An Approximate Lagrange Multiplier Rule

J. Dutta, S. R. Pattanaik
Department of Mathematics and Statistics
Indian Institute of Technology,
Kanpur, India
email : jdutta@iitk.ac.in, suvendu@iitk.ac.in
and
Michel Théra
XLIM (UMR-CNRS 6172) and Université de Limoges,
123 Avenue A. Thomas, 87060 Limoges Cedex,
France
email: michel.thera@unilim.fr

Abstract

In this paper, we show that for a large class of optimization problems, the Lagrange multiplier rule can be derived from the so-called approximate multiplier rule. In establishing the link between the approximate and the exact multiplier rule we first derive an approximate multiplier rule for a very general class of optimization problems using the approximate sum rule and the chain rule. We also provide a simple proof to the approximate chain rule based on a fundamental result in parametric optimization. In the end we derive a mixed approximate multiplier rule for an equality and inequality constrained optimization problem and outline an approach to use the mixed approximate multiplier rule in studying the computational aspect associated with such a problem.

Keywords : Lagrange multiplier rule, Locally Lipschitz functions, Subdifferential, composite optimization, nonsmooth calculus.
Mathematics Subject Classification 2000 :90C26,90C30 and 90C46

1 Introduction

Variational analysis is a part of modern analysis and is mainly motivated by optimization-related problems. A large part of the literature on this topics is devoted to the study of the properties of generalized derivatives, their associated subdifferentials and derivatives of set-valued mappings. We refer the

reader to the books by Rockafellar & Wets [19], Mordukhovich [13] [14], Borwein & Zhu [2], Schirotzek [20] and the references therein for a complete and comprehensive expository on variational analysis, as well as for applications of the key issues of this theory in finite and infinite dimension.
The main motivation behind defining subdifferentials for arbitrary functions comes from the convex case. One such subdifferential in the finite dimensional setting is the *regular subdifferential* (Fréchet subdifferential) which is a very intuitive concept in the sense that it is quite natural to define such a notion of a subdifferential when moving away from convexity (details are in Rockafellar & Wets's 1998 book). When we will present the definition of the regular subdifferential in the next section one can see that the regular subgradient is generated from the definition of the subgradient of a convex function by adding some small error term to the right hand side. This fact essentially makes the idea of the regular subdifferential so intuitive. The next natural question is whether the regular subdifferential is nonempty at each point of the effective domain of a proper extended-valued function. The answer is negative even in the case of locally Lipschitz functions. There can be points on the effective domain where the regular subdifferential is empty (see [19] for details). The problem is rectified by defining the notion of a basic subgradient which is the limit of a sequence of regular subgradients. Thus, a more robust subdifferential called the *basic subdifferential* emerged and was first introduced by Mordukhovich [11] in the context of optimal control problems. The basic subdifferential is non-empty and compact in the case of locally Lipschitz functions. Further, the basic subdifferential admits a very nice calculus which unfortunately is not true for the regular subdifferential. Of course, then one may ask why we need to study the regular subdifferential. First of all, let us note that even though the regular subdifferential does not satisfy exact calculus rules like its robust counterpart the basic subdifferential, it admits however an approximate calculus, which in many cases appears to be more easily verifiable than exact calculus rules. These approximate calculus rules lead to approximate Lagrange multiplier rules. We will also try to suggest some usefulness of the approximate multiplier rule. Our principal aim in this article is to develop an approximate chain rule for regular subdifferentials and apply it to derive approximate necessary optimality conditions in terms of the regular subdifferential. Further, we shall also demonstrate using the standard equality and inequality constrained optimization problem, that one can develop a mixed approximate necessary optimality condition which appears to be less restrictive than the

pure approximate multiplier rule but also preserves the advantages of the pure approximate multiplier rule. By a mixed approximate multiplier rule we mean that both the regular subdifferential and the basic subdifferential appear in the expression of the multiplier rule.
This article is organized as follows. In section 2 we shall describe the necessary tools that we require from nonsmooth analysis and we shall also present the approximate sum rule and the approximate chain rule followed by section 3 in which will deal with the application of the chain rule in deriving approximate necessary optimality conditions. Further in this section, we shall also demonstrate that by a sequential application of the approximate Lagrangian multiplier rules one can actually arrive at an exact multiplier rule. This explains why possibly one of the main advantages of using approximate optimality conditions in finite dimensional nonsmooth optimization is to build optimization algorithms using the approximate rules since they have more flexibility built into them. The flexibility of the approximate optimality conditions stems from the fact that in the case of the approximate multiplier rule, the regular subgradients involved are not evaluated exactly at the reference point but at some points in the neigborhood of the reference point. In most practical cases, in the neighborhood of a nonsmooth point one will find that the function is densely differentiable. Thus, the approximate optimality condition in many cases would allow us to work with gradient information, a luxury which we do not have when using the exact nonsmooth Lagrange multiplier rule. Further in section 3 we also develop a sufficient condition under which the set of Lagrange multipliers associated with the problem under consideration is nonempty and bounded.We also provide a result showing that the non-emptiness and the boundedness of the set of Lagrange multipliers guarantee that a certain qualification condition holds under appropriate conditions. In section 4 we study the mixed approximate multiplier rule for equality and inequality constrained problem and also attempt to outline the possible usefulness of studying such a multiplier rule from the computational point of view.

2 Preliminaries

We begin this section by recalling the definition of the Fréchet normal cone (or regular normal cone), the limiting Frchet normal cone or the basic normal cone, the Frchet (or regular) subdifferential and the limiting Frchet subdif-

ferential or the basic subdifferential.

Definition 1 *A vector $v \in \mathbb{R}^n$ is called a Frchet normal to C at $\bar{x}$ if*

$$\limsup_{x \to \bar{x}} \frac{\langle v, x - \bar{x} \rangle}{||x - \bar{x}||} \leq 0.$$

The set of all Fréchet normals form a cone called the Frchet normal cone to C at $\bar{x}$ and is denoted by $N_C^F(\bar{x})$. Also observe that equivalently one can show that a vector $v \in \mathbb{R}^n$ is a Frchet normal to C at $\bar{x}$ if

$$\langle v, x - \bar{x} \rangle \leq o(\|x - \bar{x}\|), \quad \forall x \in C,$$

where $\lim_{x \to \bar{x}} \dfrac{o(\|x - \bar{x}\|)}{\|x - \bar{x}\|} = 0$.

Further we would like to note that the Frchet normal cone is also referred to as the *regular normal cone* in the literature. See for example Chapter 6 in Rockafellar and Wets [19]. It appears that the term regular normal became popular due to its use in Rockafellar and Wets [19]. The Frchet normal cone is a closed and convex object but it suffers from the disadvantage that it can reduce to the trivial cone containing only the zero element, at some points on the set C. For simplicity henceforth we will always refer to the Frchet normal cone as the regular normal cone.
This difficulty is resolved by considering the notion of a limiting Frchet normal cone or the basic normal cone where every vector can be realized as a limit of a sequence of Frchet normals.

Definition 2 *A vector $v \in \mathbb{R}^n$ is an element of the limiting Frchet normal cone or the basic normal cone $N_C(\bar{x})$ to the set C at the point $\bar{x}$ if there exist sequences $\{x_k\}$ with $x_k \in C$ and $\{v_k\}$ with $x_k \to \bar{x}$, $v_k \to \bar{v}$ and $v_k \in N_C^F(x_k)$. In a more compact form this is written as*

$$N_C(\bar{x}) = \limsup_{x \to \bar{x}} N_C^F(x),$$

where as usual the symbol $\limsup$ *denotes the limit superior of sets in the sense of Kuratowski-Painlevé (see for instance the book by G. Beer [1] for the definition).*

We would like to mention that the limiting Frchet normal cone is also known as the Mordukhovich normal cone. However for simplicity in the presentation we would just refer to the limiting Frchet normal cone as the basic normal cone. It is important to note that the basic normal cone is a closed cone but need not be convex. However this cone is never trivial at any point on the boundary of the set C and is stable in the sense that its graph is closed when viewed as a set-valued map. However if $\bar{x}$ is the interior point of C then $N_C(\bar{x}) = \{0\}$. Further the set C is said to be *normally regular* at $\bar{x}$ if $N_C^F(\bar{x}) = N_C(\bar{x})$. For more details on this notion see for example Mordukhovich [13][14]. Associated with the notions of the above two types of normal cones are the Frchet and limiting Frchet or the basic subdifferential which we now define below. However we will first provide the original definition of the Frchet and the basic subdifferential and then present their equivalent representation in terms of the Frchet normal cone and the basic normal cone respectively. Our definition will involve extended-real-valued functions by which we mean functions of the form $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. The set of points where f is finite is denoted by $dom f$.

