

Multifunctional and functional analytic techniques in nonsmooth analysis

Jonathan M. BORWEIN*

*Department of Mathematics and Statistics
Simon Fraser University
Burnaby, BC, V5A 1S6
Canada*

Qiji J. ZHU†

*Department of Mathematics and Statistics
Western Michigan University
Kalamazoo, MI 49008
USA*

Abstract

These lectures center on the structure of real-valued Lipschitz functions, and their generalized derivatives on Banach spaces. We pay some attention to the role of measure and category and will try to illustrate a number of different techniques. These published notes are much more detailed and comprehensive than the lectures as given. Much of this development is based on recent joint work with Warren Moors (Wellington) and others.

The exposition will be organized around the following interwoven themes.

- Set-valued analysis. *Minimal* upper semi-continuous *multifunctions* (‘uscos and cuscos’) and related topological tools. Selections and single-valuedness. Relations to differentiability.
- Measure and category as competing notions of smallness: *Haar null sets* and generic sets in Banach spaces. Other concepts of prevalence.
- ‘Utility grade’ renorming theory and its application to the study of *viscosity subderivatives* and (partially) smooth variational principles. “Fuzzy” calculus and some equivalent reformulations.
- The structure of Lipschitz functions. Especially the calculus of *essentially smooth Lipschitz functions* and the vector-lattice algebra they generate. Chain rules, and questions of integrability and representability.
- Applications to and examples of *distance functions*. Minimality of distance functions and proximal normal formulae revisited. More general perturbation functions.
- Convex functions and related *sequences in Banach space*. How properties of given Banach spaces are reflected in the convex functions they support. Conjugates and subdifferentials of eigenvalue functions.

*Research was supported by NSERC and by the Shrum Endowment at Simon Fraser University.

†Research was supported by the National Science Foundation under grant DMS-9704203.

Notation

X – (separable) Banach spaces.

T – Topological spaces.

$\bar{\mathbf{R}}$ –the extended real line $\mathbf{R} \cup \{+\infty\}$.

$d_C(x)$ –the distance between point x and set C defined by $d_C(x) := \inf\{\|x - c\| : c \in C\}$.

$[A, B]$ –the convex hull of $A \cup B$.

φ, ψ –lower semicontinuous functions.

f, g –(locally) Lipschitz functions.

h, k –convex functions.

$S(X)$ – the unit sphere of X .

$B_r(x)$ –closed ball centered at x with radius r .

$\text{Gr}(F)$ or $\text{Graph}(F)$ –graph of a (multi)function F .

Directional derivatives: Clarke: f^0 ; Dini: $d^+ f, d^- f$; right-hand directional derivative: f'_+ .

$:=$ or \equiv : equal, equal by definition.

Contents

1	Real-valued functions and their subdifferentials	62
	1.1 Examples	63
	1.2 The convex subdifferential	65
	1.3 The Clarke generalized gradient	67
	1.4 Small sets	71
2	USCOs and CUSCOs	75
	2.1 Basic properties	75
	2.2 Minimality	77
	2.3 Maximal monotones and minimal cuscos	83
	2.4 Minimality of the Clarke generalized gradient	85
	2.5 Subgradient representation of multifunctions	88
3	Partially smooth variational principles	93
	3.1 Smooth variational principles	94
	3.2 Differentiability of the norm	94
	3.3 Partially smooth subdifferentials	95
	3.4 Partially smooth equivalent norms	98
	3.5 Partially smooth variational principles	100
	3.6 Several useful tools	101
	3.7 Applications	107
4	Essentially smooth Lipschitz functions	112
	4.1 An example	113
	4.2 Essentially strictly differentiable Lipschitz functions	114
	4.3 Stability properties for $S_e(A)$ and a chain rule	117
	4.4 Perturbation functions	122
	4.5 Distance functions	124
	4.6 Relationships between integrability, representability and smoothness	129
5	Convex functions and classifications of normed spaces	133
	5.1 Finite dimensions	134

5.2	Reflexive spaces	136
5.3	Separable spaces	138
5.4	Asplund spaces and spaces containing ℓ_1	140
5.5	Spaces containing c_0	141
6	Appendix: Convex Haar null sets	142
6.1	Separable reflexive spaces	143
6.2	Super-reflexive separable spaces	144
6.3	Other extensions	146

1 Real-valued functions and their subdifferentials

Many – perhaps most – problems arising naturally in optimization and optimal control intrinsically involve nonsmoothness. The following two simple examples illustrate how such intrinsic nonsmoothness arises even in problems with entirely smooth data.

Example 1.1 Often one wishes to deal with the maximum of two or more functions. Let $f(x) := \max(f_1, f_2)$. For the benign smooth functions on \mathbf{R} , $f_1(x) := x$ and $f_2(x) := -x$ one obtains $f(x) = |x|$, a familiar essentially nonsmooth function.

Example 1.2 Consider the very simple constrained minimization problem of minimizing $f(x)$ subject to $g(x) = a$, $x \in \mathbf{R}$. Here $a \in \mathbf{R}$ is a parameter allowing for perturbation of the constraint. In practice it is often important to know how the model responds to the perturbation a . For this we need to consider, for example, the *optimal value*

$$v(a) := \inf\{f(x) : g(x) = a\}$$

as a function of a . Consider a concrete example with two smooth functions $f(x) := 1 - \cos x$ and $g(x) := \sin(6x) - 3x$ and $a \in [-\pi/2, \pi/2]$ which corresponds to $x \in [-\pi/6, \pi/6]$. The graph of $(g(x), f(x)) : x \in [-\pi/6, \pi/6]$ is given in Figure 1.1. It is easy to see from Figure 1.1 that the optimal value function v is not smooth, in fact, not even continuous.

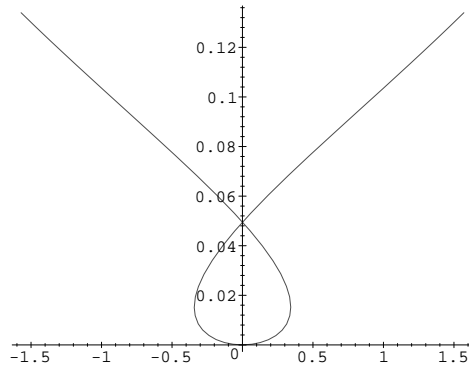


Figure 1.1: Smooth becomes Nonsmooth

In the attempt to deal with such problems, various (set-valued) generalized derivatives have been introduced to replace the nonexistent derivative for various classes of functions. The study of convex functions and their subdifferentials has a long and honourable history. The publication of Rockafellar's seminal book "Convex Analysis" [127] marked the maturity of convex analysis. Clarke's lovely work on generalized gradients of Lipschitz functions inaugurated the systematic study of nonsmooth analysis in realms beyond the class of convex functions and subsequently many other generalized derivatives were introduced. Among the most frequently used *generalized derivative* concepts are the *co-derivatives* introduced by Mordukhovich [113, 114, 115], *approximate* and *geometric subderivatives* introduced by Ioffe [87, 88], *contingent derivatives* introduced by Aubin [3], *Michel and Penot's derivatives* [109], *Treiman's B-derivatives* [144, 145], and more general objects such as Warga's *derivate containers* [151, 152] and their more recent refinement *multidifferentials* by Sussmann [137, 138]. As time has gone by the range of applications of such study has also expanded significantly. By now a reasonably consistent theory exists for arbitrary lower semicontinuous functions (see the survey paper [43]).

Despite these developments the class of locally Lipschitz functions, in particular, convex functions, merits special attention. Firstly, they represent the bulk of (computationally) tractable problems arising in practice. Secondly, they possess special structure which leads to results that are not available for more general classes of lower semicontinuous functions. Finally, this is often the place we can understand most starkly why things work or go wrong. For these reasons we will concentrate on locally Lipschitz functions and some of their most well-structured subclasses in these lectures. For similar reasons we will focus on only two generalized subdifferentials: the convex subdifferential and the Clarke generalized gradient.

1.1 Examples

The functions we are most concerned with herein are the:

- (a) (locally) Lipschitz functions (we will often denote them by f, g);
- (b) distance functions (d_C);
- (c) convex functions (we will often denote them by h, k);
- (d) differences of convex functions (*d.-c.* functions);
- (e) and of course smooth functions.

The following examples offer some motivation:

Example 1.3 Consider the following fairly general constrained minimization problem:

$$\mathcal{P} \quad \text{minimize } \{\varphi(x) : x \in D\},$$

where D is a closed subset of Banach space X and $\varphi : X \rightarrow \bar{\mathbf{R}}$ is a lower semicontinuous function. Define

$$C := \{(x, r) : \varphi(x) \leq r, x \in D\}.$$

Then \bar{x} is a solution to \mathcal{P} if and only if $(\bar{x}, \varphi(\bar{x}))$ is a solution to the following constrained minimization problem.

$$\begin{aligned} \mathcal{P}^* \quad & \text{minimize} && f(x, r) = r \\ & \text{subject to} && (x, r) \in C. \end{aligned}$$

When φ is a convex function and D a convex set we are studying a convex programming problem with a convex Lipschitz cost function. In general we may always obtain a constrained minimization problem with a Lipschitz cost function f . Then, using \mathcal{P}^* if necessary, via exact penalization we can convert problem \mathcal{P} to one without constraints.

$$\text{minimize } f + L \cdot d_C.$$

Here d_C is the distance to set C defined by

$$d_C(x) := \inf\{\|x - c\| : c \in C\}$$

and L is greater than the Lipschitz rank of f . This emphasizes how naturally and dramatically the distance function arises in optimization problems.

The special importance of the (hidden) class of differences of convex (d.-c.) functions may be limned from the following examples. A more detailed discussion of convexity follows in Lecture 1.2.

Example 1.4 Let $S(n)$ be the vector space of $n \times n$ symmetric matrices and let $A \in S(n)$. We write the n eigenvalues of A including multiplicity as $\lambda_1(A) \geq \dots \geq \lambda_n(A)$ and define the *eigenvalue map* $\lambda : S(n) \rightarrow \mathbf{R}^n$ by $\lambda(A) := (\lambda_1(A), \dots, \lambda_n(A))$. The maximum eigenvalue or more generally the k th-largest eigenvalue of a matrix is very important in many practical problems.

Define, for $x \in \mathbf{R}^n$ and $1 \leq k \leq n$, function

$$\sigma_k(x) := \max\left\{\sum_{i=1}^k x_{n_i} : 1 \leq n_1 < n_2 < \dots < n_k \leq n\right\}.$$

Then σ_k , $k = 1, \dots, n$ are convex functions. Moreover, they are invariant under coordinate permutation. Thus, we see that $\lambda_k(A)$, the k th-largest eigenvalue of a matrix $A \in S(n)$, is the composition of the difference of convex functions $\sigma_k - \sigma_{k-1}$ with λ , and as such is a d.-c. matrix function, and necessarily depends in a Lipschitz manner on the input [103].

Example 1.5 A concave function may obviously be viewed as the difference of 0 and a convex function. An important concave example is the following “barrel” function that occurs in interior point methods:

$$\ln(\det(A)) = \ln\left(\prod_{i=1}^n \lambda_i\right) = \sum_{i=1}^n \ln \lambda_i$$

where $A \in S(n)$ and $\lambda_i, i = 1, \dots, n$ are the eigenvalues of A .

Example 1.6 Let C be a closed subset of \mathbf{R}^N in the Euclidean norm. Then on $\mathbf{R}^N \setminus C$ the distance function d_C is locally a difference of convex functions. In fact, let $x_0 \in \mathbf{R}^N \setminus C$

and let $G := \{x : \|x - x_0\| < d_C(x_0)/2\}$. Then, for any $y \in C$, the function $g_y(x) = \|x - y\|$ has $2/d_C(x_0)$ -Lipschitz derivative on G . Let $K := 2/d_C(x_0)$. Then, for $y \in C$, $x \mapsto K\|x\|^2/2 - \|x - y\|$ is convex on G because the derivative of this function is monotone (as discussed below). Therefore, the function

$$V(x) := K\|x\|^2/2 - d_C(x) = \sup_{y \in C} K\|x\|^2/2 - \|x - y\|,$$

is continuous and convex on G . Thus the function $d_C(x) = K\|x\|^2/2 - V(x)$ is the difference of two convex functions on G .

A simpler argument due to Asplund uses the parallelogram law to write

$$d_C^2(x) = \|x\|^2 - \sup_{c \in C} [2\langle x, c \rangle - \|c\|^2],$$

and so to express the square of any distance function globally as the difference of two convex functions with one C^∞ .

From this example we also see that $-d_C$ may be locally written as a sum of a C^1 function and a convex function on $\mathbf{R}^N \setminus C$. The conclusion also holds in Hilbert spaces. We observe without proof that this property does not extend to a point on the boundary of the set. Details can be found in [150]. Moreover, this holds – locally – in an arbitrary smooth norm on \mathbf{R}^N .

1.2 The convex subdifferential

The theory of subdifferentials for convex functions provides prototypes for many results in nonsmooth analysis. It also provides many important examples that illuminate the boundary of nonsmooth analysis. In this lecture we recall the definition of the convex subdifferential and some of its basic properties. Phelps' lecture notes [120] are an excellent short introduction on this topic. One can find a more comprehensive account of finite-dimensional subdifferential theory in Rockafellar's classical book [127]. Many interesting non-traditional applications and developments can be found in [26].

To have the flexibility of handling indicator functions we must consider extended valued lower semicontinuous convex functions. Consider an extended-valued function $\varphi : X \rightarrow \bar{\mathbf{R}}$. The *domain* of f is defined by $\text{dom}(\varphi) := \{x \in X : \varphi(x) < +\infty\}$. Recall that φ is *lower semicontinuous* provided that, for any $x_n \rightarrow x \in \text{dom}(\varphi)$, $\liminf_n \varphi(x_n) \geq \varphi(x)$ and φ is convex provided that, for any $x, y \in X$ and any $\lambda \in (0, 1)$,

$$\varphi(\lambda x + (1 - \lambda)y) \leq \lambda\varphi(x) + (1 - \lambda)\varphi(y).$$

The *subdifferential* of φ is defined as follows. For $x \in \text{dom}(\varphi)$,

$$\partial\varphi(x) := \{x^* \in X^* : \langle x^*, y - x \rangle \leq \varphi(y) - \varphi(x), \forall y \in X\},$$

and, for $x \notin \text{dom}(\varphi)$, $\partial\varphi(x) = \emptyset$. A key fact about a lower semicontinuous convex function is that it is locally Lipschitz in the interior of its domain:

Theorem 1.7 *Let $k : X \rightarrow \bar{\mathbf{R}}$ be a lower semicontinuous convex function on a Banach space. Then, for any $x \in \text{int dom}(k)$, k is locally Lipschitz around x .*

Proof *Step 1.* k is locally bounded above around x . Define $E_n := \{x \in X : k(x) \leq n\}$. Then E_n are closed subsets of X because k is lower semicontinuous. Moreover, $\text{int dom}(k) \subset \bigcup_{n=1}^{\infty} E_n$. By the Baire category theorem 1.26 there exists an E_n such that $\text{int dom}(k) \cap \text{int } E_n \neq \emptyset$. Let us assume that $B_\eta(y) \subset \text{int dom}(k) \cap \text{int } E_n$. Take an $\alpha > 0$ small enough so that $z = (1 + \alpha)x - \alpha y \in \text{int dom}(k)$. Since k is convex $\text{int dom}(k)$ is a convex set and, therefore, $[z, B_\eta(y)] \subset \text{int dom}(k)$. For any $u \in [z, B_\eta(y)]$ there exists $\lambda \in [0, 1]$ and $v \in B_\eta(y)$ such that $u = \lambda z + (1 - \lambda)v$. Then

$$k(u) \leq \lambda k(z) + (1 - \lambda)k(v) \leq \max(k(z), n).$$

That is to say k is bounded above on $[z, B_\eta(y)]$. It remains to observe that $B_{\frac{\alpha\eta}{1+\alpha}}(x) \subset [z, B_\eta(y)]$.

Step 2. Now we can assume that k is bounded above on $B_{2\eta}(x)$, say, by $M > 0$. We show it is Lipschitz on $B_\eta(x)$. Let $y, z \in B_\eta(x)$ be distinct points. Set $\alpha := \|z - y\|$ and $u := (z - y)/\alpha$. Then $v := z + \eta u \in B_{2\eta}(x)$. Since $z = \frac{\alpha}{\alpha + \eta}v + \frac{\eta}{\alpha + \eta}y$ we have

$$k(z) = \frac{\alpha}{\alpha + \eta}k(v) + \frac{\eta}{\alpha + \eta}k(y).$$

Therefore,

$$k(z) - k(y) \leq \frac{\alpha}{\alpha + \eta}(k(v) - k(y)) \leq \frac{\alpha}{\eta}2M = \frac{2M}{\eta}\|z - y\|.$$

Interchange of y and z gives the desired result. ■

The result remains true if we replace “interior” by “core”. Thence we may show that in every Banach space:

Theorem 1.8 *Let $k : X \rightarrow \bar{\mathbf{R}}$ be a lower semicontinuous convex function and let $x \in \text{int dom}(k)$. Then $\partial k(x)$ is a nonempty, convex, and weak-star compact subset of X^* .*

Proof (Sketch) Note that $x^* \in \partial k(x)$ amounts to say that the affine function $k(x) + \langle x^*, y - x \rangle$ supports $y \rightarrow k(y)$ at $y = x$ or the hyperplane $\{(y, r) \in X \times \mathbf{R} : r - k(x) + \langle x^*, y - x \rangle = 0\}$ supports the epigraph of k , $\text{epi}(k) := \{(y, r) \in X \times \mathbf{R} : r \geq k(y)\}$ at $(x, k(x))$. We can deduce that $\partial k(x) \neq \emptyset$ from a separation argument. It follows from Theorem 1.7 that $\partial k(x)$ is bounded. Finally, one can observe that $\partial k(x)$ is weak-star closed from the definition. ■

Let f be a locally Lipschitz function. The *right-hand directional derivative* of f at x in the direction $y \in X$ is

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

if it exists, and the *directional derivative* of f at x in the direction $y \in X$ is

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

if it exists. There is a close relationship (the *max formula*) between the subdifferential and the right-hand directional derivative of a convex function.

Theorem 1.9 *Let X be a Banach space and let U be an open set of X . Suppose that k is a convex function on U and is continuous at $x \in U$. Then, for all $y \in X$,*

$$k'_+(x; y) = \max\{\langle x^*, y \rangle : x^* \in \partial k(x)\}.$$

Proof Consider the directional difference quotient

$$(k(x + \lambda y) - k(x))/\lambda.$$

It is easy to verify that this difference quotient is a decreasing function for $\lambda > 0$ and is bounded below (because k is locally Lipschitz at x by Theorem 1.7). Thus, for any $y \in X$, $k'_+(x; y)$ exists. If $x^* \in \partial k(x)$ then, for any $y \in X$ and $\lambda \in (0, 1)$,

$$\langle x^*, y \rangle \leq (k(x + \lambda y) - k(x))/\lambda.$$

Taking limits as $\lambda \rightarrow 0^+$ we have $\langle x^*, y \rangle \leq k'_+(x; y)$.

On the other hand $k'_+(x)$ is a continuous sublinear functional, so for any $y \neq 0$ we can use the *Hahn-Banach Theorem* to find a $x^* \in X^*$ such that $\langle x^*, y \rangle = k'_+(x; y)$ and $\langle x^*, z \rangle \leq k'_+(x; z)$ for all $z \in X$ (and, therefore, $\langle x^*, z \rangle \leq k(z) - k(x)$ so that $x^* \in \partial k(x)$), which completes the proof. ■

It is possible to give an entirely algebraic proof of the above [26].

1.3 The Clarke generalized gradient

Clarke's generalized gradient is a systematic extension of the subdifferential for convex functions to the class of locally Lipschitz functions. For $X = \mathbf{R}^N$ the following theorem of Rademacher gives us a convenient and natural way to define the Clarke subdifferential.

Theorem 1.10 *Any locally Lipschitz function defined on a nonempty open subset $U \subset \mathbf{R}^N$ is (Fréchet) differentiable almost everywhere on U .*

Remark 1.11 A proof of the Rademacher's theorem can be found in [71]. The proof essentially consists of two parts: (a) show that for a locally Lipschitz function f , ∇f exists a.e. and (b) show that $f'(x, h) = \langle \nabla f(x), h \rangle$ a.e. Part (a) is easy. Part (b) is relatively hard. However, for regular functions (see below) such as smooth or convex functions, the existence of the partial derivatives automatically implies the existence of the derivative.

Rademacher's theorem tells us that the gradient of a locally Lipschitz function exists on a large set. Using this property one can define the Clarke generalized gradient of a locally Lipschitz function by

$$\partial_c f(x) = \bigcap_{\varepsilon > 0} \overline{\text{co}}\{\nabla f(y) : y \in B_\varepsilon(x) \text{ and } \nabla f(y) \text{ exists}\}.$$

Remark 1.12 Later we will see that a null set may be excluded from this definition and the derivatives in this definition can be replaced by subderivatives.

Let f be a locally Lipschitz function on an open set $U \subset X$. The *Clarke directional derivative* of f at $x \in U$ in the direction $y \in X$ is

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

and is a continuous sublinear function in y . We will show that the *Clarke subdifferential* of f at $x \in U$ can be characterized by f^0 as follows

$$\partial_c f(x) \equiv \left\{ x^* \in X^* : \langle x^*, y \rangle \leq f^0(x; y) \quad \text{for all } y \in X \right\}.$$

It follows from the sublinearity of $f^0(x; y)$ in y that $\partial f(x)$ is non-empty. It is also a weak *-compact convex subset of X^* . To establish this relationship we need Lebesgue's differentiation and mean value theorem.

Theorem 1.13 *Let f be a locally Lipschitz function on an open set $U \subset \mathbf{R}$. On any non-degenerate interval $[a, b] \subset U$, there exists a Borel set M of positive measure in $[a, b]$ such that, at each point $t \in M$, f is differentiable and*

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

Then we derive:

Theorem 1.14 [112, Lemma 2.3, p.134] *Suppose a locally Lipschitz function f on an open subset U of a Banach space X is such that given $y \in X$ there exists $S \subset U$ with*

$$E \equiv \{x \in U : \mu \{t \in \mathbf{R} : x + ty \in U \setminus S\} = 0\}$$

dense in U . Then for all $x \in U$

$$f^0(x; y) = \limsup_{z \rightarrow x} \{f'(z; y) : z \in S \cap D_y\},$$

where D_y is the points where f is differentiable in direction y .

Proof It is sufficient to show that given $\varepsilon > 0$ and $\delta > 0$

$$\sup \{f'(z; y) : \|z - x\| < \delta, z \in S \cap D_y\} > f^0(x; y) - \varepsilon$$

for all $x \in U$. Now there is a $z \in E$ and $0 < \lambda < 1$ such that $\|z + \lambda y - x\| < \delta$, $\|z - x\| < \delta$ and

$$\frac{f(z + \lambda y) - f(z)}{\lambda} > f^0(x; y) - \frac{\varepsilon}{2}.$$

By Lebesgue's Theorem 1.13, there exists a set M of positive measure in $(z, z + \lambda y) \cap D_y$ such that

$$f'(t; y) > \frac{f(z + \lambda y) - f(z)}{\lambda} - \frac{\varepsilon}{2} \quad \text{for all } t \in M.$$

Then

$$\sup \{f'(z; y) : \|z - x\| < \delta, z \in S \cap D_y\} > f^0(x; y) - \varepsilon. \quad \blacksquare$$

Another very useful pair of directional derivatives are the *upper Dini derivative* of f at x in the direction $y \in X$ defined by

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

and the *lower Dini derivative* of f at x in the direction $y \in X$ defined by

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

which are locally Lipschitz functions of y . We can immediately deduce the following characterizations of the Clarke derivative.

Corollary 1.15 *For a locally Lipschitz f on an open U in a Banach space X , given $x \in U$ and $\|y\| = 1$,*

$$\begin{aligned} f^0(x; y) &= \limsup_{z \rightarrow x} f^0(z; y) \\ &= \limsup_{z \rightarrow x} d^+ f(z; y) \\ &= \limsup_{z \rightarrow x} \{f'(z; y) : z \in D_y\} \end{aligned}$$

where D_y is the set of points in U where f is differentiable in direction y .

Theorem 1.16 [50, p.63] *For a locally Lipschitz function f on an open subset U of a Euclidean space \mathbf{R}^n , given $x \in U$, unit vector $y \in \mathbf{R}^n$ and any null subset N of \mathbf{R}^n ,*

$$f^0(x; y) = \limsup_{z \rightarrow x} \{f'(z; y) : z \in D \setminus N\},$$

where D is the set of points in U where f is differentiable. Consequently,

$$\begin{aligned} \partial_c f(x) &= \{x^* \in X^* : \langle x^*, y \rangle \leq f^0(x; y) \ \forall y \in X\} \\ &= \text{co} \{\lim f'(z_n) \ \text{where } z_n \rightarrow x, z_n \in D \setminus N\}. \end{aligned}$$

Proof It follows from Fubini's Theorem (see Corollary 1.22) that for a given unit vector $y \in \mathbf{R}^n$ the subset

$$E \equiv \{x \in U : \mu \{t \in \mathbf{R} : x + ty \in U \setminus (D \setminus N)\} = 0\}$$

is full measure in \mathbf{R}^n . So $S \equiv D \setminus N$ satisfies the hypothesis of Theorem 1.14. ■

For a continuous convex function f on an open $U \ni x$, it is well known that, for any $y \in X$,

$$f'_+(x; y) = d^+ f(x; y) = f^0(x; y).$$

Thus, by Theorem 1.9, the Clarke generalized gradient coincides with the convex subdifferential for continuous convex functions. More generally, we say f is (*subdifferentially*) *regular* at x if $d^- f(x; y) = f^0(x; y)$ for all y . If f is (pseudo-)regular at x then $f'_+(x; y) = f^0(x; y)$ is convex in y and must equal $\langle \nabla f(x), y \rangle$ whenever $\nabla f(x)$ exists.

We note that the Clarke generalized gradient is not a generalization of the usual concept of Fréchet differentiability as is clarified by the following example.

Example 1.17 Let

$$f(x) = \begin{cases} x^2 \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

Then $f'(0) = 0$ while $\partial_c f(0) = [-1, 1]$.

Clarke's generalized gradient is rather a generalization of smoothness or *strict differentiability* as we now discuss. 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,

$$\left| \frac{f(z + ty) - f(z)}{t} - \langle \nabla f(x), y \rangle \right| < \varepsilon$$

whenever $0 < t < \delta$ and $\|z - x\| < \delta$ (uniformly over $y \in S(X)$).

Then $\partial_c f(x)$ is a singleton if and only if f is strictly differentiable at x . Example 1.17 shows that in general strict differentiability differs from differentiability. For continuous convex or regular functions, however, Gâteaux differentiability coincides with strict differentiability. For continuous convex functions Fréchet differentiability and strict Fréchet differentiability coincide. Consider the following example due to Rockafellar [131].

Example 1.18 Let $E \subset \mathbf{R}$ be measurable such that $0 < \mu(E \cap I) < \mu(I)$ for every nonvoid open interval $I \subset \mathbf{R}$. Let $\chi_E, \chi_{\mathbf{R} \setminus E}$ be the characteristic functions. Define $r_E : \mathbf{R} \mapsto \mathbf{R}$ by

$$r_E(x) := \int_0^x (\chi_E - \chi_{\mathbf{R} \setminus E})(s) ds.$$

Then f is Lipschitz with Lipschitz constant 1 on \mathbf{R} and $\partial_c f(x) \equiv [-1, 1]$ for every $x \in \mathbf{R}$. We call r_E *Rockafellar's function* associated with E .

Since there are many possible ways of choosing E the Clarke generalized gradients of the r_E do not provide enough information to discriminate within this class of functions. In contrast to the subdifferential for convex functions the problem here is that the Clarke generalized gradient for r_E is too "thick" in that it is an interval $[-1, 1]$ everywhere: a subdifferential of a convex function on \mathbf{R} is single valued with the exception of at most a countable number of points. This observation and Rademacher's Theorem suggest that we should try looking for Clarke generalized gradients that are only thick on small sets.

There are many different ways to view a set as small [119]. We will discuss several possibilities in the next section. We conclude this section by stating a pleasing generalization of the Rockafellar function from Example 1.18.

Theorem 1.19 [33] *Let U be a non-empty open subset of a separable Banach space X . If $\{f_1, f_2, \dots, f_n\}$ are real-valued strictly differentiable (or equivalently, continuously Gâteaux differentiable) locally Lipschitz functions defined on U , then there exists a real-valued locally Lipschitz function g defined on U such that*

$$\partial g(x) = \text{co}\{\nabla f_1(x), \nabla f_2(x), \dots, \nabla f_n(x)\}$$

for each $x \in U$. In particular, when $\{f_1, f_2, \dots, f_n\}$ are continuous linear functions on X , then

$$\partial g(x) \equiv \text{co}\{f_1, f_2, \dots, f_n\}$$

for each $x \in U$.

This theorem is a special case of a result in [33] which provides a rich source of examples. We refer to [33] for details.

Open Question Can one provide a countable extension of this result that is strong enough to produce a function whose subdifferential is identically equal to the unit disc in \mathbf{R}^2 ? In fact by Baire category methods one can show the genericity within the non-expansive mappings of those whose Clarke subdifferentials are the entire unit ball. In a separable smooth infinite dimensional space, this is true of the approximate subdifferential. In consequence, the minimal subdifferentials form a meager set. (See [40] for details.)

1.4 Small sets

There are many different ways to identify ‘small’ sets. They all have the intuitive properties that they are preserved by inclusion, translation and countable unions. Moreover the complement of a small set is “big” – at least dense. We collect basic facts about four types of small sets.

a. Countable sets. This is the most obvious concept of a small set. It is, however, often too small to use in our context.

b. Null sets. By a *Borel measure* on a topological space X we mean any measure defined on $\mathcal{B}(X)$ – the Borel subsets of X (the σ -algebra generated by the open sets). By a *Radon measure* μ on X we mean any Borel measure on X , which satisfies; (i) $\mu(K) < \infty$ for each compact subset $K \subseteq X$; (ii) $\mu(A) = \sup\{\mu(K) : K \subseteq A, K \text{ compact}\}$ for each $A \in \mathcal{B}(X)$.

We will call a Borel subset N of a Banach space X , a *Haar-null* set if there exists a (not necessarily unique) Radon probability measure p on X , such that

$$p(x + N) = 0 \text{ for each } x \in X.$$

In such a case, we shall call the measure p a *test-measure* for N . More generally, we say that an arbitrary subset $N \subseteq X$ is a *Haar-null* set if it is contained in a Borel Haar-null set.

We show now that the Haar-null sets are closed under translation and countable unions, [29], and that if N is a Haar-null set then $X \setminus N$ is dense in X .

Proposition 1.20 *Let X be a Banach space. Then we have the following:*

- (a) *Every subset of a Haar-null set in X is Haar-null;*
- (b) *If A is a Haar-null set, then so is $x + A$, for every $x \in X$;*
- (c) *If A is a Haar-null set, then there exists a test-measure for A , with compact support;*
- (d) *If A is a Haar-null set, then $X \setminus A$ is dense in X ;*
- (e) *If $\{A_j : j \in N\}$ are Haar-null sets, then so is $\bigcup\{A_j : j \in N\}$.*

Proof The proofs of (a)-(c) are obvious and left to the reader. To prove (d) it is sufficient to show there are no (non-empty) open Haar-null sets. To this end, let U be a non-empty open subset of X and suppose, to the contrary, that there does exist a test-measure p for U on X . Let A denote the support of p . As A is separable, for some $x_0 \in X$, $(x_0 + U) \cap A \neq \emptyset$. Hence, $p(x_0 + U) \geq p((x_0 + U) \cap A) > 0$; which contradicts the fact that p is a test-measure for U .

(e) Let us first observe, that without loss of generality we may assume that each set A_j is Borel. For each $j \in N$, let p_j be a test-measure for A_j on G . Let H be the smallest closed subspace of X that contains the support of each p_j . Since the support of each p_j is separable (see, Theorem 2.1 part (a) of [29]) it is not too difficult to see that H is separable. Next, let p_j^* denote the restriction of p_j to H . We know from Theorem 1 in [46] that there exists a Radon probability measure p^* on H that is a test-measure for each set of the form $\bigcup\{B_j : j \in N\}$, provided that $B_j \in \mathcal{B}(H)$ and p_j^* is a test-measure for B_j . (The measure p^* is essentially the infinite convolution of the p_j^* .) Let p be the extension, to X , of p^* . We claim that p is a test-measure for $\bigcup\{A_j : j \in N\}$. To prove this, we must show that for each $x \in X$, $p(x + \bigcup\{A_j : j \in N\}) = 0$. So let us fix $x \in X$. Then,

$$p(x + \bigcup_{j=1}^{\infty} A_j) = p^*((x + \bigcup_{j=1}^{\infty} A_j) \cap H) = p^*(\bigcup_{j=1}^{\infty} (x + A_j) \cap H).$$

However, each p_j^* is a test-measure for $(x + A_j) \cap H$ since each p_j is a test-measure for A_j and $h + ((x + A_j) \cap H) = ((h + x) + A_j) \cap H \subseteq (h + x) + A_j$ for each $h \in H$. Therefore, p^* is a test-measure for $\bigcup\{(x + A_j) \cap H : j \in N\}$, and so $p^*(\bigcup\{(x + A_j) \cap H : j \in N\}) = 0$, which gives that $p(x + \bigcup\{A_j : j \in N\}) = 0$. ■

It is instructive, using the product-measure definition

$$\mu \otimes \nu(A) := \mu \times \nu(\{(x, y) : x + y \in A\})$$

to explicitly write down the convolution, $\mu_A \otimes \mu_B$, of two test measures for A and B respectively and to check that it is a test measure for $A \cup B$.

One of the most powerful theorems in measure theory is Fubini's theorem, and it is natural to want a version that holds for Haar-null sets. Disappointingly, in [46], Christensen gives an example to show that we must banish all hope of obtaining a *full* version of Fubini's theorem. Nonetheless, he indicates (without proof) that a weaker version of Fubini's theorem does hold (in a Polish Abelian group). We present the proof of this theorem given in [29], restricted to a Banach space setting.

Theorem 1.21 *Suppose H and T are Banach spaces with T finite dimensional. Then for each Borel subset $A \subseteq H \times T$ the following are equivalent:*

- (i) $[A]_H(h)$ is a null set, for the Lebesgue measure on T , for almost all $h \in H$;
- (ii) The set A is a Haar-null set in the product space $H \times T$.

Here $[A]_H(h) := \{t \in T : (h, t) \in A\}$.

