipschitz function defined on a non-empty open subset \( A \) of a finite dimensional Banach space \( X \). If \( f \) is arc-wise essentially upper semi-smooth (arc-wise essentially lower semi-smooth) on \( A \), then \( f \in \mathcal{A}_c(A) \).

**Proof.** Suppose that \( f \) is arc-wise essentially upper semi-smooth on \( A \) (the proof for the case when \( f \) is arc-wise essentially lower semi-smooth on \( A \) is obtained by considering \( -f \)). Let \( x : (0, 1) \to \mathbb{R}^n \) be strictly differentiable almost everywhere in \((0, 1)\). Consider the mapping \( T : (0, 1) \to \mathbb{R} \) defined by \( T(t) = f^+(x(t); x^+(t)) \). Of course, it follows from Remark 2.2 that almost everywhere \( T(t) = f'(x(t); x'(t)) \). Let \( P : (0, 1) \to \mathbb{R} \) be defined by \( P(t) = f^0(x(t); x^+(t)) \). Clearly, \( T(t) \leq P(t) \) for each \( t \in (0, 1) \). We claim that \( P \) is upper semi-continuous almost everywhere in \((0, 1)\) and so Lebesgue measurable on \((0, 1)\). Indeed, \( P \) is upper semi-continuous at each point \( t_0 \in (0, 1) \) where \( t \to x^+(t) \) is continuous. It now follows from Theorem
2.1 that $P$ is approximately continuous almost everywhere in $(0, 1)$. Let $S \subseteq (0, 1)$ be the set of all points in $(0, 1)$ where $P$ is approximately continuous, $t \rightarrow x^+(t)$ is continuous and $f$ is upper semi-smooth at $x(t)$ in the direction $x'(t)$. From the hypothesis of the theorem it follows that $\mu(S) = 1$. We claim that $T(t) = P(t)$ at each point of $S$.

To prove this, consider an arbitrary point $t_0 \in S$. Let $\epsilon > 0$ be fixed, but arbitrary. We will show that $P(t_0) \leq T(t_0) + \epsilon$. Note that without loss of generality we may assume that $x^+(t_0) \neq 0$. Now, since $f$ is upper semi-smooth at $x(t_0)$ in the direction $x'(t_0)$ there exists a $0 < \delta$ such that,

$$f^+(z; x'(t_0)) \leq T(t_0) + \epsilon/3$$

for each $z$ in the non-empty open subset $V$, where $V$ equals,

$$V \equiv \{ z \in A : z = x(t_0) + sw, \; 0 < s < \delta \text{ and } \|w - x'(t_0)\| < \delta \}$$

Let

$$E(\lambda) \equiv \frac{x(t_0 + \lambda) - x(t_0)}{\lambda} - x'(t_0) \text{ for } 0 < \lambda < (1 - t_0).$$

Since $x$ is Gateaux differentiable at $t_0$ there exists a $\eta \in (0, \delta)$ such that $\|E(\lambda)\|_\infty < \delta$ for each $\lambda \in (0, \eta)$. Therefore, for each $\lambda \in (0, \eta),

$$x(t_0 + \lambda) = x(t_0) + \lambda(x'(t_0) + E(\lambda)) \in V$$

and so, $f^0(x(t_0 + \lambda); x'(t_0)) \leq T(t_0) + \epsilon/3$. Now, $f$ is locally Lipschitz so there exists an open neighbourhood $U$ of $x(t_0)$ and a constant $M > 0$ such that $|f(z) - f(y)| \leq M\|z - y\|$ for each $x, y \in U$. It follows from this, that for each $z \in U$ and each $x, y \in X$, $|f^0(z; x) - f^0(z; y)| \leq M\|x - y\|$. Hence, there exists an $r \in (0, \eta)$ such that

$$|P(t_0 + \lambda) - f^0(x(t_0 + \lambda); x'(t_0))| \leq M\|x^+(t_0 + \lambda) - x'(t_0)\| < \epsilon/3$$

for each $\lambda \in (0, r)$. From this we get that,

$$P(t_0 + \lambda) \leq T(t_0) + 2\epsilon/3$$

for each $\lambda \in (0, r)$. Also, since $P$ is approximately continuous at $t_0$ there exists a $\lambda_0 \in (0, r)$ such that,

$$P(t_0 + \lambda_0) \geq P(t_0) - \epsilon/3$$
So at last we have that, $P(t_0) \leq T(t_0) + \epsilon$. However, since $\epsilon$ was arbitrary, we have that $P(t_0) = T(t_0)$. The result now follows from Lemma 2.1 and the fact that $T(t) = f'(x(t); x'(t))$ almost everywhere in $(0,1)$. 

