BULL. AUSTRAL. MATH. SOC. 58C20, 90C48
VOL. 47 (1993) [443–455]

THE APPROXIMATE SUBDIFFERENTIAL OF COMPOSITE FUNCTIONS

A. JOURANI AND L. THIBAULT

This paper deals with the approximate subdifferential chain rule in a Banach space. It establishes specific results when the real-valued function is locally Lipschitzian and the mapping is strongly compactly Lipschitzian.

0. INTRODUCTION

In [**8**] we have proved that, under the metric regularity assumption (a general constraint qualification), a point $\overline{x}$ is a local minimum to the constrained problem

$$(\mathcal{P}) \qquad \text{minimise } g(x) \qquad \text{subject to } x \in C \text{ and } G(x) \in D$$

(where $g\colon X \to \mathbb{R}$ and $G\colon X \to Y$ are locally Lipschitz at $\overline{x}$ and C and D are two closed subsets of the Banach spaces X and Y respectively) if and only if $\overline{x}$ is a local minimum to the unconstrained problem

$$(\mathcal{P}') \qquad \text{minimise } f \circ F(x) \qquad \text{over all } x \in X$$

where $F(x) = (g(x), G(x), kd(x; C)) \in \mathbb{R} \times Y \times \mathbb{R}$ and $f(s, y, t) = s + kd(y, D) + t$. Obviously f and F are also locally Lipschitz. When Y is finite dimensional Clarke's formula says that, for $z := F(\overline{x})$,

$$(1) \qquad \partial_c(f \circ F)(\overline{x}) \subset \overline{\text{co}}\left(\bigcup_{z^* \in \partial_c f(z)} \partial_c(z^* \circ F)(\overline{x}) \right)$$

and hence, because of the convex closure operation $\overline{\text{co}}$, one cannot get directly Lagrange multipliers for problem $(\mathcal{P})$ by applying formula (1) and the well known principle $0 \in \partial_c(f \circ F)(\overline{x})$. One of the most important properties of the approximate subdifferential introduced by Mordukhovich [**9**] is that it satisfies formula (1) without the convex closure operation, that is

$$(2) \qquad \partial_A(f \circ F)(\overline{x}) \subseteq \bigcup_{z^* \in \partial_A f(z)} \partial_A(z^* \circ F)(\overline{x})$$

Received 28 May 1992

whenever X and Y are both finite dimensional (see [**3**, **9**, and **7**]) and hence by using our reduction procedure above we can easily derive Lagrange multipliers (relative to the approximate subdifferential) for problem $(\mathcal{P})$ by writing $0 \in \partial_A(f \circ F)(\overline{x})$. Moreover these multipliers are also multipliers relative to the Clarke subdifferential since the approximate subdifferential for any locally Lipschitz function is included in the Clarke subdifferential.

Ioffe [**6**] has extended formula (2) to the case where X and Y are general Banach spaces and F admit a strict prederivative with compact values. Our aim in this paper is to prove (when X and Y are Banach spaces) formula (2) for the larger class of strongly compactly Lipschitzian mappings F, a variant of the class of compactly Lipschitzian mappings introduced by the second author in [**10**]. Many results of this article are largely inspired by the papers [**2**] and [**6**] of Ioffe. Because of the importance, in our opinion, of this composition formula, and in order to make the paper self-contained we recall all the notion that we use and we give detailed proofs of the main results.

1. Preliminaries

Throughout the paper X and Y are Banach spaces and we denote by B_X, B_Y, B_X^* and B_Y^* the closed unit balls of X, Y, X^* and Y^* respectively and $B(v, s) = \{z\colon \|v - z\| \leqslant s\}$. By $\langle .;. \rangle$ we denote the canonical pairing between the space and its dual and also the inner product in any Euclidean subspace $L \subseteq X$. We also write

$$L^{\perp} = \{x^* \in X^* : \langle x^*; x\rangle = 0, \quad \forall x \in L\}.$$

If f is an extended-real-valued function on X, we write for any subset D of X

$$f_D(x) = \begin{cases} f(x), & \text{if } x \in D \\ +\infty, & \text{otherwise.} \end{cases}$$

The function

$$d^- f(x; h) = \liminf_{\substack{u \to h \\ t \downarrow 0}} t^{-1}(f(x + tu) - f(x))$$

is the lower Dini derivative of f at x and

$$\partial^- f(x) = \{x^* \in X^* : \langle x^*; h\rangle \leqslant d^- f(x; h), \quad \forall h \in X\}$$

and

$$\partial_\varepsilon^- f(x) = \{x^* \in X^* : \langle x^*; h\rangle \leqslant d^- f(x; h) + \varepsilon \|h\|, \quad \forall h \in X\}$$

are the Dini subdifferential and the Dini ε-subdifferential of f at x.

