Appl Math Optim 32:1–31 (1995)

Applied Mathematics & Optimization

Exact Barrier Function Methods for Lipschitz Programs*

G. Di Pillo and F. Facchinei

Dipartimento di Informatica e Sistemistica,
Università di Roma "La Sapienza",
Via Buonarroti 12, 00185 Roma, Italy

Communicated by F. H. Clarke

Abstract. The purpose of this paper is twofold. First we consider a class of nondifferentiable penalty functions for constrained Lipschitz programs and then we show how these penalty functions can be employed to solve a constrained Lipschitz program. The penalty functions considered incorporate a barrier term which makes their value go to infinity on the boundary of a perturbation of the feasible set. Exploiting this fact it is possible to prove, under mild compactness and regularity assumptions, a complete correspondence between the unconstrained minimization of the penalty functions and the solution of the constrained program, thus showing that the penalty functions are exact according to the definition introduced in [17]. Motivated by these results, we propose some algorithm models and study their convergence properties. We show that, even when the assumptions used to establish the exactness of the penalty functions are not satisfied, every limit point of the sequence produced by a basic algorithm model is an extended stationary point according to the definition given in [8]. Then, based on this analysis and on the one previously carried out on the penalty functions, we study the consequence on the convergence properties of increasingly demanding assumptions. In particular we show that under the same assumptions used to establish the exactness properties of the penalty functions,

* This research has been partially supported by the National Research Program on "Metodi di Ottimizzazione per le Decisioni", Ministero dell' Università e della Ricerca Scientifica e Tecnologica.

it is possible to guarantee that a limit point at least exists, and that any such limit point is a KKT point for the constrained problem.

Key Words. Constrained optimization, Nonsmooth optimization, Penalty methods, Barrier functions, Extended stationary points.

AMS Classification. 90C30, 49M30, 65K05.

1. Introduction

Nondifferentiable penalty functions for smooth nonlinear programming problems have been widely investigated, from both the algorithmic and the theoretical point of view. In fact, on the one hand, nondifferentiable penalty functions allow a constrained optimization problem to be solved by a single unconstrained minimization of the penalty function (see [1]–[3], [10], [11], [13], [16]–[20], [23], [25], [28], [33]–[35], [37], and [41]), and, in the sequential quadratic programming approach, are a basic tool for establishing global convergence (see, e.g., [24] and [38]). On the other hand it has become more and more apparent that nondifferentiable penalty functions have strong relationships with the foundation of optimization theory and, in particular, with the theory of necessary conditions of optimality (see, e.g., [2], [6], [7], [12], [21], [29]–[31], [37], and [39]). More recently these facts have led to the consideration of nondifferentiable penalty functions also in connection with nonsmooth programs [12], [13], [20], [34], [37], [39].

The aim of the paper is twofold. On the one hand we study the properties of a class of penalty functions for locally Lipschitz constrained problems. We show that, under reasonable assumptions, this class of penalty functions possesses desirable properties. On the other hand we show how these penalty functions can be used to solve a locally Lipschitz constrained problem. We consider several algorithm models and study their convergence properties under increasingly demanding assumptions, ranging from the simple local Lipschitzianity of the problem functions to various compactness and regularity conditions.

Considering in more detail the first of the two topics mentioned above, we note that most of the literature on nondifferentiable penalty functions is devoted to the study of conditions ensuring some kind of correspondence between the local (global) minimizers of the penalty function and the local (global) minimum points of the constrained problem. In general, these correspondence properties hold, under suitable constraint qualifications, only with reference to some given compact set $\mathscr{D}$ that contains the problem solutions of interest. However, it can happen that the level set of the penalty function corresponding to a given value of the penalty parameter and to a given point in $\mathscr{D}$ is not contained in this set. It follows that the sequence produced by an unconstrained algorithm in the minimization of the penalty function, for a given value of the penalty parameter, can be attracted out of the set where the correspondence is established, and this may constitute a limitation to the computational use of penalty methods [17]. On the other hand the introduction of an adjustment rule for the penalty parameter does not guarantee that the sequence of points produced is contained in a compact set, since the adjustment rule could drive the penalty parameter to zero, and unbounded sequences may still be generated. The usual way of dealing with this difficulty is simply to avoid it by assuming *a priori* that the sequence of points produced by the algorithm is bounded, or by making some equivalent assumption. Note that these kind of assumptions often "hide" strong regularity conditions.

To address these kinds of problems we introduce a class of nondifferentiable penalty functions for locally Lipschitz programs which, under suitable assumptions of compactness and regularity of the constraints, is proved to be exact according to the definition given in [17]. In particular, the functions considered incorporate a barrier term which causes the unconstrained minimizers to be located in the interior of a compact perturbation of the feasible set for every value of the penalty parameter. Exploiting this feature, we can establish a complete correspondence between local (global) minimizers of the penalty function and local (global) solutions of the constrained problem. Furthermore, due to the barrier term, the level sets of the penalty functions are compact, so that making use of simple devices in the unconstrained minimization algorithms, we can avoid unbounded sequences, thus overcoming the aforementioned difficulty. To point out the role of the barrier term, we call the exact penalty functions introduced in this paper also *exact barrier functions*. Essentially, exact barrier functions are a tool that can be used to reduce a constrained optimization problem to a *single unconstrained* minimization problem.

The idea of employing barrier terms in an exact penalty framework was first proposed in [19], but with a different perspective and with rather different results. Much closer to the approach adopted in this paper are [15] and [17], where penalty functions (both differentiable and nondifferentiable) for smooth problems are considered. We remark, however, that the results established in this paper with reference to Lipschitz programs, are sharper than those given in [17] even if specialized to the smooth case.

In the second part of the paper, instead, we analyze the behavior of algorithms for the solution of the constrained problem based on the unconstrained minimization of the barrier functions previously introduced. In particular we propose a very general algorithm model and investigate its behavior even when no compactness and regularity assumptions are made on the constrained problem (and hence, even when the assumptions for the exactness of the barrier functions are not fulfilled). In this case it is possible to show that every limit point (if any) of the sequence of points produced is a stationary point of the constrained problem in an extended sense that has been introduced in [8] and that is explained in the next section.

This analysis gives a fairly detailed picture of the behavior of the algorithm, and also clearly shows when and why certain "pathological" behaviors are possible. Based on this analysis it is then possible to add progressively various assumptions which rule out the possibility of undesirable behaviors and introduce some variants of the basic algorithm. In this stage the analysis of the properties of the barrier function carried out in the first part of the paper turns out to be crucial, since the structure of the barrier function and the assumptions employed to establish its exactness are basic to the analysis of the convergence properties of the algorithms considered. In particular, we show that, under the same assumptions used to establish the exactness of the barrier functions, it is possible to guarantee that the algorithms considered generate a sequence of points which admits a limit point at least, that every such limit point is a Karush–Kuhn–Tucker point of the original constrained problem, and that eventually the penalty parameter remains fixed, thus avoiding the numerical instabilities usually associated to sequential (interior) barrier methods.

