G. Crespi D. La Torre M.Rocca

Second-order mollified derivatives and optimization

2002/9

UNIVERSITÀ DELL'INSUBRIA
FACOLTÀ DI ECONOMIA

http://eco.uninsubria.it

Printed in Italy in March 2002
Università degli Studi dell'Insubria
Via Ravasi 2, 21100 Varese, Italy

Second–order mollified derivatives and optimization*

Giovanni P. Crespi† Davide La Torre‡ Matteo Rocca§

Abstract

The class of strongly semicontinuous functions is considered. For these functions the notion of mollified derivatives, introduced by Ermoliev, Norkin and Wets, is extended to the second order. By means of a generalized Taylor's formula, second order necessary and sufficient conditions are proved for both unconstrained and constrained optimization.

Keywords: Mollifiers, Optimization, Smooth approximations, Strong semicontinuity .

1 Introduction

In this paper we extend to the second-order the approach introduced by Ermoliev, Norkin and Wets [8] to define generalized derivatives even for discontinuous functions, which often arise in applications (see [8] for references). To deal with such applications a number of approaches have been proposed to develop a subdifferential calculus for nonsmooth and even discontinuous functions. Among the many possibilities, let us remember the notions due to Aubin [2], Clarke [5], Ioffe [13], Michel and Penot [20], Rockafellar [21], in the context of Variational Analysis. The previous approaches are based on the introduction of first-order generalized derivatives. Extensions to higher-order derivatives have been provided for instance by Hiriart-Hurruty, Strodiot and Hien Nguyen [12], Jeyakumar and Luc [14], Klatte and Tammer [15], Michel and Penot [19], Yang and Jeyakumar [26] , Yang [27]. Most of these higher-order approaches assume that the functions involved are of class $C^{1,1}$, that is once differentiable with locally Lipschitz gradient, or at least of class C^1. Anyway, another possibility, concerning the differentiation of nonsmooth functions dates back to the 30's and is related to the names of Sobolev [25], who introduced the concept of "weak derivative" and later of Schwartz [24] who generalized Sobolev's approach with the "theory of distributions". These tecniques are

*This work has been supportted by the F.A.R. 2000, of the University of Insubria.

†*Università Bocconi, I.M.Q., v.le Isonzo 25, 20137 Milano, Italia. e–mail: giovanni.crespi@unibocconi.it*

‡*Università di Milano, Dipartimento di Economia Politica e Aziendale, via Conservatorio 7, 20122 Milano, Italia. e–mail: davide.latorre@unimi.it*

§*Università dell'Insubria, Dipartimento di Economia, via Ravasi 2, 21100 Varese, Italia. e–mail: mrocca@eco.uninsubria.it*

widely used in the theory of partial differential equations, in Mathematical Physics and in related problems, but they have not been applied to deal with optimization problems involving nonsmooth functions, until the work of Ermoliev, Norkin and Wets.
The tools which allow to link the "modern" and the "ancient" approaches to Nonsmooth Analysis are those of "mollifier" and of "mollified functions". More specifically, the approach followed by Ermoliev, Norkin and Wets appeals to some of the results of the theory of distributions. They associate with a point $x \in \mathbb{R}^n$ a family of mollifiers (density functions) whose support tends toward x and converges to the Dirac function. Given such a family, say $\{\psi_\epsilon, \epsilon > 0\}$, one can define a family of mollified functions associated to a function $f : \mathbb{R}^n \to \mathbb{R}$ as the convolution of f and ψ_ϵ (mollified functions will be denoted by f_ϵ). Hence a mollified function can be viewed as an averaged function. The mollified functions possess the same regularity of the mollifiers ψ_ϵ and hence, if they are at least of class C^2, one can define first and second-order generalized derivatives as the cluster points of all possible values of first and second-order derivatives of f_ϵ. For more details one can see [8].
In this paper, section 2 recalls the notions of mollifier, of epi-convergence of a sequence of functions and some definitions introduced in [8]. Section 3 is devoted to the introduction of second-order derivatives by means of mollified functions; sections 4 and 5 deal, respectively, with second-order necessary and sufficient optimality conditions for unconstrained and constrained problems.

2 Preliminaries

To follow the approach presented in [8] we first need to introduce the notion of mollifier (see e.g. [4]):

Definition 1 *A sequence of* mollifiers *is any sequence of functions* $\psi := \{\psi_\epsilon\} : \mathbb{R}^m \to \mathbb{R}_+$, $\epsilon \downarrow 0$, *such that:*

i) $supp\, \psi_\epsilon := \{x \in \mathbb{R}^n \mid \psi_\epsilon(x) > 0\} \subseteq \rho_\epsilon cl\mathcal{B},\ \rho_\epsilon \downarrow 0,$

ii) $\displaystyle\int_{\mathbb{R}^n} \psi_\epsilon(x)dx = 1,$

where $\mathcal{B}$ *is the unit ball in* $\mathbb{R}^n$ *and* clX *means the closure of the set* X.

Although in the sequel we may consider general families of mollifiers, some examples may be useful:

Example 1 *Let* ϵ *be a positive number.*

i) The functions:

$$\psi_\epsilon(x) = \begin{cases} \frac{1}{\epsilon^m}, & \max_{1,\ldots,m} |x_i| \le \frac{\epsilon}{2} \\ 0, & otherwise \end{cases}$$

are called Steklov mollifiers.

ii) The functions:

$$\psi_\epsilon(x) = \begin{cases} \frac{C}{\epsilon^m} \exp\left(\frac{\epsilon^2}{\|x\|^2-\epsilon^2}\right), & if\ \|x\| < \epsilon \\ 0, & if\ \|x\| \geq \epsilon \end{cases}$$

with $C \in \mathbb{R}$ such that $\int_{\mathbb{R}^m} \psi_\epsilon(x)dx = 1$, are called standard mollifiers.

It is easy to check that both the previous families of functions are of class $\mathcal{C}^\infty$.

Definition 2 ([4]) *Given a locally integrable function $f : \mathbb{R}^m \to \mathbb{R}$ and a sequence of bounded mollifiers, define the functions $f_\epsilon(x)$ through the convolution:*

$$f_\epsilon(x) := \int_{\mathbb{R}^m} f(x-z)\psi_\epsilon(z)dz.$$

The sequence $f_\epsilon(x)$ is said a sequence of mollified functions*.*

In the following all the functions considered will be assumed to be locally integrable.

Remark 1 There is no loss of generality in considering $f : \mathbb{R}^m \to \mathbb{R}$. The results in this paper remain true also if f is defined on an open subset of $\mathbb{R}^m$.

Some properties of the mollified functions can be considered classical:

Theorem 1 ([4]) *Let $f \in \mathcal{C}(\mathbb{R}^m)$. Then f_ϵ converges continuosly to f, i.e. $f_\epsilon(x_\epsilon) \to f(x)$ for all $x_\epsilon \to x$. In fact f_ϵ converges uniformly to f on every compact subset of $\mathbb{R}^m$ as $\epsilon \downarrow 0$.*

The previous convergence property can be generalized.