REMARK. For any locally Lipschitz function f one has

$$d^- f(x;h) = \liminf_{t \downarrow 0} t^{-1}(f(x+th) - f(x)).$$

DEFINITION 1.1: A collection $\mathcal{L}$ of closed subspaces of X will be called admissible if

(a) every $x \in X$ belongs to some $L \in \mathcal{L}$;
(b) for any L_1, $L_2 \in \mathcal{L}$ there is an $L \in \mathcal{L}$ containing both L_1 and L_2.

EXAMPLE: The family $\mathcal{F}$ of all finite dimensional subspaces of X is an admissible one.

In all the sequel $\limsup_{x \xrightarrow{f} \overline{x}} \partial^- f_{x+L}(x)$ will denote the weak-star superior limit set, that is

$$\limsup_{x \xrightarrow{f} \overline{x}} \partial^- f_{x+L}(x) = \{x^* \in X^* : x^* = w^* - \lim_i x_i^*,\ x_i^* \in \partial^- f_{x_i+L}(x) \text{ and } x_i \xrightarrow{f} \overline{x}\}$$

where $x \xrightarrow{f} \overline{x}$ means that $x \to \overline{x}$ and $f(x) \to f(\overline{x})$.

DEFINITION 1.2: [4] Let $\mathcal{F}$ be the previous collection and f be a lower semicontinuous function on X with $|f(\overline{x})| < +\infty$. The A-approximate subdifferential of f at $\overline{x}$ is defined by

$$\partial_A f(\overline{x}) = \bigcap_{L \in \mathcal{F}} \limsup_{x \xrightarrow{f} \overline{x}} \partial^- f_{x+L}(x).$$

REMARK. The set-valued mapping $x \to \partial_A f(x)$ is upper semicontinuous in the following sense: $\partial_A f(x) = \limsup_{x \xrightarrow{f} \overline{x}} \partial_A f(x)$, (see [4]).

DEFINITION 1.3: [4] One says that X is a weak trustworthy space (WT-space) if for any two lower semicontinuous functions f^1 and f^2 on X and any $\varepsilon > 0$

$$\partial_\varepsilon^- (f^1 + f^2)(x) \subset \limsup_{\substack{x_i \xrightarrow{f^i} x \\ i=1,2}} (\partial_\varepsilon^- f^1(x_1) + \partial_\varepsilon^- f^2(x_2)).$$

EXAMPLE: Every separable space is a WT-space (see [5]).

LEMMA 1.4. *Let $T: X \to Y$ be a surjective continuous linear operator between two Banach spaces X and Y and let $M: Z \rightrightarrows Y$ be a multifunction with nonempty values where Z is a metric space. Then*

$$T^{-1}\left(\|\ \| - \limsup_{z \to \overline{z}} M(z)\right) \subset \|\ \| - \limsup_{z \to \overline{z}} T^{-1}(M(z)),$$

https://doi.org/10.1017/S0004972700015276 Published online by Cambridge University Press

where $\| \| - \limsup$ *denotes the strong superior limit set.*

PROOF: Let $Tx \in \limsup_{z \to \overline{z}} M(z)$. Without loss of generality we may assume that there is $z_n \to \overline{z}$ and $y_n \in M(z_n)$ such that $y_n \to Tx$. From the surjectivity of T one has (see for example [1]) the existence of $a \geqslant 0$ and $r > 0$ such that

$$d(x', T^{-1}(y')) \leqslant a\, d(y', Tx')$$

for all $x' \in B(x, r)$ and $y' \in B(Tx, r)$, where $d(v, D) = \inf\{\|v - v'\| : v' \in D\}$. So there is $n_0 \in \mathbb{N}$ such that for $n > n_0$ one has $y_n \in B(Tx, r)$ and

$$d(x, T^{-1}(y_n)) \leqslant a\, d(y_n, Tx),$$

and hence there is $x_n \in T^{-1}(y_n)$ such that

$$d(x, x_n) \leqslant 2a\, d(y_n, Tx)$$

which implies that $x \in \limsup_{n \to +\infty} T^{-1}(y_n) \subset \limsup_{z \to \overline{z}} T^{-1}(M(z))$. ∎

The following lemma and the next propositions will be used in Section 2.

LEMMA 1.5. *Let L be a finite dimensional subspace of X, f^1 and f^2 two lower semi-continuous functions on X and $\delta > 0$. Then*

$$\partial_\varepsilon^- (f^1 + f^2)_{x+L}(x) \subset \limsup_{\substack{x_i \xrightarrow{f^i} x \\ i=1,2}} \left(\partial_\varepsilon^- f^1_{x_1+L}(x_1) + \partial_\varepsilon^- f^2_{x_2+L}(x_2) + L^\perp\right).$$

PROOF: Let $P: L \to X$ be the imbedding operator and let $P^*: X^* \to X^*/L^\perp$ be the canonical projection. For each $h \in L$ we set $g_1(h) = f^1_{x+L}(x + Ph)$ and $g_2(h) = f^2_{x+L}(x + Ph)$. It is not difficult to see that for any $u_1, u_2 \in L$

$$d^- g_1(u_1, h) = d^- f^1_{x+Pu_1+L}(x + Pu_1, Ph) \text{ and } d^- g_2(u_2, h) = d^- f^2_{x+Pu_2+L}(x + Pu_2, Ph).$$