The remainder of this paper is organized as follows. In Section 2 we introduce the concept of an extended stationary point of a Lipschitz constrained problem, define the class of barrier functions, and recall the notion of exactness adopted. In Section 3 we discuss the assumptions employed and give sufficient conditions for their fulfillment. In Section 4 we establish the main results concerning the barrier functions, by proving that local and global minimizers of the barrier functions are local and global solutions of the constrained problem. In Section 5 we give additional results concerning the relationships of local solutions and stationary points of the constrained problem and the barrier functions. Finally, in Section 6 we describe some algorithm models based on unconstrained minimization algorithms, which include automatic adjustment rules for the penalty coefficient, and study their convergence properties.

We conclude this section by providing a list of the notation employed:

Let v be a vector in $\mathbb{R}^s$, we denote by v' its transpose and by v_+ the vector whose ith component is $\max(0, v_i)$. If K is a subset of $\{1, \ldots, s\}$, we denote by v_K the subvector with components v_i such that $i \in K$. Furthermore, for simplicity, we adopt the convention that (g, h) stands for $(g', h')'$ and analogously for other couples of vectors.

$\|\cdot\|_q: \mathbb{R}^s \to \mathbb{R}$, for $q \in [1, \infty]$ is the l_q-norm in $\mathbb{R}^s$, while $\|\cdot\|_q^0$ indicates the dual norm.

$B_q(\bar{x}; \delta) := \{x \in \mathbb{R}^s: \|x - \bar{x}\|_q \leq \delta\}$, while, with obvious notation, $B_q^0(\bar{x}; \delta) := \{x \in \mathbb{R}^s: \|x - \bar{x}\|_q^0 \leq \delta\}$.

Let $\mathscr{B}$ be a set of points in $\mathbb{R}^s$. We denote by $\mathscr{B}^c$ its complement in $\mathbb{R}^s$, by $\operatorname{co} \mathscr{B}$ its convex hull, by $\partial \mathscr{B}$ its boundary, by $\overline{\mathscr{B}}$ or by $\operatorname{cl} \mathscr{B}$ its closure, and by $N_{\mathscr{B}}(x)$ it normal cone at x. Furthermore, if $\mathscr{B}$ is closed and convex, $\operatorname{Nr} \mathscr{B}$ is its (unique) least euclidean norm point; if $\mathscr{B}$ is not convex, the least euclidean norm point is not necessarily unique, in this case $\operatorname{Nr} \mathscr{B}$ indicates any such point. Finally $\operatorname{dist}_q(x|\mathscr{B})$ denotes the distance of the point x from the set $\mathscr{B}$, i.e.,

$$\operatorname{dist}_q(x|\mathscr{B}) := \inf_{z \in \mathscr{B}} \|x - z\|_q.$$

Let $\varphi: \mathbb{R}^s \to \mathbb{R}$ be a locally Lipschitz function. $\partial\varphi(x)$ denotes the generalized gradient of Clarke. The generalized gradient of φ relative to a set $\mathscr{B}$, denoted by $\partial|_{\mathscr{B}}\varphi(x)$, is defined by

$$\partial|_{\mathscr{B}}\varphi(x) := \{\zeta \in \mathbb{R}^s: \zeta \text{ is a limit point of } \zeta_i \in \partial\varphi(x_i), x_i \to x, x_i \in \mathscr{B}\}.$$

We recall that the following relation holds for the Clarke generalized directional derivative $\varphi^0(x, d)$:

$$\varphi^0(x, d) = \max\{d'\xi, \xi \in \partial\varphi(x)\}.$$

Analogously, if Φ is a vector-valued function, i.e., $\Phi: \mathbb{R}^s \to \mathbb{R}^t$, $\partial\Phi(x)$ and $\partial|_{\mathscr{B}}\Phi(x)$, denote, respectively, the generalized jacobian and the generalized jacobian relative to $\mathscr{B}$ of Φ (see [12], [26], and [27]).

2. Problem Formulation and Basic Definitions

We consider the following Lipschitz programming problem:

$$\text{(P)} \quad \min f(x),$$
$$g(x) \le 0,$$
$$h(x) = 0,$$

where $f: \mathbb{R}^n \to \mathbb{R}$, $g: \mathbb{R}^n \to \mathbb{R}^m$, $h: \mathbb{R}^n \to \mathbb{R}^p$, $p \le n$, and we assume that f, g, and h are locally Lipschitz on $\mathbb{R}^n$.

We denote by $\mathscr{F}$ the feasible set of Problem (P):

$$\mathscr{F} := \{x \in \mathbb{R}^n: g(x) \le 0, h(x) = 0\};$$

moreover, we define the index sets:

$$I_0(x) := \{i: g_i(x) = 0\}, \qquad I_\pi(x) := \{i: g_i(x) \ge 0\}.$$

The algorithms we present in this paper are designed to locate stationary points of Problem (P) in an extended sense, so that, to a certain extent, sensible results can be obtained without making any assumption concerning the regularity or feasibility of Problem (P). Hence we adopt terminology similar to that proposed by Burke (see [5] and [8]) and say that $\bar{x} \in \mathbb{R}^n$ is an *extended stationary point* for Problem (P) if either of the following conditions is satisfied:

(i) $\bar{x} \in \mathscr{F}$ and $(\lambda, \mu) \in \mathbb{R}^{m+p}$ exists such that

$$0 \in \partial f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \,\partial g_i(\bar{x}) + \sum_{j=1}^{p} \mu_j \,\partial h_j(\bar{x}),$$

$$\lambda_i \ge 0, \quad \lambda_i g_i(\bar{x}) = 0, \quad i = 1, \dots, m;$$

(ii) $\bar{x} \in \mathscr{F}$ and $0 \ne (\lambda, \mu) \in \mathbb{R}^{m+p}$ exists such that

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

$$\lambda_i \ge 0, \quad \lambda_i g_i(\bar{x}) = 0, \quad i = 1, \dots, m;$$

(iii) $\bar{x} \notin \mathscr{F}$ and $0 \in \partial\|g_+(\bar{x}), h(\bar{x})\|_q$.

Points of type (i) are generally called Karush–Kuhn–Tucker (KKT) points, while points of type (ii) are known as Fritz John (FJ) points. It is well known that if $\bar{x}$ is a local minimum point of Problem (P), then it is either a KKT point or an FJ point.

However, FJ points cause much trouble to minimization algorithms, so that often suitable conditions (regularity conditions) are assumed to hold in order to exclude the existence of FJ points. Actually, if we do not make any regularity assumption on Problem (P), we cannot even be sure that a feasible point exists. In this event it may be useful to see Problem (P) as composed of two parts: (a) the feasibility problem and (b) the minimization problem. The feasibility problem can be written as

$$\min_{x \in \mathbb{R}^n} \|g_+(x), h(x)\|_q.$$

Hence, points satisfying (iii) can be viewed as stationary points for the feasibility problem, and are therefore relevant to Problem (P): we call them infeasible stationary points (see [5], [9], and [8] for a more complete discussion on this point of view).

We now introduce the barrier function along with some related definitions.