Definition 3 ([1], [23]) *A sequence of functions $\{f_n : \mathbb{R}^m \to \mathbb{R}\}$* epi–converges *to $f : \mathbb{R}^m \to \mathbb{R}$ at x, if:*

i) $\liminf_{n\to+\infty} f_n(x_n) \geq f(x)$ *for all $x_n \to x$;*

ii) $\lim_{n\to+\infty} f_n(x_n) = f(x)$ *for some sequence $x_n \to x$.*

The sequence $\{f_n\}$ epi–converges to f if this holds for all $x \in \mathbb{R}^m$, in which case we write $f = e - \lim f_n$.

Remark 2 It can be easily checked that when f is the epi–limit of some sequence f_n then f is lower semicontinuous. Moreover if f_n converges continuously, then also epi–converges.

Definition 4 ([8]) *A function $f : \mathbb{R}^m \to \mathbb{R}$ is said* strongly lower semicontinuous *(s.l.s.c.) at x if it is lower semicontinuous at x and there exists a sequence $x_n \to x$ with f continuous at x_n (for all n) such that $f(x_n) \to f(x)$. The function f is strongly lower semicontinuous if this holds at all x.*
The function f is said strongly upper semicontinuous *(s.u.s.c.) at x if it is upper semicontinuous at x and there exists a sequence $x_n \to x$ with f continuous at x_n (for all n) such that $f(x_n) \to f(x)$. The function f is strongly lower semicontinuous if this holds at all x.*

Proposition 1 *If $f : \mathbb{R}^m \to \mathbb{R}$ is s.l.s.c., then $-f$ is s.u.s.c. .*

Proof: It follows directly from the definitions. □

Theorem 2 ([8]) *Let $\varepsilon_n \downarrow 0$. For any s.l.s.c. funcion $f : \mathbb{R}^m \to \mathbb{R}$, and any associated sequence f_{ϵ_n} of mollified functions we have $f = e - \lim f_{\epsilon_n}$.*

Remark 3 It can be seen that, according to Remark 2, Theorem 1 follows from Theorem 2.

Theorem 3 *Let $\varepsilon_n \downarrow 0$. For any s.u.s.c. function $f : \mathbb{R}^m \to \mathbb{R}$ and any associated sequence f_{ϵ_n} of mollified functions, we have for any $x \in \mathbb{R}^m$:*

i) $\limsup_{n\to+\infty} f_{\epsilon_n}(x_n) \leq f(x)$ *for any sequence $x_n \to x$;*

ii) $\lim_{n\to+\infty} f_{\epsilon_n}(x_n) = f(x)$ *for some sequence $x_n \to x$.*

Proof: Since f is s.u.s.c., we have $-f$ s.l.s.c. and thus Theorem 2 applies:

i) for any sequence $x_n \to x$, $\liminf_{n\to+\infty} \big(- f_{\epsilon_n}(x_n)\big) \geq -f(x)$, which implies:

$$\limsup_{n\to+\infty} f_{\epsilon_n}(x_n) = - \liminf_{n\to+\infty} \big(- f_{\epsilon_n}(x_n)\big) \leq f(x);$$

ii) for some sequence $x_n \to x$, $\lim_{n\to+\infty} \big(- f_{\epsilon_n}(x_n)\big) = -f(x)$, from which we conclude:

$$\lim_{n\to+\infty} f_{\epsilon_n}(x_n) = f(x).$$

□

The following Proposition plays a crucial role in the sequel.

Proposition 2 ([24, 25]) *Whenever the mollifiers ψ_ϵ are of class $\mathcal{C}^k$, so are the associated mollified functions f_ε.*

By means of mollified functions it is possible to define generalized directional derivatives for a nonsmooth function f, which, under suitable regularity of f, coincide with Clarke's generalized derivative. Such an approach has been deepened by several authors (see e.g. [7, 8]) in the first–order case.

Definition 5 ([8]) *Let $f : \mathbb{R}^m \to \mathbb{R}$, $\epsilon_n \downarrow 0$ as $n \to +\infty$ and consider the sequence $\{f_{\epsilon_n}\}$ of mollified functions with associated mollifiers $\psi_{\epsilon_n} \in \mathcal{C}^1$. The* upper mollified derivative *of f at x in the direction $d \in \mathbb{R}^m$, with respect to (w.r.t.) the mollifiers sequence ψ_{ϵ_n} is defined as:*

$$\overline{\mathcal{D}}_{\psi} f(x,d) := \sup_{x_n \to x} \limsup_{n\to+\infty} \nabla f_{\epsilon_n}(x_n)^\top d.$$

Similarly, we might introduce the following:

Definition 6 *Let $f : \mathbb{R}^m \to \mathbb{R}$, $\epsilon_n \downarrow 0$ as $n \to +\infty$ and consider the sequence $\{f_{\epsilon_n}\}$ of mollified functions with associated mollifiers $\psi_{\epsilon_n} \in \mathcal{C}^1$. The* lower mollified derivative *of f at x in the direction $d \in \mathbb{R}^m$, w.r.t. the mollifiers sequence ψ_{ϵ_n} is defined as:*

$$\underline{\mathcal{D}}_{\psi} f(x,d) := \inf_{x_n \to x} \liminf_{n\to+\infty} \nabla f_{\epsilon_n}(x_n)^\top d.$$

In [8] it has been defined also a generalized gradient w.r.t. the mollifiers sequence ψ_{ϵ_n}, in the following way:

$$\partial_\psi f(x) := \{L := Lim \sup_{n\to+\infty} \nabla f_{\epsilon_n}(x_n), \ \ x_n \to x\}$$

i.e. the set of cluster points of all possible sequences $\{\nabla f_{\epsilon_n}(x_n)\}$ such that $x_n \to x$. Clearly (see e.g. [8]) for the above mentioned upper mollified derivative it holds:

$$\overline{\mathcal{D}}_\psi f(x;d) \geq \sup_{L\in\partial_\psi f(x)} L^\top d,$$
$$\underline{\mathcal{D}}_\psi f(x;d) \leq \inf_{L\in\partial_\psi f(x)} L^\top d.$$

This generalized gradient has been used in [7] and [8] to prove first–order necessary optimality conditions for nonsmooth optimization. The equivalence with the well–known notions of Nonsmooth Analysis is contained in the following proposition:

Proposition 3 ([8]) *Let $f : \mathbb{R}^m \to \mathbb{R}$ be locally Lipschitz at x; then $\partial_\psi f(x)$ coincides with Clarke's generalized gradient and $\overline{\mathcal{D}}_\psi f(x,d)$ coincides with Clarke's generalized derivative [5].*

Remark 4 From the previous proposition and the well–known properties of Clarke's generalized gradient, we deduce that, if f and $\psi_\varepsilon \in \mathcal{C}^1$, then $\partial_\psi f(x) = \nabla f(x)$.

Properties of these generalized derivatives and their applications to optimization problems are investigated in [7, 8]. By the way, for the aim of our paper, we will need to point out the following proposition (contained in [8]) of which we give an alternative proof.