Now, we may show that the family of functions $\mathcal{A}_e(A)$ enjoys quite remarkable closure properties.

**Theorem 2.3** Let $A$ be a non-empty open subset of $\mathbb{R}^m$ and suppose $f_1, f_2, \ldots, f_n \in \mathcal{A}_e(A)$. Let $g$ be a real-valued function locally Lipschitz function defined on a non-empty open subset $U$ of $\mathbb{R}^n$, which contains $f(A)$, where $f \equiv (f_1, f_2, \ldots, f_n)$. If $g \in \mathcal{A}_e(U)$, then $g \circ f \in \mathcal{A}_e(A)$.

**Proof.** Let $x : (0,1) \to A$ be essentially strictly differentiable on $(0,1)$ and define $f : A \to \mathbb{R}^n$, by $f \equiv (f_1, f_2, \ldots, f_n)$. Let us also define a curve $z : (0,1) \to U$, by $z(t) \equiv f(x(t))$. Our first task will be to show that $z \in \mathcal{S}_e((0,1); U)$. However, this follows almost immediately from Lemma 2.4 and the fact that $f \in \mathcal{A}_e(A)$.

Let $S_f \equiv \{ t \in (0,1) : f^0(x(t); x'(t)) = -f^0(x(t); -x'(t)) \}$ and $S_g \equiv \{ t \in (0,1) : g^0(z(t); z'(t)) = -g^0(z(t); -z'(t)) \}$. Now, define $S \equiv S_f \cap S_g$. Clearly, $\mu(S) = 1$. We claim that $(g \circ f)^0(x(t); x'(t)) = -(g \circ f)^0(x(t); -x'(t))$ at each point of $S$. To this end, consider any point $t_0 \in S$. Let us first observe that if $x'(t_0) = 0$, then we are done. So suppose that $x'(t_0) \neq 0$. Set $x_0 \equiv x(t_0)$ and $y \equiv x'(t_0)$. Note that since $y \neq 0$, we have that $f$ is strictly differentiable at $x_0$, in the direction $y$. We consider two cases:

(i) Suppose that $f'(x_0; y) = x'(t_0) \neq 0$. Then $g$ is strictly differentiable at $f(x_0) = z(t_0)$, in the direction $f'(x_0; y) = z'(t_0)$, since $t_0 \in S_g$. It now follows from Lemma 2.4 part(i) that $g \circ f$ is strictly differentiable at $x_0$, in the direction $y$; that is, $(g \circ f)^0(x(t_0); x'(t_0)) = -(g \circ f)^0(x(t_0); -x'(t_0))$.

(ii) Suppose that $f'(x_0; y) = x'(t_0) = 0$. Then by Lemma 2.4 part(ii) $g \circ f$ is strictly differentiable at $x_0$, in the direction $y$; that is, $(g \circ f)^0(x(t_0); x'(t_0)) = -(g \circ f)^0(x(t_0); -x'(t_0))$.