Let α be a positive constant and consider the set

$$\mathscr{N}_\alpha := \{x \in \mathbb{R}^n : \|g_+(x), h(x)\|_q < \alpha\}.$$

Consider the function:

$$b(x) := \alpha - \|g_+(x), h(x)\|_q.$$

We have that $b(x) > 0$ for all $x \in \mathscr{N}_\alpha$, and hence, for $x \in \mathscr{N}_\alpha$, we can define the vector functions $\hat{g}(x)$ and $\hat{h}(x)$ with components given, respectively, by

$$\hat{g}_i(x) := \frac{g_i(x)}{b(x)}, \qquad i = 1, \dots, m, \tag{1}$$

$$\hat{h}_j(x) := \frac{h_j(x)}{b(x)}, \qquad j = 1, \dots, p. \tag{2}$$

We associate to Problem (P) the following class of barrier functions:

$$Z_q(x; \varepsilon) := f(x) + \frac{1}{\varepsilon}\|\hat{g}_+(x), \hat{h}(x)\|_q,$$

where $\varepsilon > 0$. In particular, by choosing $q \in [1, \infty)$, we have

$$Z_q(x; \varepsilon) = f(x) + \frac{1}{\varepsilon}\left[\sum_{i=1}^{m} \left(\hat{g}_{+_i}(x)\right)^q + \sum_{j=1}^{p} |\hat{h}_j(x)|^q\right]^{1/q},$$

while, for $q = \infty$, we obtain the function

$$Z_\infty(x; \varepsilon) = f(x) + \frac{1}{\varepsilon} \max\left[\hat{g}_{+_1}(x), \dots, \hat{g}_{+_m}(x), |\hat{h}_1(x)|, \dots, |\hat{h}_p(x)|\right].$$

For any given $\varepsilon > 0$, consider the essentially unconstrained problem

$$\text{(U)} \quad \min Z_q(x; \varepsilon), \qquad x \in \mathscr{N}_\alpha.$$

We say that $Z_q(x; \varepsilon)$ is an *exact barrier function with respect to the set* $\mathcal{N}_\alpha$ if the following three properties are satisfied:

(P1) *for all* $\varepsilon > 0$ *and all* $x_0 \in \mathcal{N}_\alpha$, *the level set*

$$\mathcal{L}(x_0, \mathcal{N}_\alpha; \varepsilon) := \{x: x \in \mathcal{N}_\alpha, Z_q(x; \varepsilon) \leq Z_q(x_0; \varepsilon)\}$$

is compact.

A threshold value $\varepsilon^* > 0$ exists such that:

(P2) *For all* $\varepsilon \in (0, \varepsilon^*]$, *any local solution of the unconstrained Problem* (U) *is a local solution of the original constrained Problem* (P).

(P3) *For all* $\varepsilon \in (0, \varepsilon^*]$, *any global solution of Problem* (P) *is a global solution of Problem* (U), *and conversely.*

In particular these properties imply that $Z(x; \varepsilon)$ is *globally exact* according to the definition given in [17]. The interested reader should refer to [17] for a thorough discussion of this as well as other related definitions. Roughly speaking property (P1) excludes the existence of descent paths for $Z_q(x; \varepsilon)$ originating at some $x_0 \in \mathcal{N}_\alpha$ that either cross $\partial\mathcal{N}_\alpha$ or have limit points on $\partial\mathcal{N}_\alpha$. This implies that any descent method for the solution of Problem (U) generates a sequence with limit points belonging to $\mathcal{N}_\alpha$, *for any given* $\varepsilon > 0$ *and for any given starting point* $x_0 \in \mathcal{N}_\alpha$. If these limit points are local (global) solutions of Problem (U), properties (P2) and (P3) ensure that they are also local (global) solutions of Problem (P), provided that ε is sufficiently small.

It must be remarked that the notion of exactness, expressed in terms of properties (P1)–(P3), does not require that the local solutions of problem (P), which are not global solutions, correspond to local minimizers of the exact barrier function. In fact this property does not seem to be required, in practice, to give a meaning to the notion of exactness, since condition (P3) ensures that the global solutions of Problem (P) are preserved. A one-to-one correspondence between local minimizers of the two problems is a stronger property which, however, will be shown to hold for compact sets of local minimizers. To this end we say that $Z_q(x; \varepsilon)$ is *locally exact* at $\bar{x}$, a local minimum point of Problem (P), if for all ε sufficiently small, $\bar{x}$ is a local minimizer of $Z_q(x; \varepsilon)$.

In Sections 4 and 5 we investigate the properties of the barrier function $Z(x; \varepsilon)$ and show that, under suitable assumptions, it is exact with respect to the set $\mathcal{N}_\alpha$.

In Section 6, employing the analysis carried out in Sections 4 and 5, we study some algorithms which aim at solving the constrained Problem (P) by the unconstrained minimization of the barrier function $Z(x; \varepsilon)$. First we discuss the behavior of a basic algorithm scheme without making any regularity assumptions; then we introduce various regularity assumptions and study how they influence the behavior of the basic algorithm and of some of its variants.

3. Discussion of the Assumptions

In this section we introduce and comment on the two main assumptions that are invoked to establish some of the results presented later.

To introduce these assumptions we first need a definition.

Definition 1. We say that condition CQq is satisfied at x if

$$0 \notin \partial|_{\mathscr{F}^c}\|g_+(x), h(x)\|_q.$$

This regularity condition, which proved to be very useful in the context of penalization methods, was first introduced in [13] and further employed in [14] and [21]. We see in this section how conditon CQq relates to other, more usual, regularity conditions for Problem (P).

We now introduce our assumptions. They are employed in the next section to establish the exactness of $Z(x; \varepsilon)$, and, in the last section, to ensure some convergence properties of the algorithm studied there.

Assumption A1. *The set $\bar{\mathscr{N}}_\alpha$ is compact.*

Assumption A2. *Condition* CQq *is satisfied for all x in $\bar{\mathscr{N}}_\alpha$.*

Regarding Assumption A1, we observe that the existence of a compact perturbation of the feasible set is a mild requirement on the constraint functions; in particular the next proposition gives conditions ensuring that Assumption A1 is satisfied for any positive α.

Proposition 1. *Assume that one of the following conditions is satisfied:*

(i) *A function $g_i(x)$ exists such that $\lim_{\|x\|\to\infty} g_i(x) = \infty$.*
(ii) *A function $h_j(x)$ exists such that $\lim_{\|x\|\to\infty} |h_j(x)| = \infty$.*
(iii) *Index sets I and J, not both empty, exist such that the functions $g_i(x)$, $i \in I$, are convex, the functions h_j, $j \in J$, are affine, and the set $\{x: g_i(x) \leq 0, i \in I; h(x) = 0, j \in J\}$ is compact.*

Then, for any $\alpha > 0$, the set $\bar{\mathscr{N}}_\alpha$ is compact.

Proof. If either (i) or (ii) holds, the level sets of $\|g_+(x), h(x)\|_q$ are bounded and the assertion follows immediately. With regard to point (iii), note that we can write

$$\bar{\mathscr{N}}_\alpha \subseteq \left\{x \in \mathbb{R}^n: \|g_{I_+}(x), h_J(x)\|_q \leq \alpha\right\},$$

where $\|g_{I_+}(x), h_J(x)\|_q$ is a convex function. Then the assertion is a consequence of Theorem 24 of [22], which states the preservation of boundedness of convex sets given by convex inequalities, in correspondence to perturbation of the right-hand side. □

We now consider Assumption A2. We first state two results that show that the condition $0 \notin \partial|_{\mathscr{F}^c}\|g(x)_+, h(x)\|_q$ is "stable."