Proposition 4 *Let $f : \mathbb{R}^m \to \mathbb{R}$ and $x \in \mathbb{R}^m$. Then:*

i) $\overline{\mathcal{D}}_\psi f(\cdot;d)$ is upper semicontinuous (u.s.c.) at x for all $d \in \mathbb{R}^m$;

ii) $\underline{\mathcal{D}}_\psi f(\cdot;d)$ is lower semicontinuous (l.s.c.) at x for all $d \in \mathbb{R}^m$.

Proof: We can prove only i), since ii) follows with the same reasoning.
Assume $d \in \mathbb{R}^m$ is fixed. First we note that the upper semicontinuity is obviuous if $\overline{\mathcal{D}}_\psi f(x;d) = +\infty$. Otherwise, for all $K > \overline{\mathcal{D}}_\psi f(x;d)$, there exists a neighborhood $U(x)$ and an integer n_0 so that:

$$\nabla f_{\epsilon_n}(x')^\top d < K, \qquad \forall n > n_0, \ \ \forall x' \in U(x).$$

Therefore, for each $x' \in U(x)$, we have:

$$\overline{\mathcal{D}}_\psi f(x';d) = \sup_{x_n\to x'} \limsup_{n\to+\infty} \nabla f_{\epsilon_n}(x_n)^\top d \leq K,$$

which shows that $\overline{\mathcal{D}}_\psi f(\cdot;d)$ is u.s.c. indeed. □

Furthermore, we point out the following property:

Proposition 5 *$\overline{\mathcal{D}}_\psi f(x;\cdot)$ and $\underline{\mathcal{D}}_\psi f(x;\cdot)$ are positively homogeneous functions. Furthermore, if $\overline{\mathcal{D}}_\psi f(x;\cdot)$ ($\underline{\mathcal{D}}_\psi f(x;\cdot)$ respectively) is finite then it is subadditive (resp. superadditive) and hence convex (resp. convcave) as a function of the direction d.*

Proof: The positive homogeneity is trivial. Concerning the second part of the Theorem, we have, $\forall d_1,\, d_2 \in \mathbb{R}^m$:

$$\begin{aligned}\overline{\mathcal{D}}_\psi f(x;d_1+d_2) &= \sup_{x_n\to x}\limsup_{n\to+\infty} \nabla f(x_n)^\top (d_1+d_2) \leq \\ &\leq \sup_{x_n\to x}\limsup_{n\to+\infty} \nabla f(x_n)^\top d_1 + \sup_{x_n\to x}\limsup_{n\to+\infty} \nabla f(x_n)^\top d_2 = \\ &= \overline{\mathcal{D}}_\psi f(x;d_1) + \overline{\mathcal{D}}_\psi f(x;d_2),\end{aligned}$$

and hence $\overline{\mathcal{D}}_\psi f(x;\cdot)$ is subadditive. Convexity follows considering positive homogeneity and subadditivity. The proof for $\underline{\mathcal{D}}_\psi f(x;\cdot)$ is analogous. □

3 Second–order mollified derivatives

As suggested in [8], by requiring some more regularity of the mollifiers, it is possible to construct also second–order generalized derivatives.

Definition 7 *Let $f:\mathbb{R}^m\to\mathbb{R}$, $\epsilon_n\downarrow 0$ and consider the sequence of mollified functions $\{f_{\epsilon_n}\}$, obtained from a family of mollifiers $\psi_{\epsilon_n}\in\mathcal{C}^2$. We define the* second–order upper mollified derivative *of f at x in the directions d and $v\in\mathbb{R}^m$, w.r.t. to the mollifiers sequence $\{\psi_{\epsilon_n}\}$, as:*

$$\overline{\mathcal{D}}^2_\psi f(x;d,v) := \sup_{x_n\to x}\limsup_{n\to+\infty} d^\top Hf_{\epsilon_n}(x_n)v,$$

where $Hf_{\epsilon_n}(x)$ is the Hessian matrix of the function $f_{\epsilon_n}\in\mathcal{C}^2$ at the point x.

Definition 8 *Let $f:\mathbb{R}^m\to\mathbb{R}$, $\epsilon_n\downarrow 0$ and consider the sequence of mollified functions $\{f_{\epsilon_n}\}$, obtained from a family of mollifiers $\psi_{\epsilon_n}\in\mathcal{C}^2$. We define the* second–order lower mollified derivative *of f at x in the directions d and $v\in\mathbb{R}^m$, w.r.t. the mollifiers sequence $\{\psi_{\epsilon_n}\}$, as:*

$$\underline{\mathcal{D}}^2_\psi f(x;d,v) := \inf_{x_n\to x}\liminf_{n\to+\infty} d^\top Hf_{\epsilon_n}(x_n)v.$$

Proposition 6 *Let $f:\mathbb{R}^m\to\mathbb{R}$ and $x\in\mathbb{R}^m$.*

i) If $\lambda>0$, then:

$$\overline{\mathcal{D}}^2_\psi \lambda f(x;d) = \lambda\overline{\mathcal{D}}^2_\psi f(x;d);$$
$$\underline{\mathcal{D}}^2_\psi \lambda f(x;d) = \lambda\underline{\mathcal{D}}^2_\psi f(x;d).$$

Moreover, if $\lambda<0$ we get:

$$\overline{\mathcal{D}}^2_\psi \lambda f(x;d) = \lambda\underline{\mathcal{D}}^2_\psi f(x;d).$$

ii) *The maps* $(d,v) \to \overline{\mathcal{D}}^2_\psi f(x;d,v)$ *and* $(d,v) \to \underline{\mathcal{D}}^2_\psi f(x;d,v)$ *are symmetric (that is* $\overline{\mathcal{D}}^2_\psi f(x;d,v) = \overline{\mathcal{D}}^2_\psi f(x;v,d)$ *and* $\underline{\mathcal{D}}^2_\psi f(x;d,v) = \underline{\mathcal{D}}^2_\psi f(x;v,d)$*).*

iii) *The functions* $\overline{\mathcal{D}}^2_\psi f(x;d,\cdot)$ *and* $\underline{\mathcal{D}}^2_\psi f(x;d,\cdot)$ *are positively homogeneous, for any fixed* $d \in \mathbb{R}^m$*.*

iv) *If* $\overline{\mathcal{D}}^2_\psi f(x;\cdot,\cdot)$ *(*$\underline{\mathcal{D}}^2_\psi f(x;\cdot,\cdot)$ *resp.) is finite, then it is sublinear (superlinear).*

v) $\overline{\mathcal{D}}^2_\psi f(x;d,-v) = -\underline{\mathcal{D}}^2_\psi f(x;d,v)$*.*

vi) $\overline{\mathcal{D}}^2_\psi f(\cdot;d,v)$ *is upper semicontinuous (u.s.c.) at* x *for every* d*,* $v \in \mathbb{R}^m$*.*

vii) $\underline{\mathcal{D}}^2_\psi f(\cdot;d,v)$ *is lower semicontinuous (l.s.c.) at* x *for every* $d, v \in \mathbb{R}^m$*.*

