

ON PSEUDO-DIFFERENTIABILITY

ROBERTO COMINETTI

ABSTRACT. We present some new relations between the pseudo-derivatives and parabolic epiderivatives recently introduced by Rockafellar, and also its infinite dimensional counterparts. Significant extensions of the most important known results are proven, which further clarify the range of applicability of this new theory.

1. INTRODUCTION

There are nowadays several theories of generalized differentiability which have considerably enlarged the spectrum of optimization problems which can be treated by variational methods.

This renewed interest in differentiating nondifferentiable functions began in the sixties with the development of the theory of convex analysis, which is mainly associated with the names of J. J. Moreau and R. T. Rockafellar, but which is grounded, as most mathematical theories do, on many previous works, and specially in this case those of Fenchel.

In the nonconvex setting there have also been significant contributions, of which we can mention those of Pshenichnyi (1971), Clarke (1973), Warga (1976), Ioffe (1979), and Penot (1985); and which cover a large variety of functions appearing in optimization.

If the situation is quite settled in what we may call the *first order generalized analysis*, the higher order case, and particularly the second order one, is still in the making. Nevertheless, we already dispose of a variety of proposals—in addition to the already cited works of Ioffe [15] and Penot [20] which cover first order as well as higher order derivatives—specifically concerning second order derivatives, as those of Aubin [3], Auslender [4], Ben-Tal and Zowe [5, 6], Chaney [8, 9], Cominetti and Correa [10, 11], Hiriart-Urruty *et al.* [14], Hiriart-Urruty [13], Rockafellar [23], and Seeger [26] just to cite some of them.

Another approach was recently presented by Rockafellar [24] under the designation of *Pseudo-Differentiability* (see also [23 and 18]). The distinguishing

Received by the editors September 25, 1988 and, in revised form, April 4, 1989.

1980 *Mathematics Subject Classification* (1985 Revision). Primary 26B05, 49A52, 58C05, 58C20, 90C30.

Key words and phrases. Nonsmooth analysis, pseudo-differentiability, convex analysis, regularity conditions, epiconvergence, Mosco-convergence.

The author is on leave from Universidad de Chile, Casilla 170/3 Correo 3, Santiago, Chile.

feature of this approach is the way in which limits are taken for defining the “derivatives”, namely, the usual pointwise convergence is replaced by the relatively novel notion of *epiconvergence*.¹ The so-defined differentiability notions have revealed interesting properties of locally uniform approximation of the functions, which have a strong geometrical basis; and at the same time, they are still weak enough to be applicable to a large and interesting class of functions as shown in [24].

Our goal in this paper is to further explore the relations between the derivatives obtained by proceeding in this way and to present significant extensions of the most important known results in “epigraphical” differentiability [24], which further clarify the range of applicability of this theory, as they get closer to the minimal assumptions for having the results in question.

Moreover, we present our results not just in the finite dimensional case but in the more general setting of a reflexive Banach space. The only ‘difficulty’ introduced by this is at a technical level, since we must replace the epigraphical convergence by the slightly more complicated notion of Mosco-convergence, which coincides with the former in the finite dimensional case.

2. RECALLING SOME DEFINITIONS AND RESULTS

As pointed out in the introduction, pseudo-derivatives are defined in terms of Mosco-convergence (*M*-convergence hereafter) of different kinds of differential quotients. Accordingly, we shall recall in this section some basic facts on *M*-convergence (and epiconvergence), just enough so that the reader will not need to look elsewhere for definitions and results. Nevertheless, the interested reader may find a fairly complete account on the subject in [1, Attouch] and references therein.

In order to settle the terminology and notation we shall employ, let us say that for the rest of this paper $(X, \|\cdot\|)$ will be an abstract reflexive Banach space with topological dual X^* . The strong (norm) and weak topologies on X will be denoted respectively by S and W . Similarly, S^* and W^* will represent the strong and weak (or weak- $*$ since X is supposed reflexive) topologies on X^* . We denote by $\langle \cdot, \cdot \rangle$ the canonical pairing between X and X^* and by $\|\cdot\|$ the dual norm in X^* (hopefully no confusion should arise from using the same symbol to designate the norms in X and X^*). As usual we shall write ‘ \rightharpoonup ’ to indicate weak convergence and ‘ \rightarrow ’ for the strong one, both in X and X^* . Finally, we shall write ‘int’ for interior and a bar to denote closure. The closed ball of center y and radius r is denoted $B(y, r)$.

Some acquaintance with basic facts from convex analysis, particularly those concerning Fenchel-Moreau conjugacy, will be assumed. However, we shall occasionally recall some definitions in order to fix the notation. The extended real field $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ will be endowed with the customary extended

¹In this sense, Rockafellar’s approach may be compared to the work of Penot [20] who systematically considers the lower epilimit and also to that of Dolecki [12] who has used various kinds of convergences to define first order derivatives.

operations, order, and topology familiar from convex analysis.

Let us proceed by recalling what Mosco-convergence is. A sequence of functions $\varphi_n: X \rightarrow \overline{\mathbb{R}}$ is M -convergent to $\varphi: X \rightarrow \overline{\mathbb{R}}$ when for all $x \in X$ the following hold,

$$(1) \quad \begin{aligned} (a) \quad & (\forall x_n \rightarrow x) \quad \liminf_{n \rightarrow \infty} \varphi_n(x_n) \geq \varphi(x), \text{ and} \\ (b) \quad & (\exists x_n \rightarrow x) \quad \lim_{n \rightarrow \infty} \varphi_n(x_n) = \varphi(x). \end{aligned}$$

In such a case we shall write $\varphi_n \xrightarrow{M} \varphi$ or $\varphi = M\text{-}\lim \varphi_n$.

This notion of convergence enjoys many useful properties in connection with variational problems, both for applications and for theoretical questions (see [1] and the long list of references therein).

It is the infinite dimensional counterpart of yet another type of convergence, called *epiconvergence* to which is reduced when X is finite dimensional. We find it useful to recall this notion too, which can be presented in more geometric terms and may help the nonacquainted reader in understanding what these convergences are.

As the terminology suggests, epiconvergence of a family of functions from \mathbb{R}^m to $\overline{\mathbb{R}}$, amounts to the (set) convergence of the corresponding epigraphs.² Therefore we shall begin by recalling the notion of convergence of sets in the sense of Kuratowski.

The *upper limit* and *lower limit* of a family $(C_n)_{n \in \mathbb{N}}$ of subsets of \mathbb{R}^m (or more generally any topological space) are defined respectively by

$$\begin{aligned} \limsup C_n &:= \{x \in \mathbb{R}^m : (\exists J \subset \mathbb{N})(\exists y_n \in C_n), y_n \xrightarrow{n \in J} x\}, \\ \liminf C_n &:= \{x \in \mathbb{R}^m : (\exists y_n \in C_n), y_n \rightarrow x\}. \end{aligned}$$

With this notation one can say that a family of sets C_n converges to C when $C = \limsup C_n = \liminf C_n$. We shall denote this fact by $C_n \xrightarrow{K} C$.

It is easy to show that convergence of sets in the above sense is equivalent (in the finite dimensional case!) to the pointwise convergence of the associated distance functions, that is to say

$$(2) \quad C_n \xrightarrow{K} C \Leftrightarrow C \text{ is closed and } d(\cdot, C_n) \xrightarrow{\text{pw}} d(\cdot, C)$$

where " $\xrightarrow{\text{pw}}$ " denotes pointwise convergence, and $d(\cdot, S)$ the distance function

$$d(x, S) = \inf\{\|x - y\| : y \in S\}.$$

We may now define epiconvergence as follows: a sequence $(\varphi_n)_{n \in \mathbb{N}}$ of functions from \mathbb{R}^m to $\overline{\mathbb{R}}$ is said to epiconverge to a function φ if and only if $\text{epi}(\varphi_n) \xrightarrow{K} \text{epi}(\varphi)$. In such a case we shall write $\varphi_n \xrightarrow{\text{epi}} \varphi$ or $\varphi = \text{elim } \varphi_n$.

One of the most useful characterizations of epiconvergence is given in terms

²Recall that the epigraph of $f: X \rightarrow \overline{\mathbb{R}}$ is the set $\text{epi}(f) = \{(x, r) \in X \times \mathbb{R} : f(x) \leq r\}$.

of the upper and lower epilimits

$$\begin{aligned}\overline{\text{elim}}\varphi_n(x) &:= \lim_{r \downarrow 0} \limsup_{n \rightarrow \infty} \inf_{y \in B(x, r)} \varphi_n(y), \\ \underline{\text{elim}}\varphi_n(x) &:= \lim_{r \downarrow 0} \liminf_{n \rightarrow \infty} \inf_{y \in B(x, r)} \varphi_n(y).\end{aligned}$$

Namely, we have the following result.