Proposition 2. *Suppose that Condition* CQq *holds at a given point* $x \in \mathbb{R}^n$. *Then Condition* CQq *holds in a neighborhood of* x.

Proof. The assertion easily follows from the upper-semicontinuity of the point-to-set map $x \mapsto \partial|_{\mathscr{F}^c}\|g_+(x), h(x)\|_q$ (see [27]). □

Proposition 3. *Suppose that*:

(i) *An* $\bar{\alpha} > 0$ *exists such that* $\bar{\mathscr{N}}_{\bar{\alpha}}$ *is compact.*
(ii) *Condition* CQq *holds on* $\mathscr{F}$.

Then a positive α, $\alpha \leq \bar{\alpha}$ *exists such that Assumption* A2 *holds.*

Proof. Let $\bar{\alpha} > 0$ be such that $\bar{\mathscr{N}}_{\bar{\alpha}}$ is compact. We observe first that for any positive α, $\alpha \leq \bar{\alpha}$, the set $\bar{\mathscr{N}}_\alpha$ is compact and $\mathscr{F} \subseteq \bar{\mathscr{N}}_\alpha$. Suppose now that the conclusion is false. Then for any k we can find an $\alpha_k \leq \min[\bar{\alpha}, 1/k]$ and a point $x_k \in \bar{\mathscr{N}}_{\alpha_k} \subseteq \bar{\mathscr{N}}_\alpha$ such that $0 \in \partial|_{\mathscr{F}^c}\|g_+(x_k), h(x_k)\|_q$. Since $\bar{\mathscr{N}}_{\bar{\alpha}}$ is compact, we can assume, without loss of generality, that the sequence $\{x_k\}$ admits a limit point $\bar{x} \in \bar{\mathscr{N}}_{\bar{\alpha}}$; furthermore, as we have $\|g_+(x_k), h(x_k)\|_q \leq \alpha_k$, $\bar{x} \in \mathscr{F}$. However, then, by (ii) and Proposition 2, $0 \notin \partial|_{\mathscr{F}^c}\|g(x)_+, h(x)\|_q$ in a neighborhood of $\bar{x}$ and this is absurd. □

Proposition 3 gives a sufficient condition for the existence of a compact perturbation of the feasible set where Condition CQq is satisfied; now we turn to conditions ensuring that it is satisifed at a given point. We first observe that if $x \in \mathring{\mathscr{F}}$, then we always have $0 \notin \partial|_{\mathscr{F}^c}\|g(x)_+, h(x)\|_q$, in fact,

$$\partial|_{\mathscr{F}^c}\|g_+(x), h(x)\|_q = \varnothing \qquad \text{if} \quad x \in \mathring{\mathscr{F}}.$$

On the contrary, if $x \notin \mathring{\mathscr{F}}$, this condition is not trivial, however, we can give sufficient conditions for its fulfillment. To this end we consider three Mangasarian–Fromovitz-type conditions that are strictly related to Condition CQq.

Condition C1. $q \in (1, \infty)$ and

$$0 \notin [\partial(g(x), h(x)]'(\beta, \gamma),$$

where

$$\beta_i := g_{+_i}(x)^{(q-1)}, \qquad i = 1, \dots, m,$$

$$\gamma_j := \operatorname{sign}(h_j(x))|h_j(x)|^{(q-1)}, \qquad j = 1, \dots, p.$$

Condition C2.

(i) For any $(\xi_1,\dots,\xi_m,\zeta_1,\dots,\zeta_p)' \in \partial|_{\mathscr{F}^c}[g(x),h(x)]$, d exists such that

$$\xi_i' d < 0, \qquad i \in I_\pi(x),$$
$$\zeta_j' d = 0, \qquad j = 1,\dots,p.$$

(ii) $(\zeta_1,\dots,\zeta_p)$ is of maximal rank.

By alternative theorems Condition C2 can also be rewritten as:

Condition C2. For any $t_1,\dots,t_p \in \{-1,1\}$ and for any $(\xi_1,\dots,\xi_m,\zeta_1,\dots,\zeta_p)' \in \partial|_{\mathscr{F}^c}(g(x),h(x))$,

$$0 \notin \operatorname{co}\{\xi_i, i \in I_\pi(x); t_j\zeta_j, j = 1,\dots,p\}.$$

Condition C3. $d \in \mathbb{R}^n$ exists such that

(i)
$$g_i^0(x;d) < 0, \qquad i \in I_\pi(x),$$
$$h_j^0(x;d) = 0, \qquad j = 1,\dots,p.$$

(ii) $\partial h(x)$ is of maximal rank.

Condition C2 is an extension to possibly infeasible points of a condition given in [26]. If the constraints are continuously differentiable Conditions C2 and C3 are equivalent and, moreover, on the feasible set, they coincide with the well-known Mangasarian–Fromovitz constraint qualification. In the case of Lipschitz functions it is easy to verify, recalling the relationship between the generalized gradient and the generalized directional derivative, that Condition C3 implies Condition C2, and that, if x is not feasible and $q \in (1,\infty)$, Condition C2 implies Condition C1. Conditions C1–C3, or similar conditions, have already been used by various authors, see, e.g., [4] and [32].

In order to establish the relationships between these conditions and Assumption A2 we need the following result.

Proposition 4. *Let $\mathscr{A} \subseteq \mathbb{R}^m \times \mathbb{R}^p$ de defined by $A := \{(g,h) \in \mathbb{R}^m \times \mathbb{R}^p : \|g_+, h\|_q > 0\}$. Then*

(i) *For any $q \in (1,\infty)$ and $(g,h) \in \mathscr{A}$, $\|g_+,h\|_q$ is continuously differentiable and*

$$\nabla \|g_+,h\|_q = (\beta,\gamma),$$

where

$$\beta_i := \frac{(g_{+_i})^{(q-1)}}{\|g_+,h\|_q^{(q-1)}}, \qquad i = 1,\dots,m,$$

$$\gamma_j := \frac{\operatorname{sign}(h_j)|h_j|^{(q-1)}}{\|g_+,h\|_q^{(q-1)}}, \qquad j = 1,\dots,p.$$

(ii) *For any* (g, h) *such that* $\|g_+, h\|_q = 0$, *we have that*

$$0 \notin \partial|_{\mathscr{A}}\|g_+, h\|_q.$$

(iii) *For any* (g, h) *such that* $\|g_+, h\|_q = 0$, *we have that if*

$$\begin{pmatrix} \beta \\ \gamma \end{pmatrix} \in \partial\|g_+, h\|_q \qquad \textit{with} \quad \beta \in \mathbb{R}^m, \quad \gamma \in \mathbb{R}^p,$$

then $\beta_i \geq 0$ *for every* $i = 1, \ldots, m$ *and* $\beta_i = 0$ *for every* i *such that* $g_i < 0$.