Proof Let p_T be a Radon probability measure on T that is equivalent to Lebesgue measure on T , that is, for each $B \in \mathcal{B}(T)$, $p_T(B) = 0$ if, and only if, B is a null set for the Lebesgue measure on T . (Note: such a measure p_T exists since T , with the Lebesgue measure, is σ -finite.) Suppose that (i) holds. Let $A_H \equiv \{h \in H : \mu([A]_H(h)) > 0\}$. Let p be a test-measure for the set A_H . We claim that $p \otimes p_T$ is a test-measure for A . Fix $(h_0, t_0) \in H \times T$. Then,

$$\{h \in H : p_T([(h_0, t_0) + A]_H(h)) > 0\} = h_0 + A_H.$$

The result now follows from Theorem 2.1 part (f) in [29].

Suppose now that (ii) holds. Let δ_0 denote the Dirac (point-mass) measure at 0_H and define $\mu \equiv \delta_0 \otimes p_T$. Let p be a test-measure for the set A and set $A_H \equiv \{h \in H : p_T([A]_H(h)) > 0\}$. We know from [45] that A_H is Borel measurable, so it suffices to construct a test-measure for A_H . To this end, we define a Radon measure p_H on H by $p_H(B) \equiv p(B \times T)$ for each $B \in \mathcal{U}_R(H)$ (Note: $p_H(B) \equiv p(P^{-1}(B))$ where P is the natural projection of $H \times T$ onto H). Fix $h_0 \in H$, then,

$$0 = \mu \otimes p((h_0, 0_T) + A) = \int_{H \times T} \mu(A + (h_0 - h, -t)) dp(h, t).$$

Now, the mapping $(h, t) \rightarrow \mu(A + (h_0 - h, -t)) = p_T([A + (0_H, -t)]_H(h - h_0))$ is universally measurable and hence p -measurable (see, [45]) and strictly positive on $(h_0 + A_H) \times T$. Hence, $p_H(h_0 + A_H) = p((h_0 + A_H) \times T) = 0$. ■

We have been using the fact that, in finite dimensional spaces, the Haar-null sets coincide with the Lebesgue null sets. Indeed, in any locally compact group, the null sets for the *Haar measure* comprise the Haar-null sets (hence the name). Thus, we restate the following corollary.

Corollary 1.22 *Let X and F be Banach spaces with F finite dimensional. Let N be a Haar-null subset in $X \times F$. Then for almost every $b \in X$, that is, except for b in a Haar-null subset in X , the set*

$$N_b = \{z \in F : (b, z) \in N\}$$

is a Lebesgue-null subset of F .

When the space is infinite dimensional every compact subset is Haar-null since its span is not the entire space. (See the Appendix.) Further, we shall say that a property P holds *almost everywhere* in A if $\{t \in A : P(x) \text{ is not true}\}$ is a Haar-null set. Using this terminology Christensen provided the following generalization of Rademacher's theorem (see, [47]).

Theorem 1.23 [47, Theorem 7.5] *Assume that X is a separable Banach space. Let f be a real-valued locally Lipschitz function defined on a non-empty open subset A of X . Then there exists a Borel subset $D \subseteq A$ such that $A \setminus D$ is a Haar-null set and f is Gâteaux differentiable at each $x \in D$.*

That is, each real-valued locally Lipschitz function defined on a non-empty open subset of a separable Banach space, is Gâteaux differentiable almost everywhere (in its domain). As in the proof of Theorem 1.16 one can prove the following stronger result.

Theorem 1.24 [139, 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:*

$$\partial_c f(x) = \overline{co}^{w^*} \{x^* \in X^* : x^* = w^* - \lim_{x_n \rightarrow x} \nabla f(x_n), x_n \in D \setminus N\}.$$

We will see later 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.

Remark 1.25 The representations for the Clarke generalized gradient given in Theorems 1.16 and 1.24 rely on Rademacher's powerful theorem and its generalizations. It is shown in [23, 88] (see also generalizations in [38]) that if one uses subdifferentials instead of derivatives in the representation then a simpler variational argument suffices. In fact, this is one of the motivations that leads to the formulation of the smooth variational principle in [36].

c. First category sets. Let T be a metric space. A subset S of T is *nowhere dense* if \overline{S} has empty interior. A subset F is of *first category* provided that it is the union of countable many nowhere dense sets. It is easy to check that sets of first category are preserved by inclusion and countable unions. A set is of *second category* if it is not of first category. The space T is called a *Baire space* if for any set $F \subset T$ of first category, $T \setminus F$ is everywhere dense in T . The complement of any subset of first category is called a *residual* of T . A property P on T is said to hold *generically* if $\{t \in T : P(t) \text{ is true}\}$ is a residual set. While category is a topological concept, the following theorem due to Baire surprisingly links it to an analytic concept, the completeness of metric spaces, and is the foundation of many fundamental results in analysis.

Theorem 1.26 *Any complete metric space is a Baire space. More generally any \mathcal{G}_δ subset of a complete metric space may be given an equivalent complete metric and so is Baire.*

It is instructive to write down such a metric on $\mathbf{R} \setminus \mathbf{Q}$, the irrationals. The next theorem provides helpful characterizations of a Baire space.

Theorem 1.27 [44] *Let T be a topological space. Then the following are equivalent:*

1. T is a Baire space;
2. any countable intersection of open dense set is dense;
3. the complement of every first category \mathcal{F}_σ is a dense \mathcal{G}_δ ;
4. for any countable family of closed sets $\{F_n\}$ satisfying $T = \bigcup_{n=1}^\infty F_n$, $\bigcup_{n=1}^\infty \text{int } F_n$ is dense in T .

As an application of Baire category theory we show:

Theorem 1.28 [44] *Let T be a Baire space. Then a lower (upper) semicontinuous function on T is continuous at the points of a residual set.*

Proof Let $\{r_n : n \in \mathbf{N}\}$ be an enumeration of the rationals. Let $A_n \equiv \phi^{-1}(-\infty, r_n)$ and let $D_n \equiv A_n \setminus \overline{\text{int } A_n}$. Then D_n is nowhere dense and ϕ is u.s.c at each point of the complement of $\bigcup_n D_n$. Indeed, to see this suppose that $x \notin \bigcup_n D_n$, and $\alpha > \phi(x)$. Choose $r_n \in (\phi(x), \alpha)$. Since $x \notin D_n$ we must have $x \in \text{int } \overline{A_n} \subset \phi^{-1}(-\infty, r_n] \subset \phi^{-1}(-\infty, \alpha)$. ■

It is well-known that the property of being first category and measure zero are independent. For example, a Cantor set $C \subset [0, 1]$ with a positive measure is of first category. In fact, taking the union of a carefully chosen sequence Cantor sets with positive measure one can show that:

Theorem 1.29 [119] *The real line \mathbf{R} can be decomposed into the disjoint union of A and B such that A is of first category and B has (Lebesgue) measure zero.*

Although countable sets are both first category and measure zero, they are often too small to use. This leads to the concept of porosity. The formal definition is as follows.

d. Porosity. Let X be a metric space. Denote the open ball in X centered at x with radius r by $B_r(x)$. A subset E of X is *globally very porous* provided that there exists a positive constant $c > 0$ such that for every open ball $B_r(x)$, there is an open ball $B_{cr}(y) \subset B_r(x)$ such that $B_{cr}(y) \cap E = \emptyset$. A subset E of X is *σ -globally very porous* if it is a countable union of globally very porous sets.

Each globally porous set is clearly nowhere dense and each σ -globally very porous set is of the first category. In \mathbf{R}^N , we also have that if E is globally very porous then E has a Lebesgue measure 0. On the real line the Cantor set is porous (see Zajicek [153] and the references therein for more details regarding porosity). We will discuss an application of this concept later.

e. Angle small sets. Finally we discuss the concept of angle small sets due to Preiss and Zahiček [125]. For $x^* \in X^*$ and $\alpha \in (0, 1)$ define cone

$$K(x^*, \alpha) := \{x \in X : \alpha\|x\| \cdot \|x^*\| \leq \langle x^*, x \rangle\}.$$

A subset M of X is said to be *α -cone meager* (where $\alpha \in (0, 1)$) if for every $x \in M$ and $\varepsilon > 0$ there exists $z \in B_\varepsilon(x)$ and $0 \neq x^* \in X^*$ such that

$$M \cap [z + \text{int } K(x^*, \alpha)] = \emptyset.$$

The set M is said to be *angle small* if for every $\alpha \in (0, 1)$ it can be expressed as a countable union of α -cone meager sets.

It is easy to see that any angle small set is of first category. Moreover, an α -cone meager subset of \mathbf{R} can contain at most two points. Thus, a subset of \mathbf{R} is angle small if and only if it is countable.

2 USCOS and CUSCOS

The acronym *usco* (*cusco*) stands for a (convex) upper semicontinuous non-empty compact-valued multifunction (set-valued function). Such multifunctions are interesting because they describe common features of maximal monotone operators, of the convex subdifferential and of the Clarke generalized gradient. They are also perhaps the most natural extensions of continuous (single-valued) functions. Examination of cuscOs and uscos leads to serious insights into the underlying topological properties of the convex subdifferential and the Clarke generalized gradient. In a precise sense they reduce first order questions to zeroth order questions.

2.1 Basic properties

A set-valued mapping Φ from a topological space T into subsets of a topological (linear topological) space X is an *usco* (*cusco*) on T if

- (i) for each $t \in T$, $\Phi(t)$ is non-empty (convex) and compact;
- (ii) for each open subset $W \subseteq X$, $\{t \in T : \Phi(t) \subseteq W\}$ is open in T .

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

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

Proposition 2.1 [48, p.651] *Let Φ be an usco mapping from a topological space T into subsets of a topological space X and let Ω be a set-valued mapping from T into non-empty subsets of X . If $\text{Gr}(\Omega)$ is a closed subset of $T \times X$ and $\text{Gr}(\Omega) \subseteq \text{Gr}(\Phi)$, then Ω is an usco mapping on T .*

Note that the assumption $\text{Gr}(\Omega) \subseteq \text{Gr}(\Phi)$ in Proposition 2.1 cannot be dispensed with as shown by the following example.

Example 2.2 Consider the differential operator $\frac{d}{dt}$ from $C^1[0, 1]$ (viewed as a subspace of $C[0, 1]$ in the uniform norm) to $C[0, 1]$. This is a single valued unbounded linear operator and, therefore, it is not an usco. However, the graph of $\frac{d}{dt}$ is closed. In fact, suppose that $(x_n, \frac{dx_n}{dt}) \rightarrow (x, y)$ in $C^1[0, 1] \times C[0, 1]$. Then $x_n(t) = \int_0^t \frac{dx_n}{ds}(s)ds + x_n(0)$. Taking limits we have $x(t) = \int_0^t y(s)ds + x(0)$, i.e., $y = \frac{dx}{dt}$. ■

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

Proposition 2.3 *Let Ω be a densely defined set-valued mapping from a topological space T into subsets of a Hausdorff topological (separated locally convex topological) space X . If the graph of Ω is contained in the graph of an usco (cusco) mapping Φ , then there exists a unique smallest usco (cusco) containing Ω , denoted $\text{USC}(\Omega)$ ($\text{CSC}(\Omega)$), given by,*

$$\text{USC}(\Omega)(x) = \bigcap \{\overline{\Omega(V)} : V \text{ a nbhd. of } x\}$$

$$(\text{CSC}(\Omega)(x) = \bigcap \{\overline{\text{co}}\Omega(V) : V \text{ a nbhd. of } x\}.$$

Proof We shall only prove that $\text{CSC}(\Omega)$ is the smallest cusco containing Ω , as the proof that $\text{USC}(\Omega)$ is the smallest usco containing Ω , 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 Ψ , $\text{CSC}(\Psi)$ possesses a closed graph;
- (iii) If Ψ is a cusco then $\Psi = \text{CSC}(\Psi)$.

We now show that $\text{CSC}(\Omega)$ is a cusco mapping on T . 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 2.1, it is sufficient to show $\text{Dom}(\text{CSC}(\Omega)) = T$. Suppose not. Then there exists an element

$t_0 \notin \text{Dom}(\text{CSC}(\Omega))$. For each $x \in \Phi(t_0)$ choose open sets $U_x \subseteq T$ 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, Φ 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 $\text{Dom}(\text{CSC}(\Omega)) = T$; which shows that $\text{CSC}(\Omega)$ is a cusco on T .

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

$$\text{Gr}(\text{CSC}(\Omega)) \subseteq \text{Gr}(\text{CSC}(\Psi)) = \text{Gr}(\Psi)$$

on using (iii) and we are done. ■

Note: In the above Proposition, the set-valued mapping $\text{CSC}(\Omega)$ is called the *cusco generated* by Ω and $\text{USC}(\Omega)$ is called the *usco generated* by Ω .

Remark 2.4 (a) It follows from Proposition 2.3 that the Clarke subdifferential of a locally Lipschitz function and, in particular, the subdifferential of a convex continuous function are weak-* uscicos.

(b) Using the language here the well known Krasovskii solution [98] in the study of the differential equation $x'(t) = f(x(t))$ with discontinuous right hand side f is in fact the solution of the differential inclusion $x'(t) \in F(x(t))$ where $F(x) := \text{CSC}(f)(x)$.

2.2 Minimality

As shown by Example 1.18 the Clarke generalized gradient sometimes may be much too large. By contrast the convex subdifferential is always “minimal”. An usco (cusco) mapping Φ from a topological space T 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 T . It is immediate from this definition that all single-valued uscicos (uscicos) are minimal, however, there are many important examples of minimal uscicos (uscicos) which are not everywhere single-valued. We begin our study of minimal uscicos (minimal uscicos) by recalling some of their basic properties. A simple Zorn’s lemma argument (relying heavily on the compactness and non-emptiness of the image sets) establishes:

Proposition 2.5 [48, p.649] *Let Φ be an usco (cusco) mapping from a topological space T into subsets of a topological (linear topological) space X . Then, there exists a minimal usco (minimal cusco) Ψ defined on T such that $\Psi(t) \subseteq \Phi(t)$ for each $t \in T$.*

The following concise topological characterization of the minimality of uscós and cuscós turns out to be very useful.

Theorem 2.6 [78, Lemma 2.5] *A cusco (usco) Φ from a topological space T 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 T and closed and convex subset (closed subset) K in X , with $\Phi(U) \not\subseteq K$, there exists a non-empty open subset V of U such that $\Phi(V) \cap K = \emptyset$.*

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

Proposition 2.7 [93] *Suppose Ψ is a minimal usco and Φ is a cusco, both of which map from a topological space T into subsets of a separated locally convex topological space X .*

If $\Psi(t) \subseteq \Phi(t)$ for each $t \in T$, then the set-valued mapping $\Psi' : T \rightarrow 2^X$ defined by $\Psi'(t) \equiv \overline{\text{co}}\Psi(t)$ is a minimal cusco on T , and $\Psi'(t) \subseteq \Phi(t)$ for all $t \in T$.

Proof Let us show first that Ψ' is a cusco on T . It is easy to see that for each $t \in T$, $\Psi'(t)$ is non-empty, convex and compact. Let W be a non-empty open subset of X and consider the set $U \equiv \{t \in T : \Psi'(t) \subseteq W\}$. We may 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

$$\Psi(t_0) \subseteq \Psi'(t_0) + N \subseteq \Psi'(t_0) + \overline{N} \subseteq W.$$

Now, Ψ 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, $\Psi'(t_0) + \overline{N}$ is closed and convex and so $\Psi'(t) = \overline{\text{co}}\Psi(t) \subseteq \Psi'(t_0) + \overline{N} \subseteq W$ for each $t \in V$. Therefore $t_0 \in V \subseteq U$; which shows that Ψ' is a cusco on T . To see that Ψ' is a minimal cusco, we merely need to appeal to Theorem 2.6. ■

Remark 2.8 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 $\overline{\text{co}} \Psi(t)$, and so this condition is not needed when X is quasi-complete, as is any Banach space.

It is not too surprising that a minimal usco is “thick” only on a “small” set. We proceed now to make this precise. For this we need the central notion of a selection. Let Φ be a set-valued mapping from a non-empty set A into a non-empty set X . Then a function $\phi : A \rightarrow X$ is called a *selection of Φ* if $\phi(t) \in \Phi(t)$ for each $t \in A$. A Banach space is of *class (S)* [135] provided that every weak* usco from a Baire space into X^* has a selection which is generically weak* continuous. The relevance of class (S) spaces and the “thickness” of minimal uscós on small sets is given in the following.

Theorem 2.9 *A Banach space X is of class (S) if and only if every minimal weak* cusco or if merely every minimal locally bounded weak* cusco from a Baire space into X^* is generically single valued.*

Proof Suppose this property holds for weak* cuscus and let Φ be a weak* usco from a Baire space into X^* and let Ψ be a minimum weak* usco contained in Φ . By Proposition 2.7 Ψ' define by $\Psi'(t) := \overline{\text{co}}\Psi(t)$ is a minimum weak* cusco.

Now Ψ' is single-valued on generic subset G of X . Therefore, Ψ is single valued on G and any selection ψ of Ψ is weak* continuous at each point of G , showing that X is of class (S) . [We work a bit harder to show we need only consider locally bounded uscous.]

On the other hand suppose X is of class (S) . If Φ is a minimal weak* cusco from a Baire space into X^* then for any minimal weak* usco Ψ contained in Φ we have Ψ' define by $\Psi'(t) := \overline{\text{co}}\Psi(t)$ is a minimum weak* cusco. Thus, $\Psi' = \Phi$ and Φ is single valued whenever Ψ is. Thus, if X is of class (S) then Φ is generically single-valued. ■

Next we show, by variational methods, that any space with rotund dual renorm is of class (S) . This is the case in any separable, reflexive or WCG space [64]. In fact, the result is true – but much harder – if the space merely admits a smooth renorm [120].

Theorem 2.10 *Let X be a Banach space whose dual norm is strictly convex. Then X is of class (S) .*

Proof Let Φ be a minimal weak* cusco from a Baire space into X^* . By Theorem 2.9 it suffices to show Φ is generically single valued. Consider the *value function*

$$\phi(t) := \inf\{\|x^*\| : x^* \in \Phi(t)\}.$$

Then ϕ is lower semicontinuous and so, as we have seen in Theorem 1.28, is continuous on some generic set G . If $t \in G$ and $x^*, y^* \in \Phi(t)$ then it is not possible that $\|x^*\| = \|y^*\| = \phi(t)$ and $x^* \neq y^*$; for then $(x^* + y^*)/2 \in \Phi(t)$ and has a norm less than $\phi(t)$ by strict convexity, contradicting the definition of $\phi(t)$.

So if $\Phi(t)$ is not a singleton there exists $y^* \in \Phi(t)$ with $\|y^*\| > \phi(t)$. Choose $z \in X$ with $\|z\| = 1$ and $\langle y^*, z \rangle > \phi(t)$. Then

$$W := \{w^* \in X^* : \langle w^*, z \rangle > (\phi(t) + \langle y^*, z \rangle)/2\}$$

is a weak* open half space with $y^* \in W$. Since

$$\|w^*\| > (\phi(t) + \langle y^*, z \rangle)/2 > \phi(t)$$

for each $w^* \in W$ we see from continuity of ϕ at t that there is an open set U containing t such that for each $u \in U$ we have $\Phi(u) \setminus W \neq \emptyset$. We define

$$\Phi_1(v) := \begin{cases} \Phi(v) \setminus W & \text{if } v \in U, \\ \Phi(v) & \text{otherwise.} \end{cases}$$

We see that Φ_1 is a weak* cusco properly contained in Φ . That contradicts the minimality of Φ . ■

Remark 2.11 Levy and Poliquin's result [102] on the single-valuedness of premonotone multifunctions now follows directly. In fact, premonotonicity as defined in [102] implies local minimality which in turn implies global minimality.

A Banach space is an *Asplund space* if every separable subspace has a separable dual. Asplund spaces are discussed further in Lectures 4 and 5. In our setting a fundamental result discussed in [11], [27] is that:

Theorem 2.12 *A Banach space X is Asplund if and only if every locally bounded minimal weak* cusco from a Baire space into X^* is generically singleton and norm-continuous. A fortiori, Asplund spaces are class (S).*

Since a locally Lipschitz function on a Banach space is strictly Fréchet differentiable at a point if and only if its subgradient is single valued and norm-to-norm continuous at that point, [11], we recover the classical result that continuous convex functions on an Asplund space are generically (strictly) Fréchet differentiable.

Certain cuscoids such as maximal monotone operators on separable spaces are single-valued on sets with complements of “measure zero”. The following construction shows that general cuscoids do not share such a property.

Lemma 2.13 *Let T be a metric space and G_n a decreasing sequence of open subsets of T with dense intersection. Let*

$$f(t) := \sum_{n=1}^{\infty} 10^{-n} \sin(1/d(t, T \setminus G_n))$$

for $t \in G := \bigcap_{n=1}^{\infty} G_n$. Then

- (a) f is continuous on G ; and
- (b) no extension of f to T is continuous at any point of $T \setminus G$.

Proof (a) Fix $t \in G$ and $\varepsilon > 0$. Choose an integer N so that $10^{-N} < \varepsilon$ and $s \in G$. Then

$$\begin{aligned} |f(s) - f(t)| &\leq \sum_{n=N+1}^{\infty} 2 \cdot 10^{-n} \\ &\quad + \sum_{n=1}^N 10^{-n} |\sin(1/d(s, T \setminus G_n)) - \sin(1/d(t, T \setminus G_n))| \end{aligned}$$

which is less than ε if $|t - s|$ is small enough.

(b) We let N be the first integer such that $t \notin G_N$. Since G is dense for each $\delta > 0$ there are $r, s \in B(t, \delta) \cap G$ with

$$10^{-N} |\sin(1/d(r, T \setminus G_N)) - \sin(1/d(s, T \setminus G_N))| > 10^{-N}.$$

However,

$$\sum_{j=N+1}^{\infty} 10^{-j} |\sin(1/d(r, T \setminus G_j)) - \sin(1/d(s, T \setminus G_j))| < 2 \sum_{j=N+1}^{\infty} 10^{-j} < 10^{-N}/2$$

and

$$\sum_{j=1}^{N-1} 10^{-j} |\sin(1/d(r, T \setminus G_j)) - \sin(1/d(s, T \setminus G_j))| \rightarrow 0$$

as $\delta \rightarrow 0^+$ so that $|f(t) - f(s)| > 10^{-N}/2$ for some $r, s \in B(t, \delta) \cap G$. Thus f does not have an extension which is continuous at t . ■

Theorem 2.14 For f as in Lemma 2.13 define

$$\Phi(t) := [\liminf_{s \rightarrow t} \inf_{s \in G} f(s), \limsup_{s \rightarrow t} \sup_{s \in G} f(s)]$$

for each $t \in T$. The Φ is a minimal cusco which is single valued at t if and only if $t \in G$.

Proof It is easy to check that $\Phi : T \rightarrow \mathbf{R}$ is a cusco. If Φ contains a cusco Ψ then $\Psi(t)$ must contain $f(t)$ for each $t \in G$ as $\Phi(t) = \{f(t)\}$ for those points. It follows that $\Psi(t)$ contains $\lim_{s \rightarrow t} \inf_{s \in G} f(s)$ and $\lim_{s \rightarrow t} \sup_{s \in G} f(s)$ and by the convexity of $\Psi(t)$ we have $\Phi(t) \subset \Psi(t)$ as required. ■

We turn to discuss certain operations that preserves minimality. First we show that the minimality of a cusco (usco) mapping is preserved under composition with a continuous linear (continuous) function.

Theorem 2.15 Let Φ be a minimal cusco (minimal usco) from a topological space T 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 T .

Proof Clearly, $f \circ \Phi$ is a cusco (an usco) on T , so it remains to show that it is a minimal cusco (minimal usco) on T . Consider a closed and convex subset (closed subset) K of Y and an open set U in T such that $(f \circ \Phi)(U) \not\subseteq K$. Since f is continuous and linear (continuous) on X , $f^{-1}(K)$ is a closed and convex subset (closed subset) of X . Since Φ 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 $\Phi(V) \cap f^{-1}(K) = \emptyset$, by Theorem 2.6. Hence, $(f \circ \Phi)(V) \cap K = \emptyset$. ■

Then we have:

Theorem 2.16 Consider a minimal cusco (minimal usco) Φ from a topological space T into subsets of a separated locally convex topological (Hausdorff topological) space X .

- (i) Given a continuous real-valued function g defined on T , the set-valued mapping $g \cdot \Phi$ is a minimal cusco (minimal usco) on T .
- (ii) Given a continuous mapping f from T into X , the set-valued mapping $f + \Phi$ is a minimal cusco (minimal usco) on T .

Proof (i) In the case when Φ is a minimal usco, $g \cdot \Phi$ is the composition of the continuous mapping P , from $\mathbf{R} \times X$ into X defined by $P(t, x) = t \cdot x$ with the minimal usco mapping $t \rightarrow (g(t), \Phi(t))$ from T into $\mathbf{R} \times X$. Therefore by Theorem 2.15, $g \cdot \Phi$ is a minimal usco. In the case when Φ is a minimal cusco, consider the following. Let Ψ be a minimal usco whose graph is contained in $\text{Graph}(\Phi)$. By the above, $g \cdot \Psi$ is a minimal usco on T and $g(t)\Psi(t) \subseteq g(t)\Phi(t)$ for all $t \in A$. Now, by Proposition 2.5, $\overline{\text{co}}\Psi(t) = \Phi(t)$ for all $t \in A$. Therefore, $g(t)\Phi(t) = g(t)\overline{\text{co}}\Psi(t) = \overline{\text{co}}(g(t)\Psi(t))$ for all $t \in T$. So by again appealing to Proposition 2.5 we have that $t \rightarrow \overline{\text{co}}(g(t)\Psi(t))$ is a minimal cusco and so $g \cdot \Phi$ is a minimal cusco.

(ii) The mapping $f + \Phi$ 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 $t \rightarrow (f(t), \Phi(t))$ from T into $X \times X$, and so $f + \Phi$ is a minimal cusco (minimal usco) by Theorem 2.15. ■

Recently, the notion of minimality, for a set-valued mapping, has been successfully extended outside the class of cusco (usco) mappings, (see, for example, [77], [79], [112] and [96]). The key to these extensions is Theorem 2.6.

A set-valued mapping Φ from a topological space T 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 T 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 Φ from a topological space T into non-empty subsets of a topological space X is *minimal* if for any open set W in X and open set U in T 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 2.6, that a cusco (usco) mapping from a topological space T into subsets of a separated locally convex topological (Hausdorff topological) space X is a minimal cusco (usco) on T if, and only if, it is hyperplane minimal (minimal) on T .

Corollary 2.17 *Let Ω be a densely defined set-valued mapping from a topological space T into subsets of a separated locally convex topological (Hausdorff topological) space X .*

If the graph of Ω is contained in the graph of a cusco (usco) Φ , then $\text{CSC}(\Omega)$ ($\text{USC}(\Omega)$) is a minimal cusco (minimal usco) if, and only if, Ω is hyperplane minimal (minimal).

Proof The proof is a straight forward application of Theorem 2.6. ■

We conclude this section with several ‘useful’ characterizations of minimality.

Theorem 2.18 *For a cusco mapping Φ , from a topological space T into subsets of a separated locally convex topological space X , the following conditions are equivalent;*

- (i) Φ is a minimal cusco on T ;
- (ii) there exists a densely defined, hyperplane minimal selection σ of Φ such that $\text{CSC}(\sigma) = \Phi$;
- (iii) for any densely defined selection σ of Φ , $\text{CSC}(\sigma) = \Phi$;
- (iv) there exists a densely defined selection σ of Φ such that $\Phi = \text{CSC}(\sigma|_D)$ for each dense subset D of $\text{Dom}(\sigma)$.

Proof Corollary 2.17 gives us that (i) \Leftrightarrow (ii) and clearly (i) \Rightarrow (iii) and (iii) \Rightarrow (iv). So it remains to show that (iv) \Rightarrow (i).

We proceed via the characterization given in Theorem 2.6. To this end, let U be a non-empty open subset of T and suppose that $\Phi(U) \not\subseteq K$, where $K \equiv \{x \in X : f(x) \leq \alpha\}$, $\alpha \in \mathbf{R}$ and $f \in X^*$. Choose $x_0 \in \Phi(U) \setminus K$ such that $f(x_0) > \alpha + \varepsilon$, for some $\varepsilon > 0$ and set $D' \equiv \{t \in \text{Dom}(\sigma) \cap U : f(\sigma(t)) \leq \alpha + \varepsilon\}$. Clearly D' is not dense in U , because if D' were dense in U then by hypothesis $\Phi = \text{CSC}(\sigma|_{D^*})$ where $D^* \equiv D' \cup \text{Dom}(\sigma) \setminus U$, and this would imply that $\sup\{f(x) : x \in \Phi(U)\} \leq \alpha + \varepsilon$; 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 $t \in V \cap \text{Dom}(\sigma)$, $f(\sigma(t)) > \alpha + \varepsilon$, therefore,

$$\Phi(V) \cap K = \text{CSC}(\sigma|_{\text{Dom}(\sigma)})(V) \cap K = \emptyset. \quad \blacksquare$$

2.3 Maximal monotones and minimal cuscos

A multifunction $F : X \rightarrow 2^{X^*}$ is a *monotone operator* provided

$$\langle y^* - x^*, y - x \rangle \geq 0$$

for any pairs (x, x^*) and (y, y^*) in the graph of F . It is said to be *maximal monotone* provided that it is maximal in the family of monotone multifunctions ordered by the inclusion of their graphs. Again, by Zorn's lemma, every monotone mapping admits maximal extensions. By appropriately closing and convexifying a monotone operator, is easy to check that maximal monotone operators have convex images and norm- w^* closed graphs. Maximality with respect to an open $U \subset X$ is similarly defined. The subdifferential of a convex function provides a central example.

Theorem 2.19 *Let k be a continuous convex function on a Banach space X . Then the subdifferential ∂k of k is a maximal monotone operator.*

Proof It is easy to check that ∂k is monotone. To show it is maximal it suffices to show that whenever $y \in X$ and $y^* \in X^*$ are such that $y^* \notin \partial k(y)$, then there exists $x \in X$ and $x^* \in \partial k(x)$ such that $\langle y^* - x^*, y - x \rangle < 0$.

Replacing k by $k_1(x) := k(x - y) - \langle y^*, x \rangle$ we may assume that $y = 0$ and $y^* = 0$. Then 0 is not a minimum for k , so there exists a point $z \in X$ such that $k(z) < k(0)$. Consider the convex function $h(t) := k(tz)$, $0 \leq t \leq 1$. Its right-hand derivative at a point s is clearly equal $d^+ k(s; z)$ which is necessarily negative for some $\bar{t} \in (0, 1)$ (otherwise h would be an increasing function contradicting $h(0) > h(1)$). Letting $x = \bar{t}z$ we have $d^+ f(x; x) < 0$. Then, by [120, Proposition 2.24], there exists $x^* \in \partial f(x)$ such that $\langle x^*, x \rangle = d^+ f(x; x) < 0$, which completes the proof. ■

In fact, Theorem 2.19 is a special case of the following more general theorem due to Rockafellar, a short proof of which may be found in [134].

Theorem 2.20 *Let f be a proper lower semicontinuous convex function on a Banach space X . Then the subdifferential ∂f of f is a maximal monotone operator.*

This can go badly awry in an incomplete normed space.

We have seen that a convex function is locally Lipschitz in the interior of its domain and, therefore, the subdifferential is locally bounded. This is a special case of the more general fact that a monotone operator is locally bounded in the interior of its domain. We say F is *locally bounded* at $x \in \text{dom } F$ if there exist $M > 0$ and $\eta > 0$ such that $\|y^*\| \leq M$ whenever $y \in (x + \eta B_X) \cap \text{dom } F$ and $y^* \in F(y)$.

It is often possible to derive information about monotone operators from variational or convex analysis as the next result illustrates.

Theorem 2.21 (Boundedness of Monotone Operators) *Let $F : X \rightarrow 2^{X^*}$ be a monotone operator. Suppose that $x \in \text{int}(\text{dom } F)$. Then F is locally bounded at x .*

Proof By choosing any $x^* \in F(x)$ and replacing F by the monotone operator $y \rightarrow F(y + x) - x^*$, we lose no generality in assuming that $x = 0$ and that $0 \in F(0)$. Define, for $x \in X$,

$$k(x) := \sup\{\langle y^*, x - y \rangle : y \in \text{dom } F, \|y\| \leq 1, y^* \in F(y)\}.$$

As the supremum of affine continuous functions, k , is convex and lower semicontinuous. We show that $\text{dom } f$ is an absorbing set. First since $0 \in F(0)$, we must have $k \geq 0$. Second, whenever $y \in \text{dom } F$ and $y^* \in F(y)$, monotonicity implies that

$$0 \leq \langle y^* - 0, y - 0 \rangle,$$

so $k(0) \leq 0$. Thus, $k(0) = 0$. Suppose $x \in X$. By hypothesis, $\text{dom } F$ is absorbing so there exists $t > 0$ such that $F(tx) \neq \emptyset$. Choose any element $u^* \in F(tx)$. If $y \in \text{dom } F$ and $y^* \in F(y)$, then by monotonicity

$$\langle y^*, tx - y \rangle \leq \langle u^*, tx - y \rangle.$$

Consequently,

$$\begin{aligned} k(tx) &\leq \sup\{\langle u^*, tx - y \rangle : y \in \text{dom } F, \|y\| \leq 1\} \\ &< \langle u^*, tx \rangle + \|u^*\| < +\infty. \end{aligned}$$

By virtue of Theorem 1.7, k is continuous at 0 and hence there exists $\eta > 0$ such that $k(x) < 1$ for all $x \in 2\eta B_X$. Equivalently, if $x \in 2\eta B_X$, then $\langle y^*, x \rangle \leq \langle y^*, y \rangle + 1$ whenever $y \in \text{dom } F$, $\|y\| \leq 1$ and $y^* \in F(y)$. Thus, if $y \in \eta B_X \cap \text{dom } F$ and $y^* \in F(y)$, then

$$\begin{aligned} 2\eta\|y^*\| &= \sup\{\langle y^*, x \rangle : x \in 2\eta B_X\} \\ &\leq \|y^*\| \cdot \|y\| + 1 \leq \eta\|y^*\| + 1, \end{aligned}$$

so $\|y^*\| \leq 1/\eta$. ■

Remark 2.22 Note that Theorem 2.21 does not require that the domain of F be convex. Moreover, we can replace $x \in \text{int}(\text{dom } F)$ in Theorem 2.21 by $x \in \text{core}(\text{dom } F)$ with a similar proof.

There are trivial examples which show that 0 can be an absorbing point of $\text{dom } F$ but not an interior point (see [120]). The best development is to be found in Simons [134] where more complex ways of attaching a convex function to a multifunction are exploited beautifully.

The following theorem makes the link between maximal monotone multifunctions and minimal cuscus.