Theorem 2.1 (Mosco). *The following are equivalent:*

- (i) $\varphi_n \xrightarrow{\text{epi}} \varphi$,
- (ii) $\varphi(x) = \underline{\text{elim}}\varphi_n(x) = \overline{\text{elim}}\varphi_n(x)$, $(\forall x \in \mathbb{R}^m)$. \square

Having made this digression about epiconvergence, let us come back to the more general setting. As a matter of fact, in this paper we shall be concerned with “continuous” sequences of functions $(\varphi_t)_{t>0}$. In this case we shall simply say that this family M -converges to φ as t tends to 0, which we still denote by $\varphi_t \xrightarrow{M} \varphi$ or $\varphi = M\text{-}\lim \varphi_t$, when for every sequence $t_n \downarrow 0$ we have $\varphi_{t_n} \xrightarrow{M} \varphi$ in the previously defined sense. This amounts to say that for every $x \in X$ one has,

$$(3) \quad \begin{aligned} (a) \quad & (\forall x_t \rightarrow x) \quad \liminf_{t \downarrow 0} \varphi_t(x_t) \geq \varphi(x), \\ (b) \quad & (\forall t_n \downarrow 0)(\exists x_n \rightarrow x) \quad \lim_{n \rightarrow \infty} \varphi_{t_n}(x_n) = \varphi(x). \end{aligned}$$

We may now proceed to recall some results on M -convergence, of which the following one asserting the “bicontinuity” of the Fenchel-Moreau conjugacy, is one of the cornerstones.

Theorem 2.2 (Mosco). *Suppose that $(\varphi_t)_{t>0}$ and φ are closed proper convex functions. Then $\varphi_t \xrightarrow{M} \varphi$ if and only if $\varphi_t^* \xrightarrow{M} \varphi^*$. \square*

Simple examples may be constructed, even in the finite dimensional setting, to show that M -convergence and pointwise convergence do not imply each other. However some relations between both kinds of convergence do exist. Let us mention the following trivial fact.

Proposition 2.3. *The following are equivalent:*

- (a) *the M -limit and the pointwise limit of φ_t exist and coincide.*
- (b) $(\forall x \in X)(\forall x_t \rightarrow x), \liminf_{t \downarrow 0} \varphi_t(x_t) \geq \limsup_{t \downarrow 0} \varphi_t(x)$. \square

Much more subtle relations are known. We shall only retain the following finite dimensional results [25, Corollaries 2C, 2D, 3B].

Theorem 2.4 (Salinetti-Wets). *Let X be finite dimensional and suppose φ_t, φ are closed proper convex functions. Then,*

- (a) *if $\varphi_t \xrightarrow{\text{pw}} \varphi$ and $\text{aff}(\text{dom}(\varphi_t)) \subset \text{aff}(\text{dom}(\varphi))$, Then $\varphi_t \xrightarrow{\text{epi}} \varphi$.*
- (b) *if $\varphi_t \xrightarrow{\text{epi}} \varphi$ then $\varphi_t \xrightarrow{\text{pw}} \varphi$ on $\text{int}(\text{dom}(\varphi))$. \square*

Here, $\text{aff}(\cdot)$ denotes the affine hull and $\text{dom}(f) = \{x : f(x) < +\infty\}$ is the domain of f . Notice that the condition $\text{aff}(\text{dom}(\varphi_t)) \subset \text{aff}(\text{dom}(\varphi))$ in (a) of the above result holds trivially when $\text{int}(\text{dom}(\varphi)) \neq \emptyset$.

Let us also mention the following result about Mosco convergence of monotone sequences of convex functions.

Proposition 2.5. *Let $(\varphi_t)_{t>0}$ be a family of closed proper convex functions.*

- (a) *if φ_t increases as $t \downarrow 0$, then M -converges to $\sup_{t>0} \varphi_t$.*
- (b) *if φ_t decreases as $t \downarrow 0$, M -converges to $\text{cl}[\inf_{t>0} \varphi_t]$, where ‘cl’ denotes the l.s.c. closure (weak or strong since $\inf_{t>0} \varphi_t$ is convex). \square*

To end this short review, let us prove the following simple result about lower semicontinuity of M -limits which we shall need later.

Proposition 2.6. *If φ is the M -limit of a family of functions $(\varphi_t)_{t>0}$, then it is sequentially weakly lower semicontinuous (l.s.c.), that is to say,*

$$(4) \quad (\forall x \in X)(\forall x_n \rightharpoonup x) \quad \liminf_{n \rightarrow \infty} \varphi(x_n) \geq \varphi(x).$$

Proof. Let us take $x_n \rightharpoonup x$ and extract a subsequence $(x_k)_{k \in J}$ such that $\varphi(x_k) \rightarrow r := \liminf_{n \rightarrow \infty} \varphi(x_n)$. We may suppose $r < \infty$ since otherwise the result is obvious.

Let us fix a sequence $r_k \rightarrow r$ with $r_k > \varphi(x_k)$ and choose an arbitrary sequence $t_i \downarrow 0$. According to (3)(b) we may select, for each fixed $k \in J$, a sequence $x_i^k \xrightarrow{i} x_k$ with $\lim_{i \rightarrow \infty} \varphi_{t_i}(x_i^k) = \varphi(x_k)$. Then we may find an integer $i(k)$ such that for $x'_k = x_{i(k)}^k$ one has $\|x'_k - x_k\| < 1/k$ and $\varphi_{t_{i(k)}}(x'_k) < r_k$.

It follows easily that $x'_k \rightharpoonup x$, and since we may always choose $i(k)$ to be strictly increasing we also have that $\lim_{k \in J \rightarrow \infty} t_{i(k)} = 0$. We may then use (3)(a) to conclude that

$$\varphi(x) \leq \liminf_{k \in J \rightarrow \infty} \varphi_{t_{i(k)}}(x'_k) \leq \liminf_{k \in J \rightarrow \infty} r_k = r$$

which gives us the desired conclusion. \square

3. PSEUDO-DERIVATIVES

With the notion of M -convergence at hand we may proceed to introduce the first and second order pseudo-derivatives, along the same lines as in [24, Rockafellar]. As pointed out before, the only difference with respect to Rockafellar's definitions is at a technical level since we consider the infinite dimensional case by replacing the epiconvergence with M -convergence.

From now on, $f: X \rightarrow \overline{\mathbb{R}}$ will be a given fixed function, and $x \in X$ a point at which f is finite.

Definition 3.1. We say that f is *pseudo-differentiable* at x if the quotients

$$\Delta_t(v) := \frac{f(x + tv) - f(x)}{t}$$

M -converge, with $M\text{-}\lim \Delta_t(0) = 0$. The limit function is then denoted f'_x and it is called the (first order) *pseudo-derivative* of f at x .

A vector $v^* \in X^*$ is called a *pseudo-gradient* for f at x if for every $v \in X$ one has $\langle v^*, v \rangle \leq f'_x(v)$. The set of pseudo-gradients for f at x is denoted $\partial f(x)$.

The next proposition contains the basic features of (first order) pseudo-derivatives.

Proposition 3.2. *If f is pseudo-differentiable at x then,*

- (a) f'_x is sequentially weakly l.s.c., positively homogeneous and proper, and
 - (b) the pseudo-gradients of f at x are the vectors $v^* \in X^*$ such that³
- (5)
$$f(y) \geq f(x) + \langle v^*, y - x \rangle + o(\|y - x\|).$$

Proof. (a) The lower semicontinuity of f'_x is immediate from Proposition 2.6. To show the positive homogeneity let us take $v \in X$ and $\lambda > 0$, and choose $t_n \downarrow 0$ and $v_n \rightarrow v$ with $[f(x + t_n v_n) - f(x)]/t_n \rightarrow f'_x(v)$. It follows that

$$\frac{f(x + (t_n/\lambda)\lambda v_n) - f(x)}{(t_n/\lambda)} \rightarrow \lambda f'_x(v)$$

and using (3)(a) we deduce that $f'_x(\lambda v) \leq \lambda f'_x(v)$. This inequality applied with λv and $1/\lambda$ instead of v and λ gives us the converse inequality, $\lambda f'_x(v) \leq f'_x(\lambda v)$ and the conclusion follows.

To prove that f'_x is proper, we proceed by contradiction supposing the existence of $v \in X$ such that $f'_x(v) = -\infty$. Then for all $n \in \mathbb{N}$ we have $f'_x(v/n) = -\infty$ and by the previously established semicontinuity we deduce $f'_x(0) = -\infty$ contradicting the pseudo-differentiability of f at x .