Let us note that $\partial_\varepsilon^- f^1_{x+Pu_1+L}(x + Pu_1) = (P^*)^{-1}(\partial_\varepsilon^- g_1(u_1))$ since for $P^*x \in \partial_\varepsilon^- g_1(u_1)$ we have for all $h \in L$

$$\begin{aligned} \langle P^* x^*, h \rangle &\leqslant d^- g_1(u_1, h) + \varepsilon \|h\| \\ &= d^- f^1_{x+Pu_1+L}(x + Pu_1, Ph) + \varepsilon \|h\| \end{aligned}$$

which is equivalent to $x^* \in \partial_\varepsilon^- f^1_{x+Pu_1+L}(x + Pu)$. So as L is a WT-space

$$\partial_\varepsilon^- (g_1 + g_2)(0) \subset \limsup_{\substack{u_i \xrightarrow{g_i} 0 \\ i=1,2}} \left(\partial_\varepsilon^- g_1(u_1) + \partial_\varepsilon^- g_2(u_2)\right)$$

and hence becasue the surjectivity of P^* and Lemma 1.4 it follows that

$$\begin{aligned}
\partial_\varepsilon^-\left(f^1+f^2\right)_{x+L}(x) &= (P^*)^{-1}\left(\partial_\varepsilon^-(g_1+g_2)(0)\right)\\
&\subset (P^*)^{-1}\left(\|\ \| - \limsup_{\substack{u_i \xrightarrow{g_i} 0\\ i=1,2}} \left(\partial_\varepsilon^- g_1(u_1)+\partial_\varepsilon^- g_2(u_2)\right)\right)\\
&\subset \|\ \| - \limsup_{\substack{u_i \xrightarrow{g_i} 0\\ i=1,2}} (P^*)^{-1}\left(\partial_\varepsilon^- g_1(u_1)+\partial_\varepsilon^- g_2(u_2)\right)\\
&\subset \limsup_{\substack{u_i \xrightarrow{g_i} 0\\ i=1,2}} (P^*)^{-1}\left(\partial_\varepsilon^- g_1(u_1)+\partial_\varepsilon^- g_2(u_2)\right)\\
&\subset \limsup_{\substack{u_i \xrightarrow{g_i} 0\\ i=1,2}} [(P^*)^{-1}\left(\partial_\varepsilon^- g_1(u_1)\right)+(P^*)^{-1}\left(\partial_\varepsilon^- g_2(u_2)\right)+L^\perp]\\
&\subset \limsup_{\substack{u_i \xrightarrow{g_i} 0\\ i=1,2}} [\partial_\varepsilon^- f^1_{x+Pu_1+L}(x+Pu_1)+\partial_\varepsilon^- f^2_{x+Pu_2+L}(x+Pu_2)+L^\perp]\\
&\subset \limsup_{\substack{x_i \xrightarrow{f^i} x\\ i=1,2}} [\partial_\varepsilon^- f^1_{x_1+L}(x_1)+\partial_\varepsilon^- f^2_{x_2+L}(x_2)+L^\perp].
\end{aligned}$$

□

In the sequel we shall denote by $d(S;\,.)$ the distance function to a subset S of X. The notation $x \xrightarrow{S} \bar{x}$ will mean $x \longrightarrow \bar{x}$ and $x \in S$.

Proposition 1.6. [4] *Let $\mathcal{L}$ be an admissible collection of WT-subspaces of X. Then*

$$\partial_A f(\bar{x}) = \bigcap_{L\in\mathcal{L}} \limsup_{\substack{x \xrightarrow{f} \bar{x}\\ \varepsilon\downarrow 0}} \partial_\varepsilon^- f_{x+L}(x).$$

Moreover if S is a subset of X which is closed around $\bar{x}\in S$ (that is $S\cap B(\bar{x},r)$ is closed for some closed ball $B(\bar{x},r)$), then

$$\partial_A d(S;\bar{x}) = \bigcap_{L\in\mathcal{L}} \limsup_{\substack{x \xrightarrow{S} \bar{x}\\ \varepsilon\downarrow 0}} \partial_\varepsilon^- d_{x+L}(S;x).$$

Remark. Following Ioffe [2] we see that for any $\varepsilon>0$ and any $L\in\mathcal{L}$, each $x^*\in \partial_\varepsilon^- d_{x+L}(S;x)$ satisfies $\langle x^*,h\rangle \leqslant (1+\varepsilon)\|h\|$ for all $h\in L$. Therefore $x^*\in(1+\varepsilon)B_{x^*}+L^\perp$ (where B_{x^*} is the closed unit ball of X^*) and hence

$$\partial_\varepsilon^- d_{x+L}(S;x) \subset \partial_\varepsilon^- d_{x+L}(S;x)\cap(1+\varepsilon)B_X^* + L^\perp.$$

https://doi.org/10.1017/S0004972700015276 Published online by Cambridge University Press