Theorem 2.23 *Let U be an open set in the Banach space X and let Φ be a maximal monotone multifunction from U to X^* . Then Φ is a minimal w^* -cusco.*

Proof Since Φ is locally bounded and has a norm- w^* closed graph with convex images, Proposition 2.1 shows Φ is a w^* -cusco; so the only question is whether it is minimal in this family. Suppose that $\Psi : U \rightarrow 2^{X^*}$ is a w^* -cusco and that $\text{graph } \Psi \subset \text{graph } \Phi$. We show that Ψ must be maximal monotone. Observe that Ψ is monotone. It suffices to show that if $(y, y^*) \in X \times X^*$ satisfies

$$\langle y^* - x^*, y - x \rangle \geq 0 \quad \forall x \in U, x^* \in \psi(x), \tag{1}$$

then $y^* \in \psi(y)$. If not, by the separation theorem there exists $z \in X$ such that $F(y) \subset W$ where

$$W := \{z^* \in X^* : \langle z^*, z \rangle < \langle y^*, z \rangle\}.$$

Since W is w^* open and Ψ is norm to weak star upper semicontinuous, there exists a neighborhood V of y in U such that $\Psi(V) \subset W$. For $t > 0$ sufficiently small, $y + tz \in V$ and therefore $\Psi(y + tz) \subset W$. Applying (1) to any $u^* \in \Psi(y + tz)$ we get

$$0 \leq \langle y^* - u^*, y - (y + tz) \rangle = -t \langle y^* - u^*, z \rangle,$$

which implies that $\langle u^*, z \rangle \geq \langle y^*, z \rangle$, that is, u^* is not in W , a contradiction. \blacksquare

It follows that *any* monotone w^* -cusco must be maximal monotone since it lies in a maximal extension which is minimal! It also follows directly that:

Corollary 2.24 *Let U be an open set in the class (S) Banach space X and let Φ be a maximal monotone multifunction from U to X^* . Then Φ is generically single-valued.*

In the language of Section 1.4.e., a lovely result specific to monotone operators and convex functions is:

Theorem 2.25 [125] *Suppose that the Banach space X has separable dual and that $\Phi : X \rightarrow 2^{X^*}$ is monotone. Then there exists an angle small set $A \subset \text{dom}(T)$ such that T is single-valued and norm-to-norm upper semicontinuous at each point of $\text{dom}(T) \setminus A$.*

2.4 Minimality of the Clarke generalized gradient

We begin this portion 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 previous lectures to establish further properties enjoyed by those locally Lipschitz functions whose subdifferential mappings are minimal. We recapitulate a well-known characterization of the Clarke generalized directional derivative contained in Corollary 1.16.

Proposition 2.26 *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 $y \in X$,*

$$f^0(x; y) = \limsup_{z \rightarrow x} d^+ f(z; y) = \limsup_{z \rightarrow x} d^- f(z; y).$$

In order to expedite the rest of this lecture we will introduce the following definition. Let A be a non-empty 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 $x \in A$

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

(Here λ represents Lebesgue measure on \mathbf{R} .)

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

Proposition 2.27 *Let f be a locally Lipschitz function defined on a non-empty open subset A of a Banach space X . Then for each $y \in S(X)$,*

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

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

Theorem 2.28 [112, 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, $x \rightarrow \partial f(x)$, is a minimal weak* cusco on A if, and only if, for each $y \in S(X)$, one of the following conditions holds.*

- (i) *The mapping $T_y : A \rightarrow 2^{\mathbf{R}}$ defined by $T_y(x) = \{\phi(y) : \phi \in \partial f(x)\}$ is a minimal cusco.*
- (ii) *The function $D_y : A \rightarrow \mathbf{R}$ defined by $D_y(x) = d^+ f(x; y)$ 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 2.28. Let φ be a real-valued function defined on a topological space T . Then φ is *quasi lower semi-continuous* (*quasi upper semi-continuous*) on T if for each $t_0 \in T$, $\varepsilon > 0$ and open neighbourhood U of t_0 there exists a non-empty open subset V of U such that

$$\begin{aligned} \inf\{\varphi(t) : t \in V\} &> \varphi(t_0) - \varepsilon \\ (\sup\{\varphi(t) : t \in V\} &< \varphi(t_0) + \varepsilon), \end{aligned}$$

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

Theorem 2.29 *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 $D_y : P_y \rightarrow \mathbf{R}$ defined by $D_y(x) \equiv d^+ f(x; y)$ 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 2.28 part(ii) we have that the mapping, $x \rightarrow d^+ f(x; y)$, is hyperplane minimal on A and so 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 $D_y : P_y \rightarrow \mathbf{R}$ defined by $D_y(x) \equiv d^+ f(x; y)$ is quasi lower semi-continuous (quasi upper semi-continuous) on P_y .

Fix $y \in S(X)$. We will show there exists a Borel subset R_y of A , which is 1- D almost everywhere in A , in direction y , such that the mapping, $x \rightarrow d^+ f(x; y)$, is hyperplane minimal on 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

$$d^+ f(x; y) = -f'(x; -y) = -d^+ f(x; -y)$$

and so the mapping D_y , restricted to R_y , is both quasi upper and lower semi-continuous on R_y (i.e., D_y is hyperplane minimal on R_y), which completes the proof (via, Theorem 2.28 part (iii)). ■

An especially useful consequence is:

Theorem 2.30 *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 $A \setminus M$.*

Proof It follows directly from Theorem 2.6 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 2.29. To this end, fix $y \in S(X)$ and let $P_y \equiv \{x \in A : f'(x; y) \text{ exists}\}$. By Proposition 2.27, P_y is 1- D almost everywhere in A , in the direction y . We will show that the mapping $D_y : P_y \rightarrow \mathbf{R}$ defined by $D_y(x) \equiv f'(x; y) = d^+ f(x; y)$ 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 2.28 part(ii)). So we consider the case when x_0 is in 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 $V \subseteq B(x_0 + t_0 y, r) \subseteq U$ 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 $s \equiv \max\{t \in [0, 1) : x_0 + ty' \in M\}$. 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 $s_0 > s$, $x_0 + s_0 y' \notin M$. 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 $d^+ f(z; y') > 0$ for each $z \in V$, and by positive homogeneity, $d^+ f(z; t_0^{-1} y') > 0$ for each $z \in V$. However, by our choice of y' ,

$$\|t_0 y - y'\| < r < \varepsilon t_0 / K$$

and so $D_y(z) =$

$$\begin{aligned} d^+ f(z; y) &= d^+ f(z; t_0^{-1} y') + (d^+ f(z; y) - d^+ f(z; t_0^{-1} y')) \\ &\geq d^+ f(z; t_0^{-1} y') - \varepsilon > -\varepsilon \end{aligned}$$

for each $z \in V \cap P_y$. This ends the proof. ■

By far and away the most important application of Theorem 2.30 is to distance functions. Let C be a non-empty closed subset of a Banach space $(X, \|\cdot\|)$. Again recall that the *distance function* associated with the set C and the norm $\|\cdot\|$, (denoted by d_C), is defined by, $d_C(x) \equiv \inf\{\|x - c\| : c \in C\}$. We may now obtain a notable fact concerning the minimality of the Clarke subdifferential mapping of a distance function.

Theorem 2.31 *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, $x \rightarrow \partial_c d_C(x)$, is a minimal weak* cusco on $X \setminus C$.*

2.5 Subgradient representation of multifunctions

There has been some considerable work on the subgradient representation of multifunctions. An early result is due to Rockafellar, who proved that a multifunction is the subdifferential of a lower semicontinuous proper convex function if and only if the multifunction is *maximal cyclically monotone*; see [120]. Janin further showed in [92] that a multifunction is *cyclically submonotone* if and only if this multifunction is the Clarke subdifferential mapping of a lower- C^1 (locally Lipschitz) function in the sense of Rockafellar [129]. Recently Poliquin [123] proved that a multifunction is the proximal subdifferential mapping of a lower semicontinuous function bounded below by a quadratic if and only if it satisfies a monotone selection property. Finally, in one dimension, a full characterization of when a multifunction is the Clarke subgradient of a Lipschitz function may be found in [17].

In general, the problem of how to characterize the multifunction given by a Clarke generalized gradient is still open. However, for a given minimal weak* cusco a nice characterization in terms of line integrals (as in the classical Green's theorem) was derived in [32]. To state this result we need some preparation.

We define the *line integral* on $[a, b]$ of a single-valued mapping $\sigma : X \rightarrow X^*$, $\int_{[a,b]} \sigma(z) dz$, as the Lebesgue integral

$$\int_0^1 \langle \sigma(tb + (1-t)a), (b-a) \rangle dt,$$

where we implicitly assume that the function $T : [0, 1] \rightarrow \mathbf{R}$, defined by $T(t) \equiv \langle \sigma(tb + (1-t)a), (b-a) \rangle$, is Lebesgue-measurable. It can easily be seen that

$$\int_{[a,b]} \sigma(z) dz = - \int_{[b,a]} \sigma(z) dz.$$

A *polygonal path* C in X is an ordered collection of line segments $\{[a_i, a_{i+1}] : 1 \leq i \leq n-1\}$ for some positive integer n . Such a path is said to be closed when $a_1 = a_n$. We write $-C$ to denote the ordered collection of line segments $\{[a_{n-i+1}, a_{n-i}] : 1 \leq i \leq n-1\}$. The line integral $\int_C \sigma(z) dz$ on a polygonal path C is defined as

$$\int_C \sigma(z) dz \equiv \sum_{i=1}^{n-1} \int_{[a_i, a_{i+1}]} \sigma(z) dz. \tag{2}$$

Recall that a multifunction $\Phi : X \rightarrow 2^{X^*}$ is *locally bounded* if for each $x_0 \in X$ there exists a positive number L and a neighbourhood U of x_0 such that

$$\|y^*\| \leq L \tag{3}$$

for all $y^* \in \Phi(u)$ with $u \in U$. Moreover, Φ is said to be *bounded by L* if inequality (3) is satisfied by all y^* in the image of Φ .

We first provide in an arbitrary Banach space a sufficient condition for a minimal weak* cusco to be the Clarke subdifferential mapping of a Lipschitz function.

Theorem 2.32 *Assume that X is a Banach space and A is a non-empty connected open subset of X . Let $\Phi : A \rightarrow 2^{X^*}$ be a locally bounded weak* cusco. Suppose Φ possesses a selection $\sigma : A \rightarrow X^*$ such that*

$$\oint_C \sigma(z) dz \leq 0 \tag{4}$$

for every closed polygonal path C in A . Then there is some locally Lipschitz function f on A with $\partial f \subseteq \Phi$. Moreover, if Φ is minimal, then $\Phi = \partial f$.

Proof Let us note first that open connected sets are actually polygonally connected. We define $f : A \rightarrow \mathbf{R}$ by

$$f(x) \equiv \int_{\Gamma} \sigma(z) dz,$$

where Γ is some polygonal path in A from a given point $a \in A$ to $x \in A$. Observe that (4) holding for each closed polygonal path C in A implies

$$\oint_C \sigma(z) dz = 0$$

for every closed polygonal path C in A . As in the classical case, we see that f is well defined. Let us now verify that f is locally Lipschitz on A . Given any $x \in A$, since Φ is locally bounded, there exists a positive number L such that $\Phi(U) \subseteq LB(X^*)$ for some convex neighbourhood U of x contained in A . Thus, for any $y_1, y_2 \in U$, one has

$$\begin{aligned} |f(y_2) - f(y_1)| &= \left| \int_{[y_1, y_2]} \sigma(z) dz \right| \\ &= \left| \int_0^1 \langle \sigma(ty_2 + (1-t)y_1), y_2 - y_1 \rangle dt \right| \\ &\leq \int_0^1 |\langle \sigma(ty_2 + (1-t)y_1), y_2 - y_1 \rangle| dt \\ &\leq L \|y_2 - y_1\|. \end{aligned}$$

Finally, we show that $\partial f(z) \subseteq \Phi(z)$ for any $z \in X$. We assume, to obtain a contradiction, that for some z , $\partial f(z) \not\subseteq \Phi(z)$. Then we may find an $x^* \in \partial f(z)$ and an $x \in S(X)$ such that $\langle x^*, x \rangle > \alpha > \max \langle \Phi(z), x \rangle$. This implies $f^0(z; x) > \alpha > \max \langle \Phi(t), x \rangle$. By the upper semicontinuity of Φ , we can select a convex neighbourhood V of z such that $\max \langle \Phi(V), x \rangle < \alpha$. On the other hand, we may select $v \in V$ such that $f'(v; x) > \alpha$. Hence there exists a positive number ϵ such that

$$\frac{f(v + \mu x) - f(v)}{\mu} > \alpha \quad \text{if } |\mu| < \epsilon.$$

The definition of f allows us to conclude

$$f(v + \epsilon x) - f(v) = \int_{[v, v + \epsilon x]} \sigma(z) dz.$$

Then, employing the mean value theorem and Lebesgue's differentiation theorem, we obtain $\sigma(w)(\epsilon x) > \alpha \epsilon$ for some $w \in [v, v + \epsilon x]$, which gives $\sigma(w)(x) > \alpha$. This is impossible since we have $w \in V$ and $\sigma(w) \in \Phi(w) \subseteq \Phi(V)$. In consequence we have $\partial f(z) \subseteq \Phi(z)$ for all $z \in A$.

It is now immediate that $\partial f = \Phi$ whenever Φ is minimal, since ∂f is a weak* cusco. ■

We illustrate the application of Theorem 2.32 by recapturing the Hilbert space case of Rockafellar's cyclic monotonicity theorem [120] as follows. Recall that a multifunction $T : X \rightarrow 2^{X^*}$ is *cyclically monotone* if given $(x_i, x_i^*) \in \text{Gr}(T)$ ($i = 0, 1, 2, \dots, n$, where n is arbitrary), we have

$$\langle x_0^*, x_1 - x_0 \rangle + \langle x_1^*, x_2 - x_1 \rangle + \dots + \langle x_n^*, x_0 - x_n \rangle \leq 0.$$

Corollary 2.33 *Let H be a Hilbert space and $T : H \rightarrow 2^H$ be maximal and cyclically monotone. Then there exists a proper closed convex function f defined on E such that $T = \partial f$.*

Proof (Special case: $D(T) = H$) [In this case the reasoning applies in *arbitrary Banach spaces* and the function is continuous.] Recall that T is a locally bounded minimal weak* cusco on H . Let σ be any selection of T . We claim that $\int_C \sigma(x) dz \leq 0$ for any closed polygonal path C . Notice that for $[a, b] \subset H$ the function $t \rightarrow \sigma(tb + (1 - t)a)(b - a)$ is monotone, and hence Riemann integrable. Also,

$$\begin{aligned} \int_{[a,b]} \sigma(z) dz &= \sup_{0=t_1 < t_2 < \dots < t_n=1} \left\{ \sum_{i=1}^n \sigma(t_{i-1}b + (1 - t_{i-1})a)[(b - a)(t_i - t_{i-1})] \right\} \\ &\leq \sup \left\{ \sum_{i=1}^n \sigma(x_{i-1})(x_i - x_{i-1}) \right\} \end{aligned}$$

over polygonal arcs $P := \{[x_{i-1}, x_i] : 1 \leq i \leq n\}$ from a to b . Hence $\int_C \sigma(z) dz \leq 0$ for any closed polygonal path C in H because

$$\sigma(x_n)(x_0 - x_n) + \sigma(x_{n-1})(x_n - x_{n-1}) + \dots + \sigma(x_0)(x_1 - x_0) \leq 0.$$

for any finite set of points $\{x_0, x_1, \dots, x_n\} \subset H$. We see from Theorem 2.1 that the conclusion holds whenever $D(T) = H$.

(General Case) Since T is maximal monotone, so is T^{-1} . Using Minty's theorem [111] we then know that the range $R(T^{-1} + I)$ is the whole space H . This means that the domain $D(T_1)$ of the resolvent $T_1 = (T^{-1} + I)^{-1}$ is H . Observe that the cyclical monotonicity of T^{-1} follows from that of T . This implies the cyclical monotonicity of $T^{-1} + I$ and then that of T_1 . Thus applying the assertion just proved above for the special case that $D(T) = H$, we obtain a locally Lipschitz function g such that $T_1 = \partial g$. By virtue of Theorem 5 of [58] we deduce that g is convex from the monotonicity of T_1 . We thus infer that $T_1^{-1} = \partial g^*$, where g^* is the conjugate of g . This allows us to conclude that

$$T^{-1} = \partial g^* - I = \partial(g^* - \|\cdot\|^2/2)$$

by virtue of the subgradient sum rule in non-smooth analysis. We again appeal to Theorem 5 of [58] to see that $h^* = g^* - \|\cdot\|^2/2$ is convex since T^{-1} is monotone. We have determined that $T = \partial h$ with h given as the conjugate of h^* . This ends the proof of the Corollary. ■

Now we turn to characterize, in line integral terms, when a minimal cusco is the Clarke subdifferential mapping of a Lipschitz function. This requires some further discuss polygonal paths. Let X be a Banach space, and A be a nonempty open connected subset of X . Suppose also that $B \subseteq A$ is a Borel set. For a fixed $\varepsilon > 0$ we will call an ordered collection of line segments $P(\varepsilon) \equiv \{[a_i, b_i] \mid 1 \leq i \leq n - 1\}$ an ε -path from a to b provided that

$$\|a - a_1\| + \sum_{i=1}^{n-1} \|a_{i+1} - b_i\| + \|b_n - b\| < \varepsilon.$$

Such a path is said to be *closed* if $a = b$. Moreover, we write $-P(\varepsilon) = \{[b_i, a_i] \mid 1 \leq i \leq n - 1\}$. Furthermore, we say that P is a *B-admissible ε -path* from a to b if P is an ε -path from a to b and

$$\lambda(\{t \in [0, 1] : tb_i + (1 - t)a_i \notin B\}) = 0$$

for each $1 \leq i \leq n-1$. Line integrals on a ε -path are defined similarly to (2). Again, ‘ λ ’ denotes Lebesgue measure on the line.

We are now ready for our main result.

Theorem 2.34 *Let X be a separable Banach space, and let A be a non-empty open connected subset of X . Then for a minimal weak* cusco $\Phi : A \rightarrow 2^{X^*}$ to be the Clarke subdifferential mapping of a locally Lipschitz function (with uniform rank L) it is necessary and sufficient that (i) Φ be bounded by L and (ii) that there exist a Borel set $B \subset A$ with $A \setminus B$ Haar-null and a measurable selection $\sigma : B \rightarrow X^*$ such that for each $\varepsilon > 0$ and each closed B -admissible ε -path $P(\varepsilon)$ in A , one has*

$$\left| \oint_{P(\varepsilon)} \sigma(x) dx \right| < L\varepsilon. \quad (5)$$

Proof (a) “only if”. Let us assume that $\Phi = \partial f$ for some Lipschitz function f with rank L . Then it is evident that Φ is bounded by L . Now we verify (5) for every $\varepsilon > 0$ and each closed B -admissible ε -path $P(\varepsilon)$ in A . Letting

$$B \equiv \{x \in A : f \text{ is G\^ateaux differentiable at } x\},$$

we know from Theorem 1.23 that the set $N \equiv A \setminus B$ is Haar-null. We define $\sigma : B \rightarrow X^*$ by $\sigma(x) = \nabla f(x)$. It follows that the mapping σ is Borel measurable. Given any $\varepsilon > 0$, let $P = \{[a_i, b_i] : 1 \leq i \leq n\}$ be a closed B -admissible ε -path in A . Then by virtue of Lebesgue’s differentiation theorem, we obtain

$$\begin{aligned} f(b_i) - f(a_i) &= \int_0^1 \langle \nabla f(tb_i + (1-t)a_i), b_i - a_i \rangle dt \\ &= \int_{[a_i, b_i]} \sigma(z) dz. \end{aligned}$$

We may now derive (5) from the following chain of the inequalities:

$$\begin{aligned} \left| \int_{P(\varepsilon)} \sigma(x) dx \right| &= \left| \sum_{i=1}^n f(b_i) - f(a_i) \right| \\ &\leq |f(a_1) - f(b_n)| + \sum_{i=1}^{n-1} |f(a_{i+1}) - f(b_i)| \\ &\leq L\|a_1 - b_n\| + \sum_{i=1}^{n-1} L\|a_{i+1} - b_i\| < L\varepsilon. \end{aligned}$$

(b) “if”. Let us now prove the converse. We first define the function

$$f(x) \equiv \lim_{\varepsilon \rightarrow 0^+} \int_{P(\varepsilon)} \sigma(z) dz \quad \text{for all } x \in A, \quad (6)$$

where $P(\varepsilon)$ is any B -admissible ε -path in A from some fixed $a \in A$ to x . Then f is well defined. Indeed, for any $\varepsilon_1, \varepsilon_2 > 0$, let $P_i(\varepsilon_i)$ be a B -admissible ε_i -paths from a to x , $i = 1, 2$.

Then $P_1(\varepsilon_1) + (-P_2(\varepsilon_2))$ is a closed B -admissible $(\varepsilon_1 + \varepsilon_2)$ -path. By assumption we then deduce

$$\left| \int_{P_1(\varepsilon_1)} \sigma(z) dz - \int_{P_2(\varepsilon_2)} \sigma(z) dz \right| = \left| \int_{P_1(\varepsilon_1) + (-P_2(\varepsilon_2))} \sigma(z) dz \right| < L(\varepsilon_1 + \varepsilon_2).$$

Thus we see the limit on the right of (6) does exist, and f is well defined, as claimed.

Next let us verify that the function f is locally Lipschitz with rank L . Let x_0 be any point in A and let U be any convex neighborhood of x_0 contained in A . Consider any two points $x, y \in U$ and fix $\delta > 0$. By the definition of f we may choose $\delta > \varepsilon_0 > 0$ sufficiently small so that if $P(\varepsilon)$ is any B -admissible ε -path ($0 < \varepsilon < \varepsilon_0$) from a to x , then

$$\left| f(x) - \int_{P(\varepsilon)} \sigma(z) dz \right| < \delta,$$

and if $P'(\varepsilon)$ is any B -admissible ε -path from a to y , then one has

$$\left| f(y) - \int_{P'(\varepsilon)} \sigma(z) dz \right| < \delta.$$

Now suppose $P(\varepsilon) = \{[a_i, b_i] \mid 1 \leq i \leq n\}$ is a B -admissible ε -path from a to x (with $0 < \varepsilon < \varepsilon_0$). By Corollary 1.22 we may extend $P(\varepsilon)$ to a B -admissible ε -path ($P'(\varepsilon)$, say) from a to y , where $P'(\varepsilon) = \{[a_i, b_i] : 1 \leq i \leq n+1\}$. It follows that

$$\begin{aligned} |f(y) - f(x)| &\leq \left| f(y) - \int_{P'(\varepsilon)} \sigma(z) dz \right| \\ &\quad + \left| \int_{P'(\varepsilon)} \sigma(z) dz - \int_{P(\varepsilon)} \sigma(z) dz \right| \\ &\quad + \left| \int_{P(\varepsilon)} \sigma(z) dz - f(x) \right| \\ &\leq \delta + \left| \int_{[a_{n+1}, b_{n+1}]} \sigma(z) dz \right| + \delta \\ &\leq L \|a_{n+1} - b_{n+1}\| + 2\delta \\ &\leq L(\|a_{n+1} - x\| + \|x - y\| + \|y - b_{n+1}\|) \\ &\quad + 2\delta \\ &\leq L\|x - y\| + 2(L+1)\delta. \end{aligned}$$

This gives $|f(y) - f(x)| \leq L\|x - y\|$ as δ is arbitrary.

Finally, let us show that $\partial f(x) \subseteq \Phi(x)$ for all $x \in A$. Define

$$D \equiv \{x \in X : f \text{ is G\^ateaux differentiable at } x\}.$$

We know from Theorem 1.23 that $X \setminus D$ is Haar-null. So in view of Theorem 1.24 it is sufficient to show that $\nabla f(x) \in \Phi(x)$ for all $x \in D$. Assume, for the purpose of obtaining contradiction, that $\nabla f(x) \notin \Phi(x)$ for some $x \in D$. Thus, by the Mazur separation theorem, we can choose an $\alpha \in \mathbf{R}$ and a $y \in S(X)$ such that

$$\langle \nabla f(x), y \rangle > \alpha > \max\langle \Phi(x), y \rangle.$$

Using the upper semicontinuity of Φ , we can find a convex neighbourhood U of x such that $\sup\langle\Phi(U), y\rangle < \alpha$. Moreover, since f is Gâteaux differentiable at x , we may select an $\varepsilon > 0$ such that

$$\frac{f(x + \varepsilon y) - f(x)}{\varepsilon} > \alpha.$$

Let $G \equiv D \cap B$. Then we know from Proposition 3.1(iv) that $A \setminus G$ is also Haar-null. By Corollary 1.22 we can find $v \in X$ close to x such that $[v, v + \varepsilon y]$ is G -admissible and

$$\frac{f(v + \varepsilon y) - f(v)}{\varepsilon} > \alpha.$$

Moreover, we can assume without loss of generality that $[v, v + \varepsilon y] \subset U$. From the definition of f we may deduce the equality

$$f(v + \varepsilon y) - f(v) = \int_{[v, v + \varepsilon y]} \sigma(z) dz.$$

Thus, it follows from Lebesgue's differentiation theorem that there exists a $z \in [v, v + \varepsilon y] \cap G$ such that $\langle\sigma(z), \varepsilon y\rangle > \alpha\varepsilon$, which immediately gives us that $\langle\sigma(z), y\rangle > \alpha$. This is a contradiction, since $z \in U$ and $\sigma(u) \in \Phi(u) \subseteq \Phi(U)$, while $\max\langle\Phi(U), y\rangle < \alpha$. We consequently have $\partial f \subseteq \Phi$, and then $\partial f = \Phi$ due to the minimality of Φ . ■

We may now obtain more precise generalizations of Green's Theorem. It is also possible to recast the last theorem so as to reduce the explicit dependence on minimality (X. Wang, private communication).

3 Partially smooth variational principles

Differentiability issues become subtle in infinite dimensional spaces. Besides Fréchet and Gâteaux differentiability there are many other useful variants. A convenient uniform treatment of many differentiability notions is via the concept of bornology. A *bornology* β of X is a family of closed bounded and centrally symmetric subsets of X whose union is X , which is closed under multiplication by scalars and is directed upwards (that is, the union of any two members of β is contained in some member of β). We will denote by X_β^* the dual space of X endowed with the topology of uniform convergence on β -sets.

The most important bornologies are those formed by all (symmetric) bounded sets (the Fréchet bornology, denoted by F), weak compact sets (the weak Hadamard bornology, denoted by WH), compact sets (the Hadamard bornology, denoted by H) and finite sets (the Gâteaux bornology, denoted by G). Given a continuous function f on X , we say that f is β -differentiable at x and has a β -derivative $\nabla^\beta f(x)$ if $f(x)$ is finite and

$$t^{-1}(f(x + tu) - f(x) - t\langle\nabla^\beta f(x), u\rangle) \rightarrow 0$$

as $t \rightarrow 0$ uniformly in $u \in V$ for every $V \in \beta$. We say that a function f is β -smooth at x if $\nabla^\beta f : X \rightarrow X_\beta^*$ is continuous in a neighbourhood of x . It is not hard to check that a convex function f is β -smooth at x if and only if f is β -differentiable on a convex neighbourhood of x .

Now we can define β -viscosity subderivatives. We consider *extended-real-valued* functions usually lower semicontinuous and *proper* (that is to say not everywhere equal to $+\infty$) and nowhere to $-\infty$.

Definition 3.1 Let $f : X \rightarrow \bar{\mathbf{R}}$ be a lower semicontinuous function and $f(x) < +\infty$. We say f is β -viscosity subdifferentiable and x^* is a β -viscosity subderivative of f at x if there exists a locally Lipschitz concave function g such that g is β -smooth at x , $\nabla^\beta g(x) = x^*$ and $f - g$ attains a local minimum at x . We denote the set of all β -viscosity subderivatives of f at x by $D_\beta f(x)$.

Remark 3.2 By adding a constant we may always assume that the β -smooth function g in the above definition satisfies $g(x) = f(x)$.

3.1 Smooth variational principles

Theorem 3.3 (Smooth Variational Principle) *Let X be a Banach space with an equivalent β -smooth norm. Let $f : X \rightarrow \bar{\mathbf{R}}$ be a lower semicontinuous function and let constants $\varepsilon > 0$ and $\lambda > 0$ be given. Suppose that u satisfies*

$$f(u) < \varepsilon + \inf_X f.$$

Then there exists a β -smooth convex function g on X and v in X such that

- (i) *The function $x \rightarrow f(x) + g(x)$ attains a global minimum at $x = v$;*
- (ii) $\|u - v\| < \lambda$;
- (iii) $f(v) < \varepsilon + \inf_X f$;
- (iv) $\|\nabla g(v)\| < 2\varepsilon/\lambda$.

Informally, we make a small smooth, convex perturbation and obtain a nearby point which minimizes the perturbed function and leaves the function value unimpaired. An easy yet important application of the smoothvariational principle is the following density result for the β -subdifferential of a lower semicontinuous function. The proof will be given in Lecture 3.6 in a more general setting.

Theorem 3.4 *Let X be a Banach space with an equivalent β -smooth norm and let $f : X \rightarrow \bar{\mathbf{R}}$ be a lowersemicontinuous function. Then $\{(x, f(x)) : x \in \text{dom}(D_\beta f)\}$ is dense in $\text{Graph}(f)$.*

3.2 Differentiability of the norm

To apply the smooth variational principle one needs to assume the underlying spaces admit appropriately smooth renorms. Unfortunately, this is not always possible for Banach spaces arising in applications.

Example 3.5 The space ℓ^1 does not have a Fréchet smooth renorm. We need only show that the usual norm $\|\cdot\|_1$ of ℓ^1 is not Fréchet differentiable anywhere because if ℓ^1 has a Fréchet smooth renorm then Theorem 3.4 will imply that $-\|\cdot\|_1$ is at least densely subdifferentiable and, therefore, differentiable as a concave function. Consider $x = (x_n) \in \ell^1$. If for some i , $x_i = 0$ then $\|x + te_i\|_1 - \|x\|_1 = |t|$ tells us that $\|\cdot\|_1$ is not even Gâteaux differentiable at x . Now consider the case when all $x_n \neq 0$. For each $i \geq 1$ let $y^i = (0, 0, \dots, 0, -2x_i, -2x_{i+1}, \dots)$.

Then $\|y^i\|_1 \rightarrow 0$ as $i \rightarrow \infty$. Of course, the sequence $(\text{sgn } x_n)$ is our only candidate for the Fréchet differential. But

$$\left| \|x + y^i\|_1 - \|x\|_1 - \sum_{n=i}^{\infty} (y^i)_n \text{sgn } x_n \right| = \left| \sum_{n=i}^{\infty} -2|x_n| \right| = \|y^i\|_1. \quad \blacksquare$$

Example 3.6 The space ℓ^∞ does not have a Gâteaux smooth renorm. For $x = (x_n) \in \ell^\infty$, define

$$p(x) := \limsup |x_n|.$$

We show that p is sublinear and continuous, but nowhere Gâteaux differentiable. This combined with Theorem 3.4 shows that ℓ^∞ can not have a Gâteaux smooth renorm. Clearly $p(x) \leq \|x\|_\infty$ and p is a seminorm. Therefore it is convex and continuous. If $p(x) = 0$, then $x_n \rightarrow 0$. Taking $y = (1, 1, 1, \dots)$ we have

$$\frac{p(x + ty) - p(x)}{t} = |t|/t,$$

which shows that p is not Gâteaux differentiable at x . If $p(x) > 0$, we can assume that $p(x) = 1$. Choose a subsequence $(x_{n(i)})$ of x_n such that $|x_{n(i)}| \rightarrow 1$. By taking a further subsequence we can assume that the $x_{n(i)}$ have the same sign and, since $p(x) = p(-x)$, it suffices to consider the case $x_{n(i)} > 0$ for all i . Define $y_n = 0$ if either $n \neq n(i)$ or $n = n(i)$ with i odd, while $y_n = 1$ if $n = n(i)$ with i even. Then

$$t^{-1}[p(x + ty) - p(x)] = \begin{cases} 1 & \text{if } t > 0 \\ 0 & \text{if } -1 < t < 0. \end{cases} \quad \blacksquare$$

We record the smoothness of some useful Banach spaces below: ℓ^∞ has no Gâteaux (equivalently, being Lipschitz, Hadamard) smooth renorm. The supremum norm on $C[a, b]$ is nowhere weak-Hadamard and so admits no equivalent weak-Hadamard renorm. Every equivalent norm on ℓ^1 is weak-Hadamard smooth but the original norm is nowhere Fréchet smooth. There is a weak-Hadamard renorm of L^1 but no Fréchet smooth renorm. All reflexive Banach spaces possess Fréchet smooth renorms. All separable or *weakly compactly generated* (WCG) Banach spaces (those containing a densely spanning weak compact set) possess Gâteaux smooth renorms (equivalently Hadamard smooth).

3.3 Partially smooth subdifferentials

Many problems inevitably lie in large (nonsmooth or non-Fréchet) spaces, X . In such settings the ‘target’ set may be significantly smaller and so lie in a much more richly renormable space, Y .

For example, in *most* contexts existence results in control will require some measure of weak compactness of an associated lower level set, S . This set perforce lies in a weakly compactly generated and so smoothable subspace Y , and it is often the case that *only* variations in that subspace need be examined. In this spirit we define the *partially smooth subdifferential* as follows. Let Y be a Banach subspace of X . Given a function f on X , we say that f is $Y\beta$ -differentiable at x and has a $Y\beta$ -derivative $\nabla f(x)$ in X^* if $f(x)$ is finite and

$$t^{-1}(f(x + tu) - f(x) - t\langle \nabla f(x), u \rangle) \rightarrow 0$$

as $t \rightarrow 0$ uniformly in $u \in V$ for every $V \in \beta$ where β is a bornology of Y . We say that a function f is $Y\beta$ -smooth at x if $\nabla f|_Y : X \rightarrow Y_\beta^*$ is continuous in a neighbourhood of x . Note that while $\nabla f(x)$ is not uniquely determined, $\nabla f|_Y(x)$ is.

Definition 3.7 Let X be a Banach space with a closed subspace Y . Let f be a lower semicontinuous function on X . A vector $v^* \in X^*$ is a $Y\beta$ subgradient to f at x if there is a convex locally Lipschitz function g such that

1. g is $Y\beta$ smooth;
2. $-v^* \in \partial g(x)$; and
3. $f + g$ has a local minimum at x .

We call the set of all such $Y\beta$ subgradients to f at x the $Y\beta$ subdifferential of f at x , denoted by $\partial_{Y\beta} f(x)$.