(b) Suppose v^* satisfies (5) and choose $v \in X$, $t_n \downarrow 0$, and $v_n \rightarrow v$ such that $[f(x + t_n v_n) - f(x)]/t_n \rightarrow f'_x(v)$. Then from (5) we have

$$\langle v^*, v_n \rangle + \frac{o(\|t_n v_n\|)}{t_n} \leq \frac{f(x + t_n v_n) - f(x)}{t_n}$$

and going to the limit we deduce $\langle v^*, v \rangle \leq f'_x(v)$. Since v was arbitrary we conclude that v^* is a pseudo-gradient for f at x .

Conversely, let us suppose that $v^* \in \partial f(x)$. If (5) was not verified we could find $\alpha < 0$ and $y_n \rightarrow x$ such that

$$\frac{f(y_n) - f(x) - \langle v^*, y_n - x \rangle}{\|y_n - x\|} \leq \alpha.$$

³The notation $o(\|h\|^k)$ will be used in the sequel to denote a function verifying

$$\lim_{h \rightarrow 0} o(h)/\|h\|^k = 0.$$

Since X is reflexive we may extract a subsequence, still denoted y_n , such that $v_n := (y_n - x)/\|y_n - x\| \rightarrow v$ for some $v \in X$, and setting $t_n = \|y_n - x\|$ we get

$$\frac{f(x + t_n v_n) - f(x)}{t_n} \leq \langle v^*, v_n \rangle + \alpha.$$

Letting $n \rightarrow \infty$ and using (3) we finally obtain

$$f'_x(v) \leq \langle v^*, v \rangle + \alpha < \langle v^*, v \rangle$$

contradicting the fact that v^* is a pseudo-gradient. \square

It can also be seen that in the finite dimensional case one may characterize the epigraph of f'_x as the 'tangent' cone to the epigraph of f , that is to say,

$$\text{epi}(f'_x) = \lim_{t \downarrow 0} \frac{\text{epi}(f) - (x, f(x))}{t}.$$

On the other hand, in the convex case it turns out that pseudo-differentiability is equivalent to subdifferentiability in the usual sense, and that in such a case the pseudo-gradients and the subgradients coincide. Here is a more detailed description.

Proposition 3.3. *If f is convex then the following are equivalent:*

- (a) f is pseudo-differentiable at x .
- (b) the directional derivative $f'(x; \cdot)$ is lower semicontinuous at 0.
- (c) f has at least one subgradient at x .

In this case f'_x equals the l.s.c. closure of $f'(x; \cdot)$ (weak or strong since it is convex) and moreover the pseudo-gradients and subgradients of f at x coincide and verify

$$f'_x(v) = \sup \langle \partial f(x), v \rangle.$$

Proof. It is well known that in the convex case the quotients Δ_t decrease to $f'(x; \cdot)$ as $t \downarrow 0$, so that using Proposition 2.5 we see that Δ_t M -converges to $\text{cl}[f'(x; \cdot)]$, and the equivalence between (a) and (b) becomes evident.

Also, since $\text{cl}[f'(x; \cdot)]$ is a closed sublinear function, it follows from well known results of convex analysis that (b) is equivalent to the existence of a subgradient, that is, to (c).

The remaining assertions are immediate consequences of the previous facts and known results in convex analysis. \square

Let us turn now to the second order pseudo-derivatives. We shall consider here two related notions which correspond to two particular choices of second order differential quotients, and are the M -convergence equivalents of those introduced in [24].

Definition 3.4. Let f be pseudo-differentiable at x and $v^* \in \partial f(x)$. We say that f is *twice pseudo-differentiable* at x relative to v^* if the second order

differential quotients

$$\Delta_t^{v^*}(v) := \frac{f(x + tv) - f(x) - t\langle v^*, v \rangle}{t^2/2}$$

M -converge, with $M\text{-}\lim \Delta_t^{v^*}(0) = 0$. The limit function is then denoted f''_{x,v^*} and called the *second order pseudo-derivative* of f at x relative to v^* .

The second order pseudo-derivatives enjoy quite similar properties to those proved for the first order ones. The following proposition gives the basic facts.

Proposition 3.5. *Let f be twice pseudo-differentiable at x relative to the pseudo-gradient $v^* \in \partial f(x)$. Then*

- (a) f''_{x,v^*} is sequentially weakly l.s.c., proper and positively homogeneous of order two ($f''_{x,v^*}(\lambda v) = \lambda^2 f''_{x,v^*}(v)$);
- (b) f''_{x,v^*} is positive when f is convex;
- (c) $f''_{x,v^*}(v) < +\infty \Rightarrow f'_x(v) = \langle v^*, v \rangle$;
- (d) $k = \min\{f''_{x,v^*}(v) : \|v\| \leq 1\}$ is finite and attained, and for all $v \in X$ one has

$$k\|v\|^2 \leq f''_{x,v^*}(v).$$

Proof. The proof of (a) is completely analogous to the corresponding one in Proposition 3.2 and is left to the reader. Part (b) follows from the positivity of $\Delta_t^{v^*}$, and part (d) is a consequence of (a) and the weak compactness of the unit ball in X .

In order to show (c) let us take $t_n \downarrow 0$ and $v_n \rightarrow v$ such that

$$\lim_{n \rightarrow \infty} \frac{f(x + t_n v_n) - f(x) - t_n \langle v^*, v_n \rangle}{t_n^2/2} = f''_{x,v^*}(v) \in \mathbb{R}.$$

It follows that

$$\lim_{n \rightarrow \infty} \left[\frac{f(x + t_n v_n) - f(x)}{t_n} - \langle v^*, v_n \rangle \right] = 0$$

and since v_n converges to v , we deduce using (3) that $\langle v^*, v \rangle \geq f'_x(v)$. We conclude the proof by noting that the converse inequality is also satisfied since $v^* \in \partial f(x)$. \square

The second notion of twice differentiability we shall consider corresponds to the M -convergence version of the one introduced by Ben-Tal and Zowe [5, 6], and in its “epigraphical” version by Rockafellar [24].

Definition 3.6. Let f be pseudo-differentiable at x , and $v \in \text{dom}(f'_x)$. The *parabolic pseudo-derivative* of f at x in the direction v is the M -limit—whenever it exists—of the quotients

$$D_t^v(w) := \frac{f(x + tv + \frac{1}{2}t^2w) - f(x) - tf'_x(v)}{t^2/2}.$$

We shall denote it by $d^2f(x; v, \cdot)$.

In [24], a general class of functions (defined on finite dimensional spaces) is exhibited for which both f''_{x,v^*} and $d^2f(x; v, \cdot)$ do exist. Namely, they are well defined for the functions of the type $f = g \circ F$ where F is of class C^2 and g is a piecewise linear-quadratic function verifying an extra regularity condition (see [24] for details). Moreover, a duality result linking both kinds of derivatives is proved for this class of functions by studying the specific expressions that the derivatives take in such a case.

We shall extend Rockafellar's results to a larger class of functions g and moreover, without requiring the spaces to be finite dimensional. We begin by studying in the next results the case of a convex function, and we shall present the general case in the next section.

Our first result extends [24, Proposition 3.5] and further clarifies the relationship existing between the second order pseudo-derivative and the parabolic one. It gives moreover a general criteria for the existence of the latter and a formula to compute it which may be easier to handle in specific situations than the plain definition.

For the rest of this section we assume that f is *closed proper* and *convex*, with $\partial f(x) \neq \emptyset$, that is, f is pseudo-differentiable at x . We shall also use the following notation

$$\partial f(x)_v := \{w^* \in \partial f(x) : \langle w^*, v \rangle = f'_x(v)\}.$$

Theorem 3.7. *Let $v \in \text{dom}(f'_x)$ and suppose that for every $v^* \in \partial f(x)_v$, the second order pseudo-derivative f''_{x,v^*} exists and satisfies*

$$(6) \quad (H) \quad f''_{x,v^*}(v) = \lim_{t \downarrow 0} \frac{f(x + tv) - f(x) - t\langle v^*, v \rangle}{t^2/2} < \infty.$$

Then, the parabolic pseudo-derivative $d^2f(x; v, \cdot)$ exists and is given by

$$(7) \quad d^2f(x; v, w) = \sup_{v^* \in \partial f(x)_v} f''_{x,v^*}(v) + \langle v^*, w \rangle$$

and dually we have

$$(8) \quad [d^2f(x; v, \cdot)]^*(v^*) = \begin{cases} -f''_{x,v^*}(v), & \text{if } v^* \in \partial f(x)_v; \\ +\infty, & \text{otherwise.} \end{cases}$$

Proof. Setting $h(v^*) = -f''_{x,v^*}(v)$ for $v^* \in \partial f(x)_v$ and $h(v^*) = +\infty$ otherwise, and using Theorem 2.2 we see that the result is equivalent to the statement: $(D_t^v)^* \xrightarrow{M} h$.

