Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 2, No. 3, pp. 501 – 519, September, 2006

REVISITING THE LAGRANGE MULTIPLIER RULE

Joydeep Dutta

Abstract: In this article our main aim is to study the Lagrange multiplier rule associated with a very general class of optimization problem which goes much beyond the standard framework using only equality and inequality constraints. Traditionally the algorithms are developed by taking inputs from the theory. However the penalty method which is heavily used in practice has also been applied to derive the Lagrange multiplier rule associated with equality and inequality constrained problem In this article we use a penalty approximation approach due to Rockafellar and apply it to derive the Lagrange multiplier rule for a general class of optimization problems. This approach is interesting since it is inherently simple and at the same time one can figure out from the proof the qualification condition required for the problem. The other approaches to the proof of this problem is more involved and requires much more technical sophistication. On our way to the main result we will give a detailed motivation as to why such an approach is taken and its pedagogical value. We will also provide a free standing exposition to nonsmooth analysis and nonsmooth geometry that is required to derive the Lagrange multiplier rule for the problem under consideration.

Key words: *Lagrange multipliers, nonsmooth optimization, subdifferential, locally Lipschitz functions, penalty method, normal cone*

Mathematics Subject Classification: *90C30, 90C46, 49J52*

---

1 Introduction and Motivation

Lagrange multiplier rule is one of the most fundamental aspects of optimization theory. The Lagrangian multiplier rule appears in all textbooks on optimization theory and a large effort has been given by optimization researchers to make the theory more flexible for handling various types of optimization problems. Thus the obvious question therefore is : Is there any necessity to revisit the Lagrange Multiplier rule ?. The desire to do so is motivated by a paper of R. T. Rockafellar titled, *Lagrange Multipliers and Optimality* which was published in the SIAM Review in 1993. In this article Rockafellar [18] takes the reader into the fascinating world of Lagrange multipliers and shows us how our understanding about them has improved with the progress of optimization and how our views about them has evolved. In order to motivate the reader let us quote from the abstract of the above mentioned paper of Rockafellar.

> Lagrange multipliers used to be viewed as auxilliary variables introduced in a problem of constrainted minimization in order to write first-order optimality conditions formally as a system of equations. Modern applications, with their empahsis on numerical methods and more complicated side conditions than equations, have demanded a deeper understanding of the concept and how it fits into

Copyright © 2006 Yokohama Publishers http://www.ybook.co.jp

> a larger theoretical picture.
> A major line of research has been the nonsmooth geometry of one-sided tangent and normal vectors to the set of points satisfying the giving constraints Another has been the game theoretic role of multiplier vectors as a solution to the dual problem. Interpretations as generalized derivatives of the optimal value with respect to the problem parameters have also been explored. Lagrange multipliers are now being seen as arising from a general rule of subdifferentiation of a nonsmooth objective function wich allows black-and-white constraints to be replaced by penalty expressions.

Rockafellar [18] also provides a novel approach to prove the Lagrange multiplier rules by a penalty approximation scheme. This sort of approach is not new. This had been studied earlier by Hestenes [6], Mordukhovich [9], Mc Shane [15], Volin and Ostrovskii [22]. See also Polyak [16]. Though Rockafellar [18] used his result to derive the Lagrange multiplier rule for a mathematical programming problem with equality and inequality constraints his approach is important since it could be adapted to a more broader class of optimization problems which goes beyond the standard framework of inequality and equality constraints. Another important aspect of Rockafellar's approach (see [18]) is that the normal cone is used as the principle vehicle in representing optimality conditions. This approach also allows us to take a geometric view of the Lagrange multiplier itself. The utility of the normal cone as a vehicle to represent optimality conditions has been fully explored in convex analysis (see Rockafellar [17]). For example if $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth(continuously differentiable) convex function and C be a convex set in $\mathbb{R}^n$, then $\bar{x} \in C$ is a minimum of f over the convex set C if and only if

$$0 \in \nabla f(\bar{x}) + N_C(\bar{x})$$

where $N_C(\bar{x})$ denotes the well known notion of the normal cone to a convex set C at the point $\bar{x} \in C$, which is given as

$$N_C(\bar{x}) = \{v \in \mathbb{R}^n : \langle v, x - \bar{x} \rangle \leq 0, \quad \forall x \in C\}.$$

The major advantage in the convex case is that if for example C is given through convex inequality constraints i.e.

$$C = \{x \in \mathbb{R}^n : f_i(x) \leq 0, \quad i = 1, \ldots, m\},$$

then the normal cone can be explicilty calculated under some simple regularity conditions. For simplicity if we assume that each f_i is smooth and the Slater's constrainst qualification hold, i.e. there exists $\hat{x}$ such that $f_i(\hat{x}) < 0$ for all $i = 1, \ldots, m$ then we have

$$N_C(\bar{x}) = \bigcup \left\{ \sum_{i \in I(\bar{x})} y_i \nabla f_i(\bar{x}) : y_i \geq 0 \quad \forall i \in I(\bar{x}) \right\},$$

where $I(\bar{x})$ is the set of active indices. This shows us that the Lagragian multipliers in convex optimization arise when we explicitly compute the normal cone. A relevant question is whether such an optimality condition is possible if f was not convex and C was a nonconvex set. This leads us to the question of developing the notion of normal cone for an arbitary set. For a closed set $C \subseteq \mathbb{R}^n$ the notion of a normal can be developed using the idea of projection to closed set from a point outside it. This was termed as the proximal normal and the cone of all such proximal normals is said to be the proximal normal cone. This notion was first introduced in Clarke [2]. For more details see for example [3] and [4]. However here we shall present an alternative definition of proximal normal (see for example [4]) which is much easier to handle in calculations.

Definition 1.1 *Let C be a closed set in $\mathbb{R}^n$ and let $\bar{x} \in C$. Then a vector $v \in \mathbb{R}^n$ is said to be a proximal normal to C at $\bar{x}$ if there exists $\sigma > 0$ such that*

$$\langle v, x - \bar{x} \rangle \leq \sigma \|x - \bar{x}\|^2, \quad \forall x \in C.$$

The set of all such proximal normals forms a cone called the proximal normal cone to C at $\bar{x}$ which is denoted as $N_C^P(\bar{x})$.

Though the proximal normal cone is convex it has a drawback that it need not be closed. This problem can be removed by introducing a slightly broader notion of a regular normal cone.

Definition 1.2 *Let C be a subset of $\mathbb{R}^n$ and let $\bar{x} \in C$. Then a vector $v \in \mathbb{R}^n$ is said to be a regular normal to C at $\bar{x}$ if*

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

where $\dfrac{o(\|x - \bar{x}\|)}{\|x - \bar{x}\|} \to 0$ when $\|x - \bar{x}\| \to 0$. The set of all regular normals forms a cone which is called the regular normal cone and is denoted by $\hat{N}_C(\bar{x})$.