Remark 3.8 The above definition suffices for studying locally Lipschitz functions. However, when dealing with the *limiting subdifferentials* of lower semicontinuous functions one sometimes needs to have uniform control of the Lipschitz rank of the supporting functions (see [142, 143, 23, 38] for details).

The $Y\beta$ -subdifferential defined above coincide with the usual β viscosity subdifferential when restricted to the subspace Y . It is defined by using an osculating function and, therefore, is suited to variational arguments. However, outside the subspace Y , we have not heavily restricted the support function. Thus, this subdifferential could be much larger than more usual generalized subderivatives. The larger a subdifferential the more inaccurately it reflects the local behavior of a function. Therefore, it is desirable to further restrict the subdifferential relative to Y^\perp . We opt to require our subdifferential to be contained in a generalized subdifferential that has a reasonable calculus.

So, recall the following general subdifferential concept. A multifunction ∂_* is a *subdifferential* if it has the following three properties:

1. $0 \in \partial_* f(x)$ if x is a local minimum of f ;
2. $\partial_* g(x)$ coincides with the subgradient in convex analysis when g is convex and Lipschitz around x ; and
3. $0 \in \partial_* f(x) + \partial g(x)$ when $f + g$ attains a local minimum at x and g is convex and Lipschitz around x and ∂ is the usual convex subdifferential.

We define the $*Y\beta$ subdifferential as follows:

Definition 3.9 Let $Y \subset X$ be a closed subspace of X with bornology β and let ∂_* be a subdifferential. Let f be a lower semicontinuous function and $f(x) < +\infty$. We define the partial β -viscosity subdifferential of f with respect to Y at x to be

$$\partial_{*Y\beta} f(x) := \partial_* f(x) \cap \partial_{Y\beta} f(x).$$

Remark 3.10 (a) Using the generalized subdifferential ∂_* allows us flexibility. The idea of using generalized subdifferentials to unify various generalized derivatives goes back to Warga [151, 152] and Ioffe [83]. The ∂_* used here is a slight modification of the various general subdifferential concepts used in [4, 58, 141, 140]. Properties 1. and 2. are common in all the general subdifferential definitions. Property 3. as expressed here is more-or-less the weakest and it is what actually is needed when using variational arguments.

The general subdifferential ∂_* covers many specific subdifferentials. Examples are: the Clarke generalized gradient [50, 51, 130], Ioffe's Approximate and Geometric subderivatives [87, 88], the Michel-Penot subderivative [109] and Treiman's B-subderivative [144, 145].

(b) For the Fréchet subderivative we can say more. Fréchet subderivatives can be defined using either a limiting or a viscosity idea (see [63]). We observe that if X has a Y -Fréchet renorm then all the Fréchet-subderivatives are YF viscosity subderivatives. In this case the concave osculating function required in Definition 3.7 can always be built. Note that, by setting $Y = \{0\}$, or any finite dimensional subspace, we have a viscosity result in *all* spaces.

In fact, let ϕ be a Fréchet subderivative of f at x . Then by [63] there exists a C^1 function ε with $\varepsilon(0) = \varepsilon'(0) = 0$ such that

$$g(z) := f(z) - f(x) - \phi(z - x) + \varepsilon(\|x - z\|)$$

attains a local minimum at 0. Define $\delta(t) = \sup_{0 < s < t} \varepsilon'(s)$. Then δ is a nonnegative continuous increasing function and $\delta(0) = 0$. Thus $\eta(u) := \int_0^u \delta(t) dt$ is a convex C^1 function satisfying $\eta(0) = \eta'(0) = 0$. Moreover, $\eta(t) \geq \varepsilon(t), \forall t \geq 0$. Clearly,

$$z \rightarrow f(x) + \phi(z - x) - \eta(\|x - z\|)$$

is a concave local support of f at x with Fréchet derivative ϕ . This is useful as follows. Suppose X is Asplund and W is boundedly relatively weakly compact then $Y := \text{cl span}(W)$ is WCG and so Fréchet renormable. Especially this will hold for any sequence in Y . It may be possible to extend such a result to the case when X has a YF smooth bump function [63].

(c) Note that

$$D_\beta := \partial_{*X\beta}$$

is the original β -viscosity subdifferential (see [23, 41]) and

$$\partial_{*0\beta} = \partial_*$$

is the original $*$ -subdifferential. Also, by the definition, we always have $\partial_{*Y\beta} f(x) \subset \partial_* f(x)$. Thus, results in terms of $\partial_{*Y\beta}$ are more accurate than those in terms of ∂_* .

(d) The ∂_* subdifferential enables us to concentrate on the properties of various subdifferential concepts that are pertinent to variational arguments. Because of the generality of its definition, the ∂_* subdifferential may be very large and, therefore, inaccurate in reflecting the local behavior of the functions. For example, defining $\tilde{\partial} f(x) = \partial f(x)$ if f is locally convex at x and $\tilde{\partial} f(x) = X^*$ otherwise, it is easy to check that $\tilde{\partial}$ is a ∂_* subdifferential. The idea of defining $\partial_{Y\beta}$ is to make it accurate when restricted to variations in the subspace Y . ■

3.4 Partially smooth equivalent norms

In order for our partial bornological subdifferentials to have a rich calculus and be useful in analyzing nonsmooth problems we often need to assume that a Banach space X with a Banach subspace Y has a $Y\beta$ -smooth equivalent norm. We note that a “nice” equivalent norm, $\|\cdot\|_Y$ on Y always extends to an equivalent norm on X . Indeed: Suppose $\|\cdot\|_Y \leq K\|\cdot\|$ on Y . Then the standard *infimal convolution*

$$\|x\|_X := \inf_{y \in Y} \|y\|_Y + K\|x - y\|$$

is the desired norm on X . However, when $\|\cdot\|_Y$ is β -smooth this extension may not lead to a $Y\beta$ -smooth norm. Throughout this lecture, we make use of standard notions from renorming theory and refer the reader to [63] for additional details.

The following theorem establishes two useful cases when a smooth norm on a subspace Y does extend to a $Y\beta$ -smooth norm in the whole space.

Theorem 3.11 *Let X be a Banach space and let Y be a Banach subspace of X . Then X has an equivalent $Y\beta$ smooth norm provided either*

- (a) Y is separable and $\beta = UG$, the uniformly Gâteaux bornology; or
- (b) Y is weakly compactly generated and $\beta = H$, the Hadamard bornology.

Proof (a) To obtain a (uniformly) Gâteaux renorm in directions of a separable subspace Y , we may consider the norm on X whose dual on X^* is defined by

$$\| \|f\| \|^2 = \|f\|^2 + \sum_{i=1}^{\infty} f^2(y_i)/2^i$$

for $f \in X^*$, where $\{y_i : i \in \mathbf{N}\}$ is dense in S_Y .

(b) In the WCG case, we can argue as follows: there is a dense linear continuous embedding from a reflexive space Z into Y ([14, 67]). Let ι be this embedding viewed as a mapping into X . Let $\|\cdot\|_Z$ be a LUR renorm of Z^* and consider the norm whose dual on X^* is defined by

$$\| \|f\| \|^2 = \|f\|^2 + \|\iota^*(f)\|_Z^2.$$

In both cases, this norm will be Y -strictly convex in the sense that

$$\| \| \frac{f+g}{2} \| \|^2 = \frac{\| \|f\| \|^2 + \| \|g\| \|^2}{2} \Rightarrow f - g \in Y^\perp.$$

This in turn implies that the norm on X will be Y -Hadamard-smooth. ■

Moreover, in case (b) the renorm may be chosen to be uniform on a given weakly compact set W . Thus, we may work in the $H(W) := \text{Hadamard} \cup \{W\}$ -bornology (see [14]).

Remark 3.12 Every bounded relatively norm compact set lies in a separable space which thus admits an equivalent uniformly Gâteaux (UG) norm. Similarly, every bounded relatively weak compact set lies in a weakly compactly generated space which thus admits an equivalent

Gâteaux norm. Each of these classes of spaces is closed under countable unions. Every subset of a reflexive space is bounded relatively weakly compact.

By contrast, such subsets of ℓ^∞ or $L^\infty[0, 1]$ are necessarily norm separable [67]. Indeed, Lemma 2.44 in [120] notes that if E is a separable Banach space and K is a weakly compact subset of E^* then K is norm separable. ■

The previous argument is very generally applicable as Jon Vanderwerff points out to us [147]. Suppose Y has a dual norm with a reasonable rotundity property: say (UR), (LUR) or (R). Call the norm $\|\cdot\|_Y$. Then let $R: X^* \rightarrow Y^*$ be defined by $R(x^*) = x^*|_Y$. Then clearly

$$\| \|x^*\|^2 := \|x^*\|^2 + (\|R(x^*)\|_Y^*)^2$$

defines a dual norm on X . Moreover, the corresponding norm on X will be $Y\beta$ -smooth for the appropriate bornology. Thus, we have

Theorem 3.13 *Let X be a Banach space and let Y be a closed subspace of X . Suppose Y has a β -rotund dual norm. Then X has an equivalent $Y\beta$ smooth norm.*

The following results will be useful later.

Lemma 3.14 *Let X be a Banach space and let $A_i, i = 1, 2$ be convex subsets of X . Then a locally Lipschitz convex function f is uniformly differentiable at x on $A_i, i = 1, 2$ if and only if f is uniformly differentiable at x on $\text{conv}(A_1, A_2)$.*

Proof Observe that f differentiable at x uniformly on A_i amounts to, for any $\varepsilon > 0$ there exists a $\delta_i > 0$ such that

$$\sup_{h \in A_i} \frac{f(x+th) + f(x-th) - 2f(x)}{t} \leq \varepsilon$$

for $0 < t < \delta_i$. Let $\delta = \min(\delta_1, \delta_2)$. Then the convexity of f implies that

$$\sup_{h \in \text{conv}\{A_1 \cup A_2\}} \frac{f(x+th) + f(x-th) - 2f(x)}{t} \leq \varepsilon$$

for $0 < t < \delta$. This proves the “only if” part. The “if” part is trivial. ■

Now observe that, for any nonzero element $x \in X$, any norm of X is smooth at x in the directions $\pm x$. So, if $\|\cdot\|$ is a $Y\beta$ smooth norm of X and $V \in \beta$, then the norm is also uniformly differentiable at x on any set of the form $\text{conv}(V, ax)$, $a \in \mathbf{R}$, by Lemma 3.14. Thus, we have

Lemma 3.15 *Let X be a Banach space with a closed subspace Y . Then a $Y\beta$ smooth norm $\|\cdot\|$ of X is also β smooth at x on $\text{span}(Y, x)$.*

We illustrate the power of minimality by easily establishing a result on the generic partial differentiability of locally Lipschitz functions.

Theorem 3.16 *Suppose that f is locally Lipschitz on an open subset A of a Banach space X and possesses a minimal subgradient on A . When Y is a class (S) subspace of X then f is generically Y -Hadamard smooth throughout A . When Y is an Asplund subspace of X then f is generically Y -Fréchet smooth throughout A .*

Proof Let Ω_Y be the restriction of elements of ∂f to Y . As the composition of a linear operator ($R : x^* \rightarrow x^*|_Y$) and the minimal cusco ∂f , Ω_Y is a minimal cusco, by Theorem 2.15. The results of the previous lecture apply. Consider first the (S) case. By Theorem 2.9 Ω_Y is generically single valued on the open (Baire) set A . An easy application of the mean value theorem of Lebourg establishes that at each such point f is (strictly) Y -Hadamard smooth. The Asplund case follows similarly from Theorem 2.12. ■

An immediate consequence is that in any Banach space, continuous convex functions are generically Fréchet (respectively Gâteaux) differentiable with respect to any fixed Asplund (respectively class (S)) subspace.

Remark 3.17 In [74], Fabian, Zajíček and Zizler give a category argument of Asplund’s result that if a Banach space and its dual have rotund renorms respectively, then one can find a rotund renorm whose dual norm is rotund simultaneously. Their technique allows one to show, for example, that if Y is a subspace of X such that X and X^* admit Y -rotund renorms respectively, then X can be renormed to be simultaneously Y -smooth and Y -rotund in an appropriate sense.

3.5 Partially smooth variational principles

Smooth variational principles are the key to using variational arguments. To use variational arguments with respect to a smooth subspace we need the following form of the Borwein-Preiss smooth variational principle [36]. The proofs given in [36, 105] work here with slight modifications.

Theorem 3.18 Borwein-Preiss smooth variational principle *Let X be a Banach space and $Y \subset X$ a Banach subspace. Assume that X has an $Y\beta$ -smooth equivalent norm. Let $f : X \rightarrow \bar{\mathbf{R}}$ be a lower semicontinuous function and let constants $\varepsilon > 0$ and $\lambda > 0$ be given. Suppose that u satisfies*

$$f(u) < \inf_X f + \varepsilon.$$

Then there exists a locally Lipschitz $Y\beta$ -differentiable convex function g on X and v in X such that

(i) *the function*

$$x \rightarrow f(x) + g(x)$$

attains a global minimum at $x = v$;

(ii) $\|u - v\| < \lambda,$

(iii) $f(v) < \inf_X f + \varepsilon;$ and

(iv) $\|\nabla g|_Y(v)\| < \frac{2\varepsilon}{\lambda}.$

When $Y = X$ this becomes the usual form of the Borwein-Preiss smooth variational principle.

3.6 Several useful tools

We now describe several *equivalent* tools for applying the partially smooth variational principle to diverse problems that involve nonsmooth or even lower semicontinuous functions. They are a nonlocal fuzzy sum rule, a local fuzzy sum rule, a multidirectional mean value inequality and an extremal principle. We will only outline their proofs. Details can be found in [38]. We start with a *nonlocal fuzzy sum rule* that extends [155, Theorem 2.1].

Theorem 3.19 *Let X be a Banach space and $Y \subset X$ a Banach subspace. Assume that X has an equivalent $Y\beta$ -smooth norm. Suppose that $f_1, \dots, f_N : X \rightarrow \bar{\mathbf{R}}$ are lower semicontinuous functions bounded below with*

$$\liminf_{\eta \rightarrow 0} \left\{ \sum_{n=1}^N f_n(x_n) : \text{diam}(x_1, \dots, x_N) \leq \eta \right\} < +\infty.$$

Then, for any $\varepsilon > 0$, there exist $x_n, n = 1, \dots, N$, and $x_n^ \in D_{*Y\beta} f_n(x_n)$ satisfying*

$$\begin{aligned} \text{diam}(x_1, \dots, x_N) &< \varepsilon, \\ \text{diam}(x_1, \dots, x_N) \cdot \max(\|x_1^*|_Y\|, \dots, \|x_N^*|_Y\|) &< \varepsilon, \end{aligned} \quad (7)$$

and

$$\sum_{n=1}^N f_n(x_n) < \liminf_{\eta \rightarrow 0} \left\{ \sum_{n=1}^N f_n(x_n) : \text{diam}(x_1, \dots, x_N) \leq \eta \right\} + \varepsilon \quad (8)$$

such that

$$\left\| \sum_{n=1}^N x_n^*|_Y \right\| < \varepsilon.$$

Proof Assume without loss of generality that the norm $\|\cdot\|$ is $Y\beta$ smooth. Define, for any number $i > 0$,

$$w_i(y_1, \dots, y_N) := \sum_{n=1}^N f_n(y_n) + i \sum_{n,m=1}^N \|y_n - y_m\|^2$$

and $M_i := \inf w_i$. Then M_i is an increasing sequence and is bounded by

$$\liminf_{\eta \rightarrow 0} \left\{ \sum_{n=1}^N f_n(x_n) : \text{diam}(x_1, \dots, x_N) \leq \eta \right\}.$$

Let $M = \lim_{i \rightarrow \infty} M_i$. Observe that the product space X^N of N copies of X (with the Euclidean product norm) also has an equivalent $(Y^N)\beta$ -smooth norm. For each i , applying the Borwein-Preiss smooth variational principle to function w_i , we obtain that there exist a $Y^N\beta$ -smooth function ϕ_i and $x_{n,i}, n = 1, \dots, N$ such that $w_i + \phi_i$ attains a local minimum at $(x_{1,i}, \dots, x_{N,i})$, while $\|\phi_i(x_{1,i}, \dots, x_{N,i})\|$ is Lipschitz with rank less than ε/N and

$$w_i(x_{1,i}, \dots, x_{N,i}) < \inf w_i + \frac{1}{i} \leq M + \frac{1}{i}. \quad (9)$$

For each n , the function

$$y \rightarrow w_i(x_{1,i}, \dots, x_{n-1,i}, y, x_{n+1,i}, \dots, x_{N,i}) + \phi_i(x_{1,i}, \dots, x_{n-1,i}, y, x_{n+1,i}, \dots, x_{N,i})$$

attains a local minimum at $y = x_{n,i}$. Thus, for $n = 1, \dots, N$,

$$0 \in \partial_* f_n(x_{n,i}) + 2i \sum_{m=1}^N \partial_* \|\cdot\|^2(x_{n,i} - x_{m,i}) + \partial_{*x_n} \phi_i(x_{1,i}, \dots, x_{N,i}).$$

Choose

$$x_{n,i}^* = y_{n,i}^* + z_{n,i}^* \in \partial_* f_n(x_{n,i})$$

such that

$$-y_{n,i}^* \in 2i \sum_{m=1}^N \partial_* \|\cdot\|^2(x_{n,i} - x_{m,i})$$

and

$$-z_{n,i}^* \in \partial_{*x_n} \phi_i(x_{1,i}, \dots, x_{N,i}).$$

Then

$$x_{n,i}^*|_Y := -\nabla_{x_n} \phi_i|_Y(x_{1,i}, \dots, x_{N,i}) - 2i \sum_{m=1}^N \nabla \|\cdot\|^2|_Y(x_{n,i} - x_{m,i}).$$

Summing these N equalities leads to

$$\sum_{n=1}^N x_{n,i}^*|_Y = -\sum_{n=1}^N \nabla_{x_n} \phi_i|_Y(x_{1,i}, \dots, x_{N,i}) - 2i \sum_{n=1}^N \sum_{m=1}^N \nabla \|\cdot\|^2|_Y(x_{n,i} - x_{m,i}).$$

Observing that

$$\left\| -\sum_{n=1}^N \nabla_{x_n} \phi_i|_Y(x_{1,i}, \dots, x_{N,i}) \right\| \leq \varepsilon$$

and

$$\nabla \|\cdot\|^2|_Y(x_{n,i} - x_{m,i}) + \nabla \|\cdot\|^2|_Y(x_{m,i} - x_{n,i}) = 0$$

so that the double sum in the previous equality vanishes, we obtain

$$\left\| \sum_{n=1}^N x_{n,i}^*|_Y \right\| < \varepsilon.$$

By the definition of M_i we have

$$\begin{aligned} M_{i/2} &\leq w_{i/2}(x_{1,i}, \dots, x_{N,i}) \\ &= w_i(x_{1,i}, \dots, x_{N,i}) - \frac{i}{2} \sum_{n,m=1}^N \|x_{n,i} - x_{m,i}\|^2 \\ &\leq M_i + \frac{1}{i} - \frac{i}{2} \sum_{n,m=1}^N \|x_{n,i} - x_{m,i}\|^2. \end{aligned} \tag{10}$$

Rewriting (10) as

$$i \sum_{n,m=1}^N \|x_{n,i} - x_{m,i}\|^2 \leq 2(M_i - M_{1/2} + \frac{1}{i})$$

yields

$$\lim_{i \rightarrow \infty} i \sum_{n,m=1}^N \|x_{n,i} - x_{m,i}\|^2 = 0.$$

Therefore,

$$\lim_{i \rightarrow \infty} \text{diam}(x_{1,i}, \dots, x_{N,i}) = 0$$

and

$$\lim_{i \rightarrow \infty} \text{diam}(x_{1,i}, \dots, x_{N,i}) \cdot \max(\|x_{1,i}^*|_Y\|, \dots, \|x_{N,i}^*|_Y\|) = 0.$$

Thus,

$$\begin{aligned} M &\leq \liminf_{\eta \rightarrow 0} \left\{ \sum_{n=1}^N f_n(x_n) : \text{diam}(x_1, \dots, x_N) \leq \eta \right\} \\ &\leq \liminf_{i \rightarrow \infty} \sum_{n=1}^N f_n(x_{n,i}) = \liminf_{i \rightarrow \infty} w_i(x_{1,i}, \dots, x_{N,i}) \leq M \end{aligned}$$

yields

$$M = \liminf_{\eta \rightarrow 0} \left\{ \sum_{n=1}^N f_n(x_n) : \text{diam}(x_1, \dots, x_N) \leq \eta \right\}.$$

It remains to take $x_n = x_{n,i}$ and $x_n^* = x_{n,i}^*$, $n = 1, \dots, N$ for a sufficiently large i . ■

We now turn to a *local* fuzzy sum rule (see [23, 41, 65, 84, 87, 90]) which needs the following additional condition.

Definition 3.20 [41] Let $f_1, \dots, f_N : X \rightarrow \bar{\mathbf{R}}$ be lower semicontinuous functions and E a closed subset of X . We say that (f_1, \dots, f_N) is *uniformly lower semicontinuous* on E if

$$\inf_{x \in E} \sum_{n=1}^N f_n(x) \leq \liminf_{\eta \rightarrow 0} \left\{ \sum_{n=1}^N f_n(x_n) : \|x_n - x_m\| \leq \eta, \right. \\ \left. x_n, x_m \in E, n, m = 1, \dots, N \right\}.$$

We say that (f_1, \dots, f_N) is *locally uniformly lower semicontinuous* at $x \in \bigcap_{n=1}^N \text{dom}(f_n)$ if (f_1, \dots, f_N) is uniformly lower semicontinuous on a closed ball centered at x .

We now have:

Theorem 3.21 Let X be a Banach space and Y a Banach subspace of X . Assume that X has an equivalent $Y\beta$ -smooth norm and f_1, \dots, f_N lower semicontinuous functions on X . Suppose that (f_1, \dots, f_N) is locally uniformly lower semicontinuous and $\sum_{n=1}^N f_n$ attains a local minimum at \bar{x} . Then, for any $\varepsilon > 0$, there exist $x_n \in \bar{x} + \varepsilon B$ and $x_n^* \in \partial_{*Y\beta} f_n(x_n)$, $n = 1, \dots, N$, such that for each n $|f_n(x_n) - f_n(\bar{x})| < \varepsilon$,

$$\text{diam}(x_1, \dots, x_N) \cdot \max(\|x_1^*|_Y\|, \dots, \|x_N^*|_Y\|) < \varepsilon,$$

and

$$\left\| \sum_{n=1}^N x_n^*|_Y \right\| < \varepsilon.$$

Proof (Sketch) Apply Theorem 3.19 to $f_n + \|\cdot - \bar{x}\|^2 + \delta_{\bar{x}+hB_X}, n = 1, \dots, N$, for an appropriate small h . The term $\|\cdot - \bar{x}\|^2$ and the uniform lower semicontinuity assumption provide control of the location of the x'_n s in the theorem. ■

Remark 3.22 (a) Let (f_1, \dots, f_N) be lower semicontinuous functions on X and E a closed subset of X . Then the following conditions are sufficient for (f_1, \dots, f_N) to be uniformly lower semicontinuous on E :

- (i) all but one of the functions are uniformly continuous on E ;
- (ii) one of the functions has compact level sets when restricted to E ;
- (iii) X is finite dimensional and E is bounded.

(b) Without the local uniform lower semicontinuity condition the local fuzzy sum rule fail as shown by different examples constructed by Deville and Ivanov [66] and Vanderwerff and Zhu [148] independently. The following is a Hilbert space version of the example in [148].

Example 3.23 Let $X := \ell_2$ and let e_k be the standard basis. Then $x \in X$ can be uniquely represented as $x = \sum_{k=1}^{\infty} x(k)e_k$; for $P_n(x) := \sum_{k=1}^n x(k)e_k$, one has $\|P_m(x)\| \leq \|P_n(x)\|$ for $m \leq n$, in particular $\|P_n\| \leq 1$ for each n . Now $x(k) \rightarrow 0$ as $k \rightarrow \infty$ and so $\|x\|_{\infty} := \max\{|x(k)| : 1 \leq k < \infty\}$ exists. Moreover, for k_0 such that $|x(k_0)| = \|x\|_{\infty}$, we have $|x(k_0)| = \|P_{k_0+1}(x) - P_{k_0}(x)\| \leq 2\|x\|$. Thus $\|\cdot\|_{\infty}$ is Lipschitz with a Lipschitz constant 2. Define

$$F_n := \{x : \|x\| \leq 3, x(i) \geq 0 \text{ and } x(i) = 0 \text{ if } i \bmod 3 \neq 0 \text{ or } i < 3n\}.$$

Now we construct two functions

$$f_1(x) := \begin{cases} 0 & \text{if } x = 0; \\ -\frac{1}{\sqrt{n}} - \|y\|_{\infty} & \text{if } x = \frac{1}{n}e_{3n-1} + y, y \in F_n; \\ +\infty & \text{otherwise;} \end{cases}$$

and

$$f_2(x) := \begin{cases} 0 & \text{if } x = 0; \\ -\frac{1}{\sqrt{n}} - \|y\|_{\infty} & \text{if } x = \frac{1}{n}e_{3n-2} + y, y \in F_n; \\ +\infty & \text{otherwise.} \end{cases}$$

First observe that $\text{dom}(f_1) \cap \text{dom}(f_2) = 0$ by the uniqueness of basis representations, so $f_1 + f_2$ attains a minimum at 0. From the definitions it also follows that f_1 and f_2 are both bounded below by -7 since $\|\cdot\|_{\infty}$ is Lipschitz with a Lipschitz constant 2.

We now prove that f_1 is lower semicontinuous (the proof for f_2 is similar). Suppose $x_n \in \text{dom}(f_1)$ and $x_n \rightarrow x$. If $x = 0$, we may assume $x_n \neq 0$ and so $x_n = \frac{1}{k_n}e_{3k_n-1} + y_n, y_n \in F_{k_n}$. Now $k_n \rightarrow \infty$, and $y_n \rightarrow 0$ and so $-\frac{1}{\sqrt{k_n}} - \|y_n\|_{\infty} \rightarrow 0$. If $x \neq 0$, we know that $k_n \not\rightarrow \infty$. Indeed, if $k_n \rightarrow \infty$, then for each i , we have $x_n(i) \rightarrow 0$ as $n \rightarrow \infty$ because $x_n(i) = 0$ for all $i \leq 3k_n - 1$. Because the norm and pointwise limit must agree if they both exist, we conclude that x_n converges to 0 in norm. As $k_n \not\rightarrow \infty$, we know that $k_n = n_0$ for all large n . Indeed, when $n \neq m$,

$$\left\| \frac{1}{n}e_{3n-1} + y_n - \left(\frac{1}{m}e_{3m-1} + y_m \right) \right\| \geq \max\left\{ \frac{1}{n}, \frac{1}{m} \right\}$$

for $y_m \in F_m$, $y_n \in F_n$ by the monotonicity of the basis. Therefore,

$$x_n = \frac{1}{n_0} e_{3n_0-1} + y_n,$$

with $y_n \in F_{n_0}$ for all large n . This implies $y_n \rightarrow \bar{y} \in F_{n_0}$, which with the continuity of $\|\cdot\|_\infty$ implies

$$f_1(x_n) = -\frac{1}{\sqrt{n_0}} - \|y_n\|_\infty \rightarrow -\frac{1}{\sqrt{n_0}} - \|y\|_\infty = f_1(x).$$

This proves the lower semicontinuity of f_1 (and similarly of f_2).

We turn to prove that, for any $x_i \in B$ and $x_i^* \in D_F f_i(x_i)$, $\|x_1^* + x_2^*\| \geq 1$. Let g_i be the function associated to x_i^* , $i = 1, 2$, as in the definition. Now observe that $D_F f_1(0)$ is empty because

$$n[f_1(0 + \frac{1}{n} e_{3n-1}) - f_1(0)] \leq n[-\frac{1}{\sqrt{n}}] - 0 = -\sqrt{n};$$

similarly $D_F f_2(0)$ is empty. Thus we can write $x_1 = \frac{1}{m} e_{3m-1} + y_1$ and $x_2 = \frac{1}{n} e_{3n-1} + y_2$, where $y_1 = \sum_{k=m}^\infty a_k e_{3k} \in F_m$ and $y_2 = \sum_{k=n}^\infty b_k e_{3k} \in F_n$.

We will prove that $\|x_1^* + x_2^*\| \geq 1$ in the case $m \leq n$ (the proof for $m \geq n$ is similar). Let $b_{k_0} = \max_{k \geq n} \{b_k\}$, then $0 \leq b_{k_0} \leq 2\|x_2\| \leq 2$, and thus $y_2 + t e_{3k_0} \in F_n$ for $0 \leq t \leq 1$, and $\|y + t e_{3k_0}\|_\infty = \|y\|_\infty + t$. Therefore

$$\frac{g_2(x_2 + t e_{3k_0}) - g_2(x_2)}{t} \leq \frac{f_2(x_2 + t e_{3k_0}) - f_2(x_2)}{t} = -1.$$

Now, because $m \leq n$, we have $y_1 + t e_{3k_0} \in F_m$, and because $a_{3k_0} \geq 0$ we conclude that $\|y_1 + t e_{3k_0}\|_\infty \geq \|y_1\|_\infty$. Consequently

$$\frac{g_1(x_1 + t e_{3k_0}) - g_1(x_1)}{t} \leq \frac{f_1(x_1 + t e_{3k_0}) - f_1(x_1)}{t} \leq 0.$$

Therefore, $\langle x_2^*, e_{3k_0} \rangle \leq -1$ while $\langle x_1^*, e_{3k_0} \rangle \leq 0$. This shows that $\|x_1^* + x_2^*\| \geq 1$ as desired. ■

Clarke and Ledyaev have established a *multi-directional mean value inequality* that provides an estimate for

$$\lim_{\eta \rightarrow 0} \inf_{[x, U] + \eta B} f - f(x)$$

in terms of the (*proximal*) subdifferentials of f in [54]. It was shown in [38] that this mean value inequality is also valid in a Banach space with a $Y\beta$ -smooth equivalent norm where Y is any Banach subspace that contains

$$[x, U] := \{y : y = \lambda x + (1 - \lambda)u, u \in U, \lambda \in [0, 1]\}.$$

This extension of the mean value inequality is sufficient for most applications because the span of $[x, U]$ often has a smoothly extensible norm.

Theorem 3.24 *Let X be a Banach space, let U be a nonempty, closed and convex subset of X and $x \in X$ and let $f : X \rightarrow \bar{\mathbf{R}}$ be a lower semicontinuous function. Let $Y := \overline{\text{span}}\{x, U\}$.*

Suppose that X has a $Y\beta$ -smooth equivalent norm and, for some $h > 0$, f is bounded below on $[x, U] + hB_X$ and

$$\lim_{\eta \rightarrow 0} \inf_{y \in U + \eta B_X} f(y) - f(x) > r.$$

Then, for any $\varepsilon > 0$, there exist $z \in [x, U] + \varepsilon B$ and $z^* \in \partial_{*Y\beta} f(z)$ such that

$$r < \langle z^*, y - x \rangle + \varepsilon \|y - x\| \text{ for all } y \in U.$$

Further, we can choose z to satisfy

$$f(z) < \lim_{\eta \rightarrow 0} \inf_{[x, U] + \eta B_X} f + |r| + \varepsilon.$$

Proof (Sketch) We observe first that by introducing a new function $F(x, t) = f(x) - r't$ where $r' < r$, point $(x, 0)$ and set $U \times \{1\}$ one can reduce the general theorem to the special case when

$$\lim_{\eta \rightarrow 0} \inf_{y \in U + \eta B_X} f(y) - f(x) > 0,$$

and $r = -\varepsilon$ is an arbitrary negative number. To prove the special case one need only apply the nonlocal fuzzy sum rule to functions $\bar{f} = f + \delta_{[x, U] + hB_X}$ and $\delta_{[x, U]}$. ■

We turn to the *extremal principle* formulated by Mordukhovich [115, 116]. This may be viewed as a nonconvex extension of the separation theorem for convex sets. We may also view it as a geometric version of the fuzzy sum rule.

Definition 3.25 Let Ω_1 and Ω_2 be closed sets in a Banach space X and let $\bar{x} \in \Omega_1 \cap \Omega_2$. Then \bar{x} is called a *local extremal point of the set system* $\{\Omega_1, \Omega_2\}$ if there are a neighbourhood U of \bar{x} and sequences $\{a_{ik}\} \subset X$, $i = 1, 2$, such that $a_{ik} \rightarrow 0$ for $i = 1, 2$ and

$$(\Omega_1 - a_{1k}) \cap (\Omega_2 - a_{2k}) \cap U = \emptyset \quad \forall k = 1, 2, \dots$$

We say that the sets Ω_1 and Ω_2 generate a (local) *extremal system* $\{\Omega_1, \Omega_2\}$ if they have at least one locally extremal point.

Theorem 3.26 Let X be a Banach space and let Y be a Banach subspace of X . Assume that X has an equivalent $Y\beta$ -smooth norm, and Ω_1 and Ω_2 are closed sets in Y . Let $\bar{x} \in \Omega_1 \cap \Omega_2$ be a local extremal point of the system $\{\Omega_1, \Omega_2\}$. Then, for any $\varepsilon > 0$ there exist $x_n \in \Omega_n \cap \bar{x} + \varepsilon B$ and $x^* = (x_1^*, x_2^*) \in N_{*(Y \times Y)\beta}(\Omega_1 \times \Omega_2, x)$, such that

$$\|x_1^*|_Y\|, \|x_2^*|_Y\| \geq 1 - \varepsilon \text{ and } \|x_1^*|_Y + x_2^*|_Y\| < \varepsilon.$$

Proof Consider the product space $X \times X$ in the Euclidean product norm derived from the $Y\beta$ -smooth norm of X ; it is $(Y \times Y)\beta$ smooth. Denote elements of $X \times X$ by $z = (z^1, z^2)$. Let \bar{x} be a local extremal of (Ω_1, Ω_2) and $\bar{x} + hB_X \subset U$ where U is as in the definition of a local extremal point. Let $\varepsilon' > 0$ be arbitrary. Take $a \in X$ such that $\|a\| < \varepsilon'$ and $\Omega_1 \cap (\Omega_2 + a) \cap (\bar{x} + hB_X) = \emptyset$. The extremal principle follows from Theorem 3.19 applied to

$$f_1(z) := \delta_{\Omega_1 \times \Omega_2}(z) + \|z^1 - z^2 - a\|$$

and

$$f_2(z) := \varepsilon \|z - (\bar{x}, \bar{x})\|^2. \quad \blacksquare$$

Remark 3.27 It is shown in [101, 89, 157] that all the above tools are equivalent (based on previous partial results in [52, 118]). In fact, they exploit in different ways two underlying principles: (a) a “smooth variational principle” [36] and (b) a “decoupling lemma” used by Crandall and Lions in studying the uniqueness of viscosity solutions [60].