Let us begin with the case $v^* \notin \partial f(x)_v$. We first observe that for all $w^* \in X^*$ one obtains, after some straightforward calculations, the equality

$$(D_t^v)^*(w^*) = \frac{2}{t^2}[f(x) + f^*(w^*) - \langle w^*, x \rangle] + \frac{2}{t}[f'_x(v) - \langle w^*, v \rangle]$$

so that for each $v_t^* \rightarrow v^*$ it follows by taking $w^* = v_t^*$ that

$$(9) \quad \liminf_{t \downarrow 0} (D_t^v)^*(v_t^*) = +\infty = h(v^*).$$

Indeed, if v^* is not a subgradient, then we have

$$\liminf_{t \downarrow 0} [f(x) + f^*(v_t^*) - \langle v_t^*, x \rangle] \geq [f(x) + f^*(v^*) - \langle v^*, x \rangle] > 0$$

so that in such a case the quantities $\frac{2}{t^2}[f(x) + f^*(v_t^*) - \langle v_t^*, x \rangle]$ go to $+\infty$ (faster than $\frac{2}{t}[f'_x(v) - \langle v^*, v \rangle]$ could go to $-\infty$).

Similarly, if v^* is a subgradient but does not belong to $\partial f(x)_v$, then the term $\frac{2}{t^2}[f(x) + f^*(v_t^*) - \langle v_t^*, x \rangle]$ is nonnegative, while $[f'_x(v) - \langle v_t^*, v \rangle]$ converges to $[f'_x(v) - \langle v^*, v \rangle] > 0$ so that divided by t diverges to $+\infty$.

Since for all $t_n \downarrow 0$ we may take the trivial sequence $v_n^* \equiv v^*$ for obtaining (3)(b), we conclude using (9) that $M\text{-}\lim(D_t^v)^*(v^*) = h(v^*)$ for all vectors $v^* \notin \partial f(x)_v$.

Let us consider next the case $v^* \in \partial f(x)_v$. It is also a simple matter to obtain from the equality $D_t^v(w) = \Delta_t^{v^*}(v + \frac{1}{2}tw) + \langle v^*, w \rangle$ the characterization

$$(10) \quad (D_t^v)^*(w^*) = -\inf_{z \in X} \left\langle \frac{2(v^* - w^*)}{t}, z - v \right\rangle + \Delta_t^{v^*}(z).$$

In particular, for $v_t^* \rightarrow v^*$ we deduce that $(D_t^v)^*(v_t^*) \geq -\Delta_t^{v^*}(v)$ so that in view of (6) we obtain

$$(11) \quad \liminf_{t \downarrow 0} (D_t^v)^*(v_t^*) \geq -f''_{x, v^*}(v) = h(v).$$

Next we observe that for all $t > 0$, the ball $B(v^*, \sqrt{t})$ is weakly compact so that using a minimax theorem we obtain from (10)

$$(12) \quad \min_{w^* \in B(v^*, \sqrt{t})} (D_t^v)^*(w^*) = -\inf_{z \in X} \Delta_t^{v^*}(z) + \frac{2}{\sqrt{t}}\|z - v\|.$$

But in order to compute the infimum on the right it suffices to consider the z 's verifying $\Delta_t^{v^*}(z) + (2/\sqrt{t})\|z - v\| \leq \Delta_t^{v^*}(v)$, which since $\Delta_t^{v^*}(z)$ is always positive, are all included in $B(v, \alpha(t))$ where $\alpha(t) = (\sqrt{t}/2)\Delta_t^{v^*}(v)$.

We may then select $v_t^* \in B(v^*, \sqrt{t})$ and $v_t \in B(v, \alpha(t))$ which minimize the expressions in (12), and write

$$(D_t^v)^*(v_t^*) = -\Delta_t^{v^*}(v_t) - \frac{2}{\sqrt{t}}\|v_t - v\| \leq -\Delta_t^{v^*}(v_t).$$

From (6) we get $\alpha(t) \rightarrow 0$ as t tends to 0, so that $v_t \rightarrow v$ and passing to the limit in the previous inequality we get

$$\limsup_{t \downarrow 0} (D_t^v)^*(v_t^*) \leq -\liminf_{t \downarrow 0} \Delta_t^{v^*}(v_t) \leq -f''_{x, v^*}(v) = h(v)$$

which combined with (11) and in view of (3) allows us to conclude. \square

Remark. Let us notice that under the hypothesis (H), the quantities $f''_{x,v^*}(v)$ are constant for $v^* \in \partial f(x)_v$ and equal to

$$\theta(v) = \lim_{t \downarrow 0} \frac{f(x + tv) - f(x) - tf'_x(v)}{t^2/2}$$

so that formula (7) reads

$$d^2 f(x; v, w) = \theta(v) + \sup_{v^* \in \partial f(x)_v} \langle v^*, w \rangle.$$

As pointed out earlier, the existence of the first and second order pseudo-derivatives was demonstrated for a large class of functions in [24]. Let us add one more example, that of sublinear functions, which is of importance in optimization since a large class of problems can be reduced to minimizing a composition of a sublinear function with a \mathcal{C}^2 mapping.

Example 1. Let $p: X \rightarrow \mathbb{R}$ be a closed proper sublinear function. It is well known that such a function may be expressed as

$$p(x) = \sup_{v^* \in \partial p(0)} \langle v^*, x \rangle$$

so that p is pseudo-differentiable at 0 with $p'_0 = p$.

It is straightforward to verify that for all $v^* \in \partial p(0)$, the second order pseudo-derivative p''_{0,v^*} exists and is given by

$$p''_{0,v^*}(v) = \begin{cases} 0, & \text{if } \langle v^*, v \rangle = p(v), \\ +\infty, & \text{otherwise.} \end{cases}$$

Moreover, condition (H) is verified so that the parabolic derivative of p at 0 exists for all v and can be characterized as

$$d^2 p(0; v, w) = \sup_{v^* \in \partial p(0)_v} \langle v^*, w \rangle. \quad \square$$

The utility of Theorem 3.7 depends on our ability to verify the hypothesis (H). Example 1 presents one of such situations. We shall give next more general results ensuring this condition.

Proposition 3.8. *Suppose X is finite dimensional.*

(a) *if the pointwise limit $\Delta^{v^*} := \lim_{t \downarrow 0} \Delta_t^{v^*}$ exists and either $\text{dom}(\Delta^{v^*})$ has a nonvoid interior or $\text{dom}(f) \subset x + \text{dom}(\Delta^{v^*})$ then f''_{x,v^*} exists and coincides with Δ^{v^*} .*

(b) *if f''_{x,v^*} exists and $v \in \text{int}(\text{dom}(f''_{x,v^*}))$ then (H) is verified.*

Proof. Immediate from Theorem 2.4. \square

These criteria for verifying (H) are not very useful however, except in particular cases where $\partial f(x)_v$ is "small". In fact, according to Proposition 3.5(c) we

see that $\text{dom}(f''_{x,v} \cdot)$ is included in⁴ $N_{\partial f(x)}(v^*)$, which is likely to have a void interior (even when $\text{int}(\text{dom}(f)) \neq \emptyset$).

On the other hand, the assumption made in (b) is quite strong and allows us to deduce the following finite dimensional variant of Theorem 3.7.

Theorem 3.9. *Suppose X is finite dimensional. Let $v \in \text{dom}(f'_x)$ and assume that $v_0^* \in \partial f(x)$ exists such that f''_{x,v_0^*} is defined and*

$$(13) \quad v \in \text{int}(\text{dom}(f''_{x,v_0^*})).$$

Then the parabolic pseudo-derivative $d^2 f(x; v, \cdot)$ exists and is given by

$$(14) \quad d^2 f(x; v, w) = f''_{x,v_0^*}(v) + \langle v_0^*, w \rangle.$$

If moreover f''_{x,v^} exists for all $v^* \in \partial f(x)_v$ then (7) holds with the supremum being in fact attained in v_0^* .*

Proof. Proposition 3.5(c) implies that $v_0^* \in \partial f(x)_v$ so that for each $w \in X$ we may write

$$D_t^v(w) = \Delta_t^{v_0^*}(v + \tfrac{1}{2}tw) + \langle v_0^*, w \rangle.$$

It is clear that $\varphi_t := \Delta_t^{v_0^*}(v + \cdot)$ epiconverges to $\varphi := f''_{x,v_0^*}(v + \cdot)$ and that the linear maps $A_t w := \tfrac{1}{2}tw$ converge to $A_0 \equiv 0$ as $t \downarrow 0$.