Proof: *i)*, *ii)* and *iii)* are obvious from the definitions. The proof of *iv)* is similar to that of Proposition 5.
To prove *v)*, observe that we have:

$$\begin{aligned}
\overline{\mathcal{D}}^2_\psi f(x;d,-v) &= \sup_{x_n \to x} \limsup_{n\to+\infty} -d^\top Hf_{\epsilon_n}(x_n)v = \\
&= \sup_{x_n \to x} -\liminf_{n\to+\infty} d^\top Hf_{\epsilon_n}(x_n)v = \\
&= -\inf_{x_n \to x} \liminf_{n\to+\infty} d^\top Hf_{\epsilon_n}(x_n)v = \\
&= -\underline{\mathcal{D}}^2_\psi f(x;d,v).
\end{aligned}$$

The prooves of *vi)* and *vii)* are analogous to that of Proposition 4. □

In the following we will set for simplicity:

$$\overline{\mathcal{D}}^2_\psi f(x;d) := \overline{\mathcal{D}}^2_\psi f(x;d,d)$$

and:

$$\underline{\mathcal{D}}^2_\psi f(x;d) := \underline{\mathcal{D}}^2_\psi f(x;d,d).$$

Remark 5 Clearly the previous derivatives may be infinity. A sufficient condition for these derivatives to be finite is to require $f \in \mathcal{C}^{1,1}$ (that is once differentiable with locally Lipschitz partial derivatives). In fact, in this case the second–order mollified derivatives can be viewed as first–order mollified derivatives of a locally Lipschitz function and thus Proposition 3 applies.

Remark 6 It is important to underline that the second–order mollified derivatives are dependent on the specific family of mollifiers which we choose and also on the sequence ϵ_n. Practically by changing one of these choices we might obtain a different result for $\overline{\mathcal{D}}^2_\psi f(x;d,v)$. However, the results which follow hold true *for any* mollifiers sequence (provided they are at least of class $\mathcal{C}^2$) and any choice of ϵ_n. Moreover, by Proposition 4.10 in [8], we have that, if $f \in \mathcal{C}^1$, then for any choice of the sequence of mollifiers and of ϵ_n, $\overline{\mathcal{D}}^2_\psi f(x;d,v)$ coincides with:

$$\limsup_{x'\to x,\ t\downarrow 0} \frac{\nabla f(x'+td)^\top v - \nabla f(x')^\top v}{t}.$$

So, when f is of class $\mathcal{C}^{1,1}$ at x, $\underline{\mathcal{D}}^2_\psi f(x;d,v)$ coincides with the derivative introduced in [12] (thet is second–order generalized derivative in Clarke's sense). Hence when $f \in \mathcal{C}^{1,1}$ it is possible to recover results presented in [12] from those which follow.

Remark 7 One of the main advantages of considering mollified derivatives is that we need not to go through a first–order approximation to get the second–order derivative. Practically we derive both first and second–order generalized derivatives as the limit of two indipendent well defined sequences of "numbers".

Using these notions of derivatives, we shall introduce a Taylor's formula for strongly semicontinuous functions:

Theorem 4 (Lagrange theorem and Taylor's formula) *Let $f : \mathbb{R}^m \to \mathbb{R}$ be a s.l.s.c. (resp. s.u.s.c.) function and let $\epsilon_n \downarrow 0$, $t > 0$, d and $x \in \mathbb{R}^m$.*

i) If $\psi_{\epsilon_n} \in \mathcal{C}^1$ is a sequence of mollifiers, there exists a point $\xi \in [x, x + td]$ such that:

$$\begin{array}{c} f(x+td) - f(x) \leq t\overline{\mathcal{D}}_{\psi} f(\xi; d) \\ (f(x+td) - f(x) \geq t\underline{\mathcal{D}}_{\psi} f(\xi; d)) \end{array} .$$

ii) If $\psi_{\epsilon_n} \in \mathcal{C}^2$ is a sequence of mollifiers, there exists $\xi \in [x,\, x + td]$ such that:

$$\begin{array}{c} f(x+td) - f(x) \leq t\overline{\mathcal{D}}_{\psi} f(x; d) + \frac{t^2}{2}\overline{\mathcal{D}}^2_{\psi} f(\xi; d), \\ (f(x+td) - f(x) \geq t\underline{\mathcal{D}}_{\psi} f(x; d) + \frac{t^2}{2}\underline{\mathcal{D}}^2_{\psi} f(\xi; d)) \end{array}$$

assuming that the righthand sides are well defined, i.e. it does not happen the expression $+\infty - \infty$.

Proof: We prove only the second part. The proof of the first part is similar.
For any $x_n \to x$, we can easily write Taylor's formula for each mollified function:

$$f_{\epsilon_n}(x_n + td) - f_{\epsilon_n}(x_n) = t\nabla f_{\epsilon_n}(x_n)^\top d + \frac{t^2}{2} d^\top H f_{\epsilon_n}(\xi_n) d$$

where $\xi_n \in (x_n,\, x_n + td)$. Without loss of generality, we can think that $\xi_n \to \xi \in [x, x + td]$. Now, we consider the $\limsup$ as $n \to +\infty$ and the definition of $\overline{\mathcal{D}}_{\psi} f(x; d)$ and $\overline{\mathcal{D}}^2_{\psi} f(x; d)$ to get:

$$\begin{array}{c} \limsup_{n\to+\infty} f_{\epsilon_n}(x_n + td) - \limsup_{n\to+\infty} f_{\epsilon_n}(x_n) \leq \limsup_{n\to+\infty} [f_{\epsilon_n}(x_n + td) - f_{\epsilon_n}(x_n)] \leq \\ \leq t \limsup_{n\to+\infty} \nabla f_{\epsilon_n}(x_n)^\top d + \frac{t^2}{2} \limsup_{n\to+\infty} d^\top H f_{\epsilon_n}(\xi_n) d \leq t\overline{\mathcal{D}}_{\psi} f(x; d) + \frac{t^2}{2}\overline{\mathcal{D}}^2_{\psi} f(\xi; d). \end{array}$$

By the strong lower semicontinuity assumed on f, there exists a sequence $y_n \to x$ such that:

$$\lim_{n\to+\infty} f_{\epsilon_n}(y_n) = f(x).$$

Thus, recalling Theorem 2, we have, for this sequence:

$$\begin{array}{rcl} \limsup_{n\to+\infty} f_{\epsilon_n}(y_n + td) - \limsup_{n\to+\infty} f_{\epsilon_n}(y_n) & = & \limsup_{n\to+\infty} f_{\epsilon_n}(y_n + td) - \lim_{n\to+\infty} f_{\epsilon_n}(y_n) \geq \\ \liminf_{n\to+\infty} f_{\epsilon_n}(y_n + td) - \lim_{n\to+\infty} f_{\epsilon_n}(y_n) & \geq & f(x + td) - f(x), \end{array}$$

from which the thesis follows.
The other formula follows in a similar way, recalling Theorem 3 instead of Theorem 2. □

It should be clear that, for both semicontinuity of the generalized derivatives and Taylor's formula, we need some conditions to avoid "triviality" of the derivatives, such as f locally Lipschitz so that, as already seen, the first–order mollified derivative is finite, since it coincides with Clarke's derivative.