Proof. Points (i) and (ii) have been proved, for example, in [13]. Regarding point (iii) we observe that

$$\|g_+, h\|_q = \operatorname{dist}_q\{(g, h)|\mathbb{R}^m_- \times 0_p\},$$

so that the assertion follows by the known formula

$$\partial \operatorname{dist}_q\{(g, h)|\mathbb{R}^m_- \times 0_p\} = N_{\mathbb{R}^m_- \times 0_p}(g, h) \cap B^0_q((g, h); 1). \qquad \square$$

In the next two propositions we state the relationships of Conditions C1–C3 with Condition CQq (and hence with Assumption A2).

Proposition 5. *Let* $x \in \mathbb{R}^n$ *be given, then, if* $q \in (1, \infty)$ *and* $x \notin \mathscr{F}$, *Condition* C1 *implies Condition* CQq.

Proof. If $x \notin \mathscr{F}$ we have that $\partial|_{\mathscr{F}^c}\|g_+(x), h(x)\|_q = \partial\|g_+(x), h(x)\|_q$. Furthermore, taking into account that $q \in (1, \infty)$, we have, by Proposition 4(i) and by known calculus rules (see e.g., [12]), that the generalized gradient of $\partial\|g_+(x), h(x)\|_q$ is given by

$$\frac{1}{\|g_+(x), h(x)\|_q}[\partial(g(x), h(x)]'(\beta, \gamma),$$

where

$$\beta_i := \left(g_{+_i}(x)\right)^{(q-1)}, \qquad i = 1, \ldots, m,$$

$$\gamma_j := \operatorname{sign}(h_j(x))|h_j(x)|^{(q-1)}, \qquad j = 1, \ldots, p,$$

from which the assertion follows immediately. $\square$

Proposition 6. *Let* $x \in \mathbb{R}^n$ *and* $q \in [1, \infty]$ *be given, and assume that Condition C2 is satisfied at* x. *Furthermore, let us assume that one of the following conditions is satisfied*:

(i) $q \in (1, \infty)$.
(ii) $p = 0$ (*there are only inequality constraints*).
(iii) g *and* h *are continuously differentiable in a neighborhood of* x.

Then condition CQq *holds at* x.

Proof. The proposition is proved in [14]. Note that there we assumed x to be feasible, but this fact is never used in the proof, so that the stronger version stated here holds. □

Remark 1. It can be easily shown that there are cases where neither Condition C3 nor Condition C2 are satisfied and yet $0 \notin \partial|_{\mathscr{F}^c}\|g(x)_+, h(x)\|_q$. This is proved by the following example: $\mathscr{F} = \{x \in \mathbb{R}: x^3 \leq 0, x = 0\}$. In $\bar{x} = 0$ Conditions C2 and C3 do not hold because the gradient of the inequality constraint vanishes at $\bar{x}$. On the other hand we have

$$\lim_{x \uparrow 0} \nabla[(x^3)_+ + |x|] = -1, \qquad \lim_{x \downarrow 0} \nabla[(x^3)_+ + |x|] = 1,$$

so that Condition CQ_1 is satisfied at $\bar{x}$. We see, then, that Condition CQq is weaker than Conditions C2 and C3, and hence, in particular, if we assume that the problem functions are continuously differentiable, it is weaker than the Mangasarian–Fromovitz regularity condition.

Finally, in the next proposition we give conditions ensuring that Condition CQq holds on $\mathbb{R}^n$.

Proposition 7. *Suppose that*:

(i) *The functions* $g_i(x)$, $i = 1, \ldots, m$, *are convex and a point* $\bar{x} \in \mathscr{F}$ *exists such that* $g(\bar{x}) < 0$.
(ii) *The functions* $h_j(x)$, $j = 1, \ldots, p$, *are affine and* $\nabla h(x)$ *has full rank.*

Then Condition CQq *holds for any* $x \in \mathbb{R}^n$.

Proof. It can be easily verified that $\|g_+(x), h(x)\|_q$ is convex. Furthermore, its least value is 0 and it is attained if and only if $x \in \mathscr{F}$. Hence, if $x \notin \mathscr{F}$, from the convexity assumptions, we have that $0 \notin \partial\|g_+(x), h(x)\|_q$, and the thesis follows from the fact that the generalized gradient relative to a set is contained in the generalized gradient.

If $x \in \partial\mathscr{F}$, we can write

$$\xi_i'(\bar{x} - x) + g_i(x) \leq g(\bar{x}) < 0, \qquad \forall \xi_i \in \partial g_i(x), \quad i = 1, \ldots, m,$$

$$\nabla h_j(x)'(\bar{x} - x) = 0, \qquad j = 1, \ldots, p,$$

and hence, taking into account the definition of $I_0(x)$,

$$\begin{aligned} &\xi_i'(\bar{x} - x) < 0, \qquad \forall \xi_i \in \partial g_i(x), \quad i \in I_0(x), \\ &\quad \nabla h_j(x)'(\bar{x} - x) = 0, \qquad j = 1, \ldots, p. \end{aligned} \tag{3}$$

By Theorem 1.4 of [27] we have

$$\partial|_{\mathscr{F}^c}\|g_+(x), h(x)\|_q$$

$$\subseteq \mathrm{co}\left\{ \sum_{i\in I_0(x)} \beta_i \xi_i + \sum_{i=j}^{p} \gamma_i \zeta_j : (\beta, \gamma) \in \partial|_{\mathscr{A}}\|g_+, h\|_q|_{\substack{g=g(x)\\ h=h(x)}}, \right.$$

$$\left. \xi_i \in \partial g_i(x), \zeta_j = \nabla h_j(x) \right\}, \tag{4}$$

where $\mathscr{A} := \{(g, h) \in \mathbb{R}^m \times \mathbb{R}^p: \|g_+, h\|_q > 0\}$. The thesis now follows by known theorems of alternative, taking into account the definition of convex hull, Proposition 4, (3), and (4). □

We end this section by studying the connections between condition CQq and the FJ conditions for Problem (P).

Proposition 8.

(i) *Let x be a feasible point for Problem* (P) *such that Condition* CQq is not *satisfied in x, and suppose that the equality constraint h is continuously differentiable around x. Then x is an FJ point of Problem* (P).

(ii) *Let x be an infeasible point for Problem* (P). *Suppose that Condition* CQq is not *satisfied in x, then x is an infeasible stationary point of Problem* (P) *and vice versa.*