Moreover, (13) implies that $0 \in \text{int}(\text{Rg } A_0 - \text{dom}(\varphi))$ and we can therefore use [16, Theorem 8] to deduce that

$$\varphi_t \circ A_t \xrightarrow{\text{epi}} \varphi \circ A_0 \equiv f''_{x,v_0^*}(v).$$

Since the function $\langle v_0^*, \cdot \rangle$ is continuous, we may conclude (see [1, §1.2.3]) that $D_t^v(w) \xrightarrow{\text{epi}} f''_{x,v_0^*}(v) + \langle v_0^*, \cdot \rangle$, and (14) is therefore proved.

To prove the second assertion we observe that for $v^* \in \partial f(x)_v$ one has $D_t^v(w) = \Delta_t^{v^*}(v + \tfrac{1}{2}tw) + \langle v^*, w \rangle$ from which it follows, using the characterization of epiconvergence given in Theorem 2.1, that

$$d^2 f(x; v, w) \geq f''_{x,v^*}(v) + \langle v^*, w \rangle$$

and the result is then apparent. \square

Another criteria (this one holding in the infinite dimensional case as well) which helps in verifying hypothesis (H) in Theorem 3.7, is given in the next proposition. To this end let us recall that the ε -subdifferential of a convex function f at a point x is the set

$$\partial_\varepsilon f(x) := \{x^* \in X^* : f(x) - \varepsilon + \langle x^*, y - x \rangle \leq f(y), \forall y \in X\}.$$

⁴ $N_{\partial f(x)}(v^*) = \{v \in X : \langle w^*, v \rangle \leq \langle v^*, v \rangle \forall w^* \in \partial f(x)\}$ is the usual normal cone to $\partial f(x)$ at v^* . Equivalently $v \in N_{\partial f(x)}(v^*)$ if and only if $v^* \in \partial f(x)_v$.

Proposition 3.10. *Let $v^* \in \partial f(x)$ and suppose that for all $v \in X$ verifying $\langle v^*, v \rangle = f'_x(v)$ one has*

$$(15) \quad (\forall t_n \downarrow 0)(\forall \varepsilon > 0) \quad S^* \text{-} \limsup_{n \rightarrow \infty} \frac{1}{t_n} [\partial_{\varepsilon t_n^2} f(x + t_n v) - v^*] \neq \emptyset.$$

Then f''_{x,v^} exists if and only if $(\Delta_t^{v^*})_{t>0}$ converges pointwise and in such a case both limits coincide.*

Proof. When $\langle v^*, v \rangle < f'_x(v)$ one easily sees from Proposition 3.3 that there are $\varepsilon > 0$ and a weak neighborhood V of v such that

$$[f(x + tv') - f(x)]/t - \langle v^*, v' \rangle \geq f'(x; v') - \langle v^*, v' \rangle \geq \varepsilon$$

for all $t > 0$ and $v' \in V$. It follows at once that for all $v_t \rightarrow v$ we have

$$(16) \quad \lim_{t \downarrow 0} \Delta_t^{v^*}(v) = +\infty = \lim_{t \downarrow 0} \Delta_t^{v^*}(v_t).$$

Let us consider next the case $\langle v^*, v \rangle = f'_x(v)$ and take a sequence $v_t \rightarrow v$. Suppose first that $(\Delta_t^{v^*})$ converges pointwise and choose a sequence $t_n \downarrow 0$ so that $\Delta_{t_n}^{v^*}(v_{t_n}) \rightarrow \liminf_{t \downarrow 0} \Delta_t^{v^*}(v_t)$.

From (15) we may find, for each $\varepsilon > 0$, a linear function $w^* \in X^*$, a subset J of \mathbb{N} and a sequence $w_n^* \xrightarrow{n \in J} w^*$ with

$$v^* + t_n w_n^* \in \partial_{\varepsilon t_n^2} f(x + t_n v)$$

so that we may write

$$\Delta_{t_n}^{v^*}(v) \leq \Delta_{t_n}^{v^*}(v_{t_n}) - 2\varepsilon - 2\langle w_n^*, v_{t_n} - v \rangle$$

and going to the limit we obtain

$$\lim_{t \downarrow 0} \Delta_t^{v^*}(v) \leq \liminf_{t \downarrow 0} \Delta_t^{v^*}(v_t) - 2\varepsilon.$$

Letting $\varepsilon \downarrow 0$ and using (16) and Proposition 2.3 we conclude that f''_{x,v^*} exists and coincides with $\lim_{t \downarrow 0} \Delta_t^{v^*}$.

Conversely, let us suppose that f''_{x,v^*} exists. Let us choose $t_n \downarrow 0$ so that $\Delta_{t_n}^{v^*}(v) \rightarrow \limsup_{t \downarrow 0} \Delta_t^{v^*}(v)$ and take $v_n \rightarrow v$ with $\Delta_{t_n}^{v^*}(v_n) \rightarrow f''_{x,v^*}(v)$. For each $\varepsilon > 0$ we may find $w_n^* \xrightarrow{n \in J} w^*$ as before and deduce

$$\Delta_{t_n}^{v^*}(v) \leq \Delta_{t_n}^{v^*}(v_n) - 2\varepsilon - 2\langle w_n^*, v_n - v \rangle.$$

Letting $n \uparrow \infty$ and then $\varepsilon \downarrow 0$ we deduce

$$\limsup_{t \downarrow 0} \Delta_t^{v^*}(v) \leq f''_{x,v^*}(v)$$

and we conclude the proof by using Proposition 2.3 once again. \square

Let us remark that in the finite dimensional case, condition (15) may be equivalently expressed as

$$(\forall \varepsilon > 0) \quad \limsup_{t \downarrow 0} \frac{1}{t} d(\partial_{\varepsilon t^2} f(x + tv), v^*) < +\infty$$

which is obviously satisfied when f is of class $\mathcal{E}^{1,1}$.

4. CHAIN RULES FOR PSEUDO-DERIVATIVES

We shall consider in this section the pseudo-differentiability properties of the functions of the type

$$(17) \quad f = g \circ F$$

where F is of class \mathcal{E}^2 and g is a convex function which is locally lipschitzian on its domain $C = \text{dom}(g) = \{x : g(x) < +\infty\}$. We shall therefore extend the results in [24, Rockafellar] which apply only when g is *piecewise linear-quadratic* (see [24] for the precise definition), and the involved spaces are finite dimensional.

The class of extended real valued functions such as g is of main significance in mathematical programming. In fact, the infinite valuedness of the functions one is led to deal with in optimization, appears frequently as a consequence of manipulating the constraints of a problem by way of an infinite penalization associated to the violation of these. Namely, one replaces the problem

$$\text{Minimize } \{g(x) : x \in C\}$$

by the ‘unrestricted’ one

$$\text{Minimize } \{(g + \chi_C)(x) : x \in X\}$$

where $\chi_C(x) = 0$ for $x \in C$ and $\chi_C(x) = +\infty$ otherwise. The function $g + \chi_C$ is clearly convex and locally lipschitzian on its domain as soon as C is convex and $g : X \rightarrow \mathbb{R}$ is also convex and locally lipschitz.

More generally, functions of the type (17) arise when considering optimization problems with functional constraints like

$$(18) \quad \text{Minimize } \{\phi(G(x)) : x \in A \text{ and } H(x) \in B\}$$

where G and H are \mathcal{E}^2 mappings, A and B are closed convex sets and ϕ is a locally lipschitz convex function. In fact, if we define

$$\begin{aligned} g(y, z, x) &:= \phi(y) + \chi_B(z) + \chi_A(x), \\ F(x) &:= (G(x), H(x), x) \end{aligned}$$

one easily sees that (18) is equivalent to

$$\text{Minimize } \{g \circ F(x) : x \in X\}$$

where no explicit constraints appear and g and F verify the assumptions made for (17).

We shall fix for the rest of this section two reflexive Banach spaces Y and X , a \mathcal{E}^2 mapping $F: Y \rightarrow X$ and a closed proper convex function $g: X \rightarrow \overline{\mathbb{R}}$ such that $g|_C$ is locally lipschitzian, where $C = \text{dom}(g)$. We shall denote $f = g \circ F$ and study its pseudo-differentiability properties at a fixed point $y \in \text{dom}(f) = F^{-1}(C)$. We shall also denote $x = F(y)$.