3.7 Applications

A. Density of subdifferentials

First we prove a generalization of Theorem 3.4 to the partial viscosity subdifferential setting.

Theorem 3.28 (Density of Subdifferentiable Points) *Let X be a Banach space with a subspace Y and let $f : X \rightarrow \bar{\mathbf{R}}$ be a lower semicontinuous function. Suppose that X has an equivalent $Y\beta$ -smooth norm. Then $\{(x, f(x)) : x \in \text{dom}(D_{Y\beta}f)\}$ is dense in $\text{graph}(f)$.*

Proof Let $f : X \rightarrow \bar{\mathbf{R}}$ be a lower semicontinuous function, let $x \in \text{dom}(f)$ and let $\varepsilon \in (0, 1)$. Then there exists $\eta < \varepsilon$ such that $\|y - x\| < \eta$ implies that $f(y) > f(x) - \varepsilon$. Applying Theorem 3.19 to $f_1 = f + \delta_{x+B_X}$ and $f_2 = \delta_{\{x\}}$ yields that there exist x_1 and x_2 such that $\|x_1 - x_2\| < \eta$, $0 \in D_{Y\beta}f_1(x_1) + D_{Y\beta}\delta_{\{x\}}(x_2) + \eta B_{X^*}$ and

$$f_1(x_1) + \delta_{\{x\}}(x_2) < f(x) + \eta < f(x) + \varepsilon.$$

The last inequality forces $x_2 = x$ and, therefore, $x_1 \in B_\eta(x) \subset \text{int}B_\varepsilon(x)$ so that $D_{Y\beta}f_1(x_1) = D_{Y\beta}f(x_1)$. Moreover, $|f(x_1) - f(x)| < \varepsilon$. Therefore, $\{(x, f(x)) : x \in \text{dom}(D_{Y\beta}f)\}$ is dense in $\text{graph}(f)$. ■

B. Best approximation

Next we discuss a nice application to distance functions along the lines of the analysis in [14]. Distance functions are revisited in much more detail in Lecture 4.4. We recall that a norm is (sequentially) *Kadec-Klee* if the weak and norm topologies coincide (sequentially) on the boundary of the unit ball. Let W be a given weakly compact set. We define a Hadamard(W) bornology by

$$H(W) := \{S : S \text{ is compact } \cup W\}.$$

We will also slightly abuse the notation by denoting

$$\partial_{*YH(C,x)}f(x) := \{x^* : x^* \in \partial_*f(x) \cap (-\partial g(x))\},$$

where g is a convex Lipschitz function that is uniformly differentiable at x on $\text{conv}(C, x)$ with $f + g$ attaining a local minimum at x .

Theorem 3.29 *Let C be a closed relatively weakly compact subset of a Banach space. Let Y be the closed span of C . Suppose that the norm is Kadec-Klee on X . Then*

- (a) *The set of points in X at which every minimizing sequence clusters to a best approximation is dense in X .*
- (b) *If in addition, the original norm is Fréchet on Y then*

$$\partial_{*YH(C,x)}d_C(x) \subset \partial_{*YF}d_C(PC(x)),$$

where $PC(x)$ is the (set of) best approximations of x on C . In particular, in any Fréchet LUR norm on a reflexive space, this holds for all sets in the Fréchet sense with a single valued metric projection.

Proof We know, from Lecture 3.4, that Y has a $H(C)$ renorm. After using Lemma 3.15, the proof parallels that in Theorem 4.35 and in [14], and we deduce that at any of the dense set of points with $\partial_{*YH(C,x)}d_C(x)$ nonempty, all minimizing sequences actually converge in norm to a best approximation, p ; and the corresponding subgradient provides a proximal normal to C at p .

Finally, when the norm on Y is $H(C)$ -smooth, simple derivative estimates show that any member of

$$\partial_{*YH(C,x)}d_C(x)$$

must actually lie in

$$\partial_{*YH(C)}d_C(P_C(x)). \quad \blacksquare$$

A similar result holds for C being only bounded relatively weakly compact. Note that this result allows us to show that normal cone defined in terms of distance functions is always contained in the normal cone defined in terms of indicator functions. Note also that in Hilbert space we may derive:

$$D_F d_C(x) \subset \partial_\pi d_C(P_C(x)),$$

where ∂_π denotes the set of *proximal* subgradients.

C. Viscosity solutions

Consider the Hamilton-Jacobi equation

$$u + H(x, Du) = 0. \tag{11}$$

We recall the definition of viscosity solutions first [60, 61].

Definition 3.30 (Viscosity solution) A function $u : X \rightarrow \mathbf{R}$ is a *viscosity supersolution* (*viscosity subsolution*) of (11) if u is lower (upper) semicontinuous and, for every $x \in X$ and every $x^* \in D_F(u)(x)$ ($x^* \in D^F(u)(x)$):

$$u(x) + H(x, x^*) \geq 0 \quad (u(x) + H(x, x^*) \leq 0).$$

A continuous function u is called a *viscosity solution* if u is both a viscosity subsolution and a viscosity supersolution.

The following comparison and uniqueness principle for viscosity solutions is fundamental in the viscosity solution theory. We give a short proof by using the nonlocal fuzzy sum rule. This approach allows one to significantly relax the usual uniform continuity assumptions on u , v and H .

Theorem 3.31 Suppose $H : X \times X^* \rightarrow \mathbf{R}$ satisfies the following assumption

(A) for any $x_1, x_2 \in X$ and $x_1^*, x_2^* \in X^*$,

$$|H(x_1, x_1^*) - H(x_2, x_2^*)| \leq \omega(x_1 - x_2, x_1^* - x_2^*) + M \max(\|x_1^*\|, \|x_2^*\|) \|x_1 - x_2\|$$

where $M > 0$ is a constant and $\omega : X \times X^* \rightarrow \mathbf{R}$ is a continuous function with $\omega(0, 0) = 0$.

Let u be an upper semicontinuous function bounded above and v be a lower semicontinuous function bounded below. If u is a viscosity subsolution of (11) and v is a viscosity supersolution of (11), then $u \leq v$. In particular, any continuous bounded viscosity solution to (11) is unique.

Proof Let ε be an arbitrary positive number. Applying the nonlocal fuzzy sum rule of Theorem 3.19 with $f_1 = v$ and $f_2 = -u$, there exist $x_1, x_2 \in X$, $x_1^* \in D_F v(x_1)$ and $x_2^* \in D^F u(x_2)$ satisfying

- (i) $\|x_1 - x_2\| < \varepsilon$;
- (ii) $\|x_1^*\| \|x_1 - x_2\| < \varepsilon$ and $\|x_2^*\| \|x_1 - x_2\| < \varepsilon$;
- (iii) $v(x_1) - u(x_2) < \inf(v - u) + \varepsilon$;
- (iv) $\|x_1^* - x_2^*\| \leq \varepsilon$.

Since the function v is a viscosity supersolution of (11) we have

$$v(x_1) + H(x_1, x_1^*) \geq 0.$$

Similarly

$$u(x_2) + H(x_2, x_2^*) \leq 0.$$

Therefore,

$$\begin{aligned} \inf_X (v - u) &> v(x_2) - u(x_1) - \varepsilon \\ &\geq [H(x_2, x_2^*) - H(x_1, x_1^*)] - \varepsilon \\ &\geq -[\omega(x_2 - x_1, x_2^* - x_1^*) + M \max(\|x_1^*\|, \|x_2^*\|) \|x_2 - x_1\|] - \varepsilon. \end{aligned}$$

As $\varepsilon \rightarrow 0$ the right hand side converges to 0 which yields $\inf_X (v - u) \geq 0$. ■

D. Spectral functions

We end this section with a beautiful application of subdifferentials to the study of eigenvalues of symmetric matrices [104] as in Example 1.4. As before, we let $X := S(n)$ be the vector space of $n \times n$ symmetric matrices endowed with the inner product

$$\langle A, B \rangle := \text{trace}(AB), \quad \forall A, B \in X.$$

For $A \in X$, we write the n eigenvalues of A including multiplicity as $\lambda_1(A) \geq \dots \geq \lambda_n(A)$ and define the *eigenvalue map* $\lambda : X \rightarrow \mathbf{R}^n$ by

$$\lambda(A) := (\lambda_1(A), \dots, \lambda_n(A)).$$

We use $\text{Diag } y$ to denote the diagonal matrix with diagonal elements $y \in \mathbf{R}^n$. And we denote the *orthogonal matrices* by \mathcal{O} . We are interested in the extended-real-valued *spectral function* of the eigenvalues of matrices:

$$f \circ \lambda : X \rightarrow \bar{\mathbf{R}},$$

where $f : \mathbf{R}^n \rightarrow \bar{\mathbf{R}}$ is lower semicontinuous and invariant under the coordinate permutation. Note this type of function includes the maximum eigenvalue (or more generally the k th largest eigenvalue) of a matrix.

In finite dimensional Banach spaces, the *limiting subdifferential*

$$\partial_L \varphi(x) := \{x^* \in X : x^* = \lim x_n^*, x_n^* \in D_F \varphi(x_n), (x_n, \varphi(x_n)) \rightarrow (x, \varphi(x))\}$$

defined for lower semicontinuous functions often provides concise access to results via the the Fréchet subdifferential. Moreover, for a Lipschitz function f , $\partial_c f = \overline{\text{co}} \partial_L f$. A central result in in terms of the limiting subdifferential is:

Theorem 3.32 (Subdifferentials of Spectral Functions)

$$D_F(f \circ \lambda)(A) = \{U^T(\text{Diag } \mu)U : U \in \mathcal{O}, U^T(\text{Diag } \lambda(A))U = A, \mu \in D_F f(\lambda(A))\}.$$

Consequently, a limiting process leads to

$$\partial_L(f \circ \lambda)(A) = \{U^T(\text{Diag } \mu)U : U \in \mathcal{O}, U^T(\text{Diag } \lambda(A))U = A, \mu \in \partial_L f(\lambda(A))\}.$$

The proof of this result needs the following few auxiliary lemmas.

Lemma 3.33 *Let U be an orthogonal matrix and let $A \in X$. Then*

$$U^T D_F(f \circ \lambda)(A)U = D_F(f \circ \lambda)(U^T AU).$$

Proof Let $B \in D_F(f \circ \lambda)(A)$. Then, for any $C \in X$,

$$(f \circ \lambda)(A + tUDU^T) = (f \circ \lambda)(A) + t\langle B, A \rangle + o(t\|UDU^T\|).$$

Since similarity transformation does not change the eigenvalues and f is permutation invariant, we may rewrite the above equality as

$$(f \circ \lambda)(U^T AU + tD) = (f \circ \lambda)(U^T AU) + t\langle U^T BU, U^T AU \rangle + o(t\|D\|).$$

That is, $U^T BU \in U^T D_F(f \circ \lambda)(A)U$. As B was arbitrary in $D_F(f \circ \lambda)(A)$ we have shown

$$U^T D_F(f \circ \lambda)(A)U \subset D_F(f \circ \lambda)(U^T AU).$$

The reverse inclusion may be derived by replacing A with $U^T AU$ and U with U^T . ■

Lemma 3.34 *Let $f : \mathbf{R}^n \rightarrow \bar{\mathbf{R}}$ be a permutation invariant closed function. Then $b \in D_F f(a)$ if and only if $\text{Diag } b \in D_F(f \circ \lambda)(\text{Diag } a)$.*

Proof The “only if” part of this Lemma is easy. In fact, for any small vector $c \in \mathbf{R}^n$ we have

$$\begin{aligned} f(a + c) &= (f \circ \lambda)(\text{Diag } a + \text{Diag } c) \\ &\geq (f \circ \lambda)(\text{Diag } a) + \text{trace}(\text{Diag } b)(\text{Diag } c) \\ &\quad + o(\|\text{Diag } c\|) \\ &= f(a) + \langle b, c \rangle + o(\|c\|), \end{aligned}$$

whence $b \in D_F f(a)$.

By contrast, the “if” part is highly non-trivial. Since the proof relates mostly to algebra rather than nonsmooth analysis we refer the interested reader to [104, Section 5]. ■

The next key step is to establish commutativity.

Lemma 3.35 *Let $f : \mathbf{R}^n \rightarrow \bar{\mathbf{R}}$ be a permutation invariant function. Then $B \in D_F(f \circ \lambda)(A)$ implies $AB = BA$.*

Proof In view of the above characterization of $N(M_A, A)$ we need only show that B is in $N(M_A, A)$. Consider any vector T in the tangent space of M_A at A . Then there exists a sequence $(A_n) \subset M_A$ such that $A_n \rightarrow A$ and $(A_n - A)/\|A_n - A\| \rightarrow T/\|T\|$. By the definition of the subdifferential we have

$$(f \circ \lambda)(A_n) = (f \circ \lambda)(A) + \langle B, A_n - A \rangle + o(\|A_n - A\|).$$

Since $f \circ \lambda$ remains constant on M_A we obtain

$$\langle B, A_n - A \rangle = o(\|A_n - A\|).$$

Dividing by $\|A_n - A\|$ and taking limits as $n \rightarrow \infty$ yields $\langle B, T \rangle = 0$, as was to be shown. ■

We may now complete the main result.

Proof of Theorem 3.32 Let b be an arbitrary element of $D_F f(\lambda(A))$. Lemma 3.34 shows

$$\text{Diag } b \in D_F(f \circ \lambda)(\text{Diag } \lambda(A)).$$

For any orthogonal matrix U with $U^T \text{Diag } \lambda(A) U = A$, Lemma 3.33 implies that

$$U^T (\text{Diag } b) U \in D_F(f \circ \lambda)(U^T \text{Diag } \lambda(A) U) = D_F(f \circ \lambda)(A).$$

Thus,

$$\{U^T (\text{Diag } b) U : U \in \mathcal{O}, U^T (\text{Diag } \lambda(A)) U = A, b \in D_F f(\lambda(A))\} \subset D_F(f \circ \lambda)(A).$$

To prove the reverse inclusion, let B be an arbitrary element of $D_F(f \circ \lambda)(A)$. By virtue of Lemma 3.35, B and A commute and *therefore can be diagonalized simultaneously*, that is to say, there exists an orthogonal matrix U and a vector $b \in \mathbf{R}^n$ such that $U^T A U = \text{Diag } A$ and $U^T B U = \text{Diag } b$. Invoking Lemma 3.33 we have

$$\text{Diag } b \in D_F(f \circ \lambda)(\text{Diag } A),$$

whence $b \in D_F f(\lambda(A))$, by Lemma 3.34, as required. ■

Remark 3.36 When f is Lipschitz the formula in Theorem 3.32 holds for the Clarke generalized gradient, by a similar argument (see [104]).

To apply the characterization of Theorem 3.32 we need, of course, to compute the subdifferentials of f . In the case of the k -th largest eigenvalue we have to compute the subdifferentials of the k -th largest member of a vector.

Lemma 3.37 *At any point $x \in \mathbf{R}^n$,*

$$D_F \phi_k(x) = \begin{cases} \text{conv}\{e^i : x_i = \phi_k(x)\}, & \text{if } \phi_{k-1}(x) > \phi_k(x), \\ \emptyset, & \text{otherwise,} \end{cases}$$

$$\partial_L \phi_k(x) = \{y \in \text{conv}\{e^i : x_i = \phi_k(x)\} : \#(\text{supp } y) \leq \alpha\},$$

where $\alpha := 1 - k + \#\{i : x_i \geq \phi_k(x)\}$, and

$$\partial_c \phi_k(x) = \text{conv}\{e^i : x_i = \phi_k(x)\}.$$

The proof of this lemma is elementary. Details can be found in [104]. From this result we have:

Theorem 3.38 *Let $A \in X := S(n)$. Then*

$$\partial_c \lambda_k(A) = \text{conv}\{uu^T : u \in \mathbf{R}^n, \|u\| = 1, Au = \lambda_k(A)u\},$$

and

$$\partial_L \lambda_k(A) = \{B \in \partial_c \lambda_k(A) : \text{rank} B \leq \alpha\},$$

where $\alpha := 1 - k + \#\{i : \lambda_i(A) \geq \phi_k(\lambda(A))\}$.

Example 3.39 Consider a very simple case when A is the two by two unit matrix. Then

$$\partial_L \lambda_2(A) = \{uu^T : u \in \mathbf{R}^2, \|u\| = 1\}$$

and

$$\partial_c \lambda_2(A) = \text{conv}\{uu^T : u \in \mathbf{R}^2, \|u\| = 1\}.$$

Note $\partial_c \lambda_2(A)$ contains nonsingular matrices of the form $\text{Diag}(a, 1 - a)$ for $a \in (0, 1)$ while all the elements of $\partial_L \lambda_2(A)$ are singular matrices.

These formulae show the striking difference between the Clarke generalized gradient and the limiting subdifferential and hint at the greater utility in this setting of the latter. We should also observe that in the convex case we have the beautiful and puissant formula from [103]:

$$(h \circ \lambda)^* = h^* \circ \lambda.$$

In particular, for $p, q > 1$ satisfying $1/p + 1/q = 1$ we have $\|\cdot\|_p^* = \|\cdot\|_q$. Thus, if we denote the Banach space of $S(n)$ endowed with the norm $\|\cdot\|_p \circ \lambda$ by B_p then $B_p^* = B_q$. Finally, we observe that much of this beautiful theory can be lifted – with some difficulty – to the study of *trace class* or *Hilbert-Schmidt* operators on separable Hilbert space. This makes it applicable in many physically interesting contexts, see [133], and for infinite dimensional interior point methods.

4 Essentially smooth Lipschitz functions

An important notion related to the Clarke generalized gradient is (dense) representability. A real-valued locally Lipschitz function, f , defined on a non-empty open subset A of a normed linear space X is *D-representable* on A if:

- (a) $D \equiv \{x \in A : \nabla f(x) \text{ exists}\}$ is dense in A ; and
- (b) $\partial_c f = \text{CSC}(\Omega_{D^*})$, for each dense subset D^* of D , with $\Omega_{D^*} : D^* \rightarrow 2^{X^*}$ given by $\Omega_{D^*}(x) \equiv \{\nabla f(x)\}$.

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

The following result from [27] characterizes the D -representability.

Theorem 4.1 *Let f be a real-valued locally Lipschitz function defined on a nonempty open subset A of a class(S) Banach space X . Let $D \equiv \{x \in A : \nabla f(x) \text{ exists}\}$. Then f is D -representable on A if and only if $x \rightarrow \partial_c f(x)$ is minimal.*

We will see that a reasonably large class of functions possesses subdifferential mappings that are minimal. It is natural to ask whether this class of functions is closed under addition, multiplication or either of the lattice operations? And whether Clarke generalized gradients within this class of functions are *integrable*, i.e. determine a function up to a constant (on a connected set)? The answer to all these questions is ‘no’. An example is given below in Lecture 4.1.

Therefore, to “cage such monsters” we hunt for a class of functions which does have the desired property. It turns out that essentially strictly differentiable Lipschitz functions form such a class. We define and discuss the properties of such functions in Lecture 4.2. We shall see that essentially smooth functions are D -representable, integrable and obey reasonable chain rules.

4.1 An example

Example 4.2 Let C be a Cantor subset of $[0, 1]$ (symmetric about $\frac{1}{2}$) with $1 > \mu(C) > 0$, and let $\{(a_n, b_n) : n \in \mathbf{N}\}$ be an enumeration of the disjoint open intervals of $[0, 1] \setminus C$. Further, for each $n \in \mathbf{N}$, let $c_n \equiv (a_n + b_n)/2$ and $d_n \equiv (b_n - a_n)^2/2$. Now, consider the Lipschitz functions $f : (0, 1) \rightarrow [0, 1]$ and $g : (0, 1) \rightarrow [-1, 1]$ defined by

$$g(x) \equiv \begin{cases} 0 & \text{if } |x - c_n| \geq d_n \text{ for all } n. \\ 2(x - (c_n - d_n)) & \text{if } x \in (c_n - d_n, c_n - \frac{2}{3}d_n] \\ -(x - c_n) & \text{if } x \in (c_n - \frac{2}{3}d_n, c_n] \\ -2(x - c_n) & \text{if } x \in (c_n, c_n + \frac{1}{3}d_n] \\ x - (c_n + d_n) & \text{if } x \in (c_n + \frac{1}{3}d_n, c_n + d_n) \end{cases}$$

and $f(x) \equiv \int_0^x \theta(t) dt$, where

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

Then

- (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 $h \equiv f + g$ and $k \equiv f - g$, 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 $j : [0, 1] \rightarrow [0, 2]$ is defined by $j(x) \equiv f(x) + 1$, then the functions $j + g$ and $j - g$ possess minimal subdifferential mappings on $(0, 1)$, but $(j + g) \cdot (j - g)$ 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, $x \rightarrow \partial g(x)$, 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, $\partial g = CSC((\partial g))$. Therefore, we may conclude, from Corollary 2.17 that ∂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 [50], $\partial(f + g)(x) \subseteq \partial g(x)$ for each $x \in (0,1)$. However, since $x \rightarrow \partial g(x)$, 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 $\partial(f - g)(x) = (-1) \cdot \partial(g - f)(x)$ 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 ∂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,

$$|g|(x) \equiv \begin{cases} 0 & \text{if } |x - c_n| \geq d_n \text{ for all } n. \\ 2(x - (c_n - d_n)) & \text{if } x \in (c_n - d_n, c_n - \frac{2}{3}d_n] \\ -(x - c_n) & \text{if } x \in (c_n - \frac{2}{3}d_n, c_n] \\ 2(x - c_n) & \text{if } x \in (c_n, c_n + \frac{1}{3}d_n] \\ (c_n + d_n) - x & \text{if } x \in (c_n + \frac{1}{3}d_n, c_n + d_n). \end{cases}$$

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, ∂M is not generated by the derivatives chosen from $(0,1) \setminus C$ (which are dense in $[0,1]$), that is, M is not D -representable on $(0,1)$, and so M 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, $(j + g) \cdot (j - g) = j^2 - g^2$, 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 $(j^2 - g^2)'(x) = 2(j(x)j'(x) + g(x)g'(x)) = 2j(x)j'(x) \geq 4$ at each point of A . Hence, as in (4), it follows that $(j + g) \cdot (j - g)$ is not D -representable, and so $(j + g) \cdot (j - g)$ does not possess a minimal subdifferential mapping. ■

4.2 Essentially strictly differentiable Lipschitz functions

We now identify 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 [56] and [110]. In this way, we are able to generalize, in a unified manner, the various results contained in places such as [56], [112], [20], [59], [94], [126], [122] and [11] as well as [129] (at least in the case of Lipschitz functions). First we recall the following results from Lecture 1.

Theorem 4.3 [139, 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:*

$$\partial_c f(x) = \overline{co}^{w^*} \{x^* \in X^* : x^* = w^* - \lim_{x_n \rightarrow x} \nabla f(x_n), \text{ and } x_n \in D \setminus N\}$$

We then have:

Corollary 4.4 *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, $x \rightarrow \partial f(x)$, is a minimal weak* cusco on A if and only if the mapping, $x \rightarrow \nabla f(x)$ (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, $x \rightarrow \nabla f(x)$, is defined, and weak* hyperplane minimal on $D \setminus N$. Then, by Theorem 4.3, we have that $\partial f = \text{CSC}(\nabla f)$. The result now follows from Theorem 2.18 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 ([11], 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 [126], 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 4.5 *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 $x \rightarrow \nabla f(x)$ (defined on D) is norm to weak* continuous almost everywhere in D .*

Proof This follows from Theorem 4.3 and the fact that $\partial_c f(x)$ is a singleton iff f is strictly differentiable at x . ■

Corollary 4.6 [11, Corollary 3.10] *Let A be a non-empty open subset of a separable Banach space X . Then each member of $S_e(A)$ possesses a minimal subdifferential mapping.*

Proof This follows from Theorem 4.5 and Corollary 4.4. ■

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,

$$\begin{aligned} d^+ f(x; y) &\geq \limsup_{t \rightarrow 0^+} f^0(x + ty; y) \\ \left(d^- f(x; y) \leq \liminf_{t \rightarrow 0^+} -f^0(x + ty; -y) \right). \end{aligned}$$

Remark 4.7 If we define $T_y : A \rightarrow \mathbf{R}$ by, $T_y(x) \equiv \limsup_{t \rightarrow 0^+} f^0(x + ty; y)$, and $S_y : A \rightarrow \mathbf{R}$ by, $S_y(x) \equiv \liminf_{t \rightarrow 0^+} -f^0(x + ty; -y)$, 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:

$$T_y(x) = \lim_{n \rightarrow \infty} g_n(x)$$

where $g_n(x) \equiv \sup\{f^0(x + ty; y) : 0 < t \leq 1/n\}$ and that $g_n(x) \equiv \lim_{m \rightarrow \infty} f_m^n(x)$ where $f_m^n(x) \equiv \max\{f^0(x + ty; y) : 1/m \leq t \leq 1/n\}$ (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, $f^0(x; y) = d^+ f(x; y)$ and we shall say that f is *pseudo-regular at x* , if it is pseudo-regular at x , in every direction y .

Lemma 4.8 *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) sets*

$$F_y \equiv \{x \in A : f^0(x; y) > T_y(x)\}$$

and

$$E_y \equiv \{x \in A : -f^0(x; -y) < S_y(x)\}$$

have the property that for each $a \in A$, $F_y(a) \equiv \{r \in \mathbf{R} : a + ry \in F_y\}$ and $E_y(a) \equiv \{r \in \mathbf{R} : a + ry \in E_y\}$ are 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 case of $E_y(a)$ is identical). Note that without loss of generality we may assume that $F_y(a)$ is non-empty. So in this case, we define $s : F_y(a) \rightarrow \mathbf{Q}^2$ by, $s(t) \equiv (r_1, r_2)$ where $r_1 \in (T_y(a + ty), f^0(a + ty; y)) \cap \mathbf{Q}$ and $r_2 \in (t, \infty) \cap \mathbf{Q}$ is chosen so that

$$\sup\{f^0(a + ry; y) : t < r < r_2\} < r_1$$

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

Remark 4.9 If X is a separable Banach space, then for each $y \in S(X)$,

$$\{x \in A : f \text{ upper hemi-smooth but not pseudo-regular in direction } y\}$$

is contained in F_y and hence is a Haar-null set (in this case, we may take the normalized Lebesgue measure, supported on $sp\{y\}$, as a test-measure for the Borel set F_y).

Proposition 4.10 *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 $\{y_n : n \in N\}$ be a dense subset of $S(X)$. If for each $n \in N$, f is almost everywhere pseudo-regular in the direction y_n , then $f \in S_e(A)$.*

Proof For each $n \in N$, 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 our Rademacher result, $A \setminus D$ is a Haar-null set. Now, let $S \equiv \bigcap \{P_n : n \in N\} \cap D$. 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 $n \in N$. However, since both mappings, $y \rightarrow f^0(x_0; y)$, and $y \rightarrow \nabla f(x_0)(y)$, 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 4.11 *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 $f \in S_e(A)$.*

Proof The proof follows from Lemma 4.8 (and the subsequent remark) and Proposition 4.10. ■

Next, we show that each member of $S_e(A)$ is integrable.

Proposition 4.12 [11, 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 $\partial_c g(x) \subseteq \partial_c f(x)$ 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 $\partial_c f(x)$ is a singleton. The result now follows directly from Theorem 4.3. ■

4.3 Stability properties for $S_e(A)$ and a chain rule

We turn to discuss stability properties for $S_e(A)$.

A first but naïve guess might be that if $f_1, f_2, \dots, f_n \in S_e(A)$ and $g \in S_e(\mathbf{R}^n)$, then $g \circ f \in S_e(A)$, where $f \equiv (f_1, f_2, \dots, f_n)$. The following example shows that in general this is not true (when $n \geq 2$).

Example 4.13 Let X be a separable Banach space let C be a Cantor subset of \mathbf{R} with positive Lebesgue measure. Define the functions f_1, \dots, f_n on X by $f_n \in X^* \setminus \{0\}$ and $f_j \equiv 0$ for each $1 \leq j < n$. Further, we define $d_C : \mathbf{R} \rightarrow \mathbf{R}$ by $d_C \equiv \text{dist}(\cdot, C)$ and $g : \mathbf{R}^n \rightarrow \mathbf{R}$ by

$$g \equiv \text{dist}(\cdot, \{0\} \times \{0\} \times \dots \times C)$$

(the distance is with respect to the Euclidean norm on \mathbf{R}^n). Clearly, each f_j is strictly differentiable on A . Moreover, by Theorem 4.32 we have $g \in S_e(\mathbf{R}^n)$. We claim $g \circ f \notin S_e(X)$, where $f \equiv (f_1, f_2, \dots, f_n)$. To see this, observe that $g \circ f(x) = d_C(f_n(x))$. 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 [46]). Therefore, $g \circ f \notin S_e(X)$.

Despite this example $S_e(A)$ does possess very strong closure properties. We start with the following simple characterization of functions in $S_e(A)$.

Theorem 4.14 *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 $\overline{\text{sp}} B = X$. If for each $b \in B$, $f^0(x; b) = -f^0(x; -b)$ almost everywhere in A , then $f \in \mathcal{S}_e(A)$.*

Proof 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,

$$f^0(x_0; y) = -f^0(x_0; -y) \text{ for each } y \in \overline{\text{sp}} Y. \quad \blacksquare$$

In the sequel we will need to consider vector-valued functions. Let $A \subseteq \mathbf{R}$ and $x : A \rightarrow \mathbf{R}^n$ be defined by,

$$x(t) \equiv (x_1(t), x_2(t), \dots, x_n(t)) \quad \text{where } x_j : A \rightarrow \mathbf{R}.$$

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

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

where $x'(t) \equiv (x'_1(t), x'_2(t), \dots, x'_n(t))$. We shall denote by $\mathcal{A}_e(U)$, the family of all arc-wise essentially smooth functions on U .

Remark 4.15 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 \mathbf{R} .

Theorem 4.16 *Let A be a non-empty open subset of a separable Banach space X . If $f_1, f_2, \dots, f_n \in \mathcal{S}_e(A)$ and U is a non-empty open subset of \mathbf{R}^n , that contains $f(A)$, where $f \equiv (f_1, f_2, \dots, f_n)$, then for each $g \in \mathcal{A}_e(U)$, $g \circ f \in \mathcal{S}_e(A)$.*

Proof It suffices (see, Theorem 4.14) 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 . So fix $y \in S(X)$. Let D be the (G_δ) 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 $P_y \equiv \{x \in A : (g \circ f)^0(x; y) = -(g \circ f)^0(x; -y)\}$. Clearly P_y is a Borel set, in fact P_y is a G_δ set. Let H be any closed hyperplane in X such that $y \notin H$. Now consider the isomorphism $T : H \times \mathbf{R} \rightarrow X$ defined by $T(h, t) \equiv h + ty$. Let

$$H_D \equiv \{h \in H : \lambda(\{t \in \mathbf{R} : T(h, t) \in A \setminus D\}) = 0\}.$$

By our Fubini theorem 1.22 we see that $H \setminus H_D$ is a Haar-null set in H , since $A \setminus D$ is a Haar-null set in X . To show that $A \setminus P_y$ is a Haar-null set in X it thus suffices to show that for each $h \in H_D$, $\lambda(\{t \in \mathbf{R} : T(h, t) \in A \setminus P_y\}) = 0$. To this end, consider $h_0 \in H_D$. Let $A_{h_0} \equiv \{t \in \mathbf{R} : T(h_0, t) \in A\}$. If $A_{h_0} = \emptyset$ then we are done, so let us suppose that $A_{h_0} \neq \emptyset$. Define $z : A_{h_0} \rightarrow U$ by, $z(t) \equiv f(h_0 + ty)$. Since $h_0 \in H_D$, $z \in \mathcal{S}_e(A_{h_0}; \mathbf{R}^n)$. Let

$$D_f \equiv \bigcap_{j=1}^n \{t \in A_{h_0} : f_j^0(h_0 + ty; y) = -f_j^0(h_0 + ty; -y)\}$$

and

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

Now define $D_0 \equiv D_f \cap D_g$. 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 $x_0 \equiv h_0 + t_0 y$. Then

$$g^0(f(x_0); f'(x_0; y)) = -g^0(f(x_0); -f'(x_0; y))$$

as $t_0 \in D_g$ and $f_j^0(x_0; y) = -f_j^0(x_0; -y)$ for each $1 \leq j \leq n$ as $t_0 \in D_f$. It is now standard that

$$(g \circ f)^0(x_0; y) = -(g \circ f)^0(x_0; -y)$$

This completes the proof. ■

That Theorem 4.16 provides us with strong closure properties (for $\mathcal{S}_e(A)$) derives from the following proposition.

Proposition 4.17 [28] *Let U be a non-empty open subset of \mathbf{R}^n . Then*

- (a) $\mathcal{A}_e(U) \subseteq \mathcal{S}_e(U)$;
- (b) $\mathcal{A}_e(U)$ is closed under composition. That is, if $f \equiv (f_1, f_2, \dots, f_n)$ with $f_1, f_2, \dots, f_n \in \mathcal{A}_e(U)$ and $g \in \mathcal{A}_e(\mathbf{R}^n)$ then $g \circ f \in \mathcal{A}_e(U)$;
- (c) $\mathcal{A}_e(U)$ 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*) on A , if for each $x \in A$ and $y \in S(X)$

$$d^+ f(x; y) \geq \limsup_{\substack{t \rightarrow 0^+ \\ y' \rightarrow y}} d^+ f(x + ty'; y)$$

$$\left(d^- f(x; y) \leq \liminf_{\substack{t \rightarrow 0^+ \\ y' \rightarrow y}} d^- f(x + ty'; y) \right).$$

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

Corollary 4.18 *Let A be a non-empty open subset of a separable Banach space X , then $\mathcal{S}_e(A)$ 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 \mathbf{R}^2 . ■

Although in general, $\mathcal{S}_e(A)$ is not closed under composition, we have from the next theorem, that if $f \in \mathcal{S}_e(A)$ and $g \in \mathcal{S}_e(\mathbf{R})$, then $g \circ f \in \mathcal{S}_e(A)$.

Theorem 4.19 *If U is a non-empty open subset of \mathbf{R} then $\mathcal{A}_e(U) = \mathcal{S}_e(U)$.*

Proof We see from the above proposition that $\mathcal{A}_e(U) \subseteq \mathcal{S}_e(U)$ and so we need only show the reverse inclusion. To this end let $f \in \mathcal{S}_e(U)$ and let $x \in \mathcal{S}_e((0, 1); U)$. Now define, $C \equiv \{t \in U : f^0(t; 1) = -f^0(t; -1)\}$ and $D \equiv \{t \in (0, 1) : x(t) \in C \text{ or } d^+x(t) = 0\}$. We need to show that $\lambda(D) = 1$, since $f^0(x(t); x'(t)) = -f^0(x(t); -x'(t))$ 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 [136]). ■

Remark 4.20 It follows from Theorem 4.19 that the (distance) function g defined in Example 4.13 lies in $\mathcal{S}_e(\mathbf{R}^n)$ but not in $\mathcal{A}_e(\mathbf{R}^n)$ ($n \geq 2$). Thence, by translation and dilation one can show that for every non-empty open subset U of \mathbf{R}^n ($n \geq 2$), $\mathcal{A}_e(U)$ is a proper subset of $\mathcal{S}_e(U)$.