Proof. (i) We first note that we can write

$$\|g_+(x), h(x)\|_q = \|g_+(x), h_+(x), (-h)_+(x)\|_q = F \circ w(x),$$

where $w(x): \mathbb{R}^n \to \mathbb{R}^{m+2p}$ is defined by

$$w(x) := (g(x), h(x), -h(x));$$

while $F(w): \mathbb{R}^{m+2p} \to \mathbb{R}$ is given by

$$F(w) := F(g, h^1, h^2) = \|g_+, h_+^1, h_+^2\|_q.$$

We now show that

$$\partial_{\mathscr{F}^c}\|g_+(x), h(x)\|_q$$

$$\subseteq \left\{ \eta \in \mathbb{R}^n: \eta = \sum_{i=1}^{m} \beta_i \, \partial|_{\mathscr{F}^c} g_i(x) \right.$$

$$\left. + \sum_{j=1}^{p} (\gamma_i - \delta_i)\nabla h_j(x), (\beta, \gamma, \delta) \in \partial|_{\mathscr{A}} F(g(x), h(x), -h(x)) \right\}, \tag{5}$$

where, we recall, $\mathscr{A}$ indicates the subset of $\mathbb{R}^{m+2p}$ where the function F takes positive values, i.e.,

$$\mathscr{A} = \{w \in \mathbb{R}^{m+2p} : F(w) > 0\}.$$

To prove relation (5) we observe that $F(w)$ is isotone according to the definition given in [40] (i.e., $F(w_1) \le F(w_2)$ whenever $w_1 \le w_2$) so that we can apply Theorem 3.17 of [40] to conclude that, for any $y \in \mathscr{F}^c$,

$$\begin{aligned} \partial \|g_+(y), h(y)\|_q &= \partial(F \circ w)(y) \\ &\subseteq \Bigg\{ \eta \in \mathbb{R}^n : \eta = \sum_{i=1}^{m} \beta_i \, \partial g_i(y) + \sum_{j=1}^{p} (\gamma_i - \delta_i)\nabla h_j(x), \\ &\qquad (\beta, \gamma, \delta) \in \partial F(g(y), h(y), -h(y)) \Bigg\}. \end{aligned} \tag{6}$$

Hence, using (6) and the definition of generalized gradient relative to a set, we obtain (5).

Since we are supposing that Condition CQq is not satisfied at x, we have $0 \in \partial_{\mathscr{F}^c} \|g_+(x), h(x)\|_q$, which in turn, by (5) and by $\partial_{\mathscr{F}^c} g_i(x) \subseteq \partial g_i(x)$, implies

$$0 \in \sum_{i=1}^{m} \lambda_i \, \partial g_i(x) + \sum_{j=1}^{p} \mu_j \nabla h_j(x),$$

where we have set $\lambda = \beta$ and $\mu = (\gamma - \delta)$. Hence, to prove that x is an FJ point it remains to show that (a) $\lambda \ge 0$, (b) $\lambda_i g_i(x) = 0$ for $i = 1, \dots, m$, and (c) $(\lambda, \mu) \ne 0$. We then proceed to show that (a)–(c) hold. (a) and (b) follow immediately by the definition of $\lambda\ (= \beta)$ (see (5) and Proposition 4(iii)). Regarding (c), using the special structure of F (more precisely the fact that if $h^1 > 0$, then $h^2 < 0$ while if $h^2 > 0$, then $h^1 < 0$), the definition of a generalized gradient relative to a set, Proposition 4(ii), and the explicit expression of the subgradient of F at points belonging to $\mathscr{A}$ (see, e.g., [13]), it can be verified that either $\beta \ne 0$ or $\gamma \ne \delta$, so that (c) follows.

(ii) This point follows immediately by noting that if x is not feasible, then

$$\partial_{\mathscr{F}^c} \|g_+(x), h(x)\|_q = \partial \|g_+(x), h(x)\|_q. \qquad \square$$

Remark 2. Proposition 8 clearly shows that the only feasible points where condition CQq can possibly fail are the FJ points. However, we note that the set of FJ points is, in general, strictly larger than the set of points which violate condition CQq. This is shown, for example, by the example considered in the Remark 1. This example may seem exceptional since the feasible region is made up of just one point ($x = 0$) and, for any objective function, $x = 0$ is also a KKT point. Hence we also consider another example:

$$\min x_2,$$
$$g_1(x) = x_1^2 + x_2^2 - 1 \leq 0,$$
$$g_2(x) = -x_1^2 - 2x_2^2 + 1 \leq 0,$$

and the feasible point $\bar{x} = (1, 0)$. Obviously $\bar{x}$ is not a minimum point, but the FJ conditions are satisfied with $\lambda = (1, 1)$. On the other hand, it can be easily verified that $\partial|_{\mathscr{F}^c}\|g_{1+}(\bar{x}), g_{2+}(\bar{x})\|_q$ is equal to $\{(2, 0), (-2, 0)\}$ and hence Condition CQq is satisfied. Note that $\bar{x}$ is neither a minimum point nor a KKT point.

4. Exactness of $Z_q(x; \varepsilon)$

In this section we show that, under suitable assumptions, the barrier function $Z_q(x; \varepsilon)$ is exact, in the sense that properties (P1)–(P3) stated in Section 2 are satisfied. More precisely, in this section, and in the following one, we always assume that Assumptions A1 and A2 are satisfied.

With regard to property (P1) it can be easily verified that it is satisfied, by construction, by $Z_q(x; \varepsilon)$ when Assumption A1 holds. Therefore, the main concern of this section is that of establishing properties (P2) and (P3).

In the next proposition we show that a threshold value $\varepsilon^* > 0$ exists such that, for every $\varepsilon \in (0, \varepsilon^*]$, the function $Z_q(x; \varepsilon)$ has no (unconstrained) stationary points in $\mathscr{N}_\alpha \setminus \mathscr{F}$.

Proposition 9. *A threshold value $\varepsilon^* > 0$ exists such that, for every $\varepsilon \in (0, \varepsilon^*]$, the function $Z_q(x; \varepsilon)$ has no stationary points in $\mathscr{N}_\alpha \setminus \mathscr{F}$.*

Proof. We proceed by contradiction. If the assertion is false, for any integer k an $\varepsilon_k \leq 1/k$ and a point $x_k \in \mathscr{N}_\alpha \setminus \mathscr{F}$ exist such that

$$0 \in \partial Z_q(x_k; \varepsilon_k). \tag{7}$$

Since $\bar{\mathscr{N}}_\alpha$ is compact, a convergent subsequence (relabel it $\{x_k\}$) exists such that

$$\lim_{k\to\infty} x_k = \tilde{x} \in \bar{\mathscr{N}}_\alpha \setminus \mathring{\mathscr{F}}.$$

We now show that a constant $C > 0$ exists such that

$$\|\hat{\eta}\|_q \geq C, \qquad \forall \hat{\eta} \in \partial\|\hat{g}_+(x), \hat{h}(x)\|_q, \quad \forall x \in \mathscr{N}_\alpha \setminus \mathscr{F}. \tag{8}$$

If (8) were false, sequences $\{y_k\}$ and $\{\hat{\eta}_k\}$ would exist such that

$$\|\hat{\eta}_k\|_q \to 0, \qquad \hat{\eta}_k \in \partial\|\hat{g}_+(y_k), \hat{h}(y_k)\|_q, \quad y_k \in \mathscr{N}_\alpha \setminus \mathscr{F}. \tag{9}$$

However, by Proposition 2.3.14 of [12] and Theorem 3.2 of [36],

$$\partial \|\hat{g}_+(y), \hat{h}(y)\|_q \subseteq \frac{\alpha}{b(y)^2} \partial \|g_+(y), h(y)\|_q. \tag{10}$$

Since $\bar{\mathcal{N}}_\alpha$ is compact, we can assume, without loss of generality, $\{y_k\} \to \tilde{y} \in \bar{\mathcal{N}}_\alpha$. Now, noting that

$$\frac{\alpha}{b(y)^2} \geq \frac{1}{\alpha} > 0, \qquad \forall y \in \mathcal{N}_\alpha,$$

and noting that (10) implies

$$\eta_k \in \frac{\alpha}{b(y_k)^2} \partial \|g_+(y_k), h(y_k)\|_q,$$

we can conclude by (9) that

$$0 \in \partial|_{\mathcal{F}^c} \|g_+(\tilde{y}), h(\tilde{y})\|_q,$$

thus contradicting Assumption A2. Hence (8) holds.