As the reverse inclusion is obvious we obtain

$$\partial_\varepsilon^- d_{x+L}(S;x) = \partial_\varepsilon^- d_{x+L}(S;x) \cap (1+\varepsilon)B_X^* + L^\perp$$

which ensures that

$$\partial_A d(S;\overline{x}) = \bigcap_{L\in\mathcal{L}} \limsup_{\substack{x \xrightarrow{S} \overline{x} \\ \varepsilon\downarrow 0}} \partial_\varepsilon^- d_{x+L}(S;x) \cap (1+\varepsilon)B_X^*.$$

∎

Theorem 1.7. [4] *Let f be a lower semicontinuous function on X with $|f(\overline{x})| < +\infty$ and let g be a Lipschitz function at $\overline{x}$. Then*

$$\partial_A(f+g)(\overline{x}) \subset \partial_A f(\overline{x}) + \partial_A g(\overline{x}).$$

Proposition 1.8. [12] *Let F be a mapping from X into Y which is Lipschitz at $\overline{x}$. Then*

$$\|y - F(x)\| \leqslant (k+1)d(GrF; x, y)$$

for all x and y belonging to some neighbourhood of $\overline{x}$ and $F(\overline{x})$ respectively, where k is a Lipschitz constant of F at $\overline{x}$ and GrF denotes the graph of F, that is $GrF = \{(x, y) \in X \times Y : y = F(x)\}$.

2. The main result

Definition 2.1: [8] A mapping $F: X \to Y$ is said to be strongly compactly Lipschitzian at $\overline{x}$ if there exists a multifunction $R: X \rightrightarrows \mathrm{Comp}(Y)$, where $\mathrm{Comp}(Y)$ is the collection of all non void $\|\ \|$-compact subsets of Y, and a function $r: X \times X \to \mathbb{R}_+$ satisfying the following properties:

(1) $\lim\limits_{\substack{x\to\overline{x} \\ h\to 0}} r(x, h) = 0$,

(2) there is $\mu > 0$ such that for all $h \in \mu B_X$, $x \in B(\overline{x};\mu)$ and all $t \in]0, \mu[$

$$t^{-1}(F(x+th) - F(x)) \in R(h) + \|h\|\, r(x, th) B_X,$$

(3) $R(0) = \{0\}$ and R is upper semicontinuous.

Remarks.

(1) If F is strictly differentiable at $\overline{x}$, then F is strongly compactly Lipschitzian at $\overline{x}$.
(2) If Y is finite dimensional, then F is strongly compactly Lipschitzian at $\overline{x}$ if and only if it is Lipschitzian at $\overline{x}$.

Let us recall some results concerning these mappings. The proof of the following is similar to the one established by Thibault [**11**].

PROPOSITION 2.2.

(1) *Every strongly compactly Lipschitzian mapping at $\overline{x}$ is Lipschitzian at $\overline{x}$.*

(2) *The sum of two strongly compactly Lipschitzian mappings is strongly compactly Lipschitzian.*

The proof of the following result is inspired by Ioffe [**2**].

PROPOSITION 2.3. *Let $F\colon X \to Y$ be a strongly compactly Lipschitzian mapping at $\overline{x}$, $c > 0$ and $\varepsilon > 0$. Then there is $\gamma > 0$ such that for all $L \in \mathcal{F}$ there exists a ω^*-neighbourhood U of 0 in Y^* such that for all $x \in B(\overline{x};\gamma)$, $v^* \in U$ with $\|v^*\| \leqslant c$, $\sigma > 0$ and all $x^* \in \partial_\sigma^-(v^* \circ F)_{x+L}(x)$ one has*

$$x^* \in (2\sigma + \varepsilon)B_X^* + L^{\perp}.$$

PROOF: Since F is strongly compactly Lipschitzian at $\overline{x}$ there are $\mu > 0$, a multifunction $R\colon X \rightrightarrows \operatorname{Comp}(Y)$ and a function $r\colon X \times X \to \mathbb{R}_+$ satisfying the conditions (1), (2) and (3) of Definition 2.1. Let $\varepsilon > 0$ and $c > 0$ be given. The compactness of the closed unit ball B_L of L in $(X;\|\ \|)$ ensures the existence of elements $h_1, \ldots, h_p$ of B_L such that

$$B_L \subset \bigcup_{j=1}^{p} \left(h_j + \frac{\varepsilon}{2(ck+\varepsilon)}B_L\right) \tag{2.1}$$

where k is a Lipschitz constant of F at $\overline{x}$. The compactness of $R(\mu h_j)$, for $j = 1, \ldots, p$, in $(Y;\|\ \|)$ also ensures the existence of $v_1, \ldots, v_{q_j}$ in $R(\mu h_j)$ such that

$$R(\mu h_j) \subset \bigcup_{i=1}^{q_j} \left(v_i + \frac{\mu\varepsilon}{8c}B_Y\right).$$