The fact that $\hat{N}_C(\bar{x})$ is a closed is not really apparent from the definition but however it can be shown to be a polar of the Bouligand tangent cone (which is a well known object in optimization). For details see for example Rockafellar and Wets [20]. The regular normal cone has also been referred to as the Frechet normal cone in the literature. See for example Mordukhovich [10] and Vinter [21]. The drawback of the regular normal cone is that there can be points on the boundary of C where the regular normal cone can just degenerate into the trivial normal cone containing only the zero element. This problem can be overcome by the following notion of the basic normal cone which is a much more robust concept.

Definition 1.3 *Let C be subset of $\mathbb{R}^n$ and let $\bar{x} \in C$. A vector $v \in \mathbb{R}^n$ is said to be a basic normal to C at $\bar{x}$ if there exist sequences $x_k \to \bar{x}$, $(x_k \in C)$ and $v^k \to v$ with $v_k \in \hat{N}(x_k)$. The set of all basic normals form a cone called the basic normal cone to C at $\bar{x}$ and is denoted by $N_C(\bar{x})$.*

We would like to note here that we do not make any distinctions in the symbols of the basic normal cone and the normal cone in convex analysis. This has a two-fold reason. The first one is that the basic normal cone plays a very central role in non-convex optimization quite analogous to the role of the normal cone in convex optimization. The second reason is when C is convex the basic normal cone reduces to the normal cone of convex analysis and thus is a more fundamental object. The basic normal cone is closed but need not be convex and is robust in the sense that when represented as a set-valued map it has a closed graph. Further at each point on the boundary of a closed set the basic normal cone is non-trivial in the sense that it contains additional elements other than zero. The notion of the basic normal come was first introduced by Mordukhovich in 1976 (see [14]) in context of an optimal control problem. For more details see [8]. We would also like to point out that when C is a closed set then in the above definition of the basic normal cone the regular normal cone can be replaced by the proximal normal cone. This is due to the fact that when C is closed each regular normal can be realized as a limit of proximal normals. This was demonstrated in Kruger and Mordukhovich [7].

If $\bar{x} \in C$ is a local minimum of a smooth function (continuously differentiable) f over C then one has

$$0 \in \nabla f(\bar{x}) + N_C(\bar{x}).$$

Thus the basic normal cone can be used as a vehicle to study necessary optimal conditions.

Now consider the following optimization problem (P)

$$\begin{aligned} &\min f_0(x) \\ &\text{subject to} \\ f_i(x) &\leq 0, \quad i = 1, \ldots, s \\ f_i(x) &= 0, \quad i = s+1, \ldots, m \\ x &\in X \end{aligned}$$

Rockafellar [18] demonstrated that the basic normal cone plays a major role in expressing the necessary optimality conditions for the problem (P) assuming that the objective and constraints are smooth functions and the set X is closed. Apart from allowing us to have a compact representation of the optimality conditions it also sheds light on the Lagrange multipliers themselves by bringing out their essential geometric character. Let us denote by C the feasible set of the problem (P). Rockafellar [18] represented C as follows

$$C = \{x \in X : F(x) \in U\}, \tag{1}$$

where $F(x) = (f_1(x), \ldots, f_m(x))$ and the set U is given as

$$U = \{u \in \mathbb{R}^m : u_i \leq 0, \quad \text{for} \quad i = 1, \ldots s, \quad u_i = 0 \quad \text{for} \quad i = s+1, \ldots, m\}.$$

Rockafellar [18] also introduced the following set Y given as

$$Y = \{y \in \mathbb{R}^m : y_i \geq 0 \quad \text{for} \quad i = 1, \ldots, s\}.$$

The following is an interesting result from Rockafellar [18] which connects the normal cone to a point on Y with that of U.

Proposition 1.1 *At any $\bar{y} \in Y$ the normal cone $N_Y(\bar{y})$ consists of vectors $u \in \mathbb{R}^m$ such that*

$$u_i \leq 0 \quad \textit{for} \quad i = 1, \ldots, s, \quad \textit{with} \quad \bar{y}_i = 0$$

and

$$u_i = 0 \quad \textit{for} \quad i = 1, \ldots, s, \quad \textit{with} \quad \bar{y}_i > 0, \quad \textit{and} \quad i = s+1, \ldots, m.$$

Further for any $\bar{u} \in U$ the normal cone $N_U(\bar{u})$ consists of all vectors $y \in \mathbb{R}^m$ such that

$$y_i = 0 \quad \textit{for} \quad i = 1, \ldots, s, \quad \textit{with} \quad \bar{u}_i < 0$$

$$y_i \geq 0 \quad \textit{for} \quad i = 1, \ldots, s, \quad \textit{with} \quad \bar{u}_i = 0$$

and

$$y_i \quad \textit{unrestricted} \qquad \textit{for} \quad i = s+1, \ldots, m$$

Further one has $\bar{y} \in N_U(\bar{u})$ if and only if $\bar{u} \in N_Y(\bar{u})$.

In order to write down a Lagrange multiplier rule the constraints need to satisfy certain qualification conditions. By imposing a qualification condition one ensures that the multiplier associated with the gradient of the objective function remains positive. This in fact is the central theme of the well known Karush-Kuhn-Tucker (KKT) conditions. KKT conditions are in fact the Lagrangian multipler rule when inequalities are added as constraints. What could be the appropriate or rather fundamental qualification condition for the given problem (P). A natural constraint qualification condition could be as follows. Consider $\bar{x}$ to be a feasible point of (P). Then the problem (P) is said to satisfy the *basic constraint qualification* (BCQ) at $\bar{x}$ if

$$y \in N_U(F(\bar{x})), \quad \text{with} \quad 0 \in y_1 \nabla f_1(\bar{x}) + \ldots, y_m \nabla f_m(\bar{x}) + N_X(\bar{x}),$$

implies that $y = 0$.

One of the important properties of this qualification condition is that it is a robust one in the sense that it is stable under perturbations. One can show that if BCQ is satisfied at $\bar{x}$ then it is satisfied for all points in some neigbhborhood of $\bar{x}$. The term basic constraint qualification was used in Rockafellar [18] and also in Rockafellar and Wets [20]. Rockafellar [18] proved the following result

Theorem 1.1 *If $\bar{x}$ is a locally optimal solution of (P) at which the basic constraint qualification is satisfied, there must exist a vector $\bar{y} \in Y$ such that*

$$-\nabla_x L(\bar{x}, \bar{y}) \in N_X(\bar{x}), \qquad \nabla_y L(\bar{x}, \bar{y}) \in N_Y(\bar{y}).$$

Observe that $\nabla_y L(\bar{x}, \bar{y}) \in N_Y(\bar{y})$ is same as $F(\bar{x}) \in N_Y(\bar{y})$ and thus $\bar{y} \in N_U(F(\bar{x}))$. As we have mentioned earlier Rockafellar [18] used a penalty approximation scheme in order to deduce the above result. The approch has two major advantages. The first one is that this approach can be used to handle problem formats beyond (P). This has been shown in a slightly indirect way in Rockafellar and Wets [20]. For the nonsmooth case this approach has been used in Dutta [5]. In the traditional approach to develop necessary optimality conditions for the problem (P) one needs to use implicit or inverse function theorem to handle the equality constraints and the machinery of separation of convex sets to handle inequality constraints. Such a sophisticated machinery is not required while using a penalty approximation scheme. Further it provides an important insight by showing that Lagrange multipliers can be generated in a constructive way. This constructive approach to Lagrange multipliers is not possible in the traditional approach since the traditional approach simply proves the existence of multipliers.