Recalling that

$$\partial Z_q(x_k; \varepsilon_k) \subseteq \partial f(x_k) + \frac{1}{\varepsilon_k} \partial \|\hat{g}_+(x_k), \hat{h}(x_k)\|_q$$

and that the generalized gradient of a locally Lipschitz function is bounded on bounded sets, we get a contradiction from (7) and (8) for sufficiently large values of k. □

We can now state the following theorem:

Theorem 10. *A threshold value $\varepsilon^* > 0$ exists such that the function $Z_q(x; \varepsilon)$ satisfies Property* (P2).

Proof. The assertion easily follows from Proposition 9, recalling that any local minimum point of $Z_q(x; \varepsilon)$ must be a stationary point of it, and taking into account that any local minimum point of $Z_q(x; \varepsilon)$ belonging to $\mathcal{F}$ is also a local minimum point of Problem (P). □

Let us denote by $\mathcal{G}$ the set of global minimum points of Problem (P), that is, the set of points $\hat{x}$ such that

$$f(\hat{x}) \leq f(x), \qquad \forall x \in \mathcal{F}.$$

Also, we denote by $\hat{f}$ the value of $f(x)$ on $\mathcal{G}$, that is,

$$\hat{f} := f(\hat{x}), \qquad \hat{x} \in \mathcal{G}.$$

The next two theorems jointly establish property (P3).

Theorem 11. *A threshold value $\varepsilon^* > 0$ exists such that, for every $\varepsilon \in (0, \varepsilon^*]$, if $x_\varepsilon \in \mathcal{N}_\alpha$ is a global minimum point of $Z_q(x; \varepsilon)$ on $\mathcal{N}_\alpha$, we have $x_\varepsilon \in \mathcal{G}$.*

Proof. We proceed by contradiction. If the assertion is false, for any integer k an $\varepsilon_k \leq 1/k$ and a point x_k, which is a global minimum point of $Z_q(x; \varepsilon_k)$ on $\mathcal{N}_\alpha$, but does not belong to $\mathcal{G}$, exist. It follows that, for all $\hat{x} \in \mathcal{G}$, we have

$$Z_q(x_k; \varepsilon_k) \leq Z_q(\hat{x}; \varepsilon_k) = f(\hat{x}) = \hat{f}. \tag{11}$$

However, the points x_k are stationary points of $Z_q(x; \varepsilon_k)$, hence, by Proposition 9, we have that, for sufficiently large values of k,

$$x_k \in \mathcal{F} \quad \Rightarrow \quad Z_q(x_k; \varepsilon_k) = f(x_k).$$

Then, by (11), $f(x_k) \leq \hat{f}$, contradicting the assumption $x_k \notin \mathcal{G}$. □

Theorem 12. *A threshold value $\varepsilon^* > 0$ exists such that, for all $\varepsilon \in (0, \varepsilon^*]$, any $\hat{x} \in \mathcal{G}$ is a global minimum point of $Z_q(x; \varepsilon)$ on $\mathcal{N}_\alpha$.*

Proof. By construction of $Z_q(x; \varepsilon)$, if $\hat{x} \in \mathcal{G} \subseteq \mathcal{N}_\alpha$, we have

$$Z_q(\hat{x}; \varepsilon) = f(\hat{x}) = \hat{f}. \tag{12}$$

Now let ε^* be the number considered in the preceding theorem, and let $\varepsilon \in (0, \varepsilon^*]$. Then, by Theorem 11, a global minimizer x_ε of $Z_q(x; \varepsilon)$ on $\mathcal{N}_\alpha$ must satisfy $x_\varepsilon \in \mathcal{G}$, so that $Z_q(x_\varepsilon; \varepsilon) = \hat{f}$. Therefore (12) implies that $\hat{x}$ is also a global minimizer of $Z_q(x; \varepsilon)$ on $\mathcal{N}_\alpha$. □

5. Further Results

As already remarked, the notion of exactness introduced in Section 2 by means of Properties (P1)–(P3), which has been validated by the results established in the preceding section, does not require that local solutions of Problem (P) be local solutions of Problem (U). However, under suitable assumptions, this correspondence can also be established with respect to compact sets of local minimum points of Problem (P). We recall that in this section also we assume that Assumptions A1 and A2 hold.

We first prove this result for a single minimum point, then we extend it to compact sets of minimum points.

Proposition 13. *Let $\bar{x}$ be a local minimum point for Problem* (P). *Then $Z_q(x; \varepsilon)$ is locally exact at $\bar{x}$.*

Proof. By the assumptions made, we can find a $\delta > 0$ such that $B(\bar{x}, \delta) \subseteq \mathcal{N}_\alpha$ and such that $f(x) \geq f(\bar{x})$ for any point x in $B(\bar{x}, \delta) \cap \mathcal{F}$. Consider now the following problem:

$$(\mathrm{P}') \quad \min f(x) + \frac{4}{3\delta^2}(f(\bar{x}) - \hat{f})\left[(x_1 - \bar{x}_1)^2 + \cdots + (x_n - \bar{x}_n)^2 - \left(\frac{\delta}{2}\right)^2\right]_+$$

$$\text{such that} \quad g(x) \leq 0, \quad h(x) = 0,$$

where $\hat{f}$ is the least value of f over $\mathcal{F}$. It is easy to verify that $\bar{x}$ is a global solution of Problem (P′). We denote by $Z'_q(x; \varepsilon)$ the barrier function associated to Problem (P′).

By Theorem 12, an ε^* exists such that, for all $\varepsilon \in (0, \varepsilon^*]$, $\bar{x}$ is a global solution of the following (unconstrained) problem:

$$\min Z'_q(x; \varepsilon), \qquad x \in \mathcal{N}_\alpha.$$

The thesis then follows observing that, for any $x \in B(\bar{x}, \delta/2)$, $Z'_q(x; \alpha) = Z_q(x; \alpha)$. □

We can now state the following theorem:

Theorem 14. *Let $\mathcal{M}$ be a compact set of local minimum points of Problem* (P) *such that $f(x)$ takes only a finite number of values on it. Then a threshold value $\varepsilon^* > 0$ exists such that:*

(i) *For all $\varepsilon \in (0, \varepsilon^*]$, if $\hat{x} \in \mathcal{M}$, then $\hat{x}$ is a local minimum point of $Z_q(x; \varepsilon)$.*
(ii) *For all $\varepsilon \in (0, \varepsilon^*)$, if $\hat{x} \in \mathcal{M}$ is a strict local minimum point of Problem* (P), *then $\hat{x}$ is a strict local minimum point of $Z_q(x; \varepsilon)$.*