For each $j = 1, \ldots, p$ put $U_j = H_j^{\perp} + (\mu\varepsilon)/(8b)B_Y^*$, where H_j is the subspace of Y generated by $\{v_1, \ldots, v_{q_j}\}$ and where $b = \max\limits_{j=1,\ldots,p} \sup\limits_{z \in R(\mu h_j)} \|z\|$. Then for each $j = 1, \ldots, p$ U_j is a w^*-neighbourhood of 0 in Y^* and for all $v^* \in U_j$, with $\|v^*\| \leqslant c$, and $z \in R(\mu h_j)$

$$\langle v^*; z\rangle \leqslant \frac{\mu\varepsilon}{4}. \tag{2.2}$$

If we take $U = \bigcap_{j=1}^{p} U_j$ then U is a w^*-neighbourhood of 0 and satisfies relation (2.2) for all $v^* \in U$ with $\|v^*\| \leqslant c$ and $z \in \bigcup_{j=1}^{p} R(\mu h_j)$. Because of (1) of Definition 2.1 if we put $r(x) = \limsup_{h \to 0} r(x, h)$ one can get $\gamma \in]0, \mu[$ such that for all $x \in B(\overline{x}; \gamma)$

$$cr(x) \leqslant \frac{\varepsilon}{4}. \tag{2.3}$$

Let $x \in B(\overline{x}; \gamma)$, $\sigma > 0$ and $v^* \in U$ with $\|v^*\| \leqslant c$ be fixed and let $x^* \in \partial_\sigma^-(v^* \circ F)_{x+L}(x)$. Then for all $j = 1, \ldots, p$

$$\begin{aligned}\langle x^*; \mu h_j\rangle &\leqslant d^-(v^* \circ F)_{x+L}(x; \mu h_j) + \mu\sigma \|h_j\| \\ &\leqslant \liminf_{t \downarrow 0} t^{-1}\langle v^*; F(x + t\mu h_j) - F(x)\rangle + \mu\sigma\end{aligned}$$

because $\|h_j\| \leqslant 1$. As $x \in B(\overline{x}; \gamma) \subset B(\overline{x}; \mu)$ it follows that for each $j = 1, \ldots, p$

$$\langle x^*; \mu h_j\rangle \leqslant \sup_{z \in R(\mu h_j)} \langle v^*; z\rangle + \mu r(x) \|v^*\| + \mu\sigma$$

and hence by (2.2) and (2.3)

$$\langle x^*; \mu h_j\rangle \leqslant \frac{\mu\varepsilon}{4} + \frac{\mu\varepsilon}{4} + \mu\sigma$$

which implies that

$$\langle x^*; h_j\rangle \leqslant \frac{\varepsilon}{2} + \sigma. \tag{2.4}$$

But for any $h \in B_L$ where exists, by (2.1), some $j \in \{1, \ldots, p\}$ such that

$$\|h - h_j\| \leqslant \frac{\varepsilon}{2(ck + \varepsilon)}$$

which together with relation (2.4) implies that

$$\begin{aligned}\langle x^*; h\rangle &= \langle x^*; h - h_j\rangle + \langle x^*; h_j\rangle \\ &\leqslant \underline{d}(v^* \circ F)_{x+L}(x; h - h_j) + \sigma \|h - h_j\| + \frac{\varepsilon}{2} + \sigma \\ &\leqslant ck \|h - h_j\| + \sigma \|h - h_j\| + \frac{\varepsilon}{2} + \sigma \\ &\leqslant \frac{\varepsilon}{2} + \sigma + \frac{\varepsilon}{2} + \sigma \\ &= \varepsilon + 2\sigma.\end{aligned}$$

By the homogeneity of this it follows that for all $h \in L$

$$\langle x^*; h\rangle \leqslant (\varepsilon + 2\sigma)\,\|h\|$$

and hence $x^* \in (\varepsilon + 2\sigma)B_X^* + L^\perp$. ∎

In the sequel $\mathcal{F}_X$ and $\mathcal{F}_Y$ denote the families of all finite dimensional subspaces of X and Y respectively.

Some techniques in the proof of the following proposition come from Ioffe [6].

Proposition 2.4. *Let $F\colon X \to Y$ be a strongly compactly Lipschitzian mapping at $\overline{x}$ and let k be a Lipschitz constant of F over $\overline{x} + \delta B_x$ and $\overline{y} = F(\overline{x})$. Then the following assertions are equivalent:*

(i) $(x^*, -y^*) \in (k+1)\,\|y^*\|\,\partial_A d(GrF; \overline{x}, \overline{y})$
(ii) $(x^*, -y^*) \in \mathbb{R}_+\partial_A d(GrF; \overline{x}, \overline{y})$
(iii) $x^* \in \partial_A(y^* \circ F)(\overline{x})$.