It is interesting to note that Theorem 1.1 also follows from Theorem 1 in Mordukhovich [11] which is obtained by the so called method of metric approximations. The Theorem 1.1 can also be deduced from Theorem 1 in Mordukhovich [9] which is obtained via penalty function approximations. See also Corollary 7.5.1 in Mordukhovich [12].

The first step to look beyond the problem (P) is to observe the way Rockafellar has represented the feasible set of (P) in (1). Thus if we just consider U to be a closed set and $F : \mathbb{R}^n \to \mathbb{R}^m$ is a smooth vector function and X is a closed set as before then we can as well consider studying the following problem(P1)

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

It is clear that the problem (P) is contained in the class of problems (P1). In fact one may not even consider F to be smooth. See for example Rockafellar and Wets [20] or Dutta [5]. It is clear that (P1) is equivalent to the problem

$$\min f_0(x) + \delta_{\hat{C}}(x), \quad \text{subject to} \quad x \in X,$$

where $\hat{C} = \{x \in \mathbb{R}^n : F(x) \in U\}$ and $\delta_{\hat{C}}$ is the indicator function of the set $\hat{C}$ of (P1). Further observe that $\delta_{\hat{C}}(x) = \delta_U(F(x))$. Thus the problem (P1) is equivalent to the following

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

The the problem (P1) may be now thought to be embedded in the larger class of problems (P2) given as

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

where ρ is a possibly extended-valued, proper, lower-semicontinuous function on $\mathbb{R}^m$. Rockafellar [18] considered the case where f_0 and F are smooth functions and ρ was additionally convex apart from being proper and lower semicontinuous. Rockafellar [18] again used a penatly approximation scheme to deduce a Lagrange multiplier rule for the problem (P2). In this paper we focus on the problem (P2) but consider that f_0 is a locally Lipschitz function and F is a locally Lipschitz vector-valued function and ρ is only a proper lower semicontinuous function. We want to demonstrate that with requisite modifications one can actually use Rockafellar's penatly approximation scheme for the problem (P2) even without any convexity assumption on ρ. We will show in section 3 that problems of the form (P2) form a very large class of optimization problems and thus having an easy way to derive the Lagrange multiplier rule for (P2) would be beneficial and the multiplier rule for various classes of optimization problems can be easily deduced from that of (P2). Thus even from the pedagogic point of view this approach has a distinct adavantage. What we demonstrate is that even in the general context of nonsmooth optimization, the penatly approximation scheme can be used profitably and is free of any complicated mathematical machinery. Further the qualification condition required for deducing the multiplier rule naturally arises from the proof of the result. This is only possible with the penalty approximation scheme. Thus we would like to put forward the claim that the penalty approximation scheme of Rockafellar [18] as not only elegant but is a fundamental approach to the Lagrange multiplier rule.

We plan the paper as follows. In section 2 we present some basic tools and result from nonsmooth analysis which will play a key role in proving the Lagrange multiplier rule for the problem (P2). In section 3 we present tha main result and show how it can be applied to some special cases and thus demonstrate how the Lagrange multiplier rule for the problem (P2) can indeed be used to generate the Lagrangian multiplier rule for various classes of optimization problem.

2 Tools from Nonsmooth Analysis

In order to make the paper self contained as much as possible let us start from the definition of the subdifferential of a proper convex function.

Definition 2.1 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper convex function and let $x \in \mathbb{R}^n$ be a point where $f(x)$ is finite. Then the vector $\xi \in \mathbb{R}^n$ is called a subgradient of f at x if*

$$f(y) - f(x) \geq \langle \xi, y - x \rangle, \quad \forall y \in \mathbb{R}^n.$$

The set of all subgradients of f at x is denoted by $\partial f(x)$ and is known as the subdifferential of f at x. If x is a point where $f(x) = +\infty$ then we define $\partial f(x) = \emptyset$. Ofcourse it is well known that if x is a point in the interior of *domf* then $\partial f(x)$ is a non-empty convex and compact set. For more details on the subdifferential of a convex function see for example Rockafellar [17]. It is important to note that this notion of a subdifferential is essentially for a convex function since it comes out directly from the geometry of the epigraph of a convex function. For a non-convex function if one uses this notion of the subdifferential then even in very simple cases the subdifferential can be empty at a point where the function achieves a local minimum. Thus one may be motivated to define a subdifferential for a non-convex function by adding a error term to the right side of the above expression of a subgradient. This can be achieved by the following notion of a regular subgradient and a regular subdifferential.

Definition 2.2 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a given function and let x be a point where $f(x)$ is finite. The vector $v \in \mathbb{R}^n$ is called a regular subgradient of f at x if*

$$f(y) \geq f(x) + \langle v, y - x \rangle + o(||y - x||), \quad \forall y \in \mathbb{R}^n.$$

The set of all regular subgradient of f at x is known as the regular subdifferential and is denoted by $\hat{\partial} f(x)$. From the above definition of a regular subgradient it is clear that we are actually talking about a proper function. So in a natural way we set $\hat{\partial} f(x) = \emptyset$ if $f(x) = +\infty$. The regular subdifferential though simple to represent also suffers from the drawback that it can also become empty at certain points where f achieves a local minimum. A fundamental property of the regular subdifferential is that the set of points over which it is non-empty is dense in $dom f$. This allows us to use a sequential approach to have a more robust notion of the subdifferential. Such a subdifferential which we term as the basic subdifferential is defined below.

Definition 2.3 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a given function and let x be a point where $f(x)$ is finite. A vector $v \in \mathbb{R}^n$ is called a basic subgradient of f at x if there are sequences $x_k \to x$ with $f(x_k) \to f(x)$ and $v_k \to v$ with $v_k \in \hat{\partial} f(x_k)$.*

The set of all basic subgradients of f at x is known as the basic subdifferential and is denoted as $\partial f(x)$. Observe that we have used the same notation for the basic subdifferential and the subdifferential of a proper convex function. There are two-fold reasons for this. The first reason pertains to the fact that the basic subdifferential plays a central role in nonsmooth analysis just as the subdifferential of a convex function plays a central role in convex analysis. The second reason is that the basic subdifferenial reduces to the subdifferental of a convex function if f is a convex function.

Remark 2.1 The notion of the basic subdifferential was first introduced by Mordukhovich [14] and then studied in detail in [10]. It is important to note the sequential nonsmooth constructions, the basic normal cone and the basic subdifferential, arein fact a by-product of the method of metric approximations. Method of metric approximations which is a very important tool for deriving necessary optimality conditions for constrained optimization and optimal control. For more details on the method of metric approximation see for example Mordukhovich [12]. Rockafellar and Wets [20] have also studied the basic subdifferential and its properties in detail but using an approach which is much different from Mordukhovich [10]. The regular subdifferential is also referred to as the Frechet subdifferential in the literaure.