The next few lemmas leading to a chain rule are quite standard.

Lemma 4.21 *Let U be a nonempty open subset of a Banach space X and $f \equiv (f_1, f_2, \dots, f_m)$ be a locally Lipschitz mapping from U into \mathbf{R}^m . Furthermore, let g be a real-valued locally Lipschitz function defined on a non-empty open subset V of \mathbf{R}^m which contains $f(U)$.*

If f is differentiable at some point x_0 , in the direction y and 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 4.22 Using Lemma 4.21 and Proposition 2.27 we see that for each fixed $y \in X$,

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

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

Lemma 4.23 [29, Theorem 3.1] *Let f be a real-valued locally Lipschitz function defined on a non-empty open subset U of a Banach space X . If $U \setminus S$ is a Haar-null set, then for each $x \in U$,*

$$f^0(x; y) = \limsup_{\substack{z \rightarrow x \\ z \in S}} d^+ f(z; y), \quad -f^0(x; -y) = \liminf_{\substack{z \rightarrow x \\ z \in S}} d^+ f(z; y).$$

Remark 4.24 By Lemma 4.23 we see that: $f^0(x; y) = -f^0(x; -y)$ if, and only if,

$$\lim_{\substack{z \rightarrow x \\ z \in S}} f^+(z; y) \text{ exists and equals } f'(x; y).$$

We can now prove a chain rule.

Theorem 4.25 *Let U be a nonempty open subset of a Banach space X and $f \equiv (f_1, f_2, \dots, f_m)$ a locally Lipschitz mapping from U into \mathbf{R}^m . Furthermore, let g be a real-valued locally Lipschitz function defined on a non-empty open subset V of \mathbf{R}^m which contains $f(U)$.*

If f is strictly differentiable at some point $x_0 \in U$, in the direction y and

$$g^0(f(x_0); f'(x_0; y)) = -g^0(f(x_0); -f'(x_0; y))$$

then $g \circ f$ is strictly differentiable at x_0 , in the direction y .

Proof Let $D \equiv \{x \in U : (g \circ f)'(x; y) = g'(f(x); f'(x; y))\}$. It follows from Remark 4.22 that D is 1- D almost everywhere in U , in the direction y . (i) Let us first observe that $x_0 \in D$. From Remark 4.24 we see that to show $(g \circ f)^0(x_0; y) = -(g \circ f)^0(x_0; -y)$ we need only show that

$$\lim_{\substack{z \rightarrow x_0 \\ z \in D}} (g \circ f)'(z; y) \text{ exists, and equals } (g \circ f)'(x_0; y).$$

So suppose $\epsilon > 0$. By the continuity of the mapping, $x \rightarrow g^+(x; f'(x_0; y))$, at $f(x_0)$ there exists an open neighbourhood W_1 of x_0 such that

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

for each $z \in W_1$. Now, as g is locally Lipschitz there exists an open neighbourhood W_2 of x_0 and $M > 0$ such that

$$\begin{aligned} |g'(f(z); f'(z; y)) - g^+(f(z); f'(x_0; y))| \\ \leq M \|f'(z; y) - f'(x_0; y)\| \end{aligned}$$

for each $z \in D \cap W_2$. On the other hand, $x \rightarrow f^+(x; y)$, is continuous at x_0 . Thus there is an open neighbourhood W_3 of x_0 with

$$\|f'(x_0; y) - f^+(x; y)\| < \epsilon/2M$$

for each $x \in W_3$. Hence, for each $x \in D \cap W_1 \cap W_2 \cap W_3$

$$\begin{aligned} |(g \circ f)'(x_0; y) - (g \circ f)'(x; y)| &= |g'(f(x_0); f'(x_0; y)) - g'(f(x); f'(x; y))| \\ &\leq |g'(f(x_0); f'(x_0; y)) - g^+(f(x_0); f'(x_0; y))| \\ &\quad + |g^+(f(x); f'(x_0; y)) - g'(f(x); f'(x; y))| \\ &\leq \epsilon/2 + \epsilon/2 = \epsilon. \end{aligned}$$

Therefore, $\lim_{\substack{z \rightarrow x_0 \\ z \in D}} (g \circ f)'(z; y) = (g \circ f)'(x_0; y)$. ■

If all the functions f_j and g are pseudo-regular, then using Lemma 4.20, the conclusions hold almost everywhere (for a fixed direction).

Several general comments regarding essentially strictly differentiable functions follow. Our first comment concerns our choice of *null* set. Our choice of Haar-null sets (as defined by J. P. R. Christensen) as our ‘null’ sets was reasonably arbitrary (except that the larger the class of null sets, the larger $\mathcal{S}_\epsilon(A)$ becomes; in [121] 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 [7] for further information). In fact, the only properties that we required of our σ -ideal of null sets were;

- (i) no open set is a null set;
- (ii) the formula in Theorem 4.3 holds;
- (iii) a Borel set $A \subseteq H \times \mathbf{R}$ is a null set if and only if

$$\lambda(\{t \in \mathbf{R} : (h, t) \in A\}) = 0 \quad \text{for almost all } h \in H.$$

Our next 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 $x \in X$.) Now we can define the essentially smooth functions on any Banach space X , in the following manner.

We 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 $y \in S(X)$, $\{x \in A : f^0(x; y) \neq -f^0(x; -y)\}$ is a Haar-null set. Using this many of the results on \mathcal{S}_e extend to nonseparable Banach spaces, [29] and [30].

Another tool is the *separable reduction* method of [31]. Let f be a locally Lipschitz function defined on a non-empty open set A of a Banach space X . We will call a family \mathcal{F} of closed separable subspaces of a Banach space X *rich* if:

- (i) for each increasing sequence of closed separable subspaces $\{Y_n : n \in \mathbf{N}\}$ in \mathcal{F} ,

$$\overline{\cup\{Y_n : n \in \mathbf{N}\}} \in \mathcal{F};$$

- (ii) for each separable subspace Y^0 of X there exists a $Y \in \mathcal{F}$ such that $Y^0 \subseteq Y$.

For a non-empty open subset A of X we shall denote by $\mathcal{R}_e(A)$ the family of all those locally Lipschitz functions f on A for which there exists a rich family \mathcal{F} (possibly depending on f) of closed separable subspaces Y of X such that $f|_{A \cap Y} \in \mathcal{S}_e(A \cap Y)$ for each Y in \mathcal{F} with $Y \cap A \neq \emptyset$.

It follows from the separable reduction results in [31] that the members of \mathcal{R}_e share many of the same properties as \mathcal{S}_e .

4.4 Perturbation functions

We apply the results of previous lectures 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 $g : A \times T \rightarrow \mathbf{R}$ 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

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

and $t \in T$. Further, we say that an extended real-valued function f defined on A is a *sup-marginal function* if, $f(x) \equiv \sup\{g(x, t) : t \in T\}$ for some function $g : A \times T \rightarrow \mathbf{R}$.

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 $(x_\alpha)_{\alpha \in I}$ in A , converging to x , there exists a point $y \in M(x)$ and elements $y_\alpha \in M(x_\alpha)$ such that y is an accumulation point of the set $\{y_\alpha : \alpha \in I\}$, that is, $y \in \overline{\{y_\alpha : \alpha \in I\}} \setminus \{y\}$.

The following theorem unifies Theorems 6.1 and 6.2 in [56] and Proposition 2.6 in [59].

Theorem 4.26 *Let A be a non-empty open subset of a separable Banach space X and let T be a Hausdorff topological space. Let $g : A \times T \rightarrow \mathbf{R}$ be locally Lipschitz on A , uniformly in T and let $f : A \rightarrow \mathbf{R}$ be defined by*

$$f(x) \equiv \max\{g(x, t) : t \in T\}.$$

Suppose that (i) the set-valued mapping $M : A \rightarrow 2^T$, defined by, $M(x) \equiv \{t \in T : f(x) = g(x, t)\}$ is semi-continuous on A and that (ii) for each $x \in A$ and each $y \in B \cup -B$, $(r, y', t) \in \mathbf{R}^+ \times X \times T \rightarrow d^+g(x + ry', t; y)$, 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 $\overline{\text{sp}} B = X$.)

Then $f \in \mathcal{S}_e(A)$.

Proof To show that $f \in \mathcal{S}_e(A)$, 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 $\{s_n : n \in \mathbf{N}\}$ converging to 0 and any sequence $\{y_n : n \in \mathbf{N}\}$ of elements of X converging to y , we have that

$$\liminf\{d^+f(x + s_n y_n; y) : n \rightarrow \infty\} \leq d^+f(x; y).$$

Indeed, by a standard subsequence argument this will show that f is upper semi-smooth at x , in the direction y .

So let $\{s_n : n \in \mathbf{N}\}$ be a sequence of positive real numbers converging to 0 and let $\{y_n : n \in \mathbf{N}\}$ be a sequence of elements of X converging to y . For each $n \in \mathbf{N}$, we may choose $0 < \lambda_n < s_n$ such that,

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

Since M is semi-continuous on A and

$$\lim_{n \rightarrow \infty} (x + s_n y_n + \lambda_n y) = x$$

there exists a point $t \in M(x)$ and a sequence $\{t_n : n \in \mathbf{N}\}$ in T such that $t_n \in M(x + s_n y_n + \lambda_n y)$ for each $n \in \mathbf{N}$ and $t \in \overline{\{t_n : n \in \mathbf{N}\}} \setminus \{t\}$. Now, for each $n \in \mathbf{N}$, 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 $n \in \mathbf{N}$ there exists a real number s'_n such that $0 < s'_n < \lambda_n$ and

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

Therefore, for each $n \in \mathbf{N}$,

$$d^+f(x + s_n y_n; y) \leq d^+g(x + s_n y_n + s'_n y, t_n; y) + 2/n$$

Now, let $s''_n \equiv (s_n + s'_n)$ and $y'_n \equiv (s_n y_n + s'_n y)/s''_n$. Then clearly, $\lim_{n \rightarrow \infty} y'_n = y$ and $\lim_{n \rightarrow \infty} s''_n = 0$. Hence,

$$\begin{aligned} \liminf_{n \rightarrow \infty} d^+f(x + s_n y_n; y) &\leq \liminf_{n \rightarrow \infty} d^+g(x + s_n y_n + s'_n y, t_n; y) \\ &= \liminf_{n \rightarrow \infty} d^+g(x + s''_n y'_n, t_n; y) \\ &\leq d^+g(x, t; y) \\ &\leq d^+f(x; y) \quad (\text{since } t \in M(x)). \quad \blacksquare \end{aligned}$$

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

4.5 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 4.27 *Let $(X, \|\cdot\|)$ be a Banach space. If each distance function on X possesses a minimal subdifferential mapping, then the norm $\|\cdot\|$ on X is smooth.*

Proof Suppose that the norm $\|\cdot\|$ is not smooth at a point $x_0 \in S(X)$, (Note, there is no loss of generality in assuming that $x_0 \in S(X)$.) Then there exist two distinct linear functionals $x_1^* \text{ and } x_2^* \in S(X^*)$ such that $x_1^*(x_0) = x_2^*(x_0) = 1$. Let $x_3^* \equiv 1/2(x_1^* + x_2^*)$. Let $K_1 \equiv \ker(x_1^*)$, $K_2 \equiv \ker(x_2^*)$ and $K_3 \equiv \ker(x_3^*)$. 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 page 216, Example 6 part(e) of [136], (see also, Example 4.1) an example is given of an everywhere differentiable Lipschitz function $f : \mathbf{R} \rightarrow \mathbf{R}$ which is strictly increasing on \mathbf{R} and for which the set $\{x \in \mathbf{R} : f'(x) = 0\}$ is dense in \mathbf{R} . Moreover, this function f is a *strict contraction* on \mathbf{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 $x = k_x + \lambda_x z + \mu_x x_0$, where $k_x \in K_1 \cap K_2$ and $\lambda_x, \mu_x \in \mathbf{R}$. Furthermore, $\mu_x = x_3^*(x)$ and $\lambda_x = 1/2(x_1^*(x) - x_2^*(x))$, and so, both mappings, $x \rightarrow \mu_x$, and, $x \rightarrow \lambda_x$, are continuous and open on X . Let

$$C \equiv \{x \in X : \mu_x \geq f(\lambda_x)\},$$

(It is instructive to think of C as the epigraph of the real-valued function $f_* : K_3 \rightarrow \mathbf{R}$, defined by, $f_*(k + \lambda z) \equiv f(\lambda)$.) Clearly, C is a proper, non-empty closed subset of X .

We will show that, $x \rightarrow \partial_c d_C(x)$, is not a minimal weak* cusco on $X \setminus C$.

We claim that $\sigma : X \setminus C \rightarrow C$, 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:

$$\mu_{\sigma(x)} = \mu_x + (f(\lambda_x) - \mu_x) = f(\lambda_x) = f(\lambda_{\sigma(x)}) \quad (*)$$

since $\lambda_x = \lambda_{\sigma(x)}$. 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

$$\begin{aligned} T_x &\equiv \{y \in X : x_j^*(\sigma(x)) \leq x_j^*(y), j = 1 \text{ or } 2\} \\ &= \{y \in X : \mu_{\sigma(x)} + \lambda_{\sigma(x)} \leq \mu_y + \lambda_y \text{ or } \mu_{\sigma(x)} - \lambda_{\sigma(x)} \leq \mu_y - \lambda_y\} \end{aligned}$$

We will show that $C \subseteq T_x$. To this end, consider $y \in C$, then either,

$$(i) \quad f(\lambda_{\sigma(x)}) - f(\lambda_y) \leq \lambda_{\sigma(x)} - \lambda_y$$

or

$$(ii) \quad f(\lambda_{\sigma(x)}) - f(\lambda_y) \leq \lambda_y - \lambda_{\sigma(x)}$$

Case(i) $f(\lambda_{\sigma(x)}) - \lambda_{\sigma(x)} \leq f(\lambda_y) - \lambda_y$. By (*) $f(\lambda_{\sigma(x)}) = \mu_{\sigma(x)}$ and since $y \in C$, $f(\lambda_y) \leq \mu_y$. Therefore, $\mu_{\sigma(x)} - \lambda_{\sigma(x)} \leq \mu_y - \lambda_y$.

Case(ii) $f(\lambda_{\sigma(x)}) + \lambda_{\sigma(x)} \leq f(\lambda_y) + \lambda_y$. As before, we have that $f(\lambda_{\sigma(x)}) = \mu_{\sigma(x)}$ and $f(\lambda_y) \leq \mu_y$. Therefore, $\mu_{\sigma(x)} + \lambda_{\sigma(x)} \leq \mu_y + \lambda_y$. 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, since $\|x_1^*\| = 1$

$$\begin{aligned} x_1^*(y) &= x_1^*(x) + x_1^*(y - x) < x_1^*(x) + (f(\lambda_x) - \mu_x) \\ &= x_1^*(\sigma(x) - (f(\lambda_x) - \mu_x)x_0) + (f(\lambda_x) - \mu_x) \\ &= x_1^*(\sigma(x)). \end{aligned}$$

And since $\|x_2^*\| = 1$

$$\begin{aligned} x_2^*(y) &= x_2^*(x) + x_2^*(y - x) < x_2^*(x) + (f(\lambda_x) - \mu_x) \\ &= x_2^*(\sigma(x) - (f(\lambda_x) - \mu_x)x_0) + (f(\lambda_x) - \mu_x) \\ &= x_2^*(\sigma(x)). \end{aligned}$$

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,

$$\nabla d_C(x) = f'(x_0^*(x)) \cdot x_0^* - x_3^*$$

on $X \setminus C$, where $x_0^* \equiv 1/2(x_1^* - x_2^*)$.

Now, if $x \rightarrow \partial_c d_C(x)$ were a minimal weak* cusco on $X \setminus C$ then $x \rightarrow \nabla d_C(x)$ would be hyperplane minimal on $X \setminus C$, but then $x \rightarrow f'(x_0^*(x)) \cdot x_0^*$ would be hyperplane minimal on $X \setminus C$. However, since x_0^* is both continuous and open on X , this would imply that, $t \rightarrow f'(t)$, is hyperplane minimal on some non-empty open subset of \mathbf{R} (this follows from the general fact if $\Phi \circ T$ is hyperplane minimal and T is both continuous and open, then Φ is hyperplane minimal), but we know this is not true (by Example 4.2). Therefore we may conclude that $x \rightarrow \partial_c d_C(x)$ is not a minimal weak* cusco on X . ■

Remark 4.28 It is interesting to observe the following facts about the set C constructed in Theorem 4.27:

- (a) d_C is Gâteaux differentiable on $X \setminus C$;
- (b) $\sigma(x) \equiv x + (f(\lambda_x) - \mu_x)x_0$ is Lipschitz-continuous on $X \setminus C$, and this means that C is *almost convex* (see [149] or [76], p. 240);
- (c) $\partial_c d_C(x) = \partial_c f(x_0^*(x)) \cdot x_0^* - x_3^*$ on $X \setminus C$. In particular, d_C is not integrable on $X \setminus C$. For example, let

$$d_*(x) \equiv h(x_0^*(x)) - x_3^*(x),$$

where $h : \mathbf{R} \rightarrow \mathbf{R}$ is chosen so that $h - f$ is not a constant function on $x_0^*(X \setminus C)$ and $\partial_c h(t) \subseteq \partial_c f(t)$ for each $t \in \mathbf{R}$, then $\partial_c d_*(x) \subseteq \partial_c d_C(x)$ for each $x \in X \setminus C$, but $d_* - d_C$ is not a constant function on $X \setminus C$ (note, $h \equiv 0$ will do the job).

So we see then, that even in \mathbf{R}^2 there are distance functions whose Clarke subdifferential mappings are not minimal (of course there are no such examples on \mathbf{R}). However, the situation is dramatically better for smooth norms. Recall, that a normed linear space X is said to have a *uniformly Gâteaux differentiable* (UG) 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

$$\left| \frac{\|x + ty\| - \|x\|}{t} - f_x(y) \right| < \varepsilon \quad \text{for all } 0 < t < \delta(\varepsilon, y)$$

Every Hilbert space and L_p space ($1 < p < \infty$) has a uniformly Gâteaux differentiable norm. Furthermore, any separable Banach space can be equivalently renormed to have a (UG) norm, [154], as can any super-reflexive Banach space. Distance functions on (UG) spaces are very well behaved:

Proposition 4.29 [19, Theorem 8] *If the norm $\|\cdot\|$ on a Banach space X is uniformly Gâteaux differentiable, then for each non-empty closed subset C of X , $-d_C$ is regular (and hence pseudo-regular) on $X \setminus C$.*

Corollary 4.30 [11, Theorem 5.2] *Let $\|\cdot\|$ be a uniformly Gâteaux 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 Gâteaux differentiable. Therefore we may deduce the next result.

Proposition 4.31 *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 $d_C \in S_e(X)$. Indeed, since no point of ∂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 Gâteaux differentiability coincide), we immediately have a necessary condition for $d_C \in S_e(X)$, namely, ∂C must be a Lebesgue-null set.

Moreover, we have from ([19], Theorem 8) that $-d_C$ is regular on $X \setminus C \cup \text{int}C$. Therefore, if ∂C is a Lebesgue-null set then d_C is strictly differentiable almost everywhere in X , since any locally Lipschitz function which is both Gâteaux differentiable and pseudo-regular at a given point is necessarily strictly differentiable at that point. Hence we may deduce the following.

Theorem 4.32 *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, ∂C is a Lebesgue-null set.*

It is natural to ask whether the characterization given in Theorem 4.32 holds for an arbitrary separable Banach space. Unfortunately the answer is ‘no’. We do salvage the following Corollary.

Corollary 4.33 *Let $\|\cdot\|$ be a uniformly Gâteaux 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 ∂C is a Haar-null set.*

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

Example 4.34 In Example 6.2 part(b) of [11], there is an example of a closed and convex subset of $c_0(N)$, such that ∂C is not a Haar-null set (see the Appendix). 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 [108] we know that such sets exist in any separable non-reflexive space. Note also, that such sets necessarily have empty interior.

Remember 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 $\{x_n : n \in \mathbf{N}\} \subseteq S(X)$ converges to an element $x \in S(X)$ in the weak topology, then it converges to x in the norm topology). Another important result related to Theorem 3.29 regarding the minimality of the subdifferential mappings of distance functions follows:

Theorem 4.35 *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, $\overline{C \cap B(0, r)}^w$ is weak compact for $r > 0$.)*

Then d_C is D -representable on X . In particular, $\partial_c d_C$ is generated by its strict Fréchet derivatives, and the set of points in $X \setminus C$ which admit a closest point in C contains a dense and G_δ subset of $X \setminus C$.

Proof By ([124], Theorem 2.5) we know that $\partial_c d_C = \text{CSC}(\Omega_D)$, where

$$D \equiv \{x \in X : d_C \text{ is Fréchet differentiable at } x\}$$

and $\Omega_D : D \rightarrow 2^{X^*}$ is defined by, $\Omega_D(x) \equiv \{\nabla d_C(x)\}$. Hence, to show that $\partial_c d_C$ is a minimal weak* cusco on X , we need only show by, Corollary 2.17 and Theorem 2.31 that Ω_D is hyperplane minimal on $D \setminus C$. To this end, we consider the following set-valued mapping $p_C : D \setminus C \rightarrow 2^C$ 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 [14] that for each $x \in D \setminus C$ we have that $d_C(x) = \lim_{n \rightarrow \infty} \nabla d_C(x)(x - z_n)$ for any *minimizing sequence* $\{z_n : n \in \mathbf{N}\} \subseteq C$ such that $\lim_{n \rightarrow \infty} \|z_n - x\| = d_C(x)$. 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 $\{z_n : n \in \mathbf{N}\}$ be any sequence in C such that $\lim_{n \rightarrow \infty} \|z_n - x_0\| = d_C(x_0)$. Since the sequence $\{z_n : n \in \mathbf{N}\}$ is bounded there exists a point $z \in X$ and a subsequence $\{z_{n_k} : k \in \mathbf{N}\}$ of $\{z_n : n \in \mathbf{N}\}$ such that $\text{weak-lim}_{n \rightarrow \infty} z_{n_k} = z$. Since every norm on X is lower semi-continuous, with respect to the weak topology on X , we have that, $\|z - x_0\| \leq \lim_{k \rightarrow \infty} \|z_{n_k} - x_0\| = d_C(x_0)$. However, by above we have

$$\|z - x_0\| \geq \nabla d_C(x_0)(z - x_0) = \lim_{k \rightarrow \infty} \nabla d_C(x_0)(z_{n_k} - x_0) = d_C(x_0)$$

(note that since d_C is Lipschitz-1, $\|\nabla d_C(x_0)\| \leq 1$). Hence, $\|z - x_0\| = d_C(x_0)$. Now, since the norm on X is Kadec-Klee and $\lim_{k \rightarrow \infty} \|z_{n_k} - x_0\| = \|z - x_0\|$ we have that $\{z_{n_k} : k \in \mathbf{N}\}$ 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 $\{x_n : n \in \mathbf{N}\} \subseteq D \setminus C$ and $\{z_n : n \in \mathbf{N}\} \subseteq C$ such that $\{x_n : n \in \mathbf{N}\}$ converges to x and $z_n \in p_C(x_n)$ for each $n \in \mathbf{N}$, $\{z_n : n \in \mathbf{N}\}$ 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 $\{x_n : n \in \mathbf{N}\}$ be a sequence in $D \setminus C$ which converges to x . Further, let $\{z_n : n \in \mathbf{N}\}$ be a sequence in C such that $z_n \in p_C(x_n)$ for each $n \in \mathbf{N}$. Now,

$$\begin{aligned} d_C(x) &\leq \liminf_{n \rightarrow \infty} \|z_n - x\| \leq \limsup_{n \rightarrow \infty} \|z_n - x\| \\ &\leq \lim_{n \rightarrow \infty} \|z_n - x_n\| + \lim_{n \rightarrow \infty} \|x_n - x\| \end{aligned}$$

$$\begin{aligned} &= \lim_{n \rightarrow \infty} d_C(x_n) \\ &= d_C(x). \end{aligned}$$

Therefore, $\lim_{n \rightarrow \infty} \|z_n - x\| = d_C(x)$. Now, by repeating the argument above, we obtain a subsequence $\{z_{n_k} : k \in \mathbf{N}\}$ of $\{z_n : n \in \mathbf{N}\}$ which converges to some point $z \in p_C(x)$ (in the norm topology). This completes part (i).

(ii) We show that Ω_D is hyperplane minimal on $D \setminus C$. Let $x \in D \setminus C$ and let z be any element of $p_C(x)$, then for each y

$$\begin{aligned} \nabla d_C(x)(y) &= \lim_{\lambda \rightarrow 0} \frac{d_C(x + \lambda y) - d_C(x)}{\lambda} \\ &\leq \lim_{\lambda \rightarrow 0} \frac{\|(x + \lambda y) - z\| - \|x - z\|}{\lambda} \\ &= \|x - z\|'(y). \end{aligned}$$

Since $y \rightarrow \|x - z\|'(y)$ is linear on X , we must have $\{\nabla d_C(x)\} = \partial_c \|x - z\|$. Furthermore, since z was arbitrary in $p(x)$ we must have that $\partial_c \|x - p(x)\| = \{\nabla d_C(x)\}$. Therefore, $x \rightarrow \partial_c \|x - p_C(x)\|$ is a single-valued weak* usco on $D \setminus C$ (since it is the composition of two usco mappings) and hence hyperplane minimal on $D \setminus C$. This completes the proof. ■

A set C is *densely proximal* 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 (see Section 5) shows that every closed set is densely proximal.

Corollary 4.36 *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 proximal and the corresponding distance function possesses a minimal subdifferential mapping.*

Proof If X is reflexive and $\|\cdot\|$ is a smooth Kadec-Klee norm, then S is Asplund and it follows from above, that each non-empty closed subset C of X is densely proximal and the corresponding distance function d_C possesses a minimal subdifferential mapping. Conversely, if each non-empty closed subset of X is densely proximal then by [97], X is reflexive and $\|\cdot\|$ is a Kadec-Klee norm. However, by Theorem 4.27, if each distance function possesses a minimal subdifferential mapping, then the norm $\|\cdot\|$ is smooth on X . ■

We define a *proximal normal selection* ρ_D on a subset D of $D(C)$, by setting $\rho_D(x) \equiv f_{x-p(x)}$ (where $f_{x-p(x)}$ is any element of $\partial_c \|x - p(x)\|$ for some nearest point $p(x)$) when $x \in D \setminus C$ and setting $\rho_D(x) \equiv 0$. In [11] it is shown that $\rho_D(x) \in \partial_c d_C(x)$ for each $x \in D(C)$.

Theorem 4.37 *Let C be a non-empty closed subset of a smooth Banach space X . Suppose that C is densely proximal. Then d_C is D -representable on X if, and only if, for each proximal normal selection ρ_D , $\text{CSC}(\rho_D) = \partial_c d_C$.*

Proof This follows directly from Theorem 2.18 part (iv). ■

We now recover the best possible proximal normal formula.

Corollary 4.38 (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 $D(C)$,*

$$\partial_c d_C(x) = \overline{\text{co}}\{x^* \in X^* : x^* = w - \lim_{x_n \rightarrow x} \rho_D(x_n), x_n \in D\}.$$

It follows from Corollary 4.36 and Theorem 4.37 that we cannot weaken the hypothesis in Corollary 4.38 and still have a ‘Proximal Normal Formula’ holding for all non-empty closed subsets of X .

We end this discussion by stating a recent result by Zajicek [153] as promised in Section 1.4 (d). It again advertises how well distance functions behave in (UG) spaces.

Theorem 4.39 *Let X be a Banach space with a uniformly Gâteaux differentiable norm. Then, for every closed set C , d_C is Gâteaux differentiable at all points of $X \setminus C$ except those which belong to a σ -globally very porous set.*

Corollary 4.40 *Let X be a Banach space with a norm which is simultaneously strictly convex and uniformly Gâteaux differentiable. Then, for every closed set C , the set of points $x \in X$ at which the metric projection $p(x)$ into C is multivalued is σ -globally very porous.*

4.6 Relationships between integrability, representability and smoothness

We can see from Example 4.2 part (2) that minimality of the Clarke subdifferential mapping, equivalently D -representability, is not enough to guarantee integrability. So we begin this part of the lecture 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 \mathbf{R} , integrability does imply D -representability (see Corollary 1.3 in [17]), in fact on \mathbf{R} , integrability implies strict differentiability, almost everywhere.

However, we will show next, that in general, integrability does not imply D -representability.

Example 4.41 *Let f be a real-valued Rockafellar function defined on \mathbf{R} such that $\partial_c f \equiv [0, 1]$. Let*

$$C \equiv \text{epi}(f) = \{(x, y) \in \mathbf{R}^2 : f(x) \leq y\}.$$

Next, consider the distance function d_C defined on \mathbf{R}^2 by the l_1 norm and the set C . Then d_C is integrable on \mathbf{R}^2 , but not D -representable on \mathbf{R}^2 , in fact d_C is not even densely strictly differentiable on \mathbf{R}^2 .

Proof Suppose that g is a real-valued locally Lipschitz function defined on \mathbf{R}^2 such that $\partial_c g(x, y) \subseteq \partial_c d_C(x, y)$ for each $(x, y) \in \mathbf{R}^2$. Now, $\partial_c d_C(x, y) = \{0\}$ on $\text{int } C$, and so $\partial_c g(x, y) = \{0\}$ on $\text{int } C$. But $\text{int } C$ is connected, therefore g is constant on $\text{int } C$, and so constant on C , that is, $g|_C \equiv c_1$ 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 4.27 for a more detailed explanation). Therefore,

$$\partial_c d_C(x, y) = \partial_c f(x) \times \{-1\} = [0, 1] \times \{-1\}$$

on $\mathbf{R}^2 \setminus C$. Let x_0 be an element of \mathbf{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, $y \rightarrow g(x_0, y)$,

we have that $g(x_0, y) = c_1 + (f(x_0) - y)$ for each $y \leq f(x_0)$. Hence, $g(x, y) = (f(x) - y) + c_1 = d_C(x, y) + c_1$ on $\mathbf{R}^2 \setminus C$. 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 \mathbf{R}^2 . ■

Remark 4.42 It is *very* interesting to observe that d_C is not integrable on $\mathbf{R}^2 \setminus C$. Indeed, let $f_1(x) \equiv x - f(x)$, then $\partial_c f_1 = \partial_c f$ on \mathbf{R} , and so $\partial_c g_1(x, y) = \partial_c f(x) \times \{-1\}$ on $\mathbf{R}^2 \setminus C$, where $g_1(x, y) \equiv f_1(x) - y$ on $\mathbf{R}^2 \setminus C$. Hence, 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 $f|_U$ 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 $f \in \mathcal{S}_e(A)$ then f is not only integrable on A , but also hereditarily integrable on A .

As integrability is not a hereditary property (with respect to open sets) one cannot expect a characterization in terms of a local differentiability property, (as was given for D -representability). We give next, a sufficient condition for a Lipschitz function to be integrable. It is noteworthy that this condition is expressed by 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 the point x , such that $V \cap A$ is connected.

Theorem 4.43 *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 $f|_U$ is integrable on U and the connected components of U are locally finite, then f is integrable on $\text{int } \overline{U} \cap A$; and*
- (b) *if $f|_U$ is hereditarily integrable on U and U is a finite union of locally connected open subsets, then f is hereditarily integrable on $\text{int } \overline{U} \cap A$.*

Proof (a) Suppose that g is a real-valued locally Lipschitz function defined on $B \equiv \text{int } \overline{U} \cap A$ such that $\partial_c g(x) \subseteq \partial_c (f|_B)(x)$ 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 $\{\overline{U}_\gamma : \gamma \in \Gamma\}$ is also locally finite (and $\overline{U} = \bigcup \{\overline{U}_\gamma : \gamma \in \Gamma\}$). Let x_0 be an arbitrary point of B . We need to show that $\partial_c (f|_B - g)(x_0) = \{0\}$. We may choose an open neighbourhood V of x_0 (contained in A) so that $\{\gamma \in \Gamma : \overline{U}_\gamma \cap V \neq \emptyset\} = \{\gamma_j : 1 \leq j \leq n\}$ and $x_0 \in \overline{U}_{\gamma_j}$ 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 $c_j \in \mathbf{R}$ such that $g(x) = f(x) + c_j$ for all $x \in U_{\gamma_j}$. It follows now, from the continuity of f and g that $g(x) = f(x) + c_j$ for each $x \in \overline{U}_{\gamma_j} \cap A$. 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 $\overline{U} \cap V$ (and so on $B \cap V$). This shows that $\partial_c (f|_B - g)(x_0) = \{0\}$, and the mean value theorem finishes the result.

(b) Let B' be an arbitrary non-empty open subset of $B \equiv \text{int } \overline{U} \cap A$ and suppose that g is a real-valued locally Lipschitz function defined on B' such that $\partial_c g(x) \subseteq \partial_c f|_{B'}(x)$ for each $x \in B'$. Let x_0 be an arbitrary point in B' . We need to show that $\partial_c (f|_{B'} - g)(x_0) =$

$\{0\}$. Let $\{U_j : 1 \leq j \leq n\}$ denote the locally connected open subset of U (note that $\overline{U} = \bigcup\{\overline{U}_j : 1 \leq j \leq n\}$). We may choose an open neighbourhood V of x_0 (contained in B') so that $\{j : \overline{U}_j \cap V \neq \emptyset\} = \{j : x_0 \in \overline{U}_j\} = \Delta$. Now, for each $j \in \Delta$ there exists an open neighbourhood V_j of x_0 (contained in V) such that $V_j \cap U_j$ is connected. Since $f|_U$ is hereditarily integrable on U for each $j \in \Delta$, there exists a $c_j \in \mathbf{R}$ 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 $x \in \overline{U}_j \cap V_j \subseteq \overline{U}_j \cap \overline{V}_j$. In particular, we have that $g(x_0) = f(x_0) + c_j$ for each $j \in \Delta$. Hence, $g(x) = f(x) + c_1$ on $\bigcap\{V_j : j \in \Delta\} \cap \overline{U}$ (and so on $\bigcap\{V_j : j \in \Delta\} \cap B'$). This shows that $\partial_c(f|_{B'} - g)(x_0) = \{0\}$ ■

These apparently harmless observations provide us with a technique for constructing integrable functions which are not essentially smooth.