Proof: Since F is strongly compactly Lipschitzian at $\overline{x}$, there are a multifunction $R\colon X \rightrightarrows \mathrm{Comp}(Y)$, a function $x\colon X \times X \to \mathbb{R}_+$ with $\lim\limits_{\substack{x \to \overline{x}\\ h' \to 0}} r(x, h') = 0$ and $s > 0$ such that for all $x \in B(x, s)$, $t \in \,]0, s[$ and $h \in B_X$

$$t^{-1}(F(x + t(sh)) - F(x)) \in R(sh) + s\,\|h\|\,r(x, t(sh))B_Y. \tag{2.4.1}$$

Let $L \in \mathcal{F}_X$. Then the closed unit ball B_L of L is a compact subset of $(X, \|\ \|)$ and from the upper semicontinuity property of R the set $R(sB_L)$ is a compact subset of $(Y, \|\ \|)$. Put $V_L = cl_Y[\mathrm{vect}(R(sB_L))]$. Then V_L is a separable subspace of Y and for all $M \in \mathcal{F}_Y$ the subspace $\overline{M + V_L} = cl_Y[M + V_L]$ of Y is also separable and hence the family $\{L \times \overline{M + V}_L\}_{L \times M \in \mathcal{F}_X \times \mathcal{F}_Y}$ is an admissible family of WT-subspaces of $X \times Y$ (see [5] and the example following Definition 1.3).

(1) (i) $\to$ (ii): this implication is obvious.

(2) (ii) $\to$ (iii): let $(x^*, -y^*) \in \mathbb{R}_+\partial_A d(GrF; \overline{x}, \overline{y})$. Then there exists $(u^*, -v^*) \in \partial_A d(GrF; \overline{x}, \overline{y})$ such that $(x^*, -y^*) = \lambda(u^*, -v^*)$, with $\lambda \geqslant 0$. Then by the remark following Proposition 1.6 for each $L \in \mathcal{F}_X$ and each $M \in \mathcal{F}_Y$ there are nets $\varepsilon_i \downarrow 0$ with $\varepsilon_i < 1$, $x_i \to \overline{x}$ and $(u_i^*, v_i^*) \xrightarrow{w^*} (u^*, v^*)$ such that

$$(u_i^*, -v_i^*) \in \left(\partial^-_{\varepsilon_i} d_{(x_i, F(x_i)) + Lx\overline{M+V_L}}(GrF; x_i, F(x_i))\right) \cap (1 + \varepsilon_i)B_X^*.$$

For any fixed $\varepsilon > 0$ there exists i_0 and $\alpha > 0$ such that $x_i \in \overline{x} + (\delta/2)B_X$ and $r(x_i, h') < 1/2\varepsilon$ for all $i > i_0$ and $\|h'\| \leqslant \alpha$, since $\lim\limits_{\substack{x \to \overline{x}\\ h' \to 0}} r(x, h') = 0$. Let $i > i_0$ be $\|h\| = 1$ and let $t_n \downarrow 0$ be such that

$$t_n^{-1}\langle v_i^*; F(x_i + t_n(sh)) - F(x_i)\rangle \to d^-(v_i^* \circ F)_{x_i + L}(x_i; sh).$$

From (2.4.1) there are $a_n \in R(sh) \subset V_L$ and $b_n \in sr(x_i, t_n(sh))B_Y$ such that

$$t_n^{-1}(F(x_i + t_n(sh)) - F(x_i)) = a_n + b_n. \tag{2.4.2}$$

Note that $\|b_n\| < s\varepsilon$ for n large enough. As $R(sh)$ is a compact set we may assume that $a_n \to a \in V_L$. Thus

$$\begin{aligned}\langle u_i^*; sh\rangle &\leqslant \langle v_i^*; a\rangle + \varepsilon_i(\|sh\| + \|a\|) + \liminf_{t\downarrow 0} t^{-1} d(GrF; x_i + tsh,\, F(x_i) + ta)\\ &\leqslant \langle v_i^*; a\rangle + \varepsilon_i(\|sh\| + \|a\|) + \liminf_{t\to+\infty} f_n^{-1} d(GrF; x_i + t_nsh,\, F(x_i) + t_na)\\ &\leqslant \langle v_i^*; a\rangle + \varepsilon_i(\|sh\| + \|a\|) + s\varepsilon\end{aligned}$$

because by (2.4.2) $(x_i + t_nsh,\, F(x_i) + t_n(a_n + b_n)) \in GrF$ and $\|b_n\| < \varepsilon s$. As $a_n \to a$, $\|b_n\| < s\varepsilon$, $\|v_i^*\| \leqslant 2$, $\|a\| \leqslant (k+\varepsilon)s$ and $\lim_{n\to\infty} \langle v_i^*; a_n + b_n\rangle = d^-(v_i^* \circ F)_{x_i+L}(x_i, sh)$ one has

$$\langle u_i^*; h\rangle \leqslant d^-(v_i^* \circ F)_{x_i+L}(x_i; h) + \varepsilon_i(k + 1 + \varepsilon) + 3\varepsilon.$$