Definition 3 *Let f be a given extended-real-valued function and let $\bar{x}$ be point where f is finite. Then the Frchet subdifferential (or the regular subdifferential) of f at $\bar{x}$ is given as*

$$\partial^F f(\bar{x}) = \{v \in \mathbb{R}^n : \liminf_{x \to \bar{x}, \quad x \neq \bar{x}} \frac{f(x) - f(\bar{x}) - \langle v, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq 0\}.$$

We would like to mention here that throughout the paper we will use the term regular subdifferential instead of Frchet subdifferential. The term regular subdifferential also seems to have become popular due to its use in Rockafellar and Wets [**?**]. Every element of the regular subdifferential is termed as a regular subgradient. If $\bar{x}$ is a point where $f(\bar{x}) = \infty$ then we set $\partial^F f(\bar{x}) = \emptyset$. In fact one can show that an element v is a regular subgradient of f at $\bar{x}$ iff

$$f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle + o(\|x - \bar{x}\|) \tag{2.1}$$

where $\dfrac{o(\|x - \bar{x}\|)}{\|x - \bar{x}\|} \to 0$ when $x \to \bar{x}$. Observe that (2.1) represents the move away from convexity in an interesting way. One can interpret the term $o(||x - \bar{x}||)$ as an error term which has been added to the right hand side of the subgradient inequality used for convex functions. Equivalently one can

represent the regular subdifferential in terms of the regular normal cone in the following way

$$\partial^F f(\bar{x}) = \{v \in \mathbb{R}^n : (v, -1) \in N^F_{\text{epi f}}(\bar{x}, f(\bar{x}))\}. \tag{2.2}$$

Further it is easy to observe that if f is differentiable then $\partial^F f(x) = \{\nabla f(x)\}$.

Definition 4 *Let f be a given extended-real-valued function and let $\bar{x}$ be a point where f is finite. Then an element $v \in \mathbb{R}^n$ is said to be a limiting Frchet subgradient or the basic subgradient of f at $\bar{x}$ if there exists a sequence $v^k \to v$ and sequence $x^k \to \bar{x}$,with $x^k \in C$, and $f(x^k) \to f(\bar{x})$ such that $v^k \in \partial^F f(x^k)$ for each k. The collection of limiting Frchet sugradients or basic subgradients of f at $\bar{x}$ is denoted by $\partial f(\bar{x})$ and is known as the limiting Frchet subdifferential or the basic subdifferential of f at $\bar{x}$.*

We would however like to mention that for the economy of the presentation we will always refer to the limiting subdifferential as the basic subdifferential. It is also important to note that the basic subdifferential is also known as the Mordukhovich subdifferential. Further noting the representation of the Frchet subdifferential in terms of the Frchet normal cone as given by (2.2) we can immediately write down the following representation of the basic subdifferential in terms of the basic normal cone.

$$\partial f(\bar{x}) = \{v \in \mathbb{R}^n : (v, -1) \in N_{epif}(\bar{x}, f(\bar{x}))\}.$$

For more details and the associated calculus rules of the basic subdifferential see for example [13] or [19]. The basic subdifferential is always non-empty at each point for a locally Lipschitz function and coincides with the subdifferential of a convex function if f is convex. Further it is always closed though need not be convex. For any point $\bar{x}$ where f is finite it is clear that $\partial^F f(\bar{x}) \subseteq \partial f(\bar{x})$. At points where f is not finite one defines the basic subdifferential as the empty set. Further the convex hull of $\partial f(\bar{x})$ denoted by $\text{co}(\partial f(\bar{x}))$ coincides with the Clarke subdifferential $\partial^C f(\bar{x})$ of f at $\bar{x}$ when f is locally lipschitz. See [4] for details on the Clarke subdifferential.

Another important subdifferential notion is that of the asymptotic subdifferential which corresponds to the horizontal component of the basic normal cone to the epigraph of the function under consideration at the point of reference. This is given by the following definition.

Definition 5 *Let f be a given extended-real-valued function and let $\bar{x}$ be point where f is finite. Then the asymptotic subdifferential of f at $\bar{x}$ is given as*

$$\partial^\infty f(\bar{x}) = \{v \in \mathbb{R}^n : (v, 0) \in N_{epif}(\bar{x}, f(\bar{x}))\}.$$

In some sense the asymptotic subdifferential measures how much a function has deviated from being locally Lipschitz since if f is locally Lipschitz at $\bar{x}$ then $\partial^\infty f(\bar{x}) = \{0\}$. In other words it means that the basic normal cone to the epigraph of a locally Lipschitz function never has a non-trivial horizontal component. The asymptotic subdifferential is also known in the literature as the singular subdifferential and was introduced in 1980 by Mordukhovich [12].

Let us now mention some symbols that we will use throughout the paper. First of all we will denote by $\mathbb{B}_X$ the unit ball in the finite dimensional space X. For any $\bar{x} \in X$ we shall denote by $\mathbb{B}_\delta(\bar{x})$ the ball of radius δ with center at $\bar{x}$. Further if $G : \mathbb{R}^n \to \mathbb{R}^m$ is a given function and $y \in \mathbb{R}^m$ then the function $(yG) : \mathbb{R}^n \to \mathbb{R}$ is given by

$$(yG)(x) = \langle y, G(x) \rangle.$$

We will now present the sum rule which will also play an important role in proving the approximate chain rule for regular subdifferentials. We present the result as given in Ngai and Théra [16].

Theorem 1 (Approximate Sum Rule) *Let $f_i : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}, i = 1, \ldots, n$ be lower semicontinuous functions. Let $\bar{x} \in dom f_1 \cap \ldots \cap f_n$. Let*

$$\xi \in \partial^F (f_1 + \cdots + f_n)(\bar{x}).$$

Then for any $\epsilon > 0$ and any neighborhood V of zero in $\mathbb{R}^n$ there exists points

$$(x_i, f_i(x_i)) \in (\bar{x}, f_i(\bar{x})) + \epsilon \mathbb{B}_{\mathbb{R}^n \times \mathbb{R}},$$

such that

$$\xi \in \partial^F f_1(x_1) + \cdots + \partial^F f_n(x_n) + V.$$

Remark 1 In most of the applications one usually sets $V = \epsilon \mathbb{B}_{\mathbb{R}^n}$. Further it is important to note that in Theorem 1 when one changes the vector ξ then the points x_i, $i = 1, \ldots, m$, also get changed in general. It is important to explain at this point why we call this result "the approximate sum rule" instead of the usual "fuzzy sum rule" introduced by Ioffe [8] and established in Asplund spaces by Fabian [7], for Fréchet subdifferentials of lower semicontinuous functions. First of all, let us note that we can write the rule for any choice of $\epsilon > 0$. The other aspect which motivates this terminology is the fact that we move away from the exact point of reference. As we will explain later on, in many applications this fact would actually allow us to work with gradient information though we are in reality dealing with a nonsmooth problem.

Further let us note that in the above results the inclusions at the end of theorem can also be given as

$$x_i \in \bar{x} + \epsilon \mathbb{B}_{\mathbb{R}^n}$$

and

$$f(x_i) \in f(\bar{x}) + \epsilon \mathbb{B}_{\mathbb{R}}.$$

The above relations hold since $\mathbb{B}_{\mathbb{R}^n \times \mathbb{R}} \subset \mathbb{B}_{\mathbb{R}^n} \times \mathbb{B}_{\mathbb{R}}$.

After the approximate sum rule, the next natural question is whether one can have a chain rule for the regular subdifferential. Consider the extended-real-valued function f given as

$$f(x) = g \circ G(x),$$

where g is a lower semicontinuous extended-real-valued mapping and $G : \mathbb{R}^n \to \mathbb{R}^m$ be a smooth mapping. Then it has been shown for example in Rockfaller & Wets [19] that for any $\bar{x} \in dom f$ one has

$$\hat{\partial} f(\bar{x}) \supset \nabla G(\bar{x})^T \hat{\partial} g(G(\bar{x}).$$

However if $\nabla G(\bar{x})$ has full rank, then equality holds in the above expression, that is

$$\hat{\partial} f(\bar{x}) = \nabla G(\bar{x})^T \hat{\partial} g(G(\bar{x})).$$

So what we have is an exact chain rule with the mild condition that the Jacobian of the inner map G at $\bar{x}$ has full rank. The next question obviously is that whether such an exact rule holds if the inner function G is no longer differentiable. It has been shown for example in Theorem 10.49 in Rockafellar & Wets [19] that if the inner function G is locally Lipschitz then one has the following inclusion satisfied,

$$\hat{\partial} f(\bar{x}) \supset \bigcup_{y \in \hat{\partial} g(F(\bar{x}))} \hat{\partial}(yF)(\bar{x}).$$

On the other hand it is important to note that from the point of view of optimization the opposite inclusion is more meaningful. Of course the opposite inclusion holds if we replace the regular subdifferential by the basic subdifferential, but it also important to note that from the computational view point it is simpler to work with a regular subdifferential. So the question now boils down to the following: what sort of condition actually holds for the regular subdifferential of a composite function if we want the reverse inclusion to hold? This leads us to the approximate chain rule which we present below.