The basic subdifferential is in general a closed set but need not be convex. However if f is locally Lipschitz then the basic subdifferential is non-empty and compact. Further if f is strictly differentiable (which includes smooth functions) then the basic subdifferential reduces to a singleton set containing just the derivative. On the other hand the regular subdifferential is always a convex set. The basic subdifferential admits very good calculus rules and that makes it an important vehicle to express necessary optimality conditions in nonsmooth optimization. Moreover when f is locally Lipschitz then ∂f is locally bounded and has a closed graph and thus is upper-semicontinuous as a set-valued map. Further the regular subdifferential and the basic subdifferential is also related in a very fundamental way with the regular normal cone and the basic normal cone. One has

$$\hat{\partial} f(x) = \{v \in \mathbb{R}^n : (v, -1) \in \hat{N}_{epif}(x, f(x))\}$$

and

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

In fact the above expressions of the regular and basic subdifferential can be considered as an equivalent definition of these objects.

It has been shown for example in Mordukhovich [8] that the basic normal cone to nonconvex sets in general contains some non-vertical components and some horizontal components. One can contsruct very simple examples of this fact by considering the set to be an epigraph of a non-Lipschitz function of a real variable. The non-vertical components correspond to the basic subdifferential which is seen from the above representation of the subdifferential. The horizontal components of the basic normal cone correspond to what is known as the horizontal subdifferential of f at x which is denoted as $\partial^\infty f(x)$ and is given as

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

When f is Lipschitz around x then $\partial^\infty f(x) = \{0\}$. Thus the asymptotic subdifferential in a certain sense measures the extent to which a function has moved away from Lipschitzianness. Further the normal cone is related in an interesting way with the basic subdifferential and the asymptotic subdifferential.

$$\partial \delta_C(x) = N_C(x), \quad \partial^\infty \delta_C(x) = N_C(x),$$

We will now collect some results on the normal cones and subdifferentials that we have just defined which would be relevant for deducing the necessary optimality condition for the problem (P2).

Lemma 2.1 *Let $f, g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be two given lower-semicontinuous functions which are finite at x. Assume that f is Lipschitz around x. Then one has*

$$\partial(f + g)(x) \subseteq \partial f(x) + \partial g(x).$$

Further if f is a smooth function or strictly differentiable function then one has

$$\partial(f + g)(x) = \nabla f(x) + \partial g(x).$$

For more details on the sum rule see for example Mordukhovich [8], [10] and Rockafellar and Wets [20]. An immediate consequence of the sum rule is the following necessary optimality condition. If $\bar{x}$ be a local mimimum of a locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ over a closed subset C of $\mathbb{R}^n$ then one has

$$0 \in \partial f(\bar{x}) + N_C(\bar{x}).$$

Let us also introduce the following notational simpification. Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a given vector function and let $y \in \mathbb{R}^m$. Then the function $(yF) : \mathbb{R}^n \to \mathbb{R}$ is given as

$$(yF)(x) = \langle y, F(x) \rangle.$$

Lemma 2.2 *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a vector-valued locally Lipschitz function and let $g : \mathbb{R}^m \to \mathbb{R}$ be a locally Lipschitz function. Consider the function $f(x) = g \circ F(x)$. Then we have*

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

We will now present a very important result on regular subgradients from Rockafellar [18].

Lemma 2.3 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a given function which is finite at x. Then $v \in \mathbb{R}^n$ is a regular subgradient of f at x if and only if on some neighborhood U of x there is a smooth function h such that $\nabla h(x) = v$ such that $h(y) \leq f(y)$ for all y in U and $h(x) = f(x)$.*

3 A General Multiplier Rule

In this section we will be concerned with the problem (P2), i.e.

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

Our aim in this section is to develop a sharp Lagarngian multiplier rule for the problem (P2). The necessary optimality condition that we shall derive will be given in terms of the basic subdifferential and the basic normal cone. In Rockafellar [19] the program (P2) is termed as an extended nonlinear programming problem. Rockafellar [19] assumes that the vector function F is smooth, the set X is a non-empty polyhedral set and function ρ is convex and admits the following representation

$$\rho(u) = \sup_{y \in Y} \{\langle y, u \rangle - k(y)\},$$

where Y is non-empty polyhedral set in $\mathbb{R}^m$ and k is a smooth function which convex on Y. In fact Rockafellar [19] showed that there are various important class of optimization problems which can be modelled as the problem (P2) which satisfies the above mentined assumptions. For example if we set $f_0(x) = 0$ for all $x \in \mathbb{R}^n$ then the problem (P2) becomes the well known composite optimization problem and in particular if ρ is the max function then (P2) represents a minimax problem. As we mentioned in section 1 we intend to assume that f_0 and F are locally Lipschitz and X is a closed set while ρ is a possibly extended-valued proper lower-semicontinuous function. Our approach will be to use a penalty approximation scheme as used in Rockafellar [18] along with the relevant techniques of nonsmooth analysis. Before we present our main result we present the following two lemmas which will be needed in the sequel.

Lemma 3.1 *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz vector-valued function. Let there exist a sequence $y_k \to 0$ and $x_k \to \bar{x}$ such that*

$$w_k \in \partial(y_k F)(x_k), \quad \forall k.$$

Then the sequence $\{w_k\}$ is bounded and every convergent subsequence of $\{w_k\}$ converges to zero and thus $w_k \to 0$.

Proof. Consider a fixed k. Then there exists sequences $w_k^n \to w_k$ and $x_k^n \to x_k$ such that

$$w_k^n \in \hat{\partial}(y_k F)(x_k^n).$$

Thus by the definition of a regular subdifferential we have

$$(y_k F)(x) - (y_k F)(x_k^n) \geq \langle w_k^n, x - x_k^n \rangle + o(||x - x_k^n||). \tag{2}$$

For any given $d \in \mathbb{R}^n$ we can choose a sequence $t_k^n \downarrow 0$ such that for n sufficiently large one has x_k^n and $x_k^n + t_k^n d$ very close to x_k. Noting the fact that F is locally Lipschitz it is clear that the sequence

$$\left\{ \frac{F(x_k^n + t_k^n d) - F(x_k^n)}{t_k^n} \right\}$$

has a cluster point θ_k as $n \to \infty$. Now using (2) we have

$$\left\langle y_k, \frac{F(x_k^n + t_k^n d) - F(x_k^n)}{t_k^n} \right\rangle \geq \langle w_k^n, d \rangle + \frac{o(t_k^n)}{t_k^n}.$$

Hence as $n \to \infty$ (considering subsequences) we have

$$\langle y_k, \theta_k \rangle \geq \langle w_k, d \rangle. \tag{3}$$

Observe that the above fact holds for each k.