4 Unconstrained Optimization

In this section we wish to give second–order necessary and sufficient conditions for unconstrained optimization problems of the form:

$$P_1) \qquad \min_{x\in\Omega} f(x)$$

where $f : \mathbb{R}^m \to \mathbb{R}$ and $\Omega \subseteq \mathbb{R}^m$ is an open set.

Some first–order necessary optimality conditions have already been studied in [8], under very weak hypotheses on f.

Theorem 5 ([8]) *Let $f : \mathbb{R}^m \to \mathbb{R}$ be s.l.s.c. and assume that x_0 is a local solution of problem P_1). Then for any sequence of mollifiers, we have:*

$$0 \in \partial_\psi f(x_0).$$

Theorem 6 *Let x_0 be a local solution of problem P_1) with $f : \mathbb{R}^m \to \mathbb{R}$ s.l.s.c.. Then the following conditions hold:*

i) $\overline{\mathcal{D}}_\psi f(x_0; d) \geq 0, \quad \forall d \in \mathbb{R}^m$.

ii) $\overline{\mathcal{D}}^2_\psi f(x_0; d) \geq 0, \quad \forall d \in \mathbb{R}^m$ *such that* $\overline{\mathcal{D}}_\psi f(x_0; d) = 0$.

Proof:

i) By Theorem 5 we know $0 \in \partial_\psi f(x_0)$. Thus, by definition:

$$\overline{\mathcal{D}}_\psi f(x_0; d) \geq 0.$$

ii) Let d be such that $\overline{\mathcal{D}}_\psi f(x_0; d) = 0$ and apply Taylor's formula to get:

$$f(x_0 + td) - f(x_0) \leq \frac{t^2}{2}\overline{\mathcal{D}}^2_\psi f(\xi; d),$$

for $t > 0$, $d \in \mathbb{R}^m$ and $\xi \in [x_0, x_0 + td]$. Fot t "small enough", since x_0 is a local minimizer we obtain $f(x_0 + td) - f(x_0) \geq 0$ and hence, using the upper semicontinuity of $\overline{\mathcal{D}}^2_\psi f(\cdot; d)$:

$$0 \leq \limsup_{t\downarrow 0} 2\frac{f(x_0 + td) - f(x_0)}{t^2} \leq \limsup_{t\downarrow 0} \overline{\mathcal{D}}^2_\psi f(\xi; d) \leq$$

$$\limsup_{x\to x_0} \overline{\mathcal{D}}^2_\psi f(x; d) \leq \overline{\mathcal{D}}^2_\psi f(x_0; d),$$

so that condition *ii*) is proved.

□

Theorem 7 *Let $f : \mathbb{R}^m \to \mathbb{R}$ be s.u.s.c.. Moreover, assume that at x_0 the following conditions hold for any $d \in S^1$ (the unit sphere in $\mathbb{R}^m$):*

i) $\underline{\mathcal{D}}_\psi f(x_0; d) > 0$ *implies that there exist a real number $\alpha(d) > 0$ and a neighborhood of the direction d, $U(d)$ such that:*

$$\underline{\mathcal{D}}_\psi f(x_0 + td'; d') > 0, \quad \forall t \in (0, \alpha(d)), \quad \forall d' \in U(d).$$

ii) $\underline{\mathcal{D}}_\psi f(x_0; d) = 0$ *implies that there exist a real number $\alpha(d) > 0$ and a neighborhood of d, $U(d)$, such that $\underline{\mathcal{D}}_\psi f(x_0; d') \geq 0$ and $\underline{\mathcal{D}}^2_\psi f(x_0 + td'; d') > 0$, $\forall t \in (0, \alpha(d))$ and $\forall d' \in U(d)$.*

Then x_0 is a (strict) local solution of P_1).

Proof: By contradiction, assume that there exists x_n such that: $f(x_n) - f(x_0) \leq 0$. It can be easily written $x_n = x_0 + t_n d_n$, $d_n \in S^1$, $d_n \to d \in S^1$, $t_n \downarrow 0$.

i) If $\underline{\mathcal{D}}_\psi f(x_0; d) > 0$, then:

$$0 \geq f(x_0 + t_n d_n) - f(x_0) \geq t_n \underline{\mathcal{D}}_\psi f(\xi_n; d_n),$$

where $\xi_n \in [x_0, x_0 + t_n d_n]$. Which contradict the hypothesis.

ii) If $\underline{\mathcal{D}}_\psi f(x_0; d) = 0$, then we have the following contradiction:

$$0 \geq f(x_0 + t_n d_n) - f(x_0) \geq t_n \underline{\mathcal{D}}_\psi f(x_0; d_n) + \frac{t_n^2}{2} \underline{\mathcal{D}}^2_\psi f(\xi_n; d_n) > 0,$$

for n sufficiently large.

□

5 Constrained Optimization

In this section we give second–order necessary and sufficient optimality conditions for constrained optimization problems. We begin considering the following problem:

$$P_2) \qquad \min_{x \in K} f(x)$$

where $K \subseteq \mathbb{R}^m$.

Definition 9 *The cone of feasible directions of the set K at x is given by:*

$$F(K, x) := \{d \in \mathbb{R}^m \mid \exists \alpha > 0 : \ \forall t \in [0, \alpha], \ x + td \in K\}.$$

Definition 10 *The set:*

$$T(K, x_0) := \{d \in \mathbb{R}^m \mid \exists \{d_n\} \to d, \ \exists \{t_n\} \downarrow 0 \ : \ x_0 + t_n d_n \in K\}$$

is called the Bouligand tangent cone to the set K at x.

Clearly we have the following inclusion:

$$F(K,x) \subseteq T(K,x).$$

Theorem 8 *Let $f : \mathbb{R}^m \to \mathbb{R}$ be s.l.s.c.. If $x_0 \in K$ is a local solution of problem P_2), then:*

i) $\overline{\mathcal{D}}_\psi f(x_0; d) \geq 0$, $\forall d \in F(K, x_0)$,

ii) $\overline{\mathcal{D}}^{\,2}_\psi f(x_0; d) \geq 0$, $\forall d \in F(K, x_0)$ *such that* $\overline{\mathcal{D}}_\psi f(x_0; d) = 0$.