Theorem 2 *Let $G : \mathbb{R}^n \to \mathbb{R}^m$ be a vector-valued locally Lipschitz function and $g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. Let $\bar{u} \in \mathbb{R}^n$ be a given point. Define f as follows*

$$f(u) = g \circ G(u).$$

Let $\xi \in \partial^F f(\bar{u})$ where $G(\bar{u}) \in domg$. Then, for any $\epsilon > 0$ there exist vectors $\eta \in \mathbb{R}^m$, $u_1 \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ with

$$(u_1, G(u_1)) \in (\bar{u}, G(\bar{u})) + \epsilon \mathbb{B}_{\mathbb{R}^n \times \mathbb{R}^m},$$

$$\begin{aligned} y &\in G(\bar{u}) + \epsilon \mathbb{B}_{\mathbb{R}^m}, \\ g(y) &\in g(G(\bar{u})) + \epsilon \mathbb{B}_{\mathbb{R}} \end{aligned}$$

and

$$\eta \in \partial^F g(y) + \epsilon \mathbb{B}_{\mathbb{R}^m}$$

such that

$$\xi \in \partial^F (\eta G)(u_1) + \epsilon \mathbb{B}_{\mathbb{R}^n}.$$

Proof : Reminding that as customary the indicator function of a set C is the function defined by $\delta_C(x) = 0$ if $x \in C$, and $\delta_C(x) = +\infty$ otherwise, we set $F(x,u) = g(x) + \delta_{\mathrm{Gr}\ G}(u,x)$, where Gr G is the graph of the mapping G. We observe that F as defined is a proper lower semicontinuous extended-real-valued function which satisfies

$$f(u) = g \circ G(u) = \min_x F(x,u).$$

Observe that for a fixed u the value of $F(x,u)$ is either $g(G(u))$ or infinity and thus $domF$ is exactly $(G(u),u)$. Hence the minimum value of $F(x,u)$ is $g(G(u))$. Let consider $\xi \in \partial^F f(\bar{u})$. It is simple to observe that $(0,\xi) \in \partial^F F(G(\bar{u}),\bar{u})$. Observe that $\xi \in \partial^F f(\bar{u})$ implies that for any $u \in \mathbb{R}^n$

$$f(u) - f(\bar{u}) \geq \langle \xi, u - \bar{u} \rangle + o(||u - \bar{u}||). \tag{2.3}$$

Noting that $f(u) \leq F(x,u)$ for all $x \in \mathbb{R}^m$ and the fact that $f(\bar{u}) = F(G(\bar{u}),\bar{u})$ we have from (2.3)

$$F(x,u) - F(G(\bar{u}),\bar{u}) \geq \langle (0,\xi), (x,u) - (G(\bar{u}),\bar{u}) \rangle + o(||u - \bar{u}||) - ||x - G(\bar{u})||^2.$$

Hence

$$F(x,u) - F(G(\bar{u}),\bar{u}) \geq \langle (0,\xi), (x,u) - (G(\bar{u}),\bar{u}) \rangle + o(||(x - G(\bar{u}), u - \bar{u})||).$$

Observing that $F(x,u) = g(x) + \phi(x,u)$ with $\phi(x,u) = \delta_{\mathrm{Gr}\ G}(u,x)$, we can write

$$\partial^F(G(\bar{u}),\bar{u}) = \partial^F(g + \phi)(G(\bar{u}),\bar{u}).$$

By applying the approximate sum rule (Theorem 1) we have for any $\varepsilon > 0$ the existence of $u_1 \in \mathbb{R}^n$, $w_1 \in \mathbb{R}^m$ and $y \in \mathbb{R}^m$ such that

$$(0,\xi) \in \partial^F g(y) \times \{0\} + \partial^F \phi(w_1,u_1) + \epsilon(\mathbb{B}_{\mathbb{R}^m} \times \mathbb{B}_{\mathbb{R}^n}), \tag{2.4}$$

where

$$\begin{aligned} y &\in G(\bar{u}) + \epsilon \mathbb{B}_{\mathbb{R}^m}, \\ g(y) &\in g(G(\bar{u})) + \epsilon \mathbb{B}_{\mathbb{R}} \end{aligned}$$

and

$$(w_1,u_1) \in (G(\bar{u}),\bar{u}) + \epsilon \mathbb{B}_{\mathbb{R}^m \times \mathbb{R}^n}.$$

However it is important to note that (2.4) is valid only if all the sets on the right hand side of (2.4) are non-empty. This is true only if $w_1 = G(u_1)$ since $\phi(x,u)$ is only finite for points of the type $(G(u),u)$. Let us recall that by definition, the regular subdifferential is an empty set for all points not in the effective domain. Hence we have

$$(G(u_1),u_1) \in (G(\bar{u}),\bar{u}) + \epsilon\mathbb{B}_{\mathbb{R}^m\times\mathbb{R}^n}.$$

Thus we have

$$(0,\xi) \in (\gamma,0) + (-\eta,\xi_1) + \epsilon(\mathbb{B}_{\mathbb{R}^m} \times \mathbb{B}_{\mathbb{R}^n}),$$

where

$$(\gamma,0) \in \partial^F g(y) \times \{0\}$$

and

$$(\xi_1,-\eta) \in \quad N^F_{\mathrm{Gr}\ G}(u_1,G(u_1)) = \partial^F \delta_{\mathrm{Gr}\ G}(u_1,G(u_1)).$$

Now by using a coordinate-wise representation we have,

$$\xi \in \xi_1 + \epsilon\mathbb{B}_{\mathbb{R}^n} \tag{2.5}$$

and

$$\eta \in \gamma + \epsilon\mathbb{B}_{\mathbb{R}^m} \subset \partial^F g(y) + \epsilon\mathbb{B}_{\mathbb{R}^m}.$$

Since $(\xi_1,-\eta) \in N^F_{\mathrm{Gr}\ G}(u_1,G(u_1))$, from the definition of the regular normal cone one has one has for all $(u,G(u)) \in \mathrm{Gr}\ G$

$$\langle(\xi_1,-\eta),(u-u_1,G(u)-G(u_1))\rangle \leq o(||(u-u_1,G(u)-G(u_1))||).$$

Since G is locally Lipschitz one has

$$\langle(\xi_1,-\eta),(u-u_1,G(u)-G(u_1))\rangle \leq o(||u-u_1||).$$

Hence we deduce that

$$(\eta G)(u) - (\eta G)(u_1) \geq \langle\xi_1,u-u_1\rangle + o(||u-u_1||).$$

Therefore, we have $\xi_1 \in \partial^F(\eta G)(u_1)$. Now by applying (2.5) we have

$$\xi \in \partial^F(\eta G)(u_1) + \epsilon\mathbb{B}_{\mathbb{R}^n},$$

and the proof is complete. □

Remark 2 The chain rule for regular subdifferentials have been also studied in infinite dimension. In this paper we simply want to present an elementary proof of the approximate chain rule in finite dimension. The key to the proof is to start with the construction of a simple parametric optimization problem and then use the approximate sum rule to conclude. For details of the proof of the approximate chain rule for Fréchet subdifferentials in Fréchet smooth Banach spaces, we refer to Borwein & Zhu [2]. The proof in the setting of Asplund spaces is given in Ngai & Théra [16]. In the recent past an important contribution to the literature of approximate or fuzzy calculus rules is due to Mordukhovich and Shao [15]. In [15] these authors developed approximate or fuzzy calculus rules for Fréchet coderivatives and ε-Fréchet coderivatives. Using these rules they also established approximate sum rule and approximate chain rule for the Fréchet derivative. If we only focus on subdifferentials, then a central theme of their work is to estimate each element of the ε-Fréchet subdifferential of the sum of two lower-semicontinuous functions or composition of two functions through an approximate or fuzzy sum rule and chain rule respectively. It is further important to note that their results are based on certain fuzzy qualification conditions which are new even in finite dimensions. Since the Fréchet subdifferential is a subset of the ε-Fréchet subdifferential for any $\varepsilon > 0$, the usual approximate or fuzzy calculus rules for the Fréchet subdifferential naturally follows. In fact it appears that the results in Mordukhovich and Shao [15] definitely play an important role in the study of ε-minimization.