As $x_k \to \bar{x}$, for k sufficiently large, x_k is very close to $\bar{x}$. Let us choose k in such a way that x_k is in a neighborhood U of $\bar{x}$ over which F has Lipschitz rank L. Now for that particular k we can choose n sufficiently large so that both $x_k^n + t_k^n d$ and x_k^n are in a neighborhood V of x_k with $V \subset U$. Thus by choosing k and n both sufficiently large such that $x_k^n + t_k^n d$ and x_k^n are in a neigborhood of $\bar{x}$ in which F has a Lipschitz rank L. Thus we have from the locally Lipschitz property of F ,

$$||F(x_k^n + t_k^n d) - F(x_k^n)|| \leq L||t_k^n d||.$$

Thus for k sufficiently large for which the above mentioned conditions are satisfied one has

$$||\theta_k|| \leq L||d||$$

as $n \to \infty$. Since the vector d is fixed it is clear from the above expression that $\{\theta_k\}$ is bounded and thus we can extract a convergence subsequence from it which converges say to θ^* . Further from (3) we have

$$\langle w_k, d \rangle \leq ||y_k|| ||\theta_k||, \quad \forall k$$

Let us now assume that the sequence $\{w_k\}$ is bounded and let w^* is a cluster of $\{w_k\}$. Thus passing to the limit in the above expression shows that $\langle w^*, d \rangle \leq 0$. Since d can be chosen arbitarily one has $w^* = 0$. In fact we have proved that if $\{w_k\}$ is bounded then every convergent subsequence of $\{w_k\}$ converges to the zero. Thus $w_k \to 0$.

Further we shall now claim that the sequence $\{w_k\}$ can never be unbounded. Assume that $\{w_k\}$ is unbounded. Hence we can construct the sequence $\{v_k\}$, where $v_k = \dfrac{w_k}{||w_k||}$ which is bounded and thus we can extract a convergent subsequence converging to v^* and $||v^*|| = 1$. Now using the positive homogeneity of the basic subdifferential we have

$$v_k \in \partial \left(\frac{y_k}{||w_k||} F \right) (x_k).$$

Now since $\dfrac{y_k}{||w_k||} \to 0$ and $x_k \to \bar{x}$ using the argument of the previous part of the lemma we can show that $v^* = 0$ which is a contradiction. This proves the result. □

Remark 3.1 Though the above lemma (Lemma 3.1) has been deduced using techniques essentially suited for finite dimensions we would like to note the the result of the above lemma can also be deduced from the proof of Theorem 5.2 in Mordukhovich and Shao [13]. See also Lemma 3.27 in Mordukhovich [8]. It is important to note that Mordukhovich and Shao [13] has studied sequential nonsmooth analysis in Asplund spaces.

Lemma 3.2 *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a vector-valued locally Lipschitz function and let X be a closed set. Let $y_k \to \bar{y}$ and $x_k \to \bar{x}$ be such that*

$$v_k \in \partial(y_k F)(x_k) + N_X(x_k).$$

If $v_k \to v^*$ then

$$v^* \in \partial(\bar{y}F)(\bar{x}) + N_X(\bar{x}).$$

Proof. Using the sum rule (Lemma 2.1) for the basic subdifferential we have

$$\partial(y_k F)(x_k) \subseteq \partial(\bar{y}F)(x_k) + \partial((y_k - \bar{y})F)(x_k).$$

Thus we have

$$\partial(y_k F)(x_k) + N_X(x_k) \subseteq \partial(\bar{y}F)(x_k) + \partial((y_k - \bar{y})F)(x_k) + N_X(x_k).$$

Thus we have for each k,

$$v_k \in \partial(\bar{y}F)(x_k) + \partial((y_k - \bar{y})F)(x_k) + N_X(x_k).$$

Thus for each k there exists $u_k \in \partial(\bar{y}F)(x_k)$, $w_k \in \partial((y_k - \bar{y})F)(x_k)$ and $z_k \in N_X(x_k)$ such that

$$v_k = u_k + w_k + z_k$$

Thus one has

$$v_k - u_k - w_k \in N_X(x_k).$$

Now by the local boundednes of $\partial(\bar{y}F)$ we can show that the sequence $\{u_k\}$ is bounded and has a cluster point u^*. Since $\partial(\bar{y}F)$ has a closed graph we have $u^* \in \partial(\bar{y}F)(\bar{x})$. Further from the hypothesis of the theorem by using Lemma 3.1 we deduce that $w_k \to 0$ and using the fact that basic normal cone mapping has a closed graph we deduce that

$$v^* - u^* \in N_X(\bar{x}).$$

This proves the result. □

We will now present our main result

Theorem 3.1 *Let us consider the problem (P2) where f_0 is a locally Lipschitz function, F is a vector-valued locally Lipschitz function, ρ is a possibly extended-valued proper lower semicontinuous function and X is a closed subset of $\mathbb{R}^n$. Let $\bar{x}$ be a locally optimal solution of (P2). Further assume that the following qualification condition (Q) holds at $\bar{x}$:*

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

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

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

Proof. By the local optimality of $\bar{x}$ we mean that there exists a compact neighborhood V of $\bar{x}$ such that

$$f_0(x) + \rho(F(x)) \geq f_0(\bar{x}) + \rho(F(\bar{x})), \quad \forall x \in X \cap V.$$

In fact without loss of generality we can consider X to be compact and $\bar{x}$ is a global minimum of $f_0 + \rho(F(.))$ over X. Further we may replace the objective function in (P2) by the function

$$\tilde{f}_0(x) = f_0(x) + \rho(F(x)) + \varepsilon(||x - \bar{x}||)^2.$$

Thus we may now consider without loss of generality that $\bar{x}$ to be a unique solution of (P2). We may further without loss of generality consider *domρ* to be a compact set large enough so as to contain $F(x)$ for each $x \in X$. This can be done by if we redefine ρ such that it takes up the value $+\infty$ outside a compact set which is sufficiently large enough to contain the image $F(x)$ for each x. This is possible since the compactness of X and the continuity of F shows that $F(X)$ is a compact set. Thus we can now say that the function $f_0(x)+\rho(u)$ has a finite minimum value on $X \times \mathbb{R}^n$. Let us now consider the problem $(\hat{P})$ given as

$$\min \hat{f}(x, u) = f_0(x) + \rho(u), \quad \text{subject to} \quad F(x) - u = 0, \quad (x, u) \in \ \times \mathbb{R}^n.$$

It is clear that $(\bar{x}, \bar{u}) = (\bar{x}, F(\bar{x}))$ is the unique solution $(\hat{P})$. Now for a sequence of values $\varepsilon_k \downarrow 0$ consider the following penalized problems $(\hat{P}^k)$ given as

$$\min \hat{f}^k(x, u) = f_0(x) + \rho(u) + \frac{1}{2\varepsilon_k}(||F(x) - u||)^2, \quad \text{subject to} \quad (x, u) \in X \times \mathbb{R}^n.$$

These problems have a solution (x_k, u_k) for each k since the level sets of $\hat{f}^k(x, u)$ in $X \times \mathbb{R}^n$ are closed and bounded. This is precisely because of the continuity of f_0 and F, the lower semicontinuity of ρ and the compactness of X and *domρ*. Let μ be the minimum value of $f_0(x) + \rho(u)$. Thus we have

$$\begin{aligned} \mu + \frac{1}{2\varepsilon_k}(||F(x_k) - u_k||)^2 &\leq f_0(x_k) + \rho(u_k) + \frac{1}{2\varepsilon_k}(||F(x_k) - u_k||)^2 \\ &= \hat{f}^k(x_k, u_k) \leq \hat{f}^k(\bar{x}, \bar{u}) = f_0(\bar{x}) + \rho(\bar{u}). \end{aligned}$$

Since X and *domρ* are compact sets it is easy to see that the sequence $\{(x_k, u_k)\}$ is a bounded sequence and thus we can extract a convergent subsequence which converges to say $(\hat{x}, \hat{u})$. Further it is clear from the previous calculations that

$$\mu + \frac{1}{2\varepsilon_k}(||F(x_k) - u_k||)^2 \leq f_0(\bar{x}) + \rho(\bar{u}).$$

This implies that

$$(||F(x_k) - u_k||)^2 \leq 2\varepsilon_k(f_0(\bar{x}) + \rho(\bar{u}) - \mu).$$

As $k \to \infty$ we have $||F(\hat{x}) - \hat{u}|| = 0$. This shows that $F(\hat{x}) = \hat{u}$.