Proof: Let x_0 be a local minimizer of f over K. For t "sufficiently small" and $d \in F(K, x_0)$, we have:

$$0 \leq \frac{f(x_0 + td) - f(x_0)}{t} \leq \overline{\mathcal{D}}_\psi f(\xi; d),$$

for some $\xi \in [x_0, x_0 + td]$. Taking lim sup as $t \downarrow 0$ of both members and recalling the upper semicontinuity of $\overline{\mathcal{D}}_\psi f(\cdot; d)$, we obtain:

$$0 \leq \limsup_{t \downarrow 0} \overline{\mathcal{D}}_\psi f(\xi; d) \leq \overline{\mathcal{D}}_\psi f(x_0; d).$$

To prove condition ii), let $d \in F(K, x_0)$ be such that $\overline{\mathcal{D}}_\psi f(x_0; d) = 0$. Using Theorem 4, we have, for $t > 0$ "sufficiently small":

$$0 \leq f(x_0 + td) - f(x_0) \leq +\frac{t^2}{2} \overline{\mathcal{D}}^{\,2}_\psi f(\xi; d),$$

where $\xi \in [x_0, x_0 + td]$. Dividing by t^2, taking lim sup for $t \downarrow 0$ and using the upper semicontinuity of $\overline{\mathcal{D}}^{\,2}_\psi f(\cdot; d)$, the thesis follows. □

Theorem 9 *Let $f : \mathbb{R}^m \to \mathbb{R}$ be s.u.s.c., $x_0 \in K$ and assume that $\forall d \in T(K, x_0) \cap S^1$ one of the following conditions holds:*

i) if $\underline{\mathcal{D}}_\psi f(x_0; d) > 0$, then there exist a real number $\alpha(d) > 0$ and a neighborhood of the direction d, say $U(d)$ such that:

$$\underline{\mathcal{D}}_\psi f(x_0 + td'; d') > 0, \quad \forall t \in (0, \alpha(d)), \quad \forall d' \in U(d) \cap S^1;$$

ii) if $\underline{\mathcal{D}}_\psi f(x_0; d) = 0$, then there exist a real number $\alpha(d) > 0$ and a neighborhood of the direction d, say $U(d)$, such that, for each $t \in (0, \alpha(d))$ and for each $d' \in U(d) \cap S^1$ we have $\underline{\mathcal{D}}_\psi f(x_0; d') \geq 0$ and $\underline{\mathcal{D}}^{\,2}_\psi f(x_0 + td'; d') > 0$.

Then x_0 is a local solution of problem P_2).

Proof: Ab assurdo, let assume there exists a feasible sequence $\{x_n\} \to x_0$ such that $f(x_n) < f(x_0)$. It can be easily written, without loss of generality $x_n = x_0 + t_n d_n$, $d_n \in S^1$, $d_n \to d \in S^1$, $t_n \downarrow 0$, and hence $d \in T(K, x_0)$.

i) If $\underline{\mathcal{D}}_\psi f(x_0; d) > 0$, then, as $n \to +\infty$:

$$0 \geq f(x_0 + t_n d_n) - f(x_0) \geq t_n \underline{\mathcal{D}}_\psi f(\xi_n; d_n),$$

with $\xi_n \in [x_0, x_0 + t_n d_n]$, which is trivially a contradiction.

ii) If $\underline{\mathcal{D}}_\psi f(x_0; d) = 0$, then:

$$0 \geq f(x_0 + t_n d_n) - f(x_0) \geq t_n \underline{\mathcal{D}}_\psi f(x_0; d_n) + \frac{t^2}{2} \underline{\mathcal{D}}^2_\psi f(\xi_n; d_n),$$

with $\xi_n \in [x_0, x_0 + t_n d_n]$, which is again a contradiction.

□

Now we deal with the following constrained optimization problem:

$$P_3) \qquad \begin{array}{ll} & \min f(x) \\ \text{s.t.} & \\ & g_i(x) \leq 0, \quad i = 1, \ldots, r \end{array}$$

where $f, g_i : \mathbb{R}^m \to \mathbb{R}$. We will define the set of active constraints at a point x_0 as the index set $I(x_0) : \{i = 1, \ldots, r \,:\, g_i(x_0) = 0\}$ and the feasible set as $\Gamma := \{x \in \mathbb{R}^m \,:\, g_i(x) \leq 0,\ i = 1, \ldots, r\}$.
Concerning this problem, we will first investigate first–order conditions expressed by means of mollified derivatives.

Lemma 1 (Generalized Abadie Lemma) *Let f and g_i, $i \in I(x_0)$ be s.l.s.c., g_i, $i, \notin I(x_0)$ be u.s.c. and assume that $x_0 \in \Gamma$ is a local solution of problem $P_3)$.*
Then $\nexists d \in \mathbb{R}^m$ such that:

$$\begin{cases} \overline{\mathcal{D}}_\psi f(x_0; d) < 0 \\ \overline{\mathcal{D}}_\psi g_i(x_0; d) < 0, \quad i \in I(x_0) \end{cases}$$

Proof: Since x_0 is a local solution of $P_3)$, we can easily check that, $\forall d \in \mathbb{R}^m$, $\nexists \alpha(d) > 0$ such that $\forall t \in (0, \alpha(d))$:

$$\begin{cases} \overline{\mathcal{D}}_\psi f(x_0 + td; d) < 0 \\ \overline{\mathcal{D}}_\psi g_i(x_0 + td; d) < 0, \quad i \in I(x_0) \end{cases}$$

Infact, if for some d such an $\alpha(d)$ would exist, from Theorem 4 we would get, $\forall t \in (0, \alpha(d))$:

$$f(x_0 + td) < f(x_0)$$

and

$$g_i(x_0 + td) < 0, \qquad i \in I(x_0).$$

Since g_i, $i \notin I(x_0)$ are u.s.c. we obtain also, for t "small enough", $g_i(x_0 + td) < 0$, $i \notin I(x_0)$. This fact contradicts that $x_0 \in \Gamma$ is a local solution of $P_3)$.
Hence, for any fixed $d \in \mathbb{R}^m$ one can find a sequence $t_n \downarrow 0$ such that for all n it holds or $\overline{\mathcal{D}}_\psi f(x_0 + t_n d; d) \geq 0$ either $\overline{\mathcal{D}}_\psi g_{\bar{i}}(x_0 + t_n d; d) \geq 0$, for some fixed $\bar{i} \in I(x_0)$.
Recalling that the first–order upper mollified derivative is u.s.c. we obtain that either $\overline{\mathcal{D}}_\psi f(x_0; d) \geq 0$ or $\overline{\mathcal{D}}_\psi g_{\bar{i}}(x_0; d) \geq 0$ and hence we get the thesis. □

Theorem 10 (Generalized F. John Conditions) *Let f, g_i, $i \in I(x_0)$ be s.l.s.c. and g_i, $i \notin I(x_0)$ be u.s.c.. Assume that $x_0 \in \Gamma$ is a local solution of problem P_3) and that $\overline{\mathcal{D}}_{\psi} f(x_0; \cdot)$ and $\overline{\mathcal{D}}_{\psi} g_i(x_0; \cdot)$, $i \in I(x_0)$ are finite.*
Then there exist scalars $\tau \geq 0$, $\lambda_i \geq 0, i \in I(x_0)$, not all zero, such that:

$$\tau \overline{\mathcal{D}}_{\psi} f(x_0; d) + \textstyle\sum_{i \in I(x_0)} \lambda_i \overline{\mathcal{D}}_{\psi} g_i(x_0; d) \geq 0, \quad \forall d \in \mathbb{R}^m. \tag{1}$$

Proof: From the previous Lemma we know that the system:

$$\begin{cases} \overline{\mathcal{D}}_{\psi} f(x_0; d) < 0 \\ \overline{\mathcal{D}}_{\psi} g_i(x_0; d) < 0 \quad i \in I(x_0) \end{cases}$$

has no solution. Since the first–order upper mollified derivatives are convex (Proposition 5), from a well known Theorem of the alternative ([3] Theorem 7.1.2), we obtain the thesis. □

Remark 8 *Of course a relevant question is which conditions would ensure $\tau > 0$ (or equivalently $\tau = 1$) in formula (1). It can be easily seen that this is the case if the following generalized Slater–type constraint qualification condition holds:*

$$\exists \overline{d} \in \mathbb{R}^m \qquad \textit{such that} \qquad \overline{\mathcal{D}}_{\psi} g_i(x_0; \overline{d}) < 0, \quad i \in I(x_0).$$

Now we prove necessary and sufficient second–order optimality conditions for problem P_3).