Thus for $E(\varepsilon, i) = \varepsilon_i(k + 1 + \varepsilon) + 3\varepsilon$ one has

$$u_i^* \in \partial^-_{E(\varepsilon,i)}(v_i^* \circ F)_{x_i+L}(x_i).$$

If we write $v_i^* \circ F = (v_i^* - v^*)\circ F + v^* \circ F$ we can get, by Lemma 1.5, some nets $u_i \to \overline{x}$, $v_i \to \overline{x}$, $z_i^* \in \partial^-_{E(\varepsilon,i)}(v^* \circ F)_{u_i+L}(u_i)$ and $q_i^* \in [\partial^-_{E(\varepsilon,i)}((v_i^* - v^*)\circ F)_{v_i+L}(v_i) + L^\perp]$ such that $z_i^* + q_i^* \xrightarrow{w^*} u^*$. But $v_i^* - v^* \xrightarrow{w^*} O$ and $(v_i^* - v^*)_i$ is bounded and hence by Proposition 2.3 one has the existence of $i(\varepsilon) > i_0$ such that $q_i^* \in (\varepsilon + 2E(\varepsilon, i))B_Y^* + L^\perp$ for all $i > i(\varepsilon)$. Thus

$$\begin{aligned}u^* &\in \limsup_{\substack{i\\ \varepsilon\downarrow 0}} \left(\partial^-_{E(\varepsilon,i)}(v^* \circ F)_{u_i+L}(u_i) + (\varepsilon + 2E(\varepsilon, i))B_X^* + L^\perp\right)\\ &\subset \limsup_{\substack{x\to\overline{x}\\ \varepsilon\downarrow 0}} \left(\partial^-_\varepsilon(v^* \circ F)_{x+L}(x) + L^\perp\right)\end{aligned}$$

and hence

$$\begin{aligned}u^* &\in \bigcap_{L\in\mathcal{F}_X} \limsup_{\substack{x\to\overline{x}\\ \varepsilon\downarrow 0}} \left(\partial^-_\varepsilon(v^* \circ F)_{x+L}(x) + L^\perp\right)\\ &= \limsup_{\substack{x\to\overline{x}\\ \varepsilon\downarrow 0\\ L\in\mathcal{F}_X}} \left(\partial^-_\varepsilon(v^* \circ F)_{x+L}(x) + L^\perp\right)\\ &= \limsup_{\substack{x\to\overline{x}\\ \varepsilon\downarrow 0\\ L\in\mathcal{F}_X}} \partial^-_\varepsilon(v^* \circ F)_{x+L}(x)\\ &= \partial_A(v^* \circ F)(\overline{x}).\end{aligned}$$

https://doi.org/10.1017/S0004972700015276 Published online by Cambridge University Press

It follows that $x^* \in \lambda\partial_A(v^* \circ F)(\overline{x}) = \partial_A(\lambda v^* \circ F)(\overline{x}) = \partial_A(y^* \circ F)(\overline{x})$.

(3) (iii) $\to$ (i): Let $x^* \in \partial_A(y^* \circ F)(\overline{x})$. Then for each $L \in \mathcal{F}_X$ there are nets $x_i \to \overline{x}$, $\varepsilon_i \downarrow 0$ and $x_i^* \xrightarrow{w^*} x^*$ such that $x_i^* \in \partial^-_{\varepsilon_i}(y^* \circ F)_{x_i+L}(x_i)$ which means that for all $h \in L$

$$\begin{aligned}\langle x_i^*; h\rangle &\leqslant \liminf_{\substack{t\downarrow 0\\ h'\to h}} t^{-1}\Big((y^* \circ F)_{x_i+L}(x_i + th') - (y^* \circ F)_{x_i+L}(x_i)\Big) + \varepsilon_i \|h\| \\ &= \liminf_{t\downarrow 0} t^{-1}\langle y^*; F(x_i + th) - F(x_i)\rangle + \varepsilon_i \|h\|\end{aligned}$$

because $y^* \circ F$ is Lipschitz at $\overline{x}$. This and Proposition 1.8 imply that for all $M \in \mathcal{F}_Y$, $h \in L$ and $v \in Y$

$$\langle x_i^*; h\rangle - \langle y^*; v\rangle \leqslant (k+1)\|y^*\| \, d^- d_{(x_i, F(x_i))+L\times M}(GrF; x_i, F(x_i); h, v) + \varepsilon_i(\|h\| + \|v\|)$$

which gives by Proposition 1.6 that $(x^*, -y^*) \in (k+1)\|y^*\| \, \partial_A d(GrF; \overline{x}, F(\overline{x}))$. ∎

THEOREM 2.5. *Let $F\colon X \to Y$ be a strongly compactly Lipschitzian mapping at $\overline{x}$ and let $f\colon Y \to \overline{\mathbb{R}}$ be a Lipschitz function at $\overline{y} = F(\overline{x})$. Then*

$$\partial_A(f \circ F)(\overline{x}) \subset \bigcup_{y^* \partial_A f(\overline{y})} \partial_A(y^* \circ F)(\overline{x}).$$