3 An approximate Multiplier Rule

In this section we will focus our attention on the following problem $(\mathcal{P}_2)$ which is given as follows

$$\min f_0(x) + \rho(F(x)), \quad \text{subject to} \quad x \in X.$$

where $\rho : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is a proper lower semicontinuous function, $f_0 : \mathbb{R}^n \to \mathbb{R}$ is a locally Lipschitz function, $F : \mathbb{R}^n \to \mathbb{R}^m$ is a vector-valued locally Lipschitz function and X is closed subset of $\mathbb{R}^n$. Our aim is to develop an approximate Largange multiplier rule for the problem $(\mathcal{P}_2)$ and also demonstrate that applying the approximate multiplier rule sequentially one can arrive at the exact Lagrange multiplier rule. This may open up an way to use the approximate optimality conditions in devising algorithms

for nonsmooth optimization. Further we will also see by their very nature the approximate optimality conditions will allow us to use the derivative information of points in the neighborhood of a reference point.

This class of problems has been studied in Rockafellar [17], [18], Rockafellar & Wets [19] and recently by Dutta [6]. Let us notice that Problem $(\mathcal{P}_2)$ covers a large class of optimization problems depending on the particular form of the function ρ. For example, if we consider the problem of minimizing f_0 over the set $C := \{x \in X : F(x) \in U\}$, where $U \subset \mathbb{R}^m$, then by choosing $\rho = \delta_U$ we observe that this problem belongs to the class of problems defined by $(\mathcal{P}_2)$. Thus by choosing an appropriate ρ, we can get various different classes of optimization problems. For example, if we choose $f_0(x) = 0$ for all x and ρ to be the *max function*, then we get the classical minmax problem. Thus problem $(\mathcal{P}_2)$ covers a very large class of optimization problems. Another important class of optimization problem which is not included among the classical optimization problems is the *extended linear-quadratic programming problem* given in Rockafellar and Wets [19]. The extended linear-quadratic programming problem is given by

$$\min\langle c, x\rangle + \frac{1}{2}\langle x, Cx\rangle + \theta_{Y,B}(b - Ax), \quad \text{subject to} \quad x \in X,$$

where C is a $n \times n$ symmetric positive semidefinite matrix and B is a $m \times m$ symmetric positive semidefinite matrix, A is a $m \times n$ matrix , $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$. The set $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ are non-empty and polyhedral and the function $\theta_{Y,B}$ is given as

$$\theta_{Y,B}(u) = \sup_{y \in Y} \left\{ \langle y, u\rangle - \frac{1}{2}\langle y, By\rangle \right\}.$$

For more details on the extended linear-quadratic programming problem see Rockafellar and Wets [19]. A good source for a very detailed study of optimality conditions for various classes of optimization problems including mathematical programming under equilibrium constraints and multiobjective optimization, is Chapter 5 in Mordukhovich [14]. For example in Theorem 5.28 in [14], Mordukhovich studies the approximate optimality conditions in terms of Frchet subdifferential for equality and inequality constrained mathematical programming problems with lower-semicontinuous objective and constraint functions. This is however beyond the framework we study here which heavily depends on the Lipschitz assumption. Though problem $(\mathcal{P}_2)$ is not explicitly

mentioned in Chapter 5 of Mordukhovich [14] but the exact necessary optimality condition for any problem in the format of $(\mathcal{P}_2)$ can be obtained usin g the methods laid out in Mordukhovich [13] and Rockafellar and Wets [19]. The proof methodology in order to derive necessary optimality conditions in Mordukhovich largely relies on the extremal principle (see for example Theorem 5.21 in [14]) while in our case we depend on the approximate optimality conditions. Thus in the approach of Mordukhovich the extremal principle is the unifying theme while we want to present a unified approach to a large class of optimization problems through approximate optimality conditions. We essentially show that for problems at least in the format of $(\mathcal{P}_2)$ with f_0 and F locally Lipschitz and ρ lower-semicontinuous and extended-valued a the gap between the exact optimality condition and the approximate optimality condition can be closed and thus hinting towards a possible use of the approximate necessary optimality conditions to devise algorithms for nonsmooth optimzation problems, as we had mentio ned above. Hence the approximate optimality conditions act as a unifying theme in our approach to the derivation of necessary optimality condition for a large class of optimization problems. Our approach is essentially finite dimensional as we heavily depend on properties of sequences in finite dimension. On the other Mordukhovich [13] and [14] studies the problems in infinite dimensions. Thus it would be interesting to see whether the approach that we take through approximate optimality conditions in finite dimensions can be extended to infinite dimensions or not.

Theorem 3 *Let us consider problem $(\mathcal{P}_2)$ with the assumptions on the functions as stated above. Suppose $\bar{x}$ is a local minimizer of $(\mathcal{P}_2)$. Then for any $\epsilon > 0$, there exist $\eta \in \mathbb{R}^m$ and $x_1, x_2, x_3, x_4 \in \mathbb{R}^n$ satisfying the following conditions:*

1. $0 \in \hat{\partial} f_0(x_1) + \hat{\partial}(\eta F)(x_3) + N_X^F(x_4) + \epsilon \mathbb{B}_{\mathbb{R}^n}$;
2. $(x_2, \rho(F(x_2)) \in (\bar{x}, \rho(F(\bar{x}))) + \epsilon \mathbb{B}_{\mathbb{R}^n \times \mathbb{R}}$;
3. $(x_3, F(x_3)) \in (x_2, F(x_2)) + \epsilon \mathbb{B}_{\mathbb{R}^n \times \mathbb{R}^m}$
4. $(x_1, f_0(x_1)) \in (\bar{x}, f_0(\bar{x})) + \epsilon \mathbb{B}_{\mathbb{R}^n \times \mathbb{R}}$;
5. $x_4 \in \bar{x} + \epsilon \mathbb{B}_{\mathbb{R}^n}$;
6. $\eta \in \hat{\partial} \rho(y) + \epsilon \mathbb{B}_{\mathbb{R}^m}$;

7. $y \in F(x_2) + \epsilon \mathbb{B}_{\mathbb{R}^m}$;

8. $\rho(y) \in \rho(F(x_2)) + \epsilon \mathbb{B}_{\mathbb{R}}$.

Proof : Since $\bar{x}$ is a local minimizer of the problem we have

$$0 \in \hat{\partial}(f_0 + \rho \circ F + \delta_X)(\bar{x}). \tag{3.1}$$

Then by applying the approximate sum rule we get for any $\epsilon > 0$

$$0 \in \hat{\partial} f_0(x_1) + \hat{\partial}(\rho \circ F)(x_2) + N_X^F(x_4) + \frac{\epsilon}{2}\mathbb{B}_{\mathbb{R}^n}, \tag{3.2}$$

where

$$\begin{aligned} (x_1, f_0(x_1)) &\in (\bar{x}, f_0(\bar{x})) + \epsilon \mathbb{B}_{\mathbb{R}^n \times \mathbb{R}} \\ (x_2, \rho(F(x_2))) &\in (\bar{x}, \rho(F(\bar{x}))) + \epsilon \mathbb{B}_{\mathbb{R}^n \times \mathbb{R}} \\ x_4 &\in \bar{x} + \epsilon \mathbb{B}_{\mathbb{R}^n}. \end{aligned}$$

By applying the approximate chain rule in (3.2) we get

$$\begin{aligned} 0 &\in \hat{\partial} f_0(x_1) + \hat{\partial}(\eta F)(x_3) + N_X^F(x_4) + \frac{\epsilon}{2}\mathbb{B}_{\mathbb{R}^n} + \frac{\epsilon}{2}\mathbb{B}_{\mathbb{R}^n} \\ 0 &\in \hat{\partial} f_0(x_1) + \hat{\partial}(\eta F)(x_3) + N_X^F(x_4) + \epsilon \mathbb{B}_n, \end{aligned}$$

where

$$\begin{aligned} (x_3, F(x_3)) &\in (x_2, F(x_2)) + \epsilon \mathbb{B}_{\mathbb{R}^n \times \mathbb{R}^m} \\ \eta &\in \hat{\partial}\rho(y) + \epsilon \mathbb{B}_{\mathbb{R}^m} \\ y &\in F(x_2) + \epsilon \mathbb{B}_{\mathbb{R}^m} \\ \rho(y) &\in \rho(F(x_2)) + \epsilon \mathbb{B}_{\mathbb{R}}. \end{aligned}$$