We have already proved that $f_0(\bar{x}) + \rho(\bar{u}) \geq \hat{f}^k(x_k, u_k)$ for all k. This shows that

$$f_0(\bar{x}) + \rho(\bar{u}) \geq \limsup_{k\to\infty} \hat{f}^k(x_k, u_k).$$

This shows that

$$f_0(\bar{x}) + \rho(\bar{u}) \geq \limsup_{k\to\infty}(f(x_k) + \rho(u_k) - (||F(x_k) - u_k||)^2).$$

Now noting that ρ is lower-semicontinuous and that $(\hat{x}, \hat{u})$ is a cluster point of $\{(x_k, u_k)\}$ and the fact that $F(\hat{x}) = \hat{u}$ and that limit supremum is bigger than the limit infimum (as $k \to \infty$) we have

$$f_0(\bar{x}) + \rho(\bar{u}) \geq f_0(\hat{x}) + \rho(\hat{u}).$$

Thus $(\hat{x}, \hat{u})$ is a solution to problem $(\hat{P})$. Since $(\bar{x}, \bar{u})$ is a unique solution of $(\hat{P})$ we have $\bar{x} = \hat{x}$ and $\bar{u} = \hat{u}$. Hence without loss of generality we can say that $x_k \to \bar{x}$ and $u_k \to \bar{u}$. Further from the above calculations one can derive that $\rho(u_k) \to \rho(\bar{u})$ (or else the minimum will change). Since (x_k, u_k) is an optimal solution of $(\hat{P}^k)$ it implies that x_k minimizes $\hat{f}^k(x, u_k)$ over $x \in X$ and u_k minimizes $\hat{f}^k(x_k, u)$ over $u \in \mathbb{R}^m$. Using the first fact we immediately have the following necessary optimality condition

$$0 \in \partial_x \hat{f}^k(x_k, u_k) + N_X(x_k),$$

where the x in the subscript signifies that the subdifferentiation is taken with respect to x. Further using Lemma 2.1 and Lemma 2.2 we have

$$\partial_x \hat{f}^k(x_k, u_k) \subseteq \partial f_0(x_k) + \partial(y_k F)(x_k),$$

where $y_k = \dfrac{F(x_k) - u_k}{\varepsilon_k}$. Thus we have

$$0 \in \partial f_0(x_k) + \partial(y_k F)(x_k) + N_X(x_k).$$

Now using the second fact that u_k minimizes $\hat{f}^k(x_k, u)$ over $u \in \mathbb{R}^m$ we have

$$f(x_k) + \rho(u_k) + \frac{1}{2\varepsilon_k}(||F(x_k) - u_k||)^2 \leq f(x_k) + \rho(u) + \frac{1}{2\varepsilon_k}(||F(x_k) - u||)^2.$$

Thus we have

$$\rho(u_k) + \frac{1}{2\varepsilon_k}(||F(x_k) - u_k||)^2 \leq \rho(u) + \frac{1}{2\varepsilon_k}(||F(x_k) - u||)^2.$$

This further reduces to the following

$$\rho(u_k) + \frac{1}{2\varepsilon_k}(||F(x_k) - u_k||)^2 - \frac{1}{2\varepsilon_k}(||F(x_k) - u||)^2 \leq \rho(u).$$

Let us now define the following function $h^k : \mathbb{R}^m \to \mathbb{R}$ as follows

$$h^k(u) = \rho(u_k) + \frac{1}{2\varepsilon_k}(||F(x_k) - u_k||)^2 - \frac{1}{2\varepsilon_k}(||F(x_k) - u||)^2.$$

Observe that for each k we have $h^k(u) \leq \rho(u)$ and $h^k(u_k) = \rho(u_k)$. Moreover observe that h^k is a smooth function for each k and thus by Lemma 2.3 we conclude that

$$\nabla h^k(u_k) \in \hat{\partial} \rho(u_k).$$

Further observe that

$$\nabla h^k(u_k) = \frac{F(x_k) - u_k}{\varepsilon_k} = y_k$$

This shows that $y_k \in \hat{\partial}\rho(u_k)$ for every k. Thus we have

$$y_k \in \hat{\partial}\rho(u_k) \quad \text{and} \quad 0 \in \partial f_0(x_k) + \partial(y_k F)(x_k) + N_X(x_k). \tag{4}$$

Now there arises two cases. Either the sequence $\{y_k\}$ is bounded or it is unbounded. Let us first consider that it is bounded and without loss of generality let us consider that $y_k \to \bar{y}$. Since $u_k \to \bar{u}$ and $\rho(u_k) \to \rho(\bar{u})$ we have $\bar{y} \in \partial\rho(\bar{u}) = \partial\rho(F(\bar{x}))$. Moreover using the the fact that ∂f_0 is locally bounded and $x_k \to \bar{x}$ we can conclude using Lemma 3.2 that

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

Now consider the second case, that is the sequence $\{y_k\}$ is unbounded. Let us now define a sequence $v_k = \dfrac{y_k}{||y_k||}$. Since $\{v_k\}$ is bounded we can consider without loss of generality that $v_k \to \bar{v}$ with $||\bar{v}|| = 1$. Thus using (4) we have

$$0 \in \frac{1}{||y_k||}\partial f_0(x_k) + \frac{1}{||y_k||}\partial(y_k F)(x_k) + \frac{1}{||y_k||}N_X(x_k).$$

Since the basic subdifferential is positively homogeneous we have

$$0 \in \frac{1}{||y_k||}\partial f_0(x_k) + \partial(\frac{y_k}{||y_k||}F)(x_k) + N_X(x_k).$$

Now as $k \to \infty$ we deduce from the local boundedness of ∂f_0 and Lemma 3.2 that

$$0 \in \partial(\bar{v}F)(\bar{x}) + N_X(\bar{x}).$$

Now we also have $y_k \in \hat{\partial}\rho(u_k)$. Now by the definition of a regular normal vector we have

$$(y_k, -1) \in \hat{N}_{epi\rho}(u_k, \rho(u_k)).$$

This shows that

$$\left(\frac{y_k}{||y_k||}, \frac{-1}{||y_k||}\right) \in \hat{N}_{epi\rho}(u_k, \rho(u_k)).$$