We will need throughout the following regularity assumption, which is akin to Robinson's regularity condition [21, 22],

$$(19) \quad 0 \in \text{core}[C - F(y) - DF(y)Y].$$

In fact the important consequence of (19) which we shall use in the sequel is the following:

Theorem 4.1. *If (19) is satisfied, then there exist $m \geq 0$ and a (strong) neighborhood W of y such that*

$$(20) \quad d(z, F^{-1}(C)) \leq md(F(z), C)$$

for all $z \in W$.

Proof. See for instance [21]. \square

For other equivalent statements and important consequences of condition (19) in mathematical programming see for example [7, 19, 21, 22] and references therein.

We proceed now to state and prove the pseudo-differentiability results for the general function $f = g \circ F$. We shall begin by studying the first order pseudo-derivative and the parabolic one, and later we shall tackle the more difficult case of the second order pseudo-derivative.

Theorem 4.2. *The function g is pseudo-differentiable at x . Moreover, if condition (19) is satisfied then $g \circ F$ is pseudo-differentiable at y , and we have*

$$(21) \quad (g \circ F)'_y(h) = g'_x(DF(y)h)$$

and also

$$(22) \quad \partial(g \circ F)(y) = DF(y)^* \partial g(x)$$

where $DF(y)^*$ represents the adjoint operator of $DF(y)$.

Proof. Let us choose a lipschitz constant l for $g|_C$ working in the (strong) neighborhood V of x , that is

$$(23) \quad |g(x_1) - g(x_2)| \leq l\|x_1 - x_2\| \quad \forall x_1, x_2 \in V \cap C$$

and suppose with no loss of generality that $V \subset W$ where W is the neighborhood given by the previous theorem.

From (23) we readily obtain that

$$-l\|v\| \leq g'(x; v) \quad \forall v \in X$$

from which we get $\text{cl}[g'(x; \cdot)](0) = 0$ and the pseudo-differentiability of g follows from Proposition 3.3.

Let us consider next the function $f = g \circ F$.

If $h \in Y$ and $h_t \rightarrow h$ then one has

$$(24) \quad \frac{F(y + th_t) - F(y)}{t} \rightarrow DF(y)h.$$

Indeed, for each $v^* \in X^*$ and every $t > 0$ we may find $\alpha_t \in]0, 1[$ such that

$$\left\langle v^*, \frac{F(y + th_t) - F(y)}{t} \right\rangle = \langle v^*, DF(y + t\alpha_t h_t)h_t \rangle.$$

Now, $h_t \rightarrow h$ so that $(h_t)_{t>0}$ is strongly bounded and therefore $y + t\alpha_t h_t$ converges strongly to y . Since F is \mathcal{C}^1 we can see that the right-hand side of the above equality converges to $\langle v^*, DF(y)h \rangle$, proving therefore (24).

Hence, using (24) and (3) we obtain that

$$g'_x(DF(y)h) \leq \liminf_{t \downarrow 0} \Delta_t(h_t)$$

where Δ_t is taken as in Definition 3.1.

Let us take next $t_n \downarrow 0$ and construct a sequence $h_n \rightarrow h$ with

$$\limsup_{n \rightarrow \infty} \Delta_{t_n}(h_n) \leq g'_x(DF(y)h).$$

It suffices to consider the case $g'_x(v) < \infty$, where we have set $v = DF(y)h$. Then if we choose $v_n \rightarrow v$ with $[g(x + t_n v_n) - g(x)]/t_n \rightarrow g'_x(v)$, for all n large enough we must have $x + t_n v_n \in C$ and we may use the previous theorem to write

$$d(y + t_n h, F^{-1}(C)) \leq m \|F(y + t_n h) - x - t_n v_n\| = o(t_n)$$

which allows us to select $h_n \rightarrow h$ with $y + t_n h_n \in F^{-1}(C)$.

Now, $v'_n = [F(y + t_n h_n) - F(y)]/t_n \rightarrow v$ and $x + t_n v'_n \in C$ so that for all n large enough we have from (23)

$$\Delta_{t_n}(h_n) \leq \frac{g(x + t_n v'_n) - g(x)}{t_n} + l \|v'_n - v_n\|$$

and passing to the limit we obtain $\limsup_{n \rightarrow \infty} \Delta_{t_n}(h_n) \leq g'_x(DF(y)h)$ as required.

The previous arguments prove that f is pseudo-differentiable and that (21) holds. Moreover, a well-known result about support functions allows us to deduce

$$\partial(g \circ F)(y) = \overline{DF(y)^* \partial g(x)}.$$

But since $C - x \subset \text{dom}(g'_x)$, we get from (19) and Baire's lemma that $0 \in \text{int}[\text{Rg } DF(y) - \text{dom}(g'_x)]$ and consequently $DF(y)^* \partial g(x)$ is closed (see for instance [3, Chapter I, §5, Theorem 5]) so that (22) holds. \square

For the parabolic pseudo-derivative we may prove the following result, whose proof parallels very closely the previous one.

Theorem 4.3. Suppose (19) holds. Let $h \in \text{dom}(g \circ F)'_y$ and suppose that the parabolic pseudo-derivative $d^2g(F(y); DF(y)h, \cdot)$ exists. Then we also have that $d^2(g \circ F)(y; h, \cdot)$ exists and is given by

$$(25) \quad d^2(g \circ F)(y; h, k) = d^2g(F(y); DF(y)h, DF(y)k + D^2F(y)hh).$$

Proof. Let us set $v = DF(y)h$ and $w = DF(y)k + D^2F(y)hh$. We claim that for all $k_t \rightarrow k$ one has

$$(26) \quad w_t := \frac{F(y + th + \frac{1}{2}t^2k_t) - F(y) - tDF(y)h}{t^2/2} \rightarrow w.$$

Indeed, similarly to the proof of the previous theorem, for each $v^* \in X^*$ and each $t > 0$ one may find $\alpha_t \in]0, 1[$ such that

$$\langle v^*, w_t \rangle = \langle v^*, DF(y)k_t + D^2F(y_t)hh \rangle + t \langle v^*, D^2F(y_t)(h + \frac{t}{4}k_t) \rangle$$

where $y_t = y + \alpha_t(th + \frac{1}{2}t^2k_t)$. Since $k_t \rightarrow k$ and $F \in \mathcal{C}^2$ we get $y_t \rightarrow y$ and $D^2F(y_t) \rightarrow D^2F(y)$, and also that the second term on the right-hand side above tends to zero, from which (26) follows easily.

Combining (26) with (21) we get

$$d^2g(x; v, w) \leq \liminf_{t \downarrow 0} D_t^h(k_t)$$

where D_t^h is taken as in Definition 3.6.

Let us consider next $t_n \downarrow 0$ and construct $k_n \rightarrow k$ so that

$$(27) \quad \limsup_{n \rightarrow \infty} D_{t_n}^h(k_n) \leq d^2g(x; v, w).$$

Obviously it suffices to consider the case $d^2g(x; v, w) < \infty$, so that by taking $w_n \rightarrow w$ with

$$\frac{g(x + t_nv + \frac{1}{2}t_n^2w_n) - g(x) - t_ng'_x(v)}{t_n^2/2} \rightarrow d^2g(x; v, w)$$

we must have $x + t_nv + \frac{1}{2}t_n^2w_n \in C$ for all n sufficiently large and we may write (see Theorem 4.1)

$$d(y + t_nh + \frac{1}{2}t_n^2k, F^{-1}(C)) \leq m \|F(y + t_nh + \frac{1}{2}t_n^2k) - x - t_nv - \frac{1}{2}t_n^2w_n\| = o(t_n^2)$$

which allows us to choose $k_n \rightarrow k$ with $y + t_nh + \frac{1}{2}t_n^2k_n \in F^{-1}(C)$.

Clearly one has,

$$w'_n := \frac{F(y + t_nh + \frac{1}{2}t_n^2k_n) - F(y) - t_nDF(y)h}{t_n^2/2} \rightarrow w$$

and using the lipschitzianity of g relative to its domain we can deduce for all n large enough

$$D_{t_n}^h(k_n) \leq \frac{g(x + t_nv + \frac{1}{2}t_n^2w_n) - g(x) - t_ng'_x(v)}{t_n^2/2} + l \|w'_n - w_n\|$$

which after taking upper limits yields (27) as required. \square

We come now to our final chain rule which concerns the second order pseudo-derivatives. It basically asserts that $g \circ F$ is twice pseudo-differentiable as soon as g satisfies the requirements in Theorem 3.7 and the regularity condition (19). Moreover, a duality relation similar to the one in Theorem 3.7 is shown to be valid for $g \circ F$.