This proves the result . □

An important issue that one faces at this point is what can be the possible use of the approximate multiplier rule. In nonsmooth optimization one of the major bottlenecks is the computation of a subgradient element of a given function at a given reference point. This difficulty gets more acute, when the function is nonconvex. On the other hand the approximate multiplier rule has a built in flexibility where we look at the regular subdifferential at a point in the neighborhood of the reference point but not exactly at the

reference point. Since in the above theorem the functions f and F are locally Lipschitz, by the virtue of the Rademacher theorem we know that they are densely differentiable in the sense that the set of points where the functions are not differentiable forms a set of measure zero. Thus once we consider an $\epsilon > 0$, then in the ball $B_\epsilon(\bar{x})$ the functions f and F are densely differentiable. So for the given $\epsilon > 0$, if we intend to check whether the point $\bar{x}$ satisfies the approximate optimality conditions, we might be able to do so by choosing points where f and F are differentiable. This is a huge advantage since the regular subdifferential reduces to the gradient at the points of differentiability of the function. Let us explain this in a slightly more detailed fashion.
Just for simplicity, consider that for problem $(\mathcal{P}_2)$ we have $X = \mathbb{R}^n$ and ρ is a finite-valued function. Now, let us consider that we are given a point $\bar{x} \in \mathbb{R}^n$ and we would like to know whether it is an approximate critical point of $(\mathcal{P}_2)$ or not. In other words we would like to know if corresponding to every $\epsilon > 0$ we are able to find points x_1, x_2, x_3 and x_4 in $\mathbb{R}^n$ and vectors η and y in $\mathbb{R}^m$ such that the conditions (1)-(8) are satisfied.
Though this looks formidable, the computational load can be reduced if one uses the random sampling technique used in Burke, Lewis and Overton [3]. For a locally Lipschitz function given a point and a ball around it of a certain radius which is termed as the sampling radius, [3] showed that one can actually generate a random sample of points which avoids a fixed set of measure zero almost surely. Now consider the ball $\mathbb{B}_{\epsilon_0}(\bar{x})$ and $\epsilon_0 > 0$ is small. Here ϵ_0 is called the sampling radius. Since the points of non-differentiability of Lipschitz functions forms a set of measure zero we can first begin by generating a sample of say k points corresponding to each of the functions f and F where these functions are differentiable. This means to begin with we consider two samples of size k each. Let for f the sample be

$$\mathcal{S}_f = \{x_1^1, x_1^2, \ldots, x_1^k\}$$

and for the function F, let the sample be

$$\mathcal{S}_F = \{x_3^1, x_3^2, \ldots, x_3^k\}.$$

Thus in our simplified setting we first begin by considering the pair $\{x_1^1, x_3^1\}$. For this pair we try to find points x_1^2, η and y^1 such that

1. $||\nabla f(x_1^1) + \nabla(\eta^1 F)(x_3^1)|| \leq \epsilon_0$;

2. $(x_2^1, \rho(F(x_2^1)) \in (\bar{x}, \rho(F(\bar{x}))) + \epsilon_0 \mathbb{B}_{\mathbb{R}^n \times \mathbb{R}}$;

3. $(x_3^1, F(x_3^1)) \in (x_2^1, F(x_2^1)) + \epsilon_0 \mathbb{B}_{\mathbb{R}^n \times \mathbb{R}^m}$;

4. $(x_1^1, f_0(x_1^1)) \in (\bar{x}, f_0(\bar{x})) + \epsilon_0 \mathbb{B}_{\mathbb{R}^n \times \mathbb{R}}$;

5. $\eta^1 \in \partial^F \rho(y^1) + \epsilon_0 \mathbb{B}_{\mathbb{R}^m}$;

6. $y^1 \in F(x_2^1) + \epsilon_0 \mathbb{B}_{\mathbb{R}^m}$;

7. $\rho(y^1) \in \rho(F(x_2^1)) + \epsilon_0 \mathbb{B}_{\mathbb{R}}$.

Let us note that though we had started with the points $\{x_1^1, x_3^1\}$, we could in fact randomly construct a pair by choosing one from each of the sets. For example one could choose $\{x_1^2, x_3^k\}$ to start with. The crucial part is that corresponding to the pair $\{x_1^1, x_3^1\}$, (say) one has to find points x_2^1, η and y^1 such that

$$\eta^1 \in \hat{\partial} \rho(y^1) + \epsilon_0 \mathbb{B}_{\mathbb{R}^m}$$

and

$$y^1 \in F(x_2^1) + \epsilon_0 \mathbb{B}_{\mathbb{R}^m}.$$

So again we come to a point where we need to compute the regular subgradient of the function ρ at y^1. Now in many optimization problems the function ρ is convex or locally Lipschitz. Let us now consider ρ to be locally Lipschitz. Further consider the open ball $\mathbb{B}_{\epsilon_0}(F(x_2^1))$. Thus by the virtue of Rademacher's theorem ρ is densely differentiable on $\mathbb{B}_{\epsilon_0}(F(x_2^1))$. Now by using the approach in [3], let us consider random sample of r points

$$\mathcal{S}_\rho = \{w_2^1, w_2^2, \ldots, w_2^r\},$$

at which ρ is differentiable. Then one can estimate η^1 and y^1 by choosing a η^1 which satisfies

$$\eta^1 \in \nabla \rho(w_2^i) + \epsilon_0 \mathbb{B}_{\mathbb{R}^m} \tag{3.3}$$

for some $w_2^i \in \mathcal{S}_\rho$ and also satisfying

$$||\nabla f(x_1^1) + \nabla(\eta^1 F)(x_3^1)|| \leq \epsilon_0.$$

Thus y^1 can be chosen as $w_2^i \in \mathcal{S}_\rho$ which satisfies (3.3). Now if corresponding to $\mathcal{S}_\rho$ we do not find an η we reject the pair $\{x_1^1, x_3^1\}$ and choose another pair

from $\mathcal{S}_f$ and $\mathcal{S}_F$ and repeat the procedure. If for given pair we are able to find η and y then we verify the conditions (2), (3), (4) and (7). Further if all k^2 possible pairs chosen from $\mathcal{S}_f$ and $\mathcal{S}_F$ fail to satisfy (1) to (7) as above, then we can declare that the point $\bar{x}$ is not approximately critical and move to a better point through some iterative techniques.
However we would like to mention that this procedure is computationally intensive even though we are just using gradient information. We shall show in section 4 that one can develop a different approach to the approximate optimality conditions which is much less computationally intensive and has much less conditions to verify.

Lemma 1 *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz vector-valued funcion. Let the sequence $\{x_k\}$ converge to $\bar{x}$ and let $\{\eta_k\}$ be a sequence converging to η. Further let $\xi_k \in \partial^F(\eta_k F)(x_k)$. Then the sequence ξ_k is bounded and every cluster point of ξ_k is in $\partial(\eta F)(\bar{x})$.*

Proof : Since $\partial^F(\eta_k F)(x_k) \subseteq \partial(\eta_k F)(x_k)$ we have

$$\eta_k \in \partial(\eta F)(x_k) + \partial((\eta_k - \eta)F)(x_k).$$

Thus there exists $w_k \in \partial((\eta_k - \eta)F)(x_k)$ and $v_k \in \partial(\eta F)(x_k)$ such that

$$\xi_k = v_k + w_k.$$

Using for example Lemma 2 in Jourani and Thibault [10] we conclude that $w_k \to 0$. Further since F is locally Lipschitz the subdifferential map $\partial(\eta F)$ is locally bounded and hence the sequence $\{v_k\}$ is bounded and thus it is clear that the sequence $\{\xi_k\}$ is bounded. Further since the subdifferential map $\partial(\eta F)$ is graph closed it is clear that every cluster point of ξ_k is in $\partial(\eta F)(\bar{x})$. $\square$.
For a more general version of the above lemma in infinite dimensions see Lemma 3.27 in Mordukhovich [13].

Let us consider problem $(\mathcal{P}_2)$ and let us mention the exact multiplier rule for problem $(\mathcal{P}_2)$ via the following theorem.

Theorem 4 *Suppose $\bar{x}$ is a local minimizer of problem $(\mathcal{P}_2)$. Further assume that the following qualification condition* **(CQ)** *holds at $\bar{x}$:*

$$y \in \partial^\infty \rho(F(\bar{x})) \quad \textit{with} \quad 0 \in \partial(yF)(\bar{x}) + N_X(\bar{x}) \quad \textit{implies that} \quad y = 0.$$

Then there exists $\eta \in \partial\rho(F(\bar{x}))$ such that

$$0 \in \partial f_0(\bar{x}) + \partial(\eta F)(\bar{x}) + N_X(\bar{x}). \tag{3.4}$$

For a proof of the above result see Dutta [6] and Vinter [21] and for the case when ρ is additionally convex see Rockafellar & Wets [19].