Corollary 4.44 *Let $\|\cdot\|$ be a uniformly Gâteaux differentiable norm on a separable Banach space X . Let C be a non-empty closed subset of X . Then,*

- (a) d_C is integrable if, the connected components of both $\text{int } C$ and $X \setminus C$ are locally finite; and
- (b) d_C is hereditarily integrable if, $\text{int } C$ and $X \setminus C$ are both locally connected subsets of X

Proof This follows from Theorem 4.43 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 $S_e(X)$. With a little more work, we can show the even stronger result:

Example 4.45 *There exists a compact nowhere dense subset C of \mathbf{R}^2 such that (i) d_C is D -representable; (ii) d_C is hereditarily integrable; (iii) d_C is not essentially smooth in \mathbf{R}^2 , that is $d_C \notin S_e(\mathbf{R}^2)$.*

Proof Let C_1 be a Cantor subset of $[0, 1]$ with $\mu(C_1) > 0$. Let

$$C \equiv C_1 \times C_1 \subseteq \mathbf{R}^2.$$

Let d_C be the distance function generated by the set C , with the Euclidean norm. Then by Proposition 4.31, d_C is D -representable on \mathbf{R}^2 .

To justify that d_C is hereditarily integrable it suffices by Corollary 4.44 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_∞ 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 $d_C \notin S_e(\mathbf{R}^2)$ we use again the 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 $\mathcal{I}(A)$ denote the family of all real-valued, integrable, locally Lipschitz function defined on A and let $\mathcal{M}(A)$ denote the

family of all such functions defined on A whose Clarke subdifferential mappings are minimal. It follows then, that

$$\mathcal{N}(A) \equiv \mathcal{I}(A) \cap \mathcal{M}(A)$$

is the largest class of functions that satisfy both the conditions. 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 4.46 $\mathcal{N}(\mathbf{R}^2)$ is not closed under addition, multiplication nor either of the lattice operations. (Note: we also show that $\mathcal{I}(\mathbf{R}^2)$ is not closed under addition, multiplication nor either of the lattice operations.)

Proof (a) We show first that $\mathcal{N}(\mathbf{R}^2)$ is not closed under addition. Let f be a non-integrable real-valued Lipschitz-1 function defined on \mathbf{R} such that, $x \rightarrow \partial_c f(x)$, is a minimal cusco. (Such functions exist, see Example 4.2.) Let $K_1 \equiv \{(x, y) : f(x) \leq y\}$ and $K_2 \equiv \{(x, y) : y \leq f(x)\}$. Next, consider the distance functions d_{K_1} and d_{K_2} defined on \mathbf{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_1} and $-d_{K_2}$ are integrable on \mathbf{R}^2 and D -representable on \mathbf{R}^2 . However, $d \equiv d_{K_1} + (-d_{K_2})$ is not integrable on \mathbf{R}^2 . In fact, $d(x, y) = f(x) - y$ on \mathbf{R}^2 and so $\partial_c d(x, y) = \partial_c f(x) \times \{-1\}$ on \mathbf{R}^2 . Hence for any real-valued function g defined on \mathbf{R} such that, $\partial_c g = \partial_c f$ and $g - f$ is not a constant function on \mathbf{R} , the function $G(x, y) \equiv g(x) - y$, shares that same Clarke subdifferential mapping as d . Therefore, d is not integrable on \mathbf{R}^2 .

(b) Next, we show that $\mathcal{N}(\mathbf{R}^2)$ is not closed under multiplication. Let $d_{K_1}^* \equiv d_{K_1} + 1$ and $d_{K_2}^* \equiv (-d_{K_2}) + 1$. Then $d^* \equiv d_{K_1}^* \cdot d_{K_2}^*$ is not integrable on \mathbf{R}^2 . To see this, we compute d^* ; $d^*(x, y) = (f(x) - y) + 1$ on \mathbf{R}^2 . Then as in (a) we see that d^* is not integrable on \mathbf{R}^2 .

(c) Finally, we show that $\mathcal{N}(\mathbf{R}^2)$ is not closed under the lattice operations. Let C be a Cantor subset of $[0, 1]$ with $\mu(C) > 0$. We define two sets; $C_1 \equiv \{(x, y) \in \mathbf{R}^2 : x \in C \text{ and } y \leq 0\}$ and $C_2 \equiv \{(x, y) \in \mathbf{R}^2 : x \in C \text{ and } y \geq 0\}$. Now, consider the distance functions d_{C_1} and d_{C_2} defined on \mathbf{R}^2 by the Euclidean norm and the sets C_1 and C_2 respectively. Then $d_{C^*}(x, y) \equiv \min\{d_{C_1}(x, y), d_{C_2}(x, y)\}$ is the distance function to the set $C^* \equiv \{(x, y) \in \mathbf{R}^2 : x \in C \text{ and } y \in \mathbf{R}\}$. Moreover, it is easy to see that $d_C(x, y) = d(x)$, where $d : \mathbf{R} \rightarrow \mathbf{R}$ is defined by, $d(x) \equiv \min\{|x - c| : c \in C\}$. However, d is not integrable on \mathbf{R} since d is not strictly differentiable almost everywhere on \mathbf{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 $\partial_c g = \partial_c d$ and $g - d$ is not a constant function on \mathbf{R} . Let $G(x, y) \equiv g(x)$. Clearly then, $\partial_c G = \partial_c d_{C^*}$, but $G - d_{C^*}$ is not constant on \mathbf{R}^2 . To show that $\mathcal{N}(\mathbf{R}^2)$ is not closed under ‘max’ we need only consider $-d_{C^*}$. ■

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

Theorem 4.47 (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*

$$\{x \in A : \partial_c g(x) \cap \partial_c f(x) \neq \emptyset\}$$

is dense in A .

Proof Suppose that $\{x \in A : \partial_c g(x) \cap \partial_c f(x) \neq \emptyset\}$ is dense in A . Consider the set-valued mapping $T : A \rightarrow 2^{X^*}$ defined by,

$$T(x) \equiv \partial_c g(x) \cap \partial_c f(x).$$

Since both $x \rightarrow \partial_c f(x)$ and $x \rightarrow \partial_c g(x)$ are upper semi-continuous on A (and possess compact images), T possesses non-empty weak* compact, convex images. Moreover, since the graphs of both $\partial_c f$ and $\partial_c g$ 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 $x \in A$, $T(x) \subseteq \partial_c f(x)$ and $T(x) \subseteq \partial_c g(x)$. Hence, by the minimality of $\partial_c g$ we must have that $\partial_c g = T$ (that is, $\partial_c g(x) \subseteq \partial_c f(x)$ for each $x \in A$). The result now follows from the fact that f is integrable. The converse is obvious. ■

In contrast, we have the following comment concerning integrability with respect to the *approximate subdifferential* mapping. It is possible to construct two Lipschitz functions f and g mapping from \mathbf{R}^2 into \mathbf{R} such that $\partial_c f = \partial_c g$ is minimal, while $\partial_a f$ and $\partial_a g$ differ on a set of positive measure. This cannot happen on the real-line, where $\partial_c f$ determines $\partial_a f$, [17], (here $\partial_a f$ denotes the approximate subgradient of f). On the other hand we observe that our conditions for integrability imply integrability with respect to any subdifferential mapping, $x \rightarrow \partial f_{\#}(x)$, which has the property that

$$\overline{\text{co}}^{w*} \partial f_{\#}(x) = \partial f(x)$$

for each x .

Let us also comment that in general $x \rightarrow \partial_a f(x)$ 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.

The following Venn diagram shows the relationship among classes of functions discussed in this lecture. In this Venn diagram all seven possible sets are nonempty and meaning of the symbol is as follows: **S** represents essentially strictly differentiable functions; **G** represents generically strictly differentiable functions, **M** represents functions with minimal Clarke generalized gradient and **I** represents integrable functions.

Before ending this lecture we point out that even the proximal separating function on line does not determine a function up to a constant. More precisely there are infinitely many Lipschitz functions on \mathbf{R} satisfying $\partial_p f(x) = (-1, 1)$ on a given countable set D and $\partial_p f(x) = \emptyset$ elsewhere [8, 22].

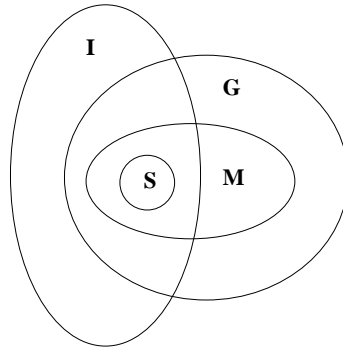


Figure 4.1: Relationship among several classes of functions

5 Convex functions and classifications of normed spaces

The aim of this lecture is to illustrate the tight connection between the sequential properties of a Banach space and the corresponding properties of the convex functions and sets which may or may not be defined on that space. The central question we address is: what convexity properties characterize the most classical classes of Banach spaces? Thus what prevails in

1. Finite dimensions ?
2. Reflexive spaces ?
3. Separable spaces ?
4. Asplund spaces and spaces not containing $\ell_1(\mathbf{N})$?
5. Spaces containing $c_0(\mathbf{N})$?

If at any point the references or discussion herein seem incomplete, the trail will be found in [21] or [12].

5.1 Finite dimensions

We begin with a compendium of standard and relatively easy results whose proofs may be pieced together from many sources.

Theorem 5.1 *The following are equivalent:*

- (i) *X is finite dimensional.*
- (ii) *Every linear functional on X is continuous.*
- (iii) *Every convex function $f : X \rightarrow \mathbf{R}$ is continuous.*
- (iv) *The closed unit ball in X is (pre-) compact.*
- (v) *The weak and norm topologies coincide on X .*

- (vi) *The weak–star and norm topologies coincide on X^* .*
- (vii) *Every (closed) convex set in X has non-empty relative interior.*
- (viii) *$A \cap R = \emptyset$, A closed and convex, R a ray $\Rightarrow A$ and R are separated by a continuous linear functional.*

In essence this result says “don’t trust finite dimensionally derived intuitions”. By comparison, a much harder and less well known set of results is:

Theorem 5.2 *The following are equivalent:*

- (i) *X is finite dimensional.*
- (ii) *Weak–star and norm convergence agree for sequences in X^* .*
- (iii) *Every continuous convex $f : X \rightarrow \mathbf{R}$ is bounded on bounded sets.*
- (iv) *For every continuous convex $f : X \rightarrow \mathbf{R}$, ∂f is bounded on bounded sets.*
- (v) *For every continuous convex $f : X \rightarrow \mathbf{R}$, any point of Gâteaux differentiability is a point of Fréchet differentiability.*

Proof (Sketch) (i) \Rightarrow (iii) or (v) is clear; (iii) \Rightarrow (iv) is easy. To see (v) \Rightarrow (ii) and (iii) \Rightarrow (ii) we proceed as follows.

Let $\|x_n^*\| = 1$ and $0 < \alpha_n \downarrow 0$. Define

$$f(x) := \sup_{n \in \mathbf{N}} \langle x_n^*, x \rangle - \alpha_n. \quad (12)$$

Then f is convex and continuous and is:

$$\text{Gâteaux differentiable at } 0 \Leftrightarrow x_n^* \xrightarrow{w^*} 0$$

and

$$\text{Fréchet differentiable at } 0 \Leftrightarrow x_n^* \xrightarrow{\|\cdot\|} 0.$$

Thus (v) \Rightarrow (ii). (See [12].) Now consider

$$f(x) := \sum_n \varphi_n(\langle x_n^*, x \rangle) \quad (13)$$

where $\varphi_n(t) := n \left(|t| - \frac{1}{2} \right)^+$. Then f is

$$\text{finite (continuous)} \Leftrightarrow x_n^* \xrightarrow{w^*} 0$$

and is

$$\text{bounded on bounded sets} \Leftrightarrow x_n^* \xrightarrow{\|\cdot\|} 0.$$

Thus (iii) \Rightarrow (ii). (See [21].) ■

Note that the sequential coincidence of weak and norm topologies characterizes the *Schur spaces* (such as $\ell_1(\mathbf{N})$; see [67]) while the sequential coincidence of weak and weak–star topologies characterizes *Grothendieck spaces* (reflexive spaces and non–reflexive spaces such as $\ell_\infty(\mathbf{N})$; see [67]).

The statements of Theorem 5.2 are equivalent in the strong sense that they are easily interderived while no “easy proof” is known of (ii) \Rightarrow (i). This is the Josephson-Nissenzweig Theorem first established in 1975, see [67]. For example, (ii) \Rightarrow (iii) follows from:

Lemma 5.3 [21] *Suppose that $f : X \rightarrow \mathbf{R}$ is continuous and convex and that $\{x_n\}$ is bounded while $f(x_n) \rightarrow \infty$. Then*

$$x_n^* \in \partial f(x_n) \Rightarrow \psi_n := \frac{x_n^*}{\|x_n^*\|} \xrightarrow{w^*} 0. \quad \blacksquare$$

Thus each such function yields a *Josephson-Nissenzweig sequence* of unit vectors w^* -convergent to 0.

5.2 Reflexive spaces

We begin with the traditional “James-Eberlein-Smulian” characterizations of reflexivity (see [62] or [82]):

Theorem 5.4 *The following are equivalent:*

- (i) X is reflexive.
- (ii) The unit ball on X is weak compact.
- (iii) Every continuous linear functional on X achieves its norm.
- (iv) If $\{C_n\}$ are non-empty nested, closed convex bounded sets, then $\bigcap_{n \in \mathbf{N}} C_n \neq \emptyset$.

One may add the less traditional:

- (v) Fenchel conjugacy is Mosco continuous for closed convex functions ([5]).

We will say that f is *coercive* if $f(x)/\|x\| \rightarrow \infty$ when $\|x\| \rightarrow \infty$. A corresponding set of subgradient characterizations given in [21] is:

Theorem 5.5 *The following are equivalent:*

- (i) X is reflexive.
- (ii) $\text{Range}(\partial f) = X^*$ for some (or all) coercive continuous convex $f : X \rightarrow \mathbf{R}$.
- (iii) $\text{Int Range}(\partial f)$ is convex for all (coercive) continuous convex $f : X \rightarrow \mathbf{R}$.

[Similar statements hold for maximal monotone operators.]

Proof (Sketch) The “key” is the construction of

$$f(x) := \max \left\{ 1/2\|x\|^2, \|x - p\| - 12 \pm \langle p^*, x \rangle \right\}, \quad (1-a)$$

where $\|p\| = 5$ and $p^* \in \partial \frac{1}{2}\|p\|^2$. Now $\text{int Range}(\partial f)$ is non-convex. Indeed, using James’ Theorem one may show that it contains $B(X^*) \pm p^*$ and that $\frac{1}{2}B(X^*)$ lies in the convex hull of $\text{Int Range}(\partial f)$ but not in $\text{Int Range}(\partial f)$. ■

Note that in any normed space $\text{Int dom}(\partial f)$ is convex. The easiest explicit example for (iii) of the previous result lies in the space $c_0(\mathbf{N})$ of null sequences endowed with the supremum norm. One may use

$$f(x) := \|x - e_1\|_\infty + \|x + e_1\|_\infty \quad (1-b)$$

where e_1 is first unit vector. Then

$$\text{Int Range}(\partial f) = \{B(\ell_1) + e_1\} \cup \{B(\ell_1) - e_1\}$$

which is far from convex. ■

We pause to indicate some relations with *nearest points*. Here as always we consider a *distance function*

$$d_C(x) := \inf_{c \in C} \|c - x\|$$

and we say C admits a nearest point if $d_C(x)$ is attained for some $x \notin C$. By James theorem $(X, \|\cdot\|)$ is reflexive iff every closed convex set in X admits nearest points. A norm is (sequentially) *Kadec-Klee* if

$$x_n \xrightarrow{w} x \text{ and } \|x_n\| \rightarrow \|x\| \Rightarrow \|x_n - x\| \rightarrow 0.$$

This is the most significant renorming property for results related to best approximation as is illustrated by:

Theorem 5.6 (Lau-Konyagin, 1976) *Every closed subset C of X densely (or generically) admits nearest points iff $(X, \|\cdot\|)$ is reflexive and has the Kadec-Klee property.*

All reflexive spaces can be renormed to be Kadec-Klee. A fundamental open isometric question is:

“When $(X, \|\cdot\|)$ is reflexive must every non-trivial closed subset admit at least one nearest point?”

This question is open even for arbitrary renorms of Hilbert space!! Counter-examples are necessarily unbounded, must fail to be weakly closed, and must lie in highly non-Kadec-Klee spaces. (See [14] for details on all these matters relating to best approximations.)

Continuing to look at reflexivity, we exhibit a striking recent characterization in a slightly specialized form. It provides a remarkable liaison between norm compactness, openness, separability and reflexivity.

Theorem 5.7 [108] *Suppose X is separable. The following are equivalent.*

- (i) X is not reflexive.
- (ii) X contains a closed convex set \tilde{C} with empty interior but such that every norm compact K lies in some additive translate of \tilde{C} .

Thus in any separable non-reflexive space there is a closed convex set C with empty interior which is not *Haar null*: meaning that no Borel probability measure can vanish on all translates of C . This was motivated by a conjecture in [35] but leaves open the following tantalizing question:

“In a reflexive space is every closed convex set with empty interior Haar null?”

This is clearly the case in finite dimensions and is true in super reflexive spaces as shown in the Appendix. As a consequence of Theorem 5.7 we also have a significant limiting example for Fréchet differentiability of Lipschitz functions.

Corollary 5.8 [35] *Let X be separable. Let*

$$d_C(x) := \inf_{c \in C} \|x - c\| \tag{1-c}$$

If $\text{int } C = \emptyset$ then d_C fails to be Fréchet differentiable at all points of C . In particular, for \tilde{C} as in Theorem 5.7, \tilde{C} is not Haar-null and so $d_{\tilde{C}}$ is not Haar almost everywhere Fréchet differentiable.

We note that if X^* is not separable there is actually a nowhere Fréchet differentiable continuous convex function on X . Since the Haar null sets are the largest class of reasonable null sets in a separable Banach space, Corollary 5.8 “rules out” studying Fréchet differentiability by measure-like techniques. By contrast, measure theoretic techniques work very well as we have seen for studying Gâteaux differentiability in a separable setting (see [35]).

5.3 Separable spaces

Many separable results continue to hold for spaces X with an (infinite dimensional) *separable quotient* Y . That is, there exists a continuous linear surjection $T : X \rightarrow Y$. There is no known instance of a space without a separable quotient. For example $\ell_2(\mathbf{N})$ is a separable quotient of $\ell_\infty(\mathbf{N})$.

Theorem 5.9 [21] *Suppose X has a separable quotient. Then there exist proper lower semi-continuous convex functions f and g with the properties that*

- (i) $\text{dom}(f) = \text{dom}(g)$ is dense in X ,
- (ii) ∂f and ∂g are both at most singleton (“almost Gâteaux”),
- (iii) $\text{dom}(\partial f) \cap \text{dom}(\partial g) = \emptyset$.

Proof (Sketch) From the existence of a separable quotient one argues that without loss X is separable. Let $\{x_n, x_n^*\}$ be an *M-basis*: meaning that $\langle x_n, x_m^* \rangle = \delta_{n,m}$ for $n, m \in \mathbf{N}$, and $\overline{\text{sp}} \{x_n\} = X$.

Then we may use

$$f(x) := \sum_{n \in \mathbf{N}} (n \langle x_n^*, x \rangle)^2 \quad \text{and} \quad g(x) := f(x - y) \tag{2-a}$$

where $y := \sum_{n \in \mathbf{N}} n^{-7/4} x_n$. ■

Before continuing we recall that the *quasi-relative interior* of C is given by

$$\text{qri}(C) := \{x \in C : T_C(x) \text{ is linear}\}.$$

Here $T_C(x)$ is the closed convex *tangent cone* generated by C at x . Equivalently, $x \in \text{qri}(C)$ iff x is a *non-support point* of C in the sense that

$$\varphi \in X^* \text{ and } \langle \varphi, x \rangle = \inf_C \varphi \Rightarrow \langle \varphi, x \rangle = \sup_C \varphi.$$

In finite dimensions it is easy to show that “qri” = “rel-int”, while

Theorem 5.10 *If X is separable every closed convex set C has non-empty quasi-relative interior; that is, C has a non-support point.*

Proof Let $\{c_n : n \in \mathbf{N}\}$ be dense in C and consider

$$\hat{c} := \sum_{n \in \mathbf{N}} \lambda_n c_n \quad \text{where} \quad \sum_{n \in \mathbf{N}} \lambda_n = 1, \quad \lambda_n > 0 \quad (2-b)$$

are chosen to ensure convergence of \hat{c} . Then

$$\langle \varphi, \hat{c} \rangle = \inf_C \varphi \Rightarrow \langle \varphi, \hat{c} \rangle = \langle \varphi, c_n \rangle \equiv \sup_C \varphi. \quad \blacksquare$$

All of this is detailed in [25]. In short, the quasi-relative interior provides a useful surrogate for the relative interior which, by Theorem 5.1, must be empty for some closed convex set as soon as X is infinite dimensional. It is conjectured that “the converse holds” to Theorem 5.10 in the sense that in any non-separable space there is a closed convex set with empty quasi-relative interior. We detail some recent partial results in this direction:

Theorem 5.11 [39] *X contains a closed convex set consisting only of support points if either*

- (1) (a) $X = Y^*$ is non-separable or (b) X^* is not weak-star separable;
- (2) X contains an uncountable biorthogonal sequence;

or

- (3) $X = C(\Gamma)$ where Γ is compact and Hausdorff and Γ contains a closed subset which is either non-separable or not a G_δ .

Proof (Sketch) With sufficient work (1) follows from (2) and actually (1) (a) is considerably deeper than (1) (b). (See [75].)

To see (2), let $\{x_\alpha, x_\beta^*\}$ be biorthogonal for $\alpha, \beta < \Omega$ (the first uncountable ordinal). Then

$$C_\Omega := \overline{\text{conv}}\{x_\alpha : \alpha < \Omega\} \quad (2-c)$$

is a *support set*: that is, it contains only support points. Indeed,

$$x_0 \in C_\Omega \Rightarrow x_0 \in \overline{\text{co}}\{x_\alpha : \alpha < \beta\}$$

for some $\beta < \Omega$. Then x_β^* properly supports C_Ω at x_0 .

In (3) the harder case uses a closed non- G_δ subset $F \subset \Gamma$. Then

$$C_F := \{f \in C(\Gamma) : f \geq 0, f|_F = 0\} \tag{2-d}$$

is a support set because

$$f|_F = 0 \Rightarrow f(t) = 0 \text{ for some } t \notin F.$$

We may now use the Tietze extension theorem to build a function g in C_F with $g(t) > 0$. Then δ_t supports C_F at t . ■

In the presence the *Continuum Hypothesis* it is shown in [39] that (2) and (3) are mutually distinct conditions. No other way of building support sets is known. Thus, the continuous function spaces for which the converse remains open form a subclass of the non-metrizable Γ which are both hereditarily separable and hereditarily normal.

Another related open question now suggests itself:

“Does every infinite dimensional space contain a closed densely spanning convex set with at least one non-support point (a quasi-relative interior point) in its boundary?”

If Y is separable (and infinite dimensional) the answer is “yes”. Indeed, let $\{y_n : n \in \mathbf{N}\}$ be dense in the unit sphere in Y . The set

$$C := \overline{\text{co}} \{\pm 2^{-n} y_n : n \in \mathbf{N}\} \tag{2-e}$$

is compact. Thus C has empty interior and so $0 \in \text{bd}(C)$; but also 0 is a non-support point of C . As another illustration of the use of separable quotients, if X has a separable quotient Y with quotient map T , then $T^{-1}(C)$ “lifts” the example to X .

5.4 Asplund spaces and spaces containing ℓ_1

Recall that X is an *Asplund space* if separable subspaces have separable duals as is the case for reflexive spaces. Equivalently, convex functions are generically Fréchet differentiable (see [67, 120]). Recall also that *Mackey convergence* in X^* is uniform convergence on weakly compact convex subsets of X and coincides with the norm topology on X^* iff X is reflexive.

Theorem 5.12 [12, 21] *The following are equivalent:*

- (i) *The space of absolutely summable sequences $\ell_1(\mathbf{N}) \not\subset X$ (isomorphically).*
- (ii) *Mackey and norm convergence agree sequentially in X^* (X is “sequentially reflexive”).*
- (iii) *Every continuous convex $f : X \rightarrow \mathbf{R}$ which is bounded on weakly compact sets is bounded on bounded sets.*
- (iv) *For every continuous convex $f : X \rightarrow \mathbf{R}$, any point of weak Hadamard differentiability is a point of Fréchet differentiability .*

Proof (Sketch) The hard step (i) \Leftrightarrow (ii) is a version of the wonderful “Rosenthal ℓ_1 theorem” (1974) given in [67].

The remainder is analogous to our finite dimensional results. As before let

$$f(x) := \sup_{n \in \mathbf{N}} \langle x_n^*, x \rangle - \alpha_n \quad (3-a)$$

Then f is convex and continuous and is:

$$\begin{aligned} \text{Gâteaux differentiable at } 0 &\iff x_n^* \xrightarrow{w^*} 0 \\ \text{weak Hadamard differentiable at } 0 &\iff x_n^* \xrightarrow{\text{Mackey}} 0 \\ \text{Fréchet differentiable at } 0 &\iff x_n^* \xrightarrow{\|\cdot\|} 0. \quad \blacksquare \end{aligned}$$

Thus in any Asplund space, somewhat surprisingly, *for convex functions* one need only establish weak Hadamard differentiability rather than the ostensibly stronger Fréchet differentiability. In contrast in any non-reflexive space there is a *non-convex* distance function with a point of weak Hadamard differentiability that is not a point of Fréchet differentiability, [21]. There is also a difference convex function with such a point.

Example $C(\Gamma)$ is Asplund iff $\ell_1 \not\subset C(\Gamma)$ but generally the Asplund class is much smaller. Correspondingly $L^1(\mu)$, μ σ -finite, admits a weak Hadamard smooth renorm (see [15]). AS we saw in Lecture three, this is useful in applications to control or optimization problems since $L^1(\mu)$ is not Asplund but convex functions are, nonetheless, generically *WH* differentiable. Moreover, any separable space with a non-separable dual with a weak Hadamard renorm, must by Theorem 5.12, contain a copy of $\ell_1(\mathbf{N})$.

Also X is Asplund iff X^* has the *Radon-Nikodym property* (RNP): every norm closed bounded convex set in X^* has a strongly exposed point (equivalently in a dual space an extreme point). Reflexive spaces have the RNP as do separable dual spaces such as ℓ_1 . \blacksquare

Theorem 5.13 [21] *The following are equivalent:*

- (i) X has the Radon-Nikodym property.
- (ii) The range of the subgradient, $\text{Range}(\partial f)$, is a generic set in X^* for each coercive lower semicontinuous convex function $f : X \rightarrow]-\infty, \infty]$.

5.5 Spaces containing c_0

The sequence space c_0 is the prototype of an Asplund space without the Radon-Nikodym property and is the home to many useful examples. For instance:

Theorem 5.14 *Let f and g be lower semicontinuous convex coercive proper convex functions.*

- (1) *Suppose X has the Radon-Nikodym property. Then the infimal convolution*

$$f \blacksquare g(x) := \inf_y \{f(y) - g(x - y)\}$$

is attained for some x .

- (2) *This fails if X contains a copy of c_0 .*

Proof (1) is not published but is fairly simple to establish while (2) is due to Edelstein and Thompson ([69]). ■

Indeed c_0 contains two closed norm balls, \overline{B}_1 and \overline{B}_2 , such that $\overline{B}_1 + \overline{B}_2$ is open. Equivalently $\|\cdot\|_1$ admits no nearest points in \overline{B}_2 and conversely. Such pairs are called *anti-proximal* (see [69]). It is easy to show that anti-proximal pairs can not be found in a space with the RNP, or more generally in a space with the slightly less arduous *convex point of continuity property* (W. Moors, private communication 1994).

The prototype of a space without the point of continuity property but which fails to admit copies of c_0 is the space of Lebesgue integrable functions $L_1[0, 1]$. It is possible to show that every separable space containing a copy of c_0 contains an anti-proximal pair. Correspondingly a slightly stronger “dual operator” version characterizes the presence of c_0 . Thus, we finish with yet another open question:

“Does an anti-proximal pair exist in L_1 ?”

6 Appendix: Convex Haar null sets

In this appendix we discuss the possible size of closed convex Haar null sets. Reasons to be interested in a notion of sets of measure zero in infinite dimensional spaces include Theorems 1.23 and 1.24.

Mankiewicz [107] gave a similar result to Theorem 1.23. Aronszajn [2] and Phelps [121] have introduced stricter notions of sets of measure zero in separable Banach spaces for which Theorem 1.23 holds. (See also [7].) The positive cone of ℓ_2 is an example of a Haar null set which is not in either of these stricter classes. By contrast, Theorem 1.24 becomes more useful the larger the excluded null set.

A result of Matoušková and Stegall characterizes reflexive spaces by an unusual criterion. We state the separable case.

Theorem 6.1 [108] *A separable Banach space E is not reflexive if and only if there exists a closed convex subset C of E with empty interior which contains some translate of each compact set in E .*

It is conjectured in [18] that *if C is a closed convex subset of a reflexive separable Banach space E and the interior of C is empty then C is a Haar null set.* One cannot dispense with the reflexivity hypothesis. Indeed:

Example 6.2 In any nonreflexive separable Banach space E there is a closed convex subset C of E with empty interior such that C is not a Haar null set.

Proof By Theorem 6.1 there is a closed convex subset C of E with empty interior such that for every compact subset K of E there is $x \in E$ such that $K + x \subset C$. It follows from tightness of Borel probability measures (Ulam’s theorem, see [68, p.176]) that C is not a Haar null set. ■

As we have seen, the motivating case for Example 6.2 is the non-negative cone, c_0^+ , in the space of null sequences, c_0 , in supremum norm [35].

We say C is a *spanning* subset of E provided E is the affine hull of C ; if C is convex this says that C contains line segments in all directions. We denote Lebesgue measure by λ . The first result is straightforward.

Proposition 6.3 *Let C be a (closed) convex subset of a separable Banach space E such that C contains no line segment in the direction d . Then C is Haar null. In fact*

$$\mu(X) := \lambda\{t \in [0, 1] : td \in X\}$$

defines a Borel probability measure such that $\mu(C + x) = 0$ for each $x \in E$.

Since sets which are not spanning are thus Haar null, we switch our attention to closed convex spanning sets. In what follows we discuss what is known in different classes of Banach spaces.

6.1 Separable reflexive spaces

We observe first that it follows that every norm compact subset of a Banach space is Haar null while they need not be Gaussian null.

Theorem 6.4 *Let E be a separable reflexive Banach space and C a closed convex spanning subset of E with empty interior. Then there are $\phi_j \in E^*$ such that $\|\phi_j\| = 1$, ϕ_j converge weakly to $0 \in E^*$ and*

$$C \subset \{x \in E : \langle \phi_j, x \rangle \leq 2^{-j} \text{ for } j = 1, 2, 3, \dots\}.$$

The sequence (ϕ_j) may be taken to be a basic sequence.

Proof Using the Bishop-Phelps theorem [9] (see [82, page 166]), choose (x_n) dense in C such that there are $f_n \in E^*$ with $\|f_n\| = 1$ which support C at x_n . Let

$$x := \sum_{n=1}^{\infty} t_n x_n$$

where $t_n > 0$ are chosen so that the series converges and $\sum_{n=1}^{\infty} t_n = 1$. Since (x_n) are dense in C we choose a subsequence (x_{n_j}) which converges to x , such that the support functionals (f_{n_j}) converge weakly to $f \in E^*$. Then for $y \in C$ we have $\langle f_{n_j}, y \rangle \leq \langle f_{n_j}, x_{n_j} \rangle$ so in the limit $\langle f, y \rangle \leq \langle f, x \rangle$. But then

$$\begin{aligned} \langle f, x \rangle &= \langle f, \sum_{n=1}^{\infty} t_n x_n \rangle = \sum_{n=1}^{\infty} t_n \langle f, x_n \rangle \\ &\leq \sum_{n=1}^{\infty} t_n \langle f, x \rangle = \langle f, x \rangle \end{aligned}$$

and since $t_n > 0$ we see that $\langle f, x_n \rangle = \langle f, x \rangle$ and by density $\langle f, y \rangle = \langle f, x \rangle$ for all $y \in C$. Since C is spanning that shows $f = 0$. Now $\langle f_{n_j}, x_{n_j} \rangle \rightarrow 0$ so we can extract a further subsequence (m_j) of (n_j) so that $\langle f_{m_j}, x_{m_j} \rangle \leq 2^{-j}$. Clearly $\phi_j := f_{m_j}$ are such that $\|\phi_j\| = 1$, ϕ_j converge weakly to $0 \in E^*$ and $C \subset \{x \in E : \langle \phi_j, x \rangle \leq 2^{-j} \text{ for } j = 1, 2, 3, \dots\}$.

By the utility-grade Bessaga-Pelczynski selection principle [67, page 42], some subsequence of (ϕ_j) is a basic sequence, and that subsequence will do. ■

6.2 Super-reflexive separable spaces

We say that a Banach space E has *variable upper estimates for basic subsequences* provided for each normalised basic sequence (u_n) in E there is a subsequence (u_{n_j}) and numbers $1 < q < \infty$ and $A > 0$ such that for each $\alpha \in \ell_q$, $\sum \alpha_j u_{n_j}$ converges and

$$\left\| \sum \alpha_j u_{n_j} \right\| \leq A \|\alpha\|_q.$$

It is not clear whether every reflexive Banach space has this property, although every super-reflexive Banach space does [91].

Theorem 6.5 *Let E be a separable reflexive Banach space such that E^* has variable upper estimates for basic subsequences and let C be a closed convex subset of E with empty interior. Then C is Haar null. In fact, there is a continuous $f_C : [0, 1] \rightarrow E$ such that*

$$\{t : f_C(t) \in C + x\}$$

has Lebesgue measure zero for every $x \in E$.

Corollary 6.6 *If C is a closed convex subset of a separable reflexive Banach space E such that E^* has variable upper estimates for basic subsequences then the boundary ∂C is a Haar null set.*

Proof If C has empty interior then the theorem is sufficient. Otherwise consider the distance function d_C which is Gâteaux differentiable except on a Haar null set N by Theorem 1.23. Since, every boundary point of C is a support point, d_C is not Gâteaux differentiable at any point of ∂C . Thus, we see that $\partial C \subset N$ and we are finished. ■

Before proving Theorem 6.5 it is convenient to give the special case where C is the positive cone for the usual order on ℓ_p , $1 < p < \infty$. The Schauder system in $C[0, 1]$ consists of the functions $s_0 := 1$, $s_1(t) := t$, $s_2(t) := 1 - |1 - 2t|$, and in general

$$s_{2^{k-1}+j}(t) := \max(0, 1 - |2j - 1 - 2^k t|)$$

for $1 \leq j \leq 2^{k-1}$.