Theorem 4.4. *Suppose (19) holds, and that for all $v \in DF(y)Y$ and every $v^* \in \partial g(x)_v$ the second order pseudo-derivative g''_{x,v^*} exists and satisfies*

$$(28) \quad g''_{x,v^*}(v) = \lim_{t \downarrow 0} \frac{g(x + tv) - g(x) - t\langle v^*, v \rangle}{t^2/2} < +\infty.$$

If the mapping $h \rightarrow D^2F(y)hh$ is continuous w.r.t. the weak topologies on Y and X (in particular if Y is finite dimensional), then the second order pseudo-derivative $(g \circ F)''_{y,h^}$ exists for all $h^* \in \partial(g \circ F)(y)$ and can be characterized as*

$$(29) \quad (g \circ F)''_{y,h^*}(h) = \begin{cases} \inf_{k \in Y} d^2(g \circ F)(y; h, k) - \langle h^*, k \rangle, \\ \quad \quad \quad h \in N_{\partial(g \circ F)(y)}(h^*); \\ +\infty, \quad \text{otherwise;} \end{cases}$$

and also

$$(30) \quad (g \circ F)''_{y,h^*}(h) = \begin{cases} \max_{v^* \in \Lambda(h^*)} g''_{x,v^*}(DF(y)h) + \langle v^*, D^2F(y)hh \rangle, \\ \quad \quad \quad h \in N_{\partial(g \circ F)(y)}(h^*); \\ +\infty, \quad \text{otherwise;} \end{cases}$$

where $\Lambda(h^*) = \{v^* \in \partial g(x) : DF(y)^*v^* = h^*\}$.

Proof. Let us fix $h^* \in \partial(g \circ F)(y)$, and let us begin by proving the otherwise parts of (29) and (30). Indeed, if $\langle h^*, h \rangle < f'_y(h)$ and $h_t \rightarrow h$ then one has

$$\liminf_{t \downarrow 0} \frac{f(y + th_t) - f(y)}{t} - \langle h^*, h_t \rangle > 0$$

from which it should be clear that

$$(31) \quad \liminf_{t \downarrow 0} \Delta_t^{h^*}(h_t) = +\infty \geq \limsup_{t \downarrow 0} \Delta_t^{h^*}(h).$$

Let us then consider the case $\langle h^*, h \rangle = f'_y(h)$. To begin with, we observe that for $h_t \rightarrow h$ and $v^* \in \Lambda(h^*)$ we may write

$$(32) \quad \begin{aligned} \Delta_t^{h^*}(h_t) &= \frac{2}{t^2} [g(x + tv_t) - g(x) - t\langle v^*, v_t \rangle] \\ &\quad + \left\langle v^*, \frac{2}{t^2} [F(y + th_t) - F(y) - tDF(y)h_t] \right\rangle \end{aligned}$$

where $v_t := [F(y + th_t) - F(y)]/t$.

It follows as in the proof of Theorem 4.2 that $v_t \rightarrow v := DF(y)h$. Moreover, using the mean value theorem we may find for each $t > 0$ an $\alpha_t \in]0, 1[$ such that with $y_t = y + t\alpha_t h_t$ one has

$$\langle v^*, F(y + th_t) - F(y) - tDF(y)h_t \rangle = \frac{t^2}{2} \langle v^*, D^2F(y_t)h_t h_t \rangle.$$

Now, it is easy to see that $y_t \rightarrow y$, and since F is of class \mathcal{C}^2 we obtain

$$\begin{aligned} \liminf_{t \downarrow 0} \left\langle v^*, \frac{F(y + th_t) - F(y) - tDF(y)h_t}{t^2/2} \right\rangle \\ = \liminf_{t \downarrow 0} \langle v^*, D^2F(y)h_t h_t \rangle \\ = \langle v^*, D^2F(y)hh \rangle \end{aligned}$$

the last equality by the weak continuity of the second derivative of F . On replacing this in (32) we deduce that for all $v^* \in \Lambda(h^*)$ one has

$$(33) \quad \liminf_{t \downarrow 0} \Delta_t^{h^*}(h_t) \geq g''_{x, v^*}(DF(y)h) + \langle v^*, D^2F(y)hh \rangle.$$

Let us fix next a sequence $t_n \downarrow 0$. Theorems 3.7 and 4.3 ensure the existence of $d^2f(y; h, \cdot)$ so that for each $k \in Y$ we may find $k_n \rightarrow k$ with $D_{t_n}^{h^*}(k_n) \rightarrow d^2f(y; h, k)$. Then, for all $m \in \mathbb{N}$ and n large enough we have $h'_n = h + \frac{1}{2}t_n k_n \in B(h, 1/m)$ and we get

$$\limsup_{n \rightarrow \infty} \inf_{h' \in B(h, 1/m)} \Delta_{t_n}^{h^*}(h') \leq d^2f(y; h, k) - \langle h^*, k \rangle.$$

Since this holds for all $k \in Y$ we deduce

$$\limsup_{n \rightarrow \infty} \inf_{h' \in B(h, 1/m)} \Delta_{t_n}^{h^*}(h') \leq \alpha := \inf_{k \in Y} d^2f(y; h, k) - \langle h^*, k \rangle$$

and consequently, if $\alpha \in \mathbb{R}$, we may find an integer $\bar{n}(m)$ so that

$$\inf_{h' \in B(h, 1/m)} \Delta_{t_n}^{h^*}(h') \leq \alpha + 1/m \quad \forall n \geq \bar{n}(m).$$

One may always choose $\bar{n}(\cdot)$ to be strictly increasing so that if we define $\bar{m}(n) = \max\{m : \bar{n}(m) \leq n\}$ we have that $\bar{m}(n)$ goes to ∞ with n and $\bar{n}(\bar{m}(n)) \leq n$. We may then find $h_n \in B(h, 1/\bar{m}(n))$ with

$$\Delta_{t_n}^{h^*}(h_n) \leq \alpha + 1/\bar{m}(n)$$

which implies that $h_n \rightarrow h$ and $\limsup_{n \rightarrow \infty} \Delta_{t_n}^{h^*}(h_n) \leq \alpha$.

Combining this last fact with (33) (and (31)) we see that everything, including the finiteness of α , boils down to showing the equality

$$\alpha = \max_{v^* \in \Lambda(h^*)} g''_{x, v^*}(DF(y)h) + \langle v^*, D^2F(y)hh \rangle.$$

Now, using (25) and defining $\Phi(z) := d^2g(F(y); DF(y)h, z + D^2F(y)hh)$; $A := DF(y)$; and $\Theta(k) := -\langle h^*, k \rangle$ we may write

$$\alpha = \inf_{k \in Y} \Phi(Ak) + \Theta(k).$$

We claim that

$$0 \in (A^* \text{dom}(\Phi^*) + \text{dom}(\Theta^*)) \quad \text{and} \quad 0 \in \text{int}(A \text{dom}(\Theta) - \text{dom}(\Phi))$$

so we can use Fenchel's duality theorem (see for example [3, Chapter IV, §4, Corollary 12]) to obtain that

$$(34) \quad \alpha = - \min_{v^* \in X^*} \Theta^*(-A^*v^*) + \Phi^*(v^*).$$

Now, trivially one has $\Theta^* = \chi_{\{-h^*\}}$ and using (8) we obtain

$$(35) \quad \Phi^*(v^*) = \begin{cases} -g''_{x,v^*}(DF(y)h) - \langle v^*, D^2F(y)hh \rangle, & \text{if } v^* \in \partial g(x)_v; \\ +\infty, & \text{otherwise;} \end{cases}$$

which replaced in (34) gives us

$$\alpha = \max_{\substack{v^* \in \partial g(x)_v \\ DF(y)^*v^* = h^*}} g''_{x,v^*}(DF(y)h) + \langle v^*, D^2F(y)hh \rangle.$$

But as soon as $v^* \in \partial g(x)$ verifies $DF(y)^*v^* = h^*$ one has automatically that $v^* \in \partial g(x)_v$ so that in fact

$$\alpha = \max_{v^* \in \Lambda(h^*)} g''_{x,v^*}(DF(y)h) + \langle v^*, D^2F(y)hh \rangle$$

as it was to be proved.

To conclude our proof we must only verify our claim in the sense that $0 \in (A^* \text{dom}(\Phi^*) + \text{dom}(\Theta^*))$ and $0 \in \text{int}(A \text{dom}(\Theta) - \text{dom}(\Phi))$.

Firstly we notice that Theorem 4.2 implies the existence of $v^* \in \partial g(x)$ such that $DF(y)^*v^* = h^*$ and, as we already noticed, we have automatically $v^* \in \partial g(x)_v$. Combining this with (35) and (28) we get $h^* \in A^* \text{dom}(\Phi^*)$ and consequently $0 \in (A^* \text{dom}(\Phi^*) + \text{dom}(\Theta^*))$.