On relaxing the hypotheses on the $f_1, f_2, \ldots, f_n$ we may still recover a satisfactory theorem:

**Theorem 2.4** Let $A$ be a non-empty open subset of a separable Banach space $X$, and suppose $f_1, f_2, \ldots, f_n \in \mathcal{S}_e(A)$. Suppose further, that $U$ is a non-empty open subset of $\mathbb{R}^n$, which contains $f(A)$, where $f \equiv (f_1, f_2, \ldots, f_n)$. Then for each $g \in \mathcal{A}_e(U)$, $g \circ f \in \mathcal{A}_e(A)$. 
A chain rule for Lipschitz functions

**Proof.** We will show first, that for each \( y \in X \), \( g \circ f \) is strictly differentiable, in the direction \( y \), almost everywhere in \( A \). Fix \( y \in X \). Let \( D \) be any Borel subset of \( A \) such that \( A \setminus D \) is a Haar-null set and \( f^0(x; z) = -f^0(x; -z) \) for each \( x \in D \) and \( z \in X \). Let \( P_y \equiv \{ x \in A : (g \circ f)^0(x; y) = -(g \circ f)^0(x; -y) \} \). Clearly, \( P_y \) is a Borel set, in fact \( P_y \) is a \( G_\delta \) set. Let \( H \) be a closed hyperplane in \( X \) such that \( y \not\in H \). Now consider the isomorphism \( T : H \times R \to X \) defined by \( T(h, t) = h + ty \). Let

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

By the remark just after Theorem 6 in [5], (see Remark 2.4 below) we see that \( H \setminus H_s \) is a Haar-null set. To show that \( A \setminus P_y \) is a Haar-null set in \( X \), it suffices (also because of the remark made after Theorem 6 in [5]) to show that for each \( h \in H_s \), \( \mu(\{ t \in R : T(h, t) \in A \setminus P_y \}) = 0 \).

To this end, consider \( h_0 \in H_s \). Let \( A_{h_0} \equiv \{ t \in R : T(h_0, t) \in A \} \). If \( A_{h_0} = \emptyset \), then we are done, so let us suppose that \( A_{h_0} \neq \emptyset \). Define \( x : A_{h_0} \to A \), by \( x(t) \equiv T(h_0, t) \) and \( z : A_{h_0} \to U \) by \( z(t) \equiv f(x(t)) \). Since \( h_0 \in H_s \), \( z \in S_c(A_{h_0}; U) \). Let

\[
S_f \equiv \{ t \in A_{h_0} : f^0(x(t); x'(t)) = -f^0(x(t); -x'(t)) \}
\]

and

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

Now, define \( S \equiv S_f \cap S_g \). Clearly, \( \mu(A_{h_0} \setminus S) = 0 \). We claim that

\[
(g \circ f)^0(x(t); x'(t)) = -(g \circ f)^0(x(t); -x'(t))
\]

at each point \( t \in S \). To see this, consider an arbitrary point \( t_0 \in S \). Set \( x_0 \equiv x(t_0) \). Note that since \( x'(t_0) = y \neq 0 \), we have that \( f \) is strictly differentiable at \( x_0 \), in the direction \( y \).

We consider two cases.

(i) If \( f'(x_0; y) = z'(t_0) \neq 0 \), then \( g \) is strictly differentiable at \( f(x_0) = z(t_0) \), in the direction \( z'(t_0) = f'(x_0; y) \), since \( t_0 \in S_y \). It now follows from Lemma 2.4 part(i) that \( g \circ f \) is strictly differentiable at \( x_0 \), in the direction \( y \); that is, \( (g \circ f)^0(x(t_0); y) = -(g \circ f)^0(x(t_0); -y) \).

(ii) Suppose that \( f'(x_0; y) = z'(t_0) = 0 \). Then by Lemma 2.4 part(ii), \( g \circ f \) is strictly differentiable at \( x_0 \), in the direction \( y \); that is, \( (g \circ f)^0(x(t_0); y) = -(g \circ f)^0(x(t_0); -y) \).
Hence, \( \mu(\{t \in A_{h_{\theta}} : T(h_{0}, t) \in A \setminus P_{y}\}) = 0 \). Let \( \{y_{n} : n \in \mathbb{N}\} \) be a dense subset of \( S(X) \), and let \( P \equiv \bigcap_{n=1}^{\infty} P_{y_{n}} \). It follows from the continuity of the mappings \( z \rightarrow (g \circ f)^{0}(x; z) \) and \( z \rightarrow -(g \circ f)^{0}(x; -z) \) that \( (g \circ f)^{0}(x; z) = -(g \circ f)^{0}(x; -z) \) for each \( z \in S(X) \) and each \( x \in P \). This completes the proof. \( \square \)

**Corollary 2.1** Let \( A \) be a non-empty open subset of \( \mathbb{R}^{m} \) then \( \mathcal{A}_{c}(A) \) and \( \mathcal{S}_{c}(A) \) are function algebras and vector lattices containing the \( C^{1} \)-functions and the convex functions.