Theorem 11 *Let f, g_i, $i \in I(x_0)$ be s.l.s.c., g_i, $i \notin I(x_0)$ be u.s.c. and assume that $x_0 \in \Gamma$ is a local solution of problem P_3). Moreover assume that $\overline{\mathcal{D}}_{\psi} f(x_0; \cdot)$ and $\overline{\mathcal{D}}_{\psi} g_i(x_0; \cdot)$, $i \in I(x_0)$ are finite.*
Then, if $\tau \geq 0$, $\lambda_i \geq 0$, $i \in I(x_0)$ satisy (1), the following condition holds:

$$\begin{aligned} &\textit{if } d \in F(\Gamma(\lambda), x_0) \textit{ is such that } \tau \overline{\mathcal{D}}_{\psi} f(x_0; d) + \textstyle\sum_{i \in I(x_0)} \lambda_i \overline{\mathcal{D}}_{\psi} g_i(x_0; d) = 0 \\ &\qquad \textit{then} \qquad \tau \overline{\mathcal{D}}^2_{\psi} f(x_0; d) + \textstyle\sum_{i \in I(x_0)} \lambda_i \overline{\mathcal{D}}^2_{\psi} g_i(x_0; d) \geq 0 \end{aligned} \tag{2}$$

where $\Gamma(\lambda) = \{x \in \Gamma \mid \sum_{i \in I(x_0)} \lambda_i g_i(x) = 0\}$.

Proof: Let $d \in F(\Gamma(\lambda), x_0)$ be such that $\tau \overline{\mathcal{D}}_{\psi} f(x_0; d) + \sum_{i \in I(x_0)} \lambda_i \overline{\mathcal{D}}_{\psi} g_i(x_0; d) = 0$ and observe that, since $\overline{\mathcal{D}}_{\psi} f(x_0; \cdot)$ and $\overline{\mathcal{D}}_{\psi} g_i(x_0; \cdot)$, $i \in I(x_0)$ are finite, we can write, for $t > 0$:

$$\begin{aligned} f(x_0 + td) - f(x_0) &\leq t \overline{\mathcal{D}}_{\psi} f(x_0; d) + \tfrac{t^2}{2} \overline{\mathcal{D}}^2_{\psi} f(\xi; d) \\ g_i(x_0 + td) - g_i(x_0) &\leq t \overline{\mathcal{D}}_{\psi} g_i(x_0; d) + \tfrac{t^2}{2} \overline{\mathcal{D}}^2_{\psi} g_i(\xi_i; d), \quad i \in I(x_0) \end{aligned}$$

where $\xi, \xi_i \in [x_0, x_0 + td]$.
Hence we have:

$$\tau f(x_0 + td) + \sum_{i \in I(x_0)} \lambda_i g_i(x_0 + td) - \tau f(x_0) - \sum_{i \in I(x_0)} \lambda_i g_i(x_0) \leq$$

$$\leq \frac{t^2}{2}[\tau \overline{\mathcal{D}}^2_\psi f(\xi; d) + \sum_{i\in I(x_0)} \lambda_i \overline{\mathcal{D}}^2_\psi g_i(\xi_i; d)].$$

For t "small enough", the lefthandside is nonnegative and hence, using the upper semicontinuity of second–order mollified derivatives:

$$\begin{aligned} 0 &\leq \limsup_{t\downarrow 0}[\tau \overline{\mathcal{D}}^2_\psi f(\xi; d) + \sum_{i\in I(x_0)} \lambda_i \overline{\mathcal{D}}^2_\psi g_i(\xi_i; d)] \leq \\ &\leq \tau \limsup_{t\downarrow 0} \overline{\mathcal{D}}^2_\psi f(\xi; d) + \sum_{i\in I(x_0)} \lambda_i \limsup_{t\downarrow 0} \overline{\mathcal{D}}^2_\psi g_i(\xi_i; d) \leq \\ &\leq \tau \overline{\mathcal{D}}^2_\psi f(x_0; d) + \sum_{i\in I(x_0)} \lambda_i \overline{\mathcal{D}}^2_\psi g_i(x_0; d), \end{aligned}$$

and so we get the thesis. □

Theorem 12 *Let f, g_i, $i \in I(x_0)$ be s.u.s.c. and $x_0 \in \Gamma$. Moreover, assume there exist scalars $\lambda_i \geq 0$, $i \in I(x_0)$ such that $\forall d \in T(\Gamma, x_0) \cap S^1$ one of the following conditions holds:*

i) If $\underline{\mathcal{D}}_\psi f(x_0; d) + \sum_{i\in I(x_0)} \lambda_i \underline{\mathcal{D}}_\psi g_i(x_0; d) > 0$, then there exist a real $\alpha(d) > 0$ and a neighborhood of the direction d, $U(d)$, so that:

$$\underline{\mathcal{D}}_\psi f(x_0 + td'; d') + \textstyle\sum_{i\in I(x_0)} \lambda_i \underline{\mathcal{D}}_\psi g_i(x_0 + td'; d') > 0 \quad \forall t \in (0, \alpha(d)),\ \forall d' \in U(d).$$

ii) If $\underline{\mathcal{D}}_\psi f(x_0; d) + \sum_{i\in I(x_0)} \lambda_i \underline{\mathcal{D}}_\psi g_i(x_0; d) = 0$, then there exist a real $\alpha(d) > 0$ and a neighborhood of the direction d, $U(d)$, so that:

$$\underline{\mathcal{D}}^2_\psi f(x_0 + td'; d') + \textstyle\sum_{i\in I(x_0)} \lambda_i \underline{\mathcal{D}}^2_\psi g_i(x_0 + td'; d') > 0 \quad \forall t \in (0, \alpha(d)),\ \forall d' \in U(d).$$

Then x_0 is a (strict) local solution of P_3).