Noting that $u_k \to \bar{u}$ and $\rho(u_k) \to \rho(\bar{u})$ as $k \to \infty$, in the limit one has

$$(\bar{v}, 0) \in \hat{N}_{epi\rho}(\bar{u}, \rho(\bar{u})).$$

This implies that $\bar{v} \in \partial^\infty\rho(\bar{u}) = \partial^\infty\rho(F(\bar{x}))$. Thus we have proved the existence of $0 \neq \bar{v} \in \partial^\infty\rho(F(\bar{x}))$ such that

$$0 \in \partial(\bar{v}F)(\bar{x}) + N_X(\bar{x}).$$

This contradicts the qualification condition (Q) and hence the result. □

Remark 3.2 It is important to note what are the advantages of the the approach to the Lagrangian multiplier rule compared to the existing approaches in the literature. It is clear that if one has to take a direct approach to prove the necessary optimality condition for the problem (P2) then one has to verify qualification conditions before applying the calculus rules. This may be quite technical since it would involve computation of horizontal subdifferentials. The present approach does not have any such technical complications. In

Rockafellar and Wets [20] the problem (P2) is studied with the assumption that ρ is convex apart from being proper and lower-semicontinuous. Rockafellar and Wets [20] begins by reformulating the problem (P2) as a composite optimization problem and then uses directly the optimality conditions and thus uses calculus rules which again needs verification of qualification conditions. In Vinter [21] the convexity assumption on ρ is dropped and an approach similar to Rockafellar and Wets [20] is adopted. However Vinter's approach is slightly complicated since the convexity assumption on ρ is dropped. Another important aspect of the penalty approximation approach is that it generates the Lagrangian multipliers in a constructive way. Further, from the last part of the proof of the above theorem the qualification condition (Q) required for the problem (P2) comes out in a very natural way. These two important aspects are however not apparent from the approaches in Rockfellar and Wets [20] and Vinter [21].

It is important to note that Theorem 3.1 can be deduced using the method of metric approximations by using the calculus rules for the basic subdifferentials as developed in [10] and [12]. In this article as we have seen the penalty function approach is used. Thus it is instructive to see what are the differences between the method of metric approximations and the penalty function method. The main difference between the method of metric approximations and the penalty function method is in how the corresponding approximation function is constructed. The method of metric approximations involves a symmetric Euclidean distance of the cost and constraint function while the penalty function method significantly distinguishes between the cost and constraints. The method of metric approximations was indeed used in Mordukhovich [12] to build the whole nonsmooth analysis and optimization theory. In this article we have explored the potential of the penalty function method to devise a robust optimality condition for a very large class of problems.

Let us now study some consequences of the above theorem.

Corollary 3.1 *Let us consider the problem (P1) as given in section 1. Let us assume that f_0 and F are locally Lipschitz functions, X is a closed subset of $\mathbb{R}^n$ and U is a closed subset of $\mathbb{R}^m$. Assume that $\bar{x}$ is a local optimal solution (P1). Further assume that the following qualification (Q1) condition holds at $\bar{x}$:*

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

Then there exits $\bar{y} \in N_U(F(\bar{x}))$ such that

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

Proof. Observe that the problem (P1) can be equivalently formulated as problem (P2) with $\rho = \delta_U$. This has been already demonstrated in section 1. Observe that $\partial^\infty \delta_U(F(\bar{x})) = N_U(F(\bar{x})$. Thus the qualification condition (Q1) is equivalent to the qualification condition (Q2) of Theorem 3.1. The result now follows from Theorem 3.1 by noting that $\partial\delta_U(F(\bar{x})) = N_U(F(\bar{x}))$. □

Since $\bar{x}$ is a local optimal solution for (P1) then the necessary optimality condition is

$$0 \in \partial f_0(\bar{x}) + N_C(\bar{x}),$$

where $C = \{x \in : F(x) \in U\}$. Thus the Lagrange multipliers are generated when we explicitly compute $N_C(\bar{x})$. The penalty approximation scheme actually allows us to skilfully avoid the normal cone computation or one may argue that the normal cone is computed in an indirect way via the penalty approximation scheme. However it is interesting to note that using Corollary 3.1 we can also compute the basic normal cone to C at $\bar{x}$ where C is as described above.

Proposition 3.1 *Let us consider the set C given as*

$$C = \{x \in X : F(x) \in U\},$$

where F is as before a locally Lipschitz function, X is a closed subset of $\mathbb{R}^n$ and U is a closed subset of $\mathbb{R}^m$. Let $\bar{x}$ be a point in C for which the following qualification condition (Q2) holds:

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

Then

$$N_C(\bar{x}) \subseteq \bigcup \{\partial(yF)(\bar{x}) + N_X(\bar{x}) : y \in N_U(F(\bar{x}))\}.$$

Proof. Observe that from the hypothesis of the proposition the set C is a closed subset of $\mathbb{R}^n$. First consider a proximal normal vector $v \in N_C^P(\bar{x})$. Now from the definition of the proximal normal cone we have that there exist $\sigma > 0$ such that

$$\langle v, x - \bar{x}\rangle \leq \sigma ||x - \bar{x}||^2, \quad \forall x \in C.$$

Thus $\bar{x}$ is the minimum of problem

$$\min \phi(x) = -\langle v, x\rangle + \sigma||x - \bar{x}||^2, \quad \text{subject to} \quad x \in C.$$

Thus using Corollary 3.1 we have that there exists $y \in N_U(F(\bar{x}))$ (we do not change the notation) such that

$$v \in \partial(yF)(\bar{x}) + N_X(\bar{x}).$$

Let us now consider an element $v \in N_C(\bar{x})$. Then there exist sequences $v_k \to v$ and $x_k \to \bar{x}$ ($x_k \in C$) such that $v_k \in N_C^P(x_k)$. Thus for each k there exists $y_k \in N_U(F(x_k))$ such that

$$v_k \in \partial(y_k F)(x_k) + N_X(x_k).$$

Now assume that the sequence $\{y_k\}$ is bounded. Then without loss of generality we can consider that $y_k \to y$. Then using Lemma 3.2 we have

$$v \in \partial(yF)(\bar{x}) + N_X(\bar{x}).$$

Further it is clear that $y \in N_U(F(\bar{x}))$.

In a manner similar to that in Theorem 3.1 we can prove that under the given qualification condition (Q2) the sequence $\{y_k\}$ can never be unbounded. This proves the result. □

Remark 3.3 To have an equality in the above proposition we need to have some additional regularity conditions on the sets and the functions involved. First we have to assume that both X and U are normally regular at $\bar{x}$ and $F(\bar{x})$ respectively. Normal regularity means that at the given point the regular normal cone coincides with basic normal cone. Every convex set is normally regular. Further we also have to assume that the function $(yF)(\bar{x})$ is regular for all $y \in N_U(F(\bar{x}))$. Regularity of a given function at a given point means that the regular subdifferential at that point coincides with the basic subdifferential at that point. For more details on normal regularity of sets and regularity of functions see for example Mordukhovich [8]. The equality in the above proposition then follows in a natural way from the regularity notions. We skip the proof here.