**Remark 2.4** The remark in [5] just after Theorem 6 says:

If \( H \) is an arbitrary Abelian Polish group and \( T \) is a locally compact Abelian group (for example \( (\mathbb{R}, +) \)) then for any universally measurable set \( A \subseteq H \times T \) the following are equivalent.

(i) The section \( A(h) \equiv \{ t \in T : (h, t) \in A \} \) is a Haar-null set, for the Haar measure on \( T \), for almost all \( h \in H \).

(ii) The set \( A \) is a Haar-null set in the product group \( H \times T \).

Recall that a topological space is said to be Polish, if it is homeomorphic to a separable, complete metric space. A complete proof of the previous remark - essentially due to Christensen - is given in an Appendix of [4]. We should also observe that from the proof of the previous theorem, it can be shown that for almost all \( x \) in \( A \),

\[
\nabla (g \circ f)(x) = \partial g(f(x)) \cdot \nabla f(x)
\]

Here \( \partial g \) is the Clarke subgradient of \( g \). Note, however, that \( g \) may fail to be strictly differentiable at \( f(x) \), unless the range of \( \nabla f(x) \) is all of \( \mathbb{R}^{n} \).

Finally, let us observe that the class of functions \( \mathcal{A}_{c}(\mathbb{R}^{n}) \) contains many distance functions. Indeed, one can show that if \( C \) is a regular subset of \( \mathbb{R}^{n} \), then the distance function generated by this set and any smooth norm on \( \mathbb{R}^{n} \), is in this class. To see this, we need to make the following observations:

(i) For any norm on \( \mathbb{R}^{n} \), the distance function \( d_{C} \) is regular at each point of \( C \), see [1].

(ii) If the norm is smooth on \( \mathbb{R}^{n} \), then \(-d_{C}\) is regular on \( \mathbb{R}^{n} \setminus C \), see [2].
A chain rule for Lipschitz functions

We can now define a class of closed subsets of $\mathbb{R}^n$ as follows. A subset $C$ of $\mathbb{R}^n$ is arc-wise essentially smooth if for each $x \in \mathcal{S}_e((0,1); \mathbb{R}^n)$ the set

$$\{t \in (0,1) : x'(t) \in K_C(x(t)) \setminus T_C(x(t))\}$$

has measure zero. Here $K_C(x)$ denotes the contingent cone of $C$ at $x$, and $T_C(x)$ denotes the Clarke tangent cone of $C$ at $x$. Recall that $C$ is regular at $x$ if $K_C(x) = T_C(x)$, [6]. Note that all sets that are regular except on a countable set are arc-wise essentially smooth. Thus, all closed convex sets and all smooth manifolds are arc-wise essentially smooth.

**Proposition 2.3** Let $\| \cdot \|$ be a smooth norm on $\mathbb{R}^n$ and let $C$ be a non-empty closed subset of $\mathbb{R}^n$. Then the distance function generated by the norm $\| \cdot \|$ and the set $C$ is a member of $\mathcal{A}_e(\mathbb{R}^n)$ if, and only if, $C$ is arc-wise essentially smooth.

It is immediate from this Proposition and Theorem 2.3 that, unlike the family of regular sets, the family of arc-wise essentially smooth sets is closed under finite unions. We should also observe that from Proposition 2.1 it follows that the boundary of every arc-wise essentially smooth set is a Lebesgue null set. We also note that the set of Example 1.1 is a null set but is not arc-wise essentially smooth. Finally, observe that we may now provide general conditions to ensure that exact penalty functions of the form

$$f(x) + Kd_C(x)$$

will be in $\mathcal{S}_e(A)$ or in $\mathcal{A}_e(A)$.

**References**