Theorem 6.7 *Let (s_n) be the Schauder system and $g := \sum_{n=2}^{\infty} (-1/\log^2 n) s_n e_n$. For $1 < p < \infty$ the mapping g is continuous from $[0, 1]$ to ℓ_p and for each $x \in \ell_p$*

$$\lambda(\{t \in [0, 1] : g(t) \geq x\}) = 0.$$

Proof Let $g_k := \sum_{n=2^{k-1}+1}^{2^k} (-1/\log^2 n) s_n e_n$ and note that

$$\|g_k(t)\|_p \leq (\log(2^{k-1} + 1))^{-2} \leq (k - 1)^{-2}.$$

Since $\sum k^{-2}$ converges and the g_k are continuous it follows that g is continuous by the Weierstrass M -test.

Now suppose there is $x \in \ell_p$ such that

$$\lambda(\{t \in [0, 1] : g(t) \geq x\}) > 0.$$

Then there is a point $z \in (0, 1)$ of density for $A := \{t \in [0, 1] : g(t) \geq x\}$ and there is $\varepsilon > 0$ such that $\lambda(A \cap (z - \varepsilon, z + \varepsilon)) > 9\varepsilon/5$. If $t \in A$ and $s_n(t) \geq 1/5$ then $x_n \leq -1/(5 \log^2 n)$.

Let k be a positive integer, $2^{k-1} + 1 \leq n \leq 2^k$ and $x_n > -1/(5 \log^2 n)$. Then $s_n^{-1}[1/5, 1] \subset (z - \varepsilon, z + \varepsilon) \setminus A$ and $\lambda(s_n^{-1}[1/5, 1]) = 2^{3-k}/5$ so the number of such n with $\{s_n \neq 0\} \subset (z - \varepsilon, z + \varepsilon)$ is at most $2^{k-3}\varepsilon$. Therefore at least $2^k\varepsilon - 2 - 2^{k-3}\varepsilon$ of the n such that $2^{k-1} + 1 \leq n \leq 2^k$ have

$$x_n \leq -1/(5 \log^2 n) \leq -1/(5k^2 \log 2)$$

and that contradicts $x \in \ell_p$ as

$$\sum 2^{k-1}\varepsilon (5k^2 \log 2)^{-p}$$

diverges. ■

We note that $g(0) = g(1)$ so that g may be extended periodically to a function on \mathbf{R} with the same property. We leave open the question as to whether g may be made C^1 or better.

Proof of Theorem 6.5 By Proposition 6.3 we may and do assume that C is spanning. By Theorem 6.4 there are $\phi_n \in E^*$ such that $\|\phi_n\| = 1$ and ϕ_n converge weakly to $0 \in E^*$ such that $C \subset \{x \in E : \langle \phi_n, x \rangle \leq 2^{-n} \text{ for } n = 1, 2, 3, \dots\}$ and the ϕ_n are a Schauder basis of $X := \overline{\text{span}}\{\phi_n\}$. Since E^* has such that E^* has variable upper estimates for basic subsequences there is a subsequence (which we may assume to be ϕ_n itself) and numbers $1 < q < \infty$ and $A > 0$ such that if $\alpha \in \ell_q$ then $\sum \alpha_n \phi_n$ converges and

$$\|\sum \alpha_n \phi_n\| \leq A \|\alpha\|_q.$$

Define $T : \ell_q \rightarrow X := \overline{\text{span}}\{\phi_n\}$ by $T\alpha := \sum \alpha_n \phi_n$. Then T is a continuous linear mapping of ℓ_q into X with dense range. Now $T^* : X^* \rightarrow \ell_p$ is injective where $p^{-1} + q^{-1} = 1$. Also consider the natural injection $r : X \rightarrow E^*$ for which $r^* : E \rightarrow X^*$ is surjective so the Michael selection theorem or the Bartle-Graves theorem [82, page 184] gives a continuous selection $\gamma : X^* \rightarrow E$ for $(r^*)^{-1}$. Now let

$$f_C := \gamma \circ (T^*)^{-1} \circ g$$

where g is as in Theorem 6.7, and $\mu := \lambda \circ f_C^{-1}$. We need to prove that f_C is well-defined and continuous. Let (x_n) be the dual basis in X^* to (ϕ_n) . Then $(T^*)^{-1} \circ g = \sum_{n=2}^{\infty} (1/\log^2 n) s_n x_n$ and if $f_k := \sum_{n=2^{k-1}+1}^{2^k} (1/\log^2 n) s_n x_n$ then $(T^*)^{-1} \circ g = \sum_{k=1}^{\infty} f_k$. Since f_k is continuous and $\|f_k(t)\| \leq \sup_n \{\|x_n\|\}/(k-1)^2$ the M -test shows that $(T^*)^{-1} \circ g$ is continuous so f_C is continuous. Now let $y \in C$. Then

$$\begin{aligned} \langle T^* r^* y, e_n \rangle &= \langle y, r T e_n \rangle \\ &= \langle y, -\phi_n \rangle \geq -2^{-n} \end{aligned}$$

so $\langle T^* r^* y + \sum_{j=1}^{\infty} 2^{-j} e_j, e_n \rangle \geq 0$ and we have

$$T^*(r^* C) \subset -\sum_{j=1}^{\infty} 2^{-j} e_j + \ell_p^+.$$

Then for $x \in E$ we have

$$\mu(x + C) = \lambda(\{t \in [0, 1] : g(t) \in T^*(\gamma^{-1}(x + C))\})$$

$$\begin{aligned}
 &\leq \lambda(\{t \in [0, 1] : g(t) \in T^*(r^*(x + C))\}) \\
 &= \lambda(\{t \in [0, 1] : g(t) \in T^*r^*x + T^*(r^*C)\}) \\
 &\leq \lambda(\{t \in [0, 1] : g(t) \in T^*r^*x - \sum_{j=1}^{\infty} 2^{-j}e_j + \ell_p^+\}) \\
 &= 0
 \end{aligned}$$

by Theorem 6.7, since $T^*r^*x - \sum_{j=1}^{\infty} 2^{-j}e_j \in \ell_p$. ■

Corollary 6.8 *Let E be a separable super-reflexive Banach space and let C be a closed convex subset of E with empty interior. Then C is Haar null. In fact, there is a continuous $f_C : [0, 1] \rightarrow E$ such that $\{t : f_C(t) \in C + x\}$ has Lebesgue measure zero for every $x \in E$.*

Proof A theorem of James [91] (also see [6, page 243]) shows that E^* has variable upper estimates for basic subsequences (in fact here one need not take subsequences and the q doesn't depend on the basic sequence chosen) so Theorem 6.5 applies. ■

Corollary 6.9 *If C is a closed convex subset of a separable super-reflexive Banach space E then the boundary ∂C is a Haar null set.*

Thus, a closed convex set in a separable super-reflexive space (e.g. $L_p(\mu)$ for σ -finite μ) is Haar null exactly when it has void interior. Again, we conjecture this remains true in every separable reflexive space. We conclude this lecture by combining Theorem 1.24 and Theorem 6.5 to observe that for a separable super-reflexive space E and Lipschitz $f : E \rightarrow \mathbf{R}$

$$\partial f(x) = \overline{\text{co}}^*\{z^* : \nabla_G f(y) \rightarrow^* z^*, y \rightarrow x, y \notin S\}$$

where S is any closed convex set in E with empty interior, or indeed the countable union of closed convex sets with empty interior and of boundaries of arbitrary closed convex sets.

6.3 Other extensions

We may abstract the results of the previous lecture as follows. Let us say a set C in E admits a *strange normal*, g_C , if $g_C : [0, 1] \rightarrow E$ is such that g_C is continuous and $\{t : g_C(t) \in C + x\}$ has Lebesgue measure zero for every $x \in E$. By Proposition 2.1, every classical normal gives rise to a strange normal.

Theorem 6.10 *Let E and F be separable Banach spaces. Let $Q : F \rightarrow E$ be a quotient mapping (i.e., linear, continuous and onto). Suppose that $D \subset F$ and $C \subset E$ are convex closed with $Q(D) \subset C$. Then*

- (a) D is Haar null as soon as C is; moreover
- (b) if g_C is a strange normal for C and γ is a continuous selection of Q^{-1} then $g_D := \gamma \circ g_C$ is a strange normal for D .
- (c) In particular, if E is super-reflexive and C has empty interior then $Q^{-1}(C)$ possesses a strange normal and so is Haar null.

Proof (a) If μ annihilates translates of C and γ is a continuous selection of Q^{-1} then $\nu := \mu \circ \gamma^{-1}$ is a Borel measure annihilating all translates of D .

(b) One similarly checks that

$$\lambda(\{t \in [0, 1] : g_D(t) \in D + x\}) \leq \lambda(\{t \in [0, 1] : g_C(t) \in C + Qx\}) = 0.$$

■

It is known that there are non super-reflexive spaces in which the boundary of every closed convex set is null. Any space with the *Banach-Sacks property* [67] suffices.

References

- [1] Arnold V. I., *Geometrical Methods in the Theory of Ordinary Differential Equations*, 2nd ed., Springer, New York, 1988.
- [2] Aronszajn, N., Differentiability of Lipschitzian mappings between Banach spaces, *Studia Math.* **57** (1976), 147–190.
- [3] Aubin J. P., Frankowska H., *Set-valued Analysis*, Birkhäuser, Boston, 1990.
- [4] Aussel D., Corvellec J.-N., Lassonde M., Mean value property and subdifferential criteria for lower semicontinuous functions, *Trans. Amer. Math. Soc.* **347** (1995), 4147–4161.
- [5] Beer G., Borwein J. M., Mosco and slice convergence of level sets and graphs of linear functionals, *J. Math. Anal. Appl.* **175** (1993), 53–67.
- [6] Beauzamy, B., *Introduction to Banach Spaces and their Geometry*, 2nd ed., North-Holland, Amsterdam 1985.
- [7] Benyamini Y., The uniform classification of Banach spaces, in: *Texas Functional Analysis Seminar 1984-1985 (Austin, Tex.)*, Longhorn Notes, Univ. Texas Press, Austin, TX, 1985, 15–38.
- [8] Benoist J., Intégration du sous-différentiel proximal: un contre exemple, to appear in *Canad. J. Math.*
- [9] Bishop, E., Phelps, R. R., The support functionals of a convex set, in: *Convexity* (V. Klee, ed.) Proc. Sympos. Pure Math. **7**, Amer. Math. Soc., Providence, RI, 1963, 27–35.
- [10] Borwein J. M., Convex relations in analysis and optimization, in: *Generalized Concavity in Optimization and Economics* (S. Schaible and W. T. Ziemba, eds.), Academic Press, New York, 1981, 335–377.
- [11] Borwein J. M., Minimal cuscos and subgradients of Lipschitz functions, in *Fixed Point Theory and its Applications*, (J-B. Baillon and M. Thera, eds.), Pitman Res. Notes Math. Ser., Longman, Essex, 1991, 57–82.
- [12] Borwein J. M., Fabian M., On convex functions having points of Gâteaux differentiability which are not points of Fréchet differentiability, *Canad. J. Math.* **45** (1993), 1121–1134.

- [13] Borwein J. M., Fabian M., A note on regularity of sets and of distance functions in Banach space, *J. Math. Anal. Appl.* **182** (1994), 566–560.
- [14] Borwein J. M., Fitzpatrick S., Existence of nearest points in Banach spaces, *Canad. J. Math.* **61** (1989), 702-720.
- [15] Borwein J. M., Fitzpatrick S., A weak Hadamard smooth renorming of $L_1(\Omega, \mu)$, *Canad. Math. Bull.* **36** (1993), 407-413.
- [16] Borwein J. M., Fitzpatrick S., Weak* sequential compactness and bornological limit derivatives, *J. Convex Anal.* **2** (1995), 59-68 (Special Issue in Celebration of R.T.Rockafellar's 60th Birthday. Part I).
- [17] Borwein J. M., Fitzpatrick S., Characterization of Clarke subgradients among one-dimensional multifunctions, in: *Proc. Optimization Miniconference II*, Univ. Ballarat Press, 1995, 61–73.
- [18] Borwein J. M., Fitzpatrick S., Closed convex Haar null sets, unpublished manuscript, CECM Research Report 95-052 (1995).
- [19] Borwein J. M., Fitzpatrick S., Giles R., The differentiability of real functions on normed linear spaces using generalized subgradients, *J. Math. Anal. Appl.* **128** (1987) 512–534.
- [20] Borwein J. M., Fitzpatrick S., Kenderov P., Minimal convex usco and monotone operators on small sets, *Canad. J. Math.* **43** (1991), 461–476.
- [21] Borwein J. M., Fitzpatrick S., Vanderwerff J., Examples of convex functions and classifications of normed spaces, *J. Convex Anal.* **1** (1994), 61-73.
- [22] Borwein J. M., Girgensohn R., Wang X., On the Construction of Hölder and Proximal Subderivatives, CECM Research Report 97-091, to appear in *Canad. Math. Bull.*
- [23] Borwein J. M., Ioffe A., Proximal analysis in smooth spaces, *Set-Valued Anal.* **4** (1996), 1-24.
- [24] Borwein J. M., Jofré A., A nonconvex separation property in Banach spaces, CECM Research Report 97-103, to appear in *ZOR: Math. Methods Oper. Res.*
- [25] Borwein J. M., Lewis, A., Partially-finite convex programming, (I) & (II), *Math. Programming (Ser. B)* **57** (1992), 15–48, 49–84.
- [26] Borwein J. M., Lewis, A., *Convex Analysis*, Wiley, in press.
- [27] Borwein J. M., Moors W. B., Essentially smooth Lipschitz functions, *J. Funct. Anal.* **49** (1997), 305-351.
- [28] Borwein J. M., Moors W. B., A chain rule for essentially strictly differentiable functions, *SIAM J. Optim.* **8** (1998), 300-308.
- [29] Borwein J. M., Moors W. B., Null sets and essentially smooth Lipschitz functions, *SIAM J. Optim.* **8** (1998), 309-323.

- [30] Borwein J. M., Moors W. B., Lipschitz functions with minimal Clarke subdifferential mappings, in: *Proc. Optimization Miniconference III (1996)*, Univ. of Ballarat Press, 1997, 5-12,.
- [31] Borwein J. M., Moors W. B., Separable determination of integrability and minimality of the Clarke subdifferential mapping, CECM Research Report 97-102, to appear in *Proc. Amer. Math. Soc.*
- [32] Borwein J. M., Moors W. B., Shao, Y., Subgradient representation of multifunctions, CECM Research Report 97-090, *J. Austral. Math. Soc. Ser. B*, volume in honour of B. Craven and B. Mond, 1998.
- [33] Borwein J. M., Moors W. B., Wang X., Lipschitz functions with prescribed derivatives and subderivatives, *Nonlinear Anal.* **29** (1997), 53-63.
- [34] Borwein J. M., Mordukhovich B. S., Shao Y., On the equivalence of some basic principles in variational analysis, CECM Research Report 97-098.
- [35] Borwein J. M., Noll D., Second order differentiability of convex functions in Banach spaces, *Trans. Amer. Math. Soc.* **342** (1994), 43-81.
- [36] Borwein J. M., Preiss D., A smooth variational principle with applications to subdifferentiability and to differentiability of convex functions, *Trans. Amer. Math. Soc.* **303** (1987), 517-527.
- [37] Borwein J. M., Treiman J. S., Zhu Q. J., Necessary conditions for constrained optimization problems with semicontinuous and continuous data, *Trans. Amer. Math. Soc.* **350** (1998), 2409-2429.
- [38] Borwein J. M., Treiman J. S., Zhu Q. J., Partially smooth variational principles and applications, CECM Research Report 96-088 (1996), to appear in *Nonlinear Anal.*
- [39] Borwein J. M., Vanderwerff J., Banach spaces which admit support sets, *Proc. Amer. Math. Soc.* **124** (1996), 751-756.
- [40] Borwein J. M., Wang X., Lipschitz functions with maximal Clarke subgradients are generic, submitted to *Proc. Amer. Math. Soc.*
- [41] Borwein J. M., Zhu Q. J., Viscosity solutions and viscosity subderivatives in smooth Banach spaces with applications to metric regularity, *SIAM J. Control Optim.* **34** (1996), 1568-1591.
- [42] Borwein J. M., Zhu Q. J., Limiting convex examples for nonconvex subdifferential calculus, CECM Research Report, 97-097, to appear in *J. Convex Anal.*
- [43] Borwein J. M., Zhu Q. J., A survey of subdifferentials and their applications, CECM Research Report 98-105 (1998), to appear in *Nonlinear Anal.*
- [44] Choquet, G. *Lectures on Analysis* Vol. I, W. A. Benjamin, New York, 1969.
- [45] Christian B., Christensen J. P. R., Ressel P., *Harmonic Analysis on Semigroups*, Springer-Verlag, New York, 1984.

- [46] Christensen J. P. R., On sets of Haar measure zero in Abelian Polish groups, *Israel J. Math.* **13** (1972), 255–260.
- [47] Christensen J. P. R., *Topological and Borel Structure*, American Elsevier, New York, 1974.
- [48] Christensen J. P. R., Theorems of Namioka and R. E. Johnson type for upper semicontinuous and compact valued set-valued mappings, *Proc. Amer. Math. Soc.* **86** (1982), 649–655.
- [49] Clarke F. H., Necessary Conditions for Nonsmooth Problems in Optimal Control and the Calculus of Variations, Ph. D. thesis, Univ. of Washington, 1973.
- [50] Clarke F. H., Optimization and Nonsmooth Analysis, John Wiley & Sons, New York, 1983, Russian edition MIR, Moscow, 1988. Reprinted as Classics Appl. Math. **5**, SIAM, Philadelphia, PA, 1990.
- [51] Clarke F. H., *Methods of Dynamic and Nonsmooth Optimization*, CBMS-NSF Regional Conf. Ser. in Appl. Math., **57**, SIAM, Philadelphia, PA, 1989.
- [52] Clarke F. H., Lecture Notes in Nonsmooth Analysis, unpublished.
- [53] Clarke F. H., Ledyaev Yu. S., Mean value inequalities, *Proc. Amer. Math. Soc.* **122** (1994), 1075–1083.
- [54] Clarke F. H., Ledyaev Yu. S., Mean value inequalities in Hilbert space, *Trans. Amer. Math. Soc.* **344** (1994), 307–324.
- [55] Clarke F. H., Ledyaev Yu. S., Stern R. J., Wolenski P. R., *Nonsmooth Analysis and Control Theory*, Grad. Texts in Math. **178**, Springer-Verlag, New York, 1998.
- [56] Correa R., Jofré A., Tangentially continuous directional derivatives in Nonsmooth Analysis, *J. Optim. Theory Appl.* **61**(1) (1989), 1–21.
- [57] Correa R., Jofré A., Thibault L., Characterization of lower semicontinuous convex functions, *Proc. Amer. Math. Soc.* **116** (1992), 61–72.
- [58] Correa R., Jofré A., Thibault L., Subdifferential monotonicity as characterization of convex functions, *Numer. Funct. Anal. Optim.* **15** (1994), 531–535.
- [59] Correa R., Thibault L., Subdifferential Analysis of bivariate separately regular functions, *J. Math. Anal. Appl.* **148** (1990), 157–174 .
- [60] Crandall M. G., Lions P.-L., Viscosity solutions of Hamilton-Jacobi equations, *Trans. Amer. Math. Soc.* **277** (1983), 1–42.
- [61] Crandall M. G., Evans L. C., Lions P.-L., Some properties of viscosity solutions of Hamilton-Jacobi equations, *Trans. Amer. Math. Soc.* **282** (1984), 487–502.
- [62] Day M. M., *Normed Linear Spaces*, 3rd ed., Springer-Verlag, New York, 1973.

- [63] Deville R., Godefroy G., Zizler V., *Smoothness and Renormings in Banach Spaces*, Pitman Monogr. Surveys Pure Appl. Math. **64**, J. Wiley, New York, 1993.
- [64] Deville R., Godefroy G., Zizler V., A smooth variational principle with applications to Hamilton-Jacobi equations in infinite dimensions, *J. Funct. Anal.* **111** (1993), 197–212.
- [65] Deville R., Haddad E. M. E., The subdifferential of the sum of two functions in Banach spaces, I. first order case, *J. Convex Anal.* **3** (1996), 295–308.
- [66] Deville R., Ivanov M., Smooth variational principles with constraints, *Arch. Math. (Basel)* **69** (1997), 418–426.
- [67] Diestel J., *Sequences and Series in Banach Spaces*, Springer-Verlag, New York, 1984.
- [68] Dudley, R. M., *Real Analysis and Probability*, Wadsworth, Belmont, 1989.
- [69] M. Edelstein M., Thompson A. C., Some results on nearest points and support properties of convex sets in c_0 , *Pacific J. Math.* **40** (1972), 553–560.
- [70] Ekeland I., On the variational principle, *J. Math. Anal. Appl.* **47** (1974), 324–353.
- [71] Evans, L. C., Gariepy, R. F. *Measure Theory and Fine Properties of Functions*, CRC Press, Boca Raton, FL, 1992.
- [72] Fabian M., On class of subdifferentiability spaces of Ioffe, *Nonlinear Anal.* **12** (1988), 63–74.
- [73] Fabian M., Subdifferentiability and trustworthiness in the light of a new variational principle of Borwein and Preiss, *Acta Univ. Carolin. – Math. Phys.* **30** (1989), 51–56.
- [74] Fabian M., Zajíček L., Zizler V., On residuality of the set of rotund norms on a Banach space, *Math. Ann.* **258** (1982), 349–351.
- [75] Finet C., Godefroy G., Biorthogonal systems and big quotient spaces, *Contemp. Math.* **85** (1989), 87–110.
- [76] Giles J. R., *Convex Analysis with Applications in Differentiation of Convex Functions*, Res. Notes in Math. **58**, Pitman, Boston–London–Melbourne, 1982.
- [77] Giles J. R., Kenderov P. S., Moors W. B., Sciffer S. D., Generic differentiability of convex functions on the dual of a Banach space, *Pacific J. Math.* **172** (1996), 413–433.
- [78] Giles J. R., Moors W. B., A continuity property related to Kuratowski's index of non-compactness, its relevance to the Drop property, and its implications for differentiability theory, *J. Math. Anal. Appl.* **178** (1993), 247–268 .
- [79] Giles J. R., Moors W. B., Generic continuity of minimal set-valued mappings, *J. Austral. Math. Soc. Ser. A* **63** (1997), 1–25.
- [80] Helgason S., *Differential Geometry, Lie Groups, and Symmetric Spaces*, Academic Press, New York, 1978.

- [81] Hiriart-Urruty J. B., Tangent cones, generalized gradients and mathematical programming in Banach spaces, *Math. Oper. Res.* **4** (1979), 79-97.
- [82] Holmes R. B., *Geometric Functional Analysis and Its Applications*, Springer, New York, 1975.
- [83] Ioffe A. D., Nonsmooth analysis: differential calculus of nondifferentiable mappings, *Trans. Amer. Math. Soc.* **266** (1981), 1-56.
- [84] Ioffe A. D., Calculus of Dini subdifferentials of functions and contingent derivatives of set-valued maps, *Nonlinear Anal.* **8** (1984), 517-539.
- [85] Ioffe A. D., Approximate subdifferentials and applications. I: The finite dimensional theory, *Trans. Amer. Math. Soc.* **281** (1984), 389-416.
- [86] Ioffe A. D., Approximate subdifferentials and applications. II: Functions on local convex spaces, *Mathematika* **33** (1986), 111-128.
- [87] Ioffe A. D., Approximate subdifferentials and applications III: The metric theory, *Mathematika* **36** (1989), 1-38.
- [88] Ioffe A. D., Proximal analysis and approximate subdifferentials, *J. London Math. Soc.* **41** (1990), 175-192.
- [89] Ioffe A. D., Fuzzy principles and characterization of trustworthiness, preprint.
- [90] Ioffe A. D., Rockafellar T. R., The Euler and Weierstrass conditions for nonsmooth variational problems, *Calc. Var. Partial Differential Equations* **4** (1996), 59-87.
- [91] James, R. C., Super-reflexive spaces with bases, *Pacific J. Math.* **41** (1972), 409-419 .
- [92] Janin, R., Sur des multiapplications qui sont des gradients géénéralisés, *C.R. Acad. Sci. Paris Sér. I Math.* **294** (1982), 115-117.
- [93] Jokl L., Minimal convex-valued weak* usco correspondences and the Radon-Nikodym property, *Comment. Math. Univ. Carolin.* **28** (1987), 353-375.
- [94] Jofre A., Thibault L., D-representation of subdifferentials of directionally Lipschitzian functions, *Proc. Amer. Math. Soc.* **110** (1990), 117-123.
- [95] Kempisty S., Sur les fonctions quasicontinues, *Fund. Math.* **19** (1932), 184-197.
- [96] Kenderov P. S., Orihuela, On a generic factorization theorem, *Mathematika* **42** (1995), 56-66.
- [97] Konjagin S. V., On approximation property of closed sets in Banach spaces and the characterization of strongly convex spaces, *Soviet Math. Dokl.* **21** (1980), 418-422.
- [98] Krasovskii N. N., Subbotin A. I., *Game-Theoretical Control Problems*, Springer-Verlag, New York, 1990.
- [99] Kruger A. Y., Mordukhovich B. S., Extremal points and Euler equations in nonsmooth optimization, *Dokl. Akad. Nauk. BSSR* **24** (1980), 684-687 (Russian).

- [100] Larman D. G., Phelps R. R., Gâteaux differentiability of convex functions on Banach spaces, *J. London Math.Soc.* **20** (1979), 115–127.
- [101] Lassonde M., First-order rules for nonsmooth constrained optimization, preprint.
- [102] Levy A. B., Poliquin R. A., Characterizing the single-valuedness of multifunctions, *Set-Valued Anal.* **5** (1997), 351–364.
- [103] Lewis A. S., Convex analysis on the Hermitian matrices, *SIAM J. Optim.* **6** (1996), 164–177.
- [104] Lewis A. S., Nonsmooth analysis of eigenvalues, preprint.
- [105] Li Y., Shi S., A generalization of Ekeland’s ε -variational principle and of its Borwein-Preiss’ smooth variant, to appear in *J. Math. Anal. Appl.*
- [106] Lindenstrauss J., Tzafriri L., *Classical Banach Spaces II: Function Spaces*, Springer-Verlag, Berlin, 1979.
- [107] Mankiewicz, P. On the differentiability of Lipschitz mappings in Fréchet spaces, *Studia Math.* **45** (1973), 15–29.
- [108] Matouskova E., Stegall C., A characterization of reflexive spaces, *Proc. Amer. Math. Soc.* **124** (1996), 1083-1090.
- [109] Michel P., Penot J. P., Calcul sous-différentiel pour des fonctions Lipschitziennes et non-Lipschitziennes, *C. R. Acad. Sci. Paris Sér. I Math.* **298** (1985), 269–272.
- [110] Mifflin R., Semismooth and semiconvex functions in Optimization, *SIAM J. Control. Optim.* **15** (1977), 959–972 .
- [111] Minty, G., Monotone (nonlinear) operators in Hilbert spaces, *Duke Math. J.* **29** (1962), 341–346 .
- [112] Moors W. B., A characterisation of minimal subdifferential mappings of locally Lipschitz functions, *Set-Valued Anal.* **3** (1995), 129–141.
- [113] Mordukhovich B. S., Maximum principle in problems of time optimal control with nonsmooth constraints, *J. Appl. Math. Mech.* **40** (1976), 960–969.
- [114] Mordukhovich B. S., Metric approximations and necessary optimality conditions for general classes of nonsmooth extremal problems, *Soviet Math. Dokl.* **22** (1980), 526–530.
- [115] Mordukhovich B. S., *Approximation Methods in Problems of Optimization and Control*, Nauka, Moscow, 1988 (Russian; English transl. to appear in Wiley-Interscience).
- [116] Mordukhovich B. S., Complete characterization of openness, metric regularity, and Lipschitzian properties of multifunctions, *Trans. Amer. Math. Soc.* **340** (1993), 1–35.
- [117] Mordukhovich B. S., Generalized differential calculus for nonsmooth and set-valued mappings, *J. Math. Anal. Appl.* **183** (1994), 250–288.

- [118] Mordukhovich B. S., Shao Y., Extremal characterizations of Asplund spaces, *Proc. Amer. Math. Soc.* **124** (1996), 197–205.
- [119] Oxtoby J. C., *Measure and Category*, Springer Verlag, New York-Berlin-Tokyo, 1971.
- [120] Phelps R. R., *Convex Functions, Monotone Operators and Differentiability*, Lecture Notes in Math. **1364**, Springer Verlag, New York, Berlin, Tokyo, 1988, 2nd ed. 1993.
- [121] Phelps R. R., Gaussian null sets and differentiability of Lipschitz maps on Banach spaces, *Pacific J. Math.* **77** (1978), 523–531.
- [122] Poliquin R. A., Integration of subdifferentials of nonconvex functions, *Nonlinear Anal.* **17** (1991), 385–398.
- [123] Poliquin R. A., A characterization of proximal subgradient set-valued mappings, *Canad. Math. Bull.* **36** (1993), 116–122.
- [124] Preiss D., Fréchet derivatives of Lipschitzian functions, *J. Funct. Anal.* **91** (1990), 312–345.
- [125] Preiss D., Zajiček L., Stronger estimates of smallness of sets of Fréchet nondifferentiability of convex functions, Proc. 11th Winter School, *Rend. Circ. Mat. Palermo (2) Suppl.* **3** (1984), 219–223.
- [126] Qi L., The maximal normal operator space and integration of subdifferentials of nonconvex functions, *Nonlinear Anal.* **13** (1989), 1003–1011.
- [127] Rockafellar R. T., *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
- [128] Rockafellar R. T., On the maximal monotonicity of subdifferential mappings, *Pacific J. Math.* **33** (1970), 209–216 .
- [129] Rockafellar R. T., Favorable classes of Lipschitz-continuous functions in subgradient optimization, in: *Progress in Nondifferentiable Optimization* (E. Nurminski, ed.), IIASA Collaborative Proc. Ser., Internat. Inst. Appl. Systems Anal., Laxenburg, 1982, 125–144. .
- [130] Rockafellar R. T., Extensions of subgradients and its applications to optimization, *Nonlinear Anal.* **9** (1985), 665–698 .
- [131] Rockafellar R. T., Wets R. J-B., *Variational Analysis*, Springer, New York, 1997.
- [132] Rudin W., *Functional Analysis*, McGraw-Hill, New York, 1973.
- [133] Simon B., *Trace Ideals and Their Applications*, Cambridge University Press, 1979.
- [134] Simons S., The least slope of a convex function and the maximal monotonicity of its subdifferential, *J. Optim. Theory Appl.* **71** (1991), 127–136.
- [135] Stegall C., A class of topological spaces and differentiation of functions between Banach spaces, in: *Proc. Conf. on Vector Measures and Integral Representations of Operators* (W. Ruess, ed.), Vorlesungen Fachbereich Math. Univ. Essen **10**, 1983, 63–77.

- [136] Stromberg K. R., *An Introduction to Classical Real Analysis*, Wadsworth International Mathematics Series, 1981.
- [137] Sussmann H. J., A strong version of the maximum principle under weak hypotheses, in: *Proc. 33rd IEEE Conf. Decision and Control (Lake Buena Vista, FL, 1994)*.
- [138] Sussmann H. J., Multidifferential calculus: chain rule, open mapping and transversal intersection theorems, to appear in: *Proc. IFIP Conf. on Optimal Control: Theory Algorithms, and Applications (Gainesville, 1997)* (W. Hager, ed.).
- [139] Thibault L., On Generalized differentials and subdifferentials of Lipschitz vector-valued functions, *Nonlinear Anal.* **6** (1982), 1037–1053.
- [140] Thibault L., A note on the Zagrodny mean value theorem, *Optimization* **35** (1995), 127-130.
- [141] Thibault L., Zagrodny D., Integration of subdifferentials of lower semicontinuous functions on Banach spaces, *J. Math. Anal. Appl.* **189** (1995), 33–58.
- [142] Treiman J. S., Characterization of Clarke’s tangent and normal cones in finite and infinite dimensions, *Nonlinear Anal.* **7** (1983), 771–783.
- [143] Treiman J. S., Clarke’s Gradients and epsilon-subgradients in Banach spaces, *Trans. Amer. Math. Soc.* **294** (1986), 65–78.
- [144] Treiman J. S., Finite dimensional optimality conditions: B-gradients, *J. Optim. Theory Appl.* **62** (1989), 139–150.
- [145] Treiman J. S., Optimal control with small generalized gradients, *SIAM J. Control Optim.* **28** (1990), 720–732.
- [146] Valadier M., “Entraînement unilatéral, lignes de descente, fonctions lipschitziennes non pathologiques,” *C. R. Acad. Sci. Paris Sér. I. Math.* **308** (1989), 241–244.
- [147] Vanderwerff J. Private communication. September, 1996.
- [148] Vanderwerff J., Zhu Q. J., A limiting example for the local “fuzzy” sum rule in nonsmooth analysis, CECM Research Report 96-083 (1996), to appear in *Proc. Amer. Math. Soc.*
- [149] Vlasov L. P., Almost convex and Chebychev sets, *Mat. Zametki* **8** (1970), 545-550 (Russian); *Math. Notes* **8** (1970), 776–779 .
- [150] Wang X. F., Distance function on any Hilbert space are δ convex, preprint.
- [151] Warga J., Derivate containers, inverse functions, and controllability, in *Calculus of Variations and Control Theory* (D.L. Russell, ed.), Math. Res. Center Univ. Wisconsin Publ. **36**, Academic Press, New York, 1976, 13-45.
- [152] Warga J., Fat homeomorphisms and unbounded derivate containers, *J. Math. Anal. Appl.* **81** (1981), 545–560.

- [153] L. Zajíček, On differentiability properties of Lipschitz functions on a Banach space with a Lipschitz uniformly Gâteaux differentiable bump function, *Comment. Math. Univ. Carolin.* **38** (1997), 329–336 .
- [154] Zizler V., Some notes on various rotundity and smoothness properties of separable Banach spaces, *Comment. Math. Univ. Carolin.* **10** (1967), 195–206.
- [155] Zhu Q. J., The Clarke-Ledyev mean value inequality in smooth Banach spaces, *Non-linear Anal.* **32** (1998), 315-324.
- [156] Zhu Q. J., Subderivatives and their applications, in: *Proc. Internat. Conf. on Dynamical Systems and Differential Equations (Springfield, MO, 1996)*.
- [157] Zhu Q. J., The equivalence of several basic theorems for subdifferentials, CECM Research Report 97-093 (1997), to appear in *Set-Valued Analysis*.