Proof: By contradiction assume there exists a feasible sequence $\{x_n\} \to x_0$ so that $f(x_n) - f(x_0) \leq 0$. We shall write $x_n = x_0 + t_n d_n$ for some $d_n \to d \in T(\Gamma, x_0) \cap S^1$.

i) If $\underline{\mathcal{D}}_\psi f(x_0; d) + \sum_{i\in I(x_0)} \lambda_i \underline{\mathcal{D}}_\psi g_i(x_0; d) > 0$, then we would have:

$$\begin{gathered} f(x_0 + t_n d_n) - f(x_0) \geq t_n \underline{\mathcal{D}}_\psi f(\xi_n; d_n) \\ \textit{and } g_i(x_0 + t_n d_n) - g_i(x_0) \geq t_n \underline{\mathcal{D}}_\psi g_i(\xi^i_n; d_n), \quad i \in I(x_0) \end{gathered}$$

where ξ_n, $\xi^i_n \in [x_0, x_0 + t_n d_n]$. Using multipliers λ_i, we get:

$$\begin{aligned} 0 &\geq f(x_0 + t_n d_n) + \sum_{i\in I(x_0)} \lambda_i g_i(x_0 + t_n d_n) - f(x_0) - \sum_{i\in I(x_0)} \lambda_i g_i(x_0) \geq \\ &\geq t_n \underline{\mathcal{D}}_\psi f(\xi_n; d_n) + t_n \sum_{i\in I(x_0)} \lambda_i \underline{\mathcal{D}}_\psi g_i(\xi^i_n; d_n) \end{aligned}$$

which contradict the hypothesis for n large enough.

ii) If $\underline{\mathcal{D}}_\psi f(x_0;d) + \sum_{i\in I(x_0)} \lambda_i \underline{\mathcal{D}}_\psi g_i(x_0;d) = 0$, then we shall write:

$$f(x_0+t_nd_n) - f(x_0) \geq t_n\underline{\mathcal{D}}_\psi f(x_0;d_n) + \tfrac{t_n^2}{2}\underline{\mathcal{D}}_\psi^2 f(\xi_n;d_n)$$
$$\textit{and } g_i(x_0+t_nd_n) - g_i(x_0) \geq t_n\underline{\mathcal{D}}_\psi g_i(x_0;d_n) + \tfrac{t_n^2}{2}\underline{\mathcal{D}}_\psi^2 g_i(\xi_n^i;d_n), \quad i\in I(x_0)$$

where ξ_n, $\xi_n^i \in [x_0, x_0+t_nd_n]$. Using multipliers λ_i and the assumption, we get:

$$\begin{aligned} 0 \ &\geq \ f(x_0+t_nd_n) + \sum_{i\in I(x_0)} \lambda_i g_i(x_0+t_nd_n) - f(x_0) - \sum_{i\in I(x_0)} \lambda_i g_i(x_0) \geq \\ &\geq \ \frac{t_n^2}{2}\underline{\mathcal{D}}_\psi^2 f(\xi_n;d_n) + \frac{t_n^2}{2}\sum_{i\in I(x_0)} \lambda_i \underline{\mathcal{D}}_\psi^2 g_i(\xi_n^i;d_n) \end{aligned}$$

which contradict again the hypothesis for n large enough.

□

References

[1] Aubin J.P., Frankowska H., *Set–valued analysis*, Birkhauser, Basel, 1990

[2] Aubin J.P., Lipschitz behaviour of solutions to convex minimization problems, Mathematics of Operations Research, 8, 1984, 87–111

[3] Bazaraa M.S., Shetty C.M., *Foundations of Optimization*, Springer, Berlin, 1976

[4] Brezis H., *Analyse fonctionelle – Theorie et applications*, Masson editeur, Paris, 1963

[5] Clarke F.H., *Optimization and nonsmooth analysis*, SIAM, Classics in Applied Mathematics, Philadelphia, 1990

[6] Cominetti R., Correa R.: *A generalized second–order derivative in nonsmooth optimization.* SIAM J. Control Optim., 28, 1990, 789-809.

[7] Craven B.D., Nondifferentiable optimization by smooth approximations, Optimization, 17, 1, 1986, 3–17

[8] Ermoliev Y.M., Norkin V.I.,Wets R.J.B., *The minimization of semicontinuous functions: mollifier subgradient*, SIAM J. Control and Optimization, 33, 1, 1995, 149–167

[9] Ginchev I., Guerraggio A.:*Second order optimality conditions in nonsmooth unconstrained optimization.* Pliska Stud. Math. Bulgar., 12, 1998, 39-50.

[10] Guerraggio A., Luc. D.T.: *On optimality conditions for $C^{1,1}$ vector optimization problems*, Journal of Optimization Theory and Applications, 109, 3, 2001, 615–629.

[11] Hiriart-Hurruty J.B.: *Contributions a la programmation mathematique: deterministe et stocastique.* (Doctoral thesis), Univ. Clermont-Ferrand, 1977.

[12] Hiriart-Urruty J.B., Strodiot J.J., Hien Nguyen V.: *Generalized Hessian matrix and second–order optimality conditions for problems with* $C^{1,1}$ *data.* Appl. Math. Optim., 11, 1984, 43-56.

[13] Ioffe A.D., Nonsmooth Analysis: differentiable calculus of nondifferentiable mappings, *Transaction of American Mathematical Society*, 266, 1981, 1–56

[14] Jeyakumar V., Luc D.T.: *Approximate Jacobian matrices for nonsmooth continuous maps and* $\mathcal{C}^1$*–optimization*, SIAM J. Control and Optimization, 36, 5, 1998, 1815–1832

[15] Klatte D., Tammer K.: *On second–order sufficient optimality conditions for* $C^{1,1}$ *optimization problems.* Optimization, 19, 1988, 169-179.

[16] La Torre D., Rocca M.: $C^{k,1}$ *functions and divided differences.* Real Analysis Exchange, 25, 1999-2000,743-752.

[17] La Torre D., Rocca M.: $C^{1,1}$ *functions on* $\mathbb{R}^n$ *and Riemann derivatives.* Journal of Computational Analysis and Applications, to appear.

[18] Luc D.T.: *Taylor's formula for* $C^{k,1}$ *functions.* SIAM J. Optimization, 5, 1995, 659-669.

[19] Michel P., Penot J.P.: *Second-order moderate derivatives.* Nonlin. Anal., 22, 1994, 809–824.

[20] Michel P., Penot J.P.: *Calcul sous-différential pour des fonctions lischitziennes an nonlipschitziennes.* Comptes Rendus de l'Academie des Sciences Paris, 298, 1984, 269–272.

[21] Rockafellar R.T., *The theory of subgradients and its application to problems of optimization: convex and nonconvex functions*, Heldermann Verlag, Berlin, 1981

[22] Rockafellar R.T., Wets R. J–B., *Variational systems, an introduction, multifunctions and integrands*, G. Salinetti ed., Lecture Notes in Mathematics 1091, Springer–Verlag, Berlin, 1–54, 1984

[23] Rockafellar R.T., Wets R. J–B., *Variational Analysis*, Springer–Verlag, Berlin, 1998

[24] Schwartz L., *Théorie des distributions*, Hermann, Paris, 1966

[25] Sobolev S.L.: *Some applications of functional analysis in mathematical phisics*, 3rd ed., Nauka, Moscow, 1988 (In Russian)

[26] Yang X.Q., Jeyakumar V.: *Generalized second–order directional derivatives and optimization with* $C^{1,1}$ *functions.* Optimization, 26, 1992, 165-185.

[27] Yang X.Q.: *Second–order conditions in $C^{1,1}$ optimization with applications.* Numer. Funct. Anal. and Optimiz., 14, 1993, 621-632.

[28] Yang X.Q.: *On second–order directional derivatives*, Nonlinear Anal-TMA, 26, 1, 1996, 55-66.