PROOF: Since f and F are Lipschitz at $\overline{y}$ and $\overline{x}$ respectively we have (see Propositions 1.8 and 2.2) the existence of $\alpha > 0$ such that for all $x \in B(\overline{x}, \alpha)$ and $y \in B(\overline{y}, \alpha)$

$$f \circ F(x) \leqslant f(y) + k\|y - F(x)\| \quad \text{and} \quad \|y - F(x)\| \leqslant (k'+1)d(GrF; x, y)$$

where k and k' are the Lipschitz constants of f and F respectively. If we put $s(x, y) = f(y) + k(k'+1)d(GrF; x, y)$, we note that for all $x \in B(\overline{x}, \alpha)$ and $y \in B(\overline{y}, \alpha)$

$$f \circ F(x) \leqslant s(x, y) \quad \text{and} \quad f \circ F(x) = s(x, F(x)).$$

For each $(h, v) \in X \times Y$ and for all finite dimensional spaces L and M of X and Y respectively we have

$$\begin{aligned}&t^{-1}[(f \circ F)_{x+L}(x + th) - (f \circ F)_{x+L}(x)] \\ &\qquad \leqslant t^{-1}[s_{(x,F(x))+L\times M}(x + th, F(x) + tv) - s_{(x,F(x))+L\times M}(x, F(x))]\end{aligned}$$

for all t small enough and x sufficiently close to $\overline{x}$ and hence

$$d^-(f \circ F)_{x+L}(x, h) \leqslant d^- s_{(x,F(x))+L\times M}(x, F(x); h, v),$$

https://doi.org/10.1017/S0004972700015276 Published online by Cambridge University Press

which implies that

$$\partial^-(f \circ F)_{x+L}(x)X\{0\} \subset \partial^- s_{(x,\,F(x))+L\times M}(x,\, F(x)).$$

Therefore we obtain that

$$\begin{aligned}\partial_A(f \circ F)(\overline{x})X\{0\} &\subset \bigcap_{L,\,M} \limsup_{x\to\overline{x}} \partial^- s_{(x,\,F(x))+L\times M}(x,\, F(x)) \\ &\subset \bigcap_{L,\,M} \limsup_{(x,\,y)\to(\overline{x},\,\overline{y})} \partial^- s_{(x,\,y)+L\times M}(x,\, y) = \partial_A s(\overline{x},\, \overline{y}).\end{aligned}$$

So Theorem 1.7 ensures that

$$\partial_A f \circ F(\overline{x}) \times \{0\} \subset \{0\} \times \partial_A f(\overline{y}) + k(k'+1)\partial_A d(GrF; \overline{x}, \overline{y})$$

and hence it suffices to use Proposition 2.4 to complete the proof. ∎

References

[1] J.M. Borwein, 'Stability and regular pooint of inequality systems', *J. Optim. Theory Appl.* **48** (1986), 9–52.

[2] A.D. Ioffe, 'Approximate subdifferentials of non-convex functions', *Cahier # 8120.*

[3] A.D. Ioffe, 'Approximate subdifferentials and applications I: The finite dimensional theory', *Trans. Amer. Math. Soc.* **281** (1984), 389–416.

[4] A.D. Ioffe, 'Approximate subdifferentials and applications II: Functions on locally convex spaces', *Mathematika* **33** (1986), 111–128.

[5] A.D. Ioffe, 'On subdifferentiability spaces', *Ann. New York Acad. Sci.* **410** (1983), 107–120.

[6] A.D. Ioffe, 'Approximate subdifferentials and applications III: The metric theory', *Mathematika* **36** (1989), 1–38.

[7] A. Jourani and L. Thibault, 'The use of metric graphical regularity in approximate subdifferential calculus rules in finite dimensions', *Optimization* **21** (1990), 509–519.

[8] A. Jourani and L. Thibault, 'Approximations and metric regularity in mathematical programming in Banach space', *Math. Oper. Res.* (to appear).

[9] B.S. Mordukhovich, 'Nonsmooth analysis with nonconvex generalized differentials and dual maps', *Dokl. Akad. Nauk. USSR* **28** (1984), 976–979.

[10] R.T. Rockafellar, 'Extensions of subgradient calculus with applications to optimization', *Nonlinear Analysis Th. Meth. Appl.* **9** (1985), 665–698.

[11] L. Thibault, 'Subdifferentials of compactly Lipschitzian vector-valued functions', *Travaux du séminaire d'analyse convese* **8** (1978).

[12] L. Thibault, 'On subdifferentials of optimal value functions', *SIAM J. Control Optim.* **29** (1991), 1019–1036.

Université de Bourgogne
Laboratoire d'analyse numérique
B.P. 138
21004 Dijon, Cedex
France

Université de Pau
Laboratoire de Mathématiques appliquées
Avenue de l'université
64000 Pau
France