Theorem 5 *Let us consider problem $(\mathcal{P}_2)$ as given in the beginning of this section. Let $\bar{x}$ be a local minimum of $(\mathcal{P}_2)$. Assume that the qualification condition* **(CQ)** *as given in Theorem 4 holds at $\bar{x}$. Then using the approximate optimality conditions in Theorem 3 one can arrive at the exact necessary optimality conditions (3.4) in Theorem 4.*

Proof : Using Theorem 3 we can show that for every natural number $\varepsilon = \frac{1}{k}$ one can find points x_{ik}, $i = 1, \ldots, 4$ converging to $\bar{x}$ with $\rho(F(x_{2k}) \to \rho(F(\bar{x}))$, $y_k \in F(x_{2k}) + \frac{1}{k}\mathbb{B}_{\mathbb{R}^m}$ and $\eta_k \in\in \partial^F \rho(y_k) + \frac{1}{k}\mathbb{B}_{\mathbb{R}^m}$, with $\rho(y_k) \in \rho(F(x_{2k}) + \frac{1}{k}\mathbb{B}_{\mathbb{R}^m}$, $\xi_{1k} \in \partial^F f_0(x_{1k})$, $\xi_{3k} \in \partial^F(\eta_k F)(x_{3k})$, $\xi_{4k} N_X^F(x_{4k})$ and $b_k \in \mathbb{B}_{\mathbb{R}^n}$ such that

$$0 = \xi_{1k} + \xi_{3k} + \xi_{4k} + \frac{1}{k} b_k. \tag{3.5}$$

Since the function f_0 is locally Lipschitz, one may assume without loss of generality the sequence $\{\xi_{1k}\}$ converges to some $\xi_1 \in \partial f_0(\bar{x})$. To begin with let us suppose that the sequence $\{\eta_k\}$ is bounded we can assume that $\{\eta_k\}$ converges to η.Then we can use Lemma 1 to conclude without loss of generality that $\{\xi_{3k}\}$ converges to some $\xi_3 \in \partial(\eta F)(\bar{x})$. Hence using (3.5) we can conclude that the sequence $\{\xi_{4k}\}$ converges to some $\xi_4 \in N_X(\bar{x})$. Hence we deduce that

$$0 = \xi_1 + \xi_3 + \xi_4$$

as required. We also need to show that $\eta \in \partial\rho(F(\bar{x}))$. Using the fact that $y_k \in F(x_{2k}) + \frac{1}{k}\mathbb{B}_{\mathbb{R}^m}$ we obtain that $y_k \to F(\bar{x})$ Further using the fact that $\rho(y_k) \in \rho(F(x_{2k}) + \frac{1}{k}\mathbb{B}_{\mathbb{R}^m}$ we obtain that $\rho(y_k) \to \rho(F(\bar{x}))$. Hence we conclude that $\eta \in \partial\rho(F(\bar{x}))$ using the fact that $\eta_k \in\in \partial^F \rho(y_k) + \frac{1}{k}\mathbb{B}_{\mathbb{R}^m}$ by passing to the limit as $k \to \infty$.

Now let us consider when $\{\eta_k\}$ is an unbounded sequence. Then we may assume that the sequence $\{\frac{\eta_k}{||\eta_k||}\}$ converges to some v with $||v|| = 1$. Further using the fact the regular subdifferential is a subset of the basic subdifferential we can use (3.5) to conclude that

$$0 \in \partial f_0(x_{1k}) + \partial(\eta_k F)(x_{3k}) + N_X(x_{4k}) + \frac{1}{k}\mathbb{B}_{\mathbb{R}^n}.$$

Thus we have

$$0 \in \frac{1}{\| \eta_k \|}\partial f_0(x_{1k}) + \frac{1}{\| \eta_k \|}\partial(\eta_k F)(x_{3k}) + \frac{1}{\| \eta_k \|}N_X(x_{4k}) + \frac{1}{k}\frac{1}{\| \eta_k \|}\mathbb{B}_{\mathbb{R}^n}. \tag{3.6}$$

By Lemma 1 we can assume that the sequence $\theta_{3k} \in \partial(\frac{\eta_l}{||\eta_k||}F)(x_{3k})$ converges to some $\theta \in \partial(vF)(\bar{x})$. Thus from (3.6) it is clear that $-\theta \in N_X(\bar{x})$. We will violate the constraint qualification **(CQ)** if we prove that $v \in \partial^\infty\rho(F(\bar{x}))$. Now we shall show that indeed $v \in \partial^\infty\rho(F(\bar{x}))$.
Observe that

$$\eta_k \in \hat{\partial}\rho(y_k) + \frac{1}{k}\mathbb{B}_{\mathbb{R}^m}.$$

By the definition of regular normal vector we have

$$(\eta_k, -1) \in N^F_{epi(\rho)}(y_k, \rho(y_k)) + \frac{1}{k}\mathbb{B}_{\mathbb{R}^{m+1}}.$$

This shows that

$$\left(\frac{\eta_k}{\| \eta_k \|}, \frac{-1}{\| \eta_k \|}\right) \in N^F_{epi(\rho)}(y_k, \rho(y_k)) + \frac{1}{k}\frac{1}{\| \eta_k \|}\mathbb{B}_{\mathbb{R}^{m+1}}.$$

Noticing that $y_k \to F(\bar{x})$ and $\rho(y_k) \to \rho(F(\bar{x}))$, in the limit one has

$$(v, 0) \in N_{epi(\rho)}(F(\bar{x}), \rho(F(\bar{x}))).$$

This implies that $v \in \partial^\infty\rho(F(\bar{x}))$. □

Remark 3 The above theorem essentially shows that using the approximate necessary optimality condition in a sequential manner one finally converges to the exact necessary optimality condition for problem $(\mathcal{P}_2)$ and thus for a large class of optimization problems. After the proof of Theorem 3 we have attempted to outline a possible use of the approximate necessary optimality conditions to develop an algorithm for nonsmooth optimization problems. The above theorem may be considered as an attempt towards a possible justification to develop nonsmooth optimization algorithms based on the approximate multiplier rule.

If the point $\bar{x}$ is local minimum of the problem $(\mathcal{P}_2)$ then the set of Lagrange multipliers associated with $\bar{x}$ is given as follows,

$$\Lambda(\bar{x}) = \{\eta \in \partial\rho(F(\bar{x})) : 0 \in \partial f_0(\bar{x}) + \partial(\eta F)(\bar{x}) + N_X(\bar{x})\}.$$

An interesting question to ask here is under what condition the above set is bounded. Since $(\mathcal{P}_2)$ covers a large class of problems we can obtain a much general result on the boundedness of the set of Lagrange multipliers by proceeding along the lines of Theorem 3.2 in Jourani [9].

Theorem 6 *Let us consider the problem* $(\mathcal{P}_2)$*. Let* $\bar{x}$ *be a local minimum of* $(\mathcal{P}_2)$*. Let us assume that the qualification condition* **(CQ)** *holds at* $\bar{x}$*. Then* $\Lambda(\bar{x})$ *is a non-empty compact set.*

Proof : The non-emptiness of the $\Lambda(\bar{x})$ is clear from Theorem 4. Further proceeding along the lines of the last part of the proof in Theorem 5 we conclude that if $\Lambda(\bar{x})$ is not bounded then **(CQ)** is violated. □

Following the lines of Theorem 3.2 in Journai [9] it might be tempting to think that if F is smooth and X is normally regular and the set $\Lambda(\bar{x})$ is non-empty and bounded then **(CQ)** is satisfied at $\bar{x}$. However this may not work in general since in order to prove this fact we need to know whether $\partial\rho(F(\bar{x})) + \partial^\infty\rho(F(\bar{x})) \subset \partial\rho(F(\bar{x}))$. This is not true in general since the basic normal cone is not convex in general. Thus in order to develop some sort of converse to the above result we need to introduce certain notions of subdifferential and normal cone introduced by Clarke [4] (see also Clarke [5]). The closed convex hull of the basic normal cone to the set X at $\bar{x}$ is denoted by $N_X^{cl}(\bar{x})$. For the lower-semicontinuous function ρ thus one may define the Clarke subdifferential $\partial^\circ\rho(\bar{x})$ as follows,

$$\partial^\circ\rho(\bar{x}) = \{v \in \mathbb{R}^n : (v, -1) \in N^{cl}_{epi\rho}(\bar{x}, \rho(\bar{x}))\}$$

It has been shown for example in Clarke [5](Proposition 1.2) that

$$\partial^\circ\rho(\bar{x}) = \text{clco}\{\partial\rho(F(\bar{x})) + \partial^\infty\rho(\bar{x})\},$$

where clco denotes the closed convex hull of a set. We introduce the following set

$$\Lambda^\circ(\bar{x}) = \{\eta \in \partial^\circ\rho(F(\bar{x})) : 0 \in \partial f_0(\bar{x}) + \partial(\eta F)(\bar{x}) + N_X(\bar{x})\}.$$

Since $(0, 0) \in N_{epi\rho}(F(\bar{x}), \rho(F(\bar{x}))$ it is clear that $0 \in \partial^\infty\rho(F(\bar{x}))$ and hence $\Lambda(\bar{x}) \subseteq \Lambda(\bar{x})$. This fact allows us to have the following proposition.