Let us finally prove that $0 \in \text{int}(A \text{dom}(\Theta) - \text{dom}(\Phi))$. From (28) we have that $x + tv \in C$ for all t small enough. Hence, for all $z \in C$ and $\lambda > 0$ we will have for all small $t > 0$,

$$(1 - \frac{t}{2})[x + tv] + \frac{t}{2}[x + t\lambda(z - x)] = x + tv + \frac{1}{2}t^2(\lambda(z - x) - v) \in C$$

and choosing $l \geq 0$ a lipschitz constant for g near x (on C) we get

$$g(x + tv + \frac{1}{2}t^2(\lambda(z - x) - v)) \leq g(x + tv) + \frac{1}{2}lt^2\|\lambda(z - x) - v\|$$

from which we may deduce, using (28), that $\lambda(z - x) - v \in \text{dom}(d^2g(x; v, \cdot))$.

It follows that $\mathbb{R}_+(C - x) - v \subset \text{dom}(\Phi) + D^2F(y)hh$ and therefore

$$\mathbb{R}_+(C - F(y) - DF(y)Y) - D^2F(y)hh \subset \text{dom}(\Phi) - A \text{dom}(\Theta)$$

and since by (19) the left-hand side equals X , our proof is complete. \square

Remark. From (28) we can see that the term $g''_{x,v^*}(DF(y)h)$ appearing in the maximum in (30) does not depend on $v^* \in \Lambda(h^*)$ when $h \in N_{\partial(g \circ F)(y)}(h^*)$ (see also the remark following Theorem 3.7).

Thus, (30) may be rewritten as

$$(g \circ F)''_{y,h^*}(h) = \begin{cases} \Gamma(h) + \max_{v^* \in \Lambda(h^*)} \langle v^*, D^2 F(y)hh \rangle, & \text{if } h \in N_{\partial(g \circ F)(y)}(h^*); \\ +\infty, & \text{otherwise;} \end{cases}$$

where $\Gamma(h) = \lim_{t \downarrow 0} [g(F(y) + tDF(y)h) - g(F(y)) - t(g \circ F)'_y(h)]/(t^2/2)$.

Remark. To be sure that these results extend those in [24], let us mention that piecewise linear-quadratic functions always satisfy (28) (see [24, Theorem 3.1]), and that the regularity condition considered by Rockafellar

$$N_C(x) \cap \text{Ker}(DF(y)^*) = \{0\}$$

is equivalent to (19) in the finite dimensional case.

5. OPTIMALITY CONDITIONS

Just in order to illustrate how pseudo-derivatives may be used in connection with optimization problems, we present the following optimality conditions. We shall assume that all the pseudo-derivatives we shall employ do exist, without explicitly mentioning it.

Theorem 5.1. *Let $f: X \rightarrow \overline{\mathbb{R}}$ and $x \in X$ a point at which f is finite.*

(a) *Necessary Conditions. If x is a local minimum for f then*

- (i) $f'_x(v) \geq 0$ for all $v \in X$, that is, $0 \in \partial f(x)$;
- (ii) $f''_{x,0}(v) \geq 0$ for all $v \in X$;
- (iii) $f'_x(v) = 0 \Rightarrow d^2 f(x; v, w) \geq 0$ for all $w \in X$.

(b) *Sufficient Condition. If X is finite dimensional and $0 \in \partial f(x)$, then a sufficient condition for x to be a strict local minimum for f is that for each nonzero $v \in X$ one has $f''_{x,0}(v) > 0$.*

Proof. (a) The proofs of all these assertions are very similar and simple. Let us show for instance (iii). Indeed, by taking $t_n \downarrow 0$ and $w_n \rightarrow w$ such that $D^v_{t_n}(w_n) \rightarrow d^2 f(x; v, w)$, we see that for all n sufficiently large we have $f(x + t_n v + \frac{1}{2} t_n^2 w_n) \geq f(x)$ so that $D^v_{t_n}(w_n) \geq 0$ and the conclusion follows.

(b) By contradiction, suppose we may find $x_n \rightarrow x$ such that $x_n \neq x$ and $f(x_n) \leq f(x)$. With no loss of generality we may suppose that the sequence $v_n := (x_n - x)/\|x_n - x\|$ converges to some v with $\|v\| = 1$, so that defining $t_n = \|x_n - x\|$ we get

$$f''_{x,0}(v) \leq \liminf_{n \rightarrow \infty} \frac{f(x + t_n v_n) - f(x)}{t_n^2/2} \leq 0$$

which is contrary to our assumption. \square

Acknowledgments. The author expresses his sincere thanks to Professor A. Auslender for the many discussions about this article, and to an anonymous referee for constructive criticism.

REFERENCES

1. H. Attouch, *Variational convergence for functions and operators*, Pitman, New York, 1984.
2. J. P. Aubin, *Lipschitz behavior of solutions to convex minimization problems*, Math. Oper. Res. **9** (1984), 87–111.
3. J. P. Aubin and I. Ekelad, *Applied nonlinear analysis*, Wiley, 1984.
4. A. Auslender, *Stability in mathematical programming with nondifferentiable data*, SIAM J. Control Optim. **22** (1984), 239–254.
5. A. Ben-Tal and J. Zowe, *Necessary and sufficient conditions for a class of nonsmooth minimization problems*, Math. Programming **24** (1982), 70–91.
6. —, *A unified theory of first and second order conditions for extremum problems in topological vector spaces*, Math. Programming Study **19** (1982), 39–76.
7. J. M. Borwein, *Stability and regular points of inequality systems*, J. Optim. Theory Appl. **48** (1986), 9–52.
8. R. Chaney, *Second order sufficient conditions for nondifferentiable programming problems*, SIAM J. Control Optim. **20** (1982), 20–33.
9. —, *On sufficient conditions in nonsmooth optimization*, Math. Oper. Res. **7** (1982), 463–475.
10. R. Cominetti and R. Correa, *Sur une dérivée de second ordre en analyse non différentiable*, C. R. Acad. Sci. Paris Sér. I Math. **303** (1986), 861–864.
11. —, *A generalized second order derivative in nonsmooth optimization*, SIAM J. Control Optim. **28** (1990), 789–809.
12. S. Dolecki, *Tangency and differentiation: some application of convergence theory*, Ann. Mat. Pura Appl. (IV) **130** (1982), 223–255.
13. J. B. Hiriart-Urruty, *A new set valued second order derivative for convex functions*, preprint, 1985.
14. J. B. Hiriart-Urruty, J.-J. Strodiot and V. H. Nguyen, *Generalized Hessian matrix and second order optimality conditions for problems with $\mathcal{C}^{1,1}$ data*, Appl. Math. Optim. **11** (1984), 43–56.
15. A. Ioffe, *Nonsmooth analysis: Differential calculus of nondifferentiable mappings*, Trans. Amer. Math. Soc. **266** (1981), 1–55.
16. L. McLinden and R. Bergstrom, *Preservation of convergence of convex sets and functions in finite dimensions*, Trans. Amer. Math. Soc. **268** (1981), 127–142.
17. U. Mosco, *On the continuity of the Young-Fenchel transform*, J. Math. Anal. Appl. **35** (1971), 518–535.
18. J. L. Ndoutume, *Epi-convergence en calcul différentiel généralisé, Résultats de convergence et approximation*, Thesis, Université de Perpignan, 1987.
19. J. P. Penot, *On regularity conditions in mathematical programming*, Math. Programming Study **19** (1982), 167–199.
20. —, *Generalized higher order derivatives and higher order optimality conditions*, preprint, 1985.
21. S. Robinson, *Stability theorems for systems of inequalities, Part II: differentiable nonlinear systems*, SIAM J. Numer. Anal. **13** (1976), 497–513.
22. —, *Regularity and stability for convex multivalued functions*, Math. Oper. Res. **1** (1976), 130–143.

23. R. T. Rockafellar, *Maximal monotone relations and the second derivatives of nonsmooth functions*, Ann. Inst. H. Poincaré **2** (1985), 167–184.
24. ———, *First and second order pseudo-differentiability in nonlinear programming*, Trans. Amer. Math. Soc. **307** (1988), 75–108.
25. G. Salinetti and R. Wets, *On the relations between two types of convergence for convex functions*, J. Math. Anal. Appl. **60** (1977), 211–226.
26. A. Seeger, *Analyse du second ordre de problèmes non différentiables*, Thesis, Université Paul Sabatier, 1986.

DÉPARTEMENT DE MATHÉMATIQUES APPLIQUÉES, UNIVERSITÉ BLAISE PASCAL (CLERMONT II),
BP 45, 63170 AUBIÈRE, FRANCE