We will end this article by showing that the Theorem 3.1 and Proposition 3.1 can be used to provide a very simple proof of the minimax optimization problem

Theorem 3.2 *Let us consider the following problem*

$$\min \psi(x) = \max\{f_1(x), \ldots f_k(x)\} \quad \textit{subject to} \quad F(x) \in U, \quad x \in X,$$

where each $f_j : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \ldots, k$ *are locally Lipschitz functions,* $F : \mathbb{R}^n \to \mathbb{R}^m$ *be a vector-valued locally Lipschitz function,* X *is a closed subset of* $\mathbb{R}^n$ *and* U *is a closed subset of* $\mathbb{R}^m$*. Let* $\bar{x}$ *be a local minimum of the above problem. Further assume that following qualification condition holds at* $\bar{x}$ *:*

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

Then there exists scalars $\lambda_j \geq 0$, $\sum_{j=1}^{k} \lambda_j = 1$ *and* $\bar{y} \in N_U(F(\bar{x}))$ *such that*

$$0 \in \lambda_1 \partial f_1(\bar{x}) + \ldots + \lambda_k \partial f_k(\bar{x}) + \partial(\bar{y}F)(\bar{x}) + N_X(\bar{x}).$$

Proof. In the first step we observe that the minimization problem posed in the theorem can be equivalently stated as

$$\min \rho(H(x)), \quad \text{subject to} \quad x \in C,$$

where $H(x) = (f_1(x), \ldots, f_k(x))$, $\rho(u) = \max\{u_1, \ldots, u_2\}$ and $C = \{x \in X : F(x) \in U\}$. Thus we have reformulated the given problem in the form of (P2) with $f_0(x) = 0$ for all $x \in \mathbb{R}^n$. Since ρ is a finite convex function we have $\partial^\infty \rho(u) = \{0\}$ for all $u \in \mathbb{R}^k$. Thus the qualification condition (Q) holds automatically for the above problem. Now using Theorem 3.1 we have that there exists $\lambda \in \partial\rho(F(\bar{x}))$ such that

$$0 \in \partial(\lambda H)(\bar{x}) + N_C(\bar{x}).$$

It is a well known fact in convex analysis that $\partial\rho(u) = \{\lambda \in \mathbb{R}^k_+ : \sum_{j=1}^{k} \lambda_j = 1\}$ for all $u \in \mathbb{R}^k$. The result now follows by a simple application of Proposition 3.1. □

Remark 3.4 The optimality condition for the minimax problem involving equality and inequality constraints was established by using the method of metric approximations in Mordukhovich [11] while in the above theorem we treat a more general case by using Theorem 3.1 which has been proved by using the penalty function method. Thus the necessary optimality condition for a very general minimax problem can indeed be deduced via the Rockafellar's penalty function approach used in this article.

Conclusions In this paper we do not claim an original contribution but rather an exposition of the Lagrange multiplier rule at a very general level done through some research. What we demonstrate is that in the finite dimensional setting the Lagrange multiplier rule can be indeed established for a very general class of problems in a very simple manner with very less technical sophistications by using the penalty function approach. We also demonstrate that Rockafellar's [18] penalty approximation scheme is more fundamental than some of the other penalty function approaches that has been mentioned in this article. For example the approaches due to Mcshane [15], Hestenes [6] and recently due to Bertsekas [1] seems to work only for equality and inequality constraints. The important question now is whether a similar scheme can be developed in the infinite dimensional setting. If such an approach is possible then it would indeed unify the theory of optimality conditions. Another advantage of this appoach is pedagogical. Since the approach is simple it can be even taught to graduate students and thus introducing them to the Lagrange multiplier rule for a very general class of problems from which the multiplier rule for a large class of optimization problems can be very easily derived.

Acknowledgements

We are grateful to the anonymus referee whose constructive suggestions have vastly improved the presentation of the paper. A part of the work on this article was completed when the author was visiting the Department of Mathematics and XLIM at the University of Limoges, France. The author is thankful to the University of Limoges for its hospitality. This visit was possible with the financial support of the Indo-French Institute of Mathematics which is gratefully acknowledged.

References

[1] D.P. Bertsekas, *Nonlinear Programming*, Athena Science Publications, 1999.

[2] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Canadian Mathematical Society Series of Monographs and Advanced Texts, A Wiley-Interscience Publication, John Wiley and Sons, Inc., New York, 1983.

[3] F.H. Clarke, *Methods of Dynamic and Nonsmooth Optimization*, CBMS-NSF Regional Conference Series in Applied Mathematics, 57, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1989.

[4] F.H. Clarke, Yu. S. Ledyaev, R.J. Stern and P.R. Wolenski, *Nonsmooth Analysis and Control Theory*, Graduate Texts in Mathematics, 178, Springer-Verlag, New York, 1998.

[5] J. Dutta, Generalized derivatives and nonsmooth optimization, a finite dimensional tour, With discussions and a rejoinder by the author. *Top* 13 (2005) 185–314.

[6] M.R. Hestenes, *Optimization Theory. The Finite Dimensional Case*, Pure and Applied Mathematics, Wiley-Interscience (John Wiley and Sons), New York-London-Sydney, 1975.

[7] A. Kruger and B.S. Morduhovich, Extremal points and the Euler equation in nonsmooth optimization problems, *Dokl. Akad. Nauk BSSR* 24 (1980) 684–687.

[8] B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation, I*, Basic theory. Grundlehren der Mathematischen Wissenschaften, 330, Springer-Verlag, Berlin, 2006.

[9] B.S. Mordukhovich, Penatly functions and necessary optimality conditions for an extremum in nonsmooth and nonconvex optimization problems, *Russian Math Surveys*, 36 (1981) 242–243.

[10] B.S. Mordukhovich, Generalized differential calculus for nonsmooth and set-valued mappings, *J. Math. Anal. Appl.* 183 (1994) 250–288.

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

[12] B.S. Mordukhovich, *Approximation Methods in Problems of Optimization and Control*, Nauka, Moscow, 1988.

[13] B.S. Mordukhovich and Y. Shao, Nonsmooth sequential analysis in Asplund spaces, *Trans. Amer. Math. Society* 348 (1996) 1235–1280.

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

[15] E.J. McShane, The Langrange multiplier rule, *Amer. Math. Monthly* 80 (1973) 922–925.

[16] B.T. Polyak, *Introduction to Optimization*, Optimization Software, Inc, Publications Division, New York, 1987.

[17] R.T. Rockafaller, *Convex Analysis*, Princeton University Press, 1970.

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

[19] R.T. Rockafellar, Extended nonlinear programmingin *Nonlinear optimization and related topics*, G. Di Pillo and F. Giannessi (eds.), Kluwer, Dordrecht, 2000, pp. 381–399.

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

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

[22] Ju. M Volin and G.M. Ostrovskii, The method of penalty functions, and necessary conditions for optimality, *Control Systems* 9 (1971) 243–251.

Manuscript received 3 September 2005
revised 28 December 2005
accepted for publication 3 January 2006

Joydeep Dutta
Department of Mathematics and Statistics, Indian Institute of Technology, Kanpur-208106, India
E-mail address: jdutta@iitk.ac.in