Proposition 1 *Let* $\bar{x}$ *be a local minimum of* $(\mathcal{P}_2)$*. Assume that* $\Lambda(\bar{x})$ *is non-empty and* $\Lambda^\circ(\bar{x})$ *is bounded. Further assume that* F *is a smooth function and* X *is normally regular at* $\bar{x}$*. Then* **(CQ)** *holds at* $\bar{x}$*.*

Proof : On the contrary let us begin by assuming that **(CQ)** does not hold at $\bar{x}$. This implies that there exists $0 \neq \eta \in \partial^{\infty}\rho(F(\bar{x}))$ such that

$$0 \in \nabla F(\bar{x})^T \eta + N_X(\bar{x}), \tag{3.7}$$

where $\nabla F(\bar{x})$ denotes the Jacobian of F at $\bar{x}$ and T denotes the transpose of a matrix. Further since for any $\mu > 0$ we have $\mu\eta \in \partial^{\infty}\rho(F(\bar{x}))$, from (3.7) we have

$$0 \in \nabla F(\bar{x})^T (\mu\eta) + N_X(\bar{x}). \tag{3.8}$$

Since $\Lambda(\bar{x})$ is non-empty there exists $\lambda \in \Lambda(\bar{x})$ such that

$$0 \in \partial f_0(\bar{x}) + \nabla F(\bar{x})^T (\lambda) + N_X(\bar{x}). \tag{3.9}$$

Adding (3.8) and (3.9) and noting that X is normally regular at $\bar{x}$ we have

$$0 \in \partial f_0(\bar{x}) + \nabla F(\bar{x})^T (\lambda + \mu\eta) + N_X(\bar{x}).$$

It is clear that $\lambda + \mu\eta \in \partial^{\circ}\rho(F(\bar{x}))$. Hence we have from the above inclusion $\lambda + \mu\eta \in \Lambda^{\circ}(\bar{x})$. Further it is simple to observe that by making $\mu \to \infty$ we have $||\lambda + \mu\eta|| \to \infty$. This contradicts the fact that $\Lambda^{\circ}(\bar{x})$ is bounded. □

It is important to note that if $epi\rho$ is normally regular at $(F(\bar{x})), \rho(F(\bar{x}))$ then the relation $\partial\rho(F(\bar{x})) + \partial^{\infty}\rho(F(\bar{x})) \subset \partial\rho(F(\bar{x}))$ will hold. Thus in such a case when F is smooth and X is normally regular at $\bar{x}$ then the non-emptiness and boundeness of $\Lambda(\bar{x})$ is equivalent to the fact that **(CQ)** holds at $\bar{x}$.

4 A Mixed Approximate Multiplier Rule

In this section we will focus our attention on the standard equality and inequality constrained problem $(\mathcal{P})$ given below:

$$\begin{aligned} &\min f(x) \\ &\text{subject to} \\ &\quad g_i(x) \leq 0, \quad i = 1, \ldots, m \\ &\quad h_j(x) = 0, \quad j = 1, \ldots, k \end{aligned}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g_i : \mathbb{R}^n \to \mathbb{R}$, and $h_j : \mathbb{R}^n \to \mathbb{R}$ are locally Lipschitz functions. Our aim is to show that using the approximate sum rule (Theorem 1) one can actually design a very flexible necessary optimality condition. In the following theorem we present a mixed optimality condition involving the regular subdifferential and the basic subdifferential.

Theorem 7 *Consider the above problem* $(\mathcal{P})$ *with locally Lipschitz data. Let* $\bar{x}$ *be a local minimum of* $(\mathcal{P})$. *Suppose that the following qualification condition* **(Q)** *holds at* $\bar{x}$.
For $\lambda_i \geq 0$, $i \in I(\bar{x}) = \{i \in \{1, \cdots, m\}$ *such that* $g_i(\bar{x}) = 0\}$, $\lambda_i = 0$, $i \notin I(\bar{x})$ *and* $\mu_j \geq 0$ *for all* $j = 1, \ldots, k$ *with*

$$0 \in \sum_{i=1}^{m} \lambda_i \partial g_i(\bar{x}) + \sum_{j=1}^{k} \mu_j (\partial h_j(\bar{x}) \cup \partial(-h_j)(\bar{x}))$$

implies that $\lambda_i = 0$, *for all* $i = 1, \ldots, m$ *and* $\mu_j = 0$ *for all* $j = 1, \ldots, k$. *Then for all* $\epsilon > 0$, *there exist points* $x_1, x_2 \in \mathbb{B}_\epsilon(\bar{x})$ *and scalars* $\bar{\lambda}_i \geq 0 \quad i = 1, \ldots, m, \quad \bar{\mu}_j \geq 0, \quad j = 1 \ldots, k$ *such that*

(i) $0 \in \hat{\partial} f(x_1) + \sum_{i=1}^{m} \bar{\lambda}_i \partial g_i(x_2) + \sum_{j=1}^{k} \bar{\mu}_j (\partial h_j(x_2) \cup \partial(-h_j)(x_2)) + \epsilon \mathbb{B}_{\mathbb{R}^n}$,

(ii) $\bar{\lambda}_i g_i(x_2) = 0$.

Proof : Since $\bar{x}$ solves $(\mathcal{P})$, then $\bar{x}$ solves the following unconstrained problem

$$\text{Min} \quad f(x) + \delta_C(x). \tag{4.1}$$

where

$$C = \{x \in \mathbb{R}^n : g_i(x) \leq 0, \quad i = 1, \ldots, m, \quad h_j(x) = 0, j = 1, 2, \ldots, k\}.$$

Thus we have

$$0 \in \hat{\partial}(f + \delta_C)(\bar{x}). \tag{4.2}$$

Further, observe that the qualification condition **(Q)** is stable in the sense that there exists $\epsilon > 0$, such that the qualification condition **(Q)** holds at each $x \in \mathbb{B}_\epsilon(\bar{x})$. This is essentially due to fact that ∂g_i for all $i = 1, \ldots, m$ and $(\partial h_j \cup \partial(-h_j))$ for all $j = 1, \ldots, k$ are closed graph set-valued mappings. Consider such an $\epsilon > 0$, for which **(Q)** is stable and apply the approximate sum rule (Theorem 1) to obtain the existence of elements $x_1, x_2 \in \mathbb{B}_\epsilon(\bar{x})$ with $x_2 \in C$, $\| f(x_1) - f(\bar{x}) \| < \epsilon$ and

$$0 \in \hat{\partial} f(x_1) + N_C^F(x_2) + \epsilon \mathbb{B}_{\mathbb{R}^n}. \tag{4.3}$$

However since $x_2 \in \mathbb{B}_\epsilon(\bar{x})$, the qualification condition **(Q)** holds at x_2. Hence from Mordukhovich [14] (Chapter 5) we have

$$N_C^F(x_2) \subseteq \left\{ \sum_{i \in I(x_2)} \lambda_i \partial g_i(x_2) + \sum_{j=1}^{k} \mu_j (\partial h_j(x_2) \cup \partial(-h_j)(x_2)) : \lambda_i \geq 0, \mu_j \geq 0 \right\}.$$

Hence there exists scalars $\bar{\lambda}_i \geq 0, \quad i = 1, \ldots, m$ and $\bar{\mu}_j \geq 0, \quad j = 1, \ldots, k$ such that

$$0 \in \hat{\partial} f(x_1) + \sum_{i \in I(x_2)} \bar{\lambda}_i \partial g_i(x_2) + \sum_{j=1}^{k} \bar{\mu}_j (\partial h_j(x_2) \cup \partial(-h_j)(x_2)) + \epsilon \mathbb{B}_{\mathbb{R}^n}.$$

Setting $\lambda_i = 0$, $i \in I(x_2)$ we get the result. □

Let us consider the simpler situation where problem $(\mathcal{P})$ consists of only inequality constraints. Let $\bar{x}$ be the local minimum. Let the following qualification condition holds at $\bar{x}$:

$$\lambda_i \geq 0, i \in I(\bar{x}), \lambda_i = 0, i \notin I(\bar{x}) \quad \text{and} \quad 0 \in \sum_{i=1}^{m} \lambda_i \partial g_i(\bar{x}) \quad \Longrightarrow \quad \lambda_i = 0, \text{ for all } i.$$

Now consider any $\epsilon > 0$. Since $\bar{x}$ is a local minimum, then using Theorem 7 we can show the existence of $\lambda_i \geq 0$ and points $x_1, x_2 \in \mathbb{R}^n$ such that

(a) $0 \in \hat{\partial} f(x_1) + \sum_{i=1}^{m} \bar{\lambda}_i \partial g_i(x_2) + \epsilon \mathbb{B}_{\mathbb{R}^n}$;

(b) $\bar{\lambda}_i g_i(x_2) = 0, \quad \text{for all} \quad ; i = 1, 2 \ldots, m;$

(c) $x_1, x_2 \in \mathbb{B}_\epsilon(\bar{x})$.

Thus a feasible point $\bar{x}$ of $(\mathcal{P})$ is said to be an ϵ-approximate critical point if there exists $\epsilon > 0$ such that the conditions (a), (b) and (c) are satisfied. The next natural question is the following. Given a feasible point $\bar{x}$, can there be any practical way in which one can check whether $\bar{x}$ satisfies the mixed approximate multiplier rule as expressed through the conditions (a), (b) and (c) above. However in practice it is very difficult to make a check for all $\epsilon > 0$. But in practice it is sufficient to check for an $\epsilon > 0$ sufficiently small. Below we outline the possibility of such an approach. The idea is to take an $\epsilon_0 > 0$ sufficiently small and try to find points x_1, x_2 and scalars $\bar{\lambda}_i \geq 0$ such that (a), (b) and (c) are satisfied. To begin with, observe that setting $\epsilon = \epsilon_0$, the expression in (a) implies

$$0 \in \hat{\partial} f(x_1) + \sum_{i=1}^{m} \bar{\lambda}_i \partial^0 g_i(x_2) + \epsilon_0 \mathbb{B}_{\mathbb{R}^n},$$

since $\partial g_i(x_2) \subseteq \partial^0 g_i(x_2)$. Let us choose two points x_1 and $x_2 \in \mathbb{B}_{\epsilon_0}(\bar{x})$ randomly and let f be differentiable at x_1. Burke, Lewis and Overton [3]

show that such a choice can be made almost surely. Once we know that f is differentiable at x_1 we have $\hat{\partial} f(x_1) = \{\nabla f(x_1)\}$. Hence (a) can now be written as

$$0 \in \nabla f(x_1) + \sum_{i=1}^{m} \bar{\lambda}_i \partial^0 g_i(x_2) + \epsilon_0 \mathbb{B}_{\mathbb{R}^n}.$$

In their paper Burke, Lewis and Overton [3] show how one can make use of gradient information in the neighborhood of a point in order to estimate the Clarke subdifferential at that point. Let us see how to use the technique in [3] to estimate $\partial^0 g_i(x_2)$, for each i. In order to use the technique in [3], we consider a sampling radius $\delta > 0$ and consider the open ball $\mathbb{B}_\delta(x_2)$, and then choose k sample points $z_1^i, \ldots, z_k^i$ where at each $z_1^i, \ldots, z_k^i$ the function g_i is differentiable. Burke, Lewis and Overton [3] showed that the following set

$$C_{g_i}(x_2) = co\{\nabla g_i(z_1^i), \ldots, \nabla g_i(z_k^i)\}$$

can be considered as a good estimate of $\partial^0 g_i(x_2)$ in a stochastic sense if k is sufficiently large. For details see Theorem 3.2, Theorem 3.3 and Corollary 3.4 in [3]. Of course for practical purpose, due to limitations of the computing machine, one can consider an upper bound on the value of k. Thus, for each $i \in I(x_2)$, once a sample $\{z_1^i, \ldots, z_k^i\}$ is chosen from $\mathbb{B}_\delta(x_2)$, a practical way by which we can deduce $\bar{x}$ to be an ϵ_0-approximate critical point is to fixed scalars $\hat{\lambda}_i \geq 0$, $i \in I(\bar{x})$ such that

(e) $0 \in \nabla f(x_1) + \sum_{i \in I(x_2)} \hat{\lambda}_i C_{g_i}(x_2) + \epsilon_0 \mathbb{B}$,

(f) $\hat{\lambda}_i \geq 0 \quad i = 1, \ldots, m.$

The interesting thing to be observed is that we are using only gradient information and moreover the approach described above seems to be more advantageous than the approach based on the approximate multiplier rule since in the current approach we have much less conditions to verify. Of course we do not reject the point $\bar{x}$ if the first sample fails to satisfy (e) and (f) but continue the random sampling procedure further by changing the points x_1 and x_2. Further from a computational point of view, for a given ϵ_0, we can also put an upper bound to the maximum number of times we choose the points x_1 and x_2 and thus a corresponding sample of k points from the $\mathbb{B}_\delta(x_2)$ corresponding to each $i \in I(x_2)$. Further one can also change the δ when x_2 changes.

Thus essentially we need to find an element $\xi_i \in C_{g_i}(x_2)$, $i \in I(x_2)$ and $\hat{\lambda}_i \geq 0$, $i \in I(\bar{x})$ such that

$$\left\| \nabla f(x_1) + \sum_{i=I(\hat{x})} \hat{\lambda}_i \xi_i \right\| \leq \epsilon_0.$$

Since we have used the techniques of Burke, Lewis and Overton [3] in estimating the Clarke subdifferential of the constraints functions, one might argue that one can have an uniform approach by replacing $\hat{\partial} f(x_1)$ by $\partial^0 f(x_1)$ since $\hat{\partial} f(x_1) \subseteq \partial^0 f(x_1)$. However, we insist that one should keep $\hat{\partial} f(x_1)$ since it not only provides a sharper optimality condition but if x_1 is a point where f is differentiable we can replace $\hat{\partial} f(x_1)$ by $\nabla f(x_1)$, a luxury which is not available for the Clarke subdifferential and the basic subdifferential. Thus by keeping $\hat{\partial} f(x_1)$ intact we in fact eliminate computational efforts which are needed to estimate the Clarke subdifferential using the technique of Burke, Lewis and Overton [3].

Acknowledgement The authors wish to thank professors V. H. Ngai for mentioning the possibilty to avoid continuity of ρ in Theorem 13 and Boris Mordukhovich for valuable suggestions and fruitful comments.

Part of this work was done while Joydeep Dutta was visiting the laboratory XLIM (Department of Mathematics and Informatics) UMR-CNRS 6272, and he would like to thank the university of Limoges for its support and its hospitality.

Research of Michel Théra has been partially supported by Agence Nationale de la Recherche under grant ANR NT05 − 1/43040 and by the project STIC-AMSUD.

We are grateful to the two anonymous referees whose constructive suggestions have vastly improved the presentation of the paper.

References

[1] Beer, G *Topologies on Closed and Closed Convex Sets*, Mathematics and its Applications, 268, Kluwer Academic Publishers, Netherlands, 1993.

[2] Borwein, J M, ; Zhu, Q. J. *Techniques of variational analysis.* CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 20. Springer-Verlag, New York, 2005.

[3] Burke, J. V.; Lewis, A. S.; Overton, M. L. Approximating subdifferentials by random sampling of gradients. Math. Oper. Res., **27**, (2002), 567–584.

[4] Clarke, F. H., *Optimization and Nonsmooth Analysis*, Wiley-Interscience, 1983.

[5] Clarke, F. H., *Methods of Dynamic and Nonsmooth Optimization*, SIAM, Philadelphia, 1989.

[6] Dutta, J. Revisiting the Lagrange multiplier rule. Pac. J. Optim. **2**, (2006), 501–519.

[7] Fabian M., Subdifferentiability and trustwothiness in the light of a new variational principle of Borwein and Preiss,. Acta Univ. Carolinae, **30** 51–56(1989).

[8] Ioffe A., Regular Points of Lipschitz functions. Trans. Amer. Soc., **251**, (1979), 61–69.

[9] Jourani, A. , Constraint qualifications and Lagrange multipliers in non-differentiable programming problems, J. Optim. Theo. Appl, **81**, (1994), 533–548.

[10] Jourani. A and Thibault. L, A note on Frchet and approximate subdifferentials of composite functions, Bull. Austral. Math. Soc. **49**,(1994), 111–116.

[11] Mordukhovich, B. S. Maximum principle in the problem of time optimal response with nonsmooth constraints. J. Appl. Math. Mech. **40**, (1976), 960–969.

[12] Mordukhovich, B. S. Metric approximations and necessary optimality conditions for general classes of extremal problems, *Soviet Math. Dokl.*, **22**, (1980), 526-530.

[13] Mordukhovich, B. S. *Variational analysis and generalized differentiation. I. Basic theory.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 330. Springer-Verlag, Berlin, 2006.

[14] Mordukhovich, B. S. *Variational analysis and generalized differentiation. II. Applications.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 331. Springer-Verlag, Berlin, 2006.

[15] Mordukovich, B. S. and Shao, Y. Fuzzy calculus for coderivatives of multifunctions, *Nonlinear Anal.*, **29**, (1997), 605-626.

[16] H. Van Ngai; Théra, M, A fuzzy necessary optimality condition for non-Lipschitz optimization in Asplund spaces. SIAM J. Optim. **12**, (2002), 656–668.

[17] Rockafellar, R. T., Lagrange multipliers and optimality. SIAM Rev., **35** , (1993), 183–238.

[18] Rockafellar, R. T, Extended nonlinear programming. *Nonlinear optimization and related topics (Erice, 1998)*, 381–399, Appl. Optim., 36, Kluwer Acad. Publ., Dordrecht, 2000.

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

[20] Schirotzek, W., *Variational Analysis*. Springer-Verlag, Heidelberg, 2007.

[21] Vinter, R., *Optimal Control*, Birkhauser, 2000.