Proof. (i) We proceed by contradiction. If the assertion is false, for any integer k an $\varepsilon_k \leq 1/k$ and a point $x_k \in \mathcal{M}$, which is not a local minimum point of $Z_q(x; \varepsilon_k)$, exist. Since $\mathcal{M}$ is compact, a convergent subsequence, which we relabel $\{x_k\}$, exists such that

$$\lim_{k \to \infty} x_k = \bar{x} \in \mathcal{M}.$$

Since $x_k \to \bar{x}$, by the assumptions made on $\mathcal{M}$ and by the continuity of $f(x)$, a $\bar{k}$ exists such that

$$Z_q(x_k; \varepsilon_k) = f(x_k) = f(\bar{x}) = Z_q(\bar{x}; \varepsilon), \qquad \forall k \geq \bar{k}, \quad \forall \varepsilon > 0, \tag{13}$$

where the first and last equality follow by the feasibility of x_k. However, by Proposition 13 $\bar{\varepsilon}$ and δ exist such that

$$Z_q(\bar{x}; \bar{\varepsilon}) \leq Z_q(x; \bar{\varepsilon}), \qquad \forall x \in B(\bar{x}, \delta). \tag{14}$$

Then, by (13) and (14), we can write, for $k \geq \bar{k}$ large enough so that $\varepsilon_k \leq \bar{\varepsilon}$,

$$Z_q(\bar{x}; \varepsilon_k) = Z_q(\bar{x}; \bar{\varepsilon}) \leq Z_q(x; \bar{\varepsilon}) \leq Z_q(x; \varepsilon_k), \qquad \forall x \in B(\bar{x}, \delta),$$

which implies that x_k is a local minimizer of $Z_q(x; \varepsilon_k)$ for all k large enough, thus contradicting the assumption made.

(ii) If $\hat{x}$ is a strict local minimum point of Problem (P), we can find a $0 < \delta' \leq \delta$ such that

$$f(x) > f(\bar{x}), \qquad \forall x \in \mathcal{F} \cap B(\bar{x}, \delta') \quad \text{and} \quad x \neq \bar{x}.$$

Assume now that $\varepsilon \in (0, \varepsilon^*)$, $x \in B(\bar{x}, \delta')$, and $x \neq \bar{x}$. If $x \in \mathcal{F}$ we can write

$$Z_q(x; \varepsilon) = f(x) > f(\bar{x}) = Z_q(\bar{x}; \varepsilon);$$

if $x \notin \mathcal{F}$ we can write

$$Z_q(x; \varepsilon) > Z_q(x; \varepsilon^*) \geq Z_q(\bar{x}; \varepsilon^*) = f(\bar{x}) = Z_q(\bar{x}; \varepsilon);$$

so that the proof is complete. □

From the computational point of view we have that unconstrained minimization algorithms yield unconstrained stationary points of the function $Z_q(x; \varepsilon)$. Therefore, it is also of interest to give conditions under which every stationary point of $Z_q(x; \varepsilon)$ is a KKT point of Problem (P). In Proposition 9 it has been shown that there are no stationary points of Problem (U) which do not belong to $\mathcal{F}$ for sufficiently small values of ε. The following proposition gives a condition that ensures that stationary points of Problem (U) belonging to $\mathcal{F}$ are also KKT points of Problem (P).

Proposition 15. *Let $\bar{x}$ be a stationary point of Problem* (U) *and suppose that $\bar{x} \in \mathcal{F}$. Then, if $h(x)$ is continuously differentiable at $\bar{x}$, $\bar{x}$ is a KKT point of Problem* (P).

Proof. By assumption,

$$0 \in \partial Z_q(\bar{x}; \varepsilon). \tag{15}$$

However, by Theorem 2.3.9 of [12], we have that

$$\partial Z_q(\bar{x}; \varepsilon) \subseteq \partial f(\bar{x}) + \frac{1}{\varepsilon b(\bar{x})} \operatorname{co}\left\{ \sum_{i=1}^{m} \beta_i \, \partial g_i(\bar{x}) + \sum_{j=1}^{p} \gamma_j \nabla h_j(\bar{x}) \right\}, \tag{16}$$

where $(\beta, \gamma) \in \partial \|g_+, h\|_q$, with $g = g(\bar{x})$ and $h = h(\bar{x})$. By (15) and (16), we can say that an integer s and s positive numbers t_r, $r = 1, \ldots, s$, with $\Sigma_{r=1}^{s} t_r = 1$, exist such that

$$0 = \varphi + \frac{1}{\varepsilon b(\bar{x})} \left[\sum_{i=1}^{m} \left(\sum_{r=1}^{s} t_r \beta_i^r \xi_i^r \right) + \sum_{j=1}^{p} \left(\sum_{r=1}^{s} t_r \gamma_j^r \nabla h_j(\bar{x}) \right) \right], \tag{17}$$

where $\varphi \in \partial f(\bar{x})$, $\xi_i^r \in \partial g_i(\bar{x})$, and $(\beta, \gamma) \in \partial \|g_+, h\|_q$, with $g = g(\bar{x})$ and $h = h(\bar{x})$. Let $\hat{I}$ be the index set defined by

$$\hat{I} := \left\{ i \in \{1, \dots, m\} \colon \sum_{r=1}^{q} t_r \beta_i^r \neq 0 \right\},$$

and let

$$R_i := \sum_{r=1}^{q} t_r \beta_i^r, \qquad S_j := \sum_{r=1}^{q} t_r \gamma_j^r, \qquad i = 1, \dots, m, \quad j = 1, \dots, p.$$

By Proposition 4(iii), if $i \notin \hat{I}$, $\beta_i^r = 0$, $r = 1, \dots, s$, so we can write (17) in the following form:

$$0 = \varphi + \frac{1}{\varepsilon b(\bar{x})} \left[\sum_{i \in \hat{I}} \left(R_i \sum_{r=1}^{s} \frac{t_r \beta_i^r}{R_i} \xi_i^r \right) + \sum_{j=1}^{p} S_j \nabla h_j(\bar{x}) \right]. \tag{18}$$

We now note that, by Proposition 4(ii), $\sum_{r=1}^{s} (t_r \beta_i^r / R_i) \xi_i^r$ is a convex combination of elements of $\partial g_i(\bar{x})$ so that, by the convexity of the generalized gradient, we can write

$$0 = \varphi + \frac{1}{\varepsilon b(\bar{x})} \left[\sum_{i \in \hat{I}} R_i \xi_i + \sum_{j=1}^{p} S_j \nabla h_j(\bar{x}) \right], \tag{19}$$

where $\xi_i \in \partial g_i(\bar{x})$, $\forall i \in \hat{I}$. By Proposition 4, $R_i > 0$, $\forall i \in \hat{I}$, and $\hat{I} \subset I_0(\bar{x})$; hence the thesis follows from (19) taking

$$\lambda_i = \frac{1}{\varepsilon b(\bar{x})} R_i, \qquad \forall i \in \hat{I},$$

$$\lambda_i = 0, \qquad \forall i \notin \hat{I},$$

$$\mu_j = \frac{1}{\varepsilon b(\bar{x})} S_j, \qquad j = 1, \dots, p. \qquad \square$$

It does not appear easy to relax the assumption that h be continuously differentiable in the preceding proposition. In fact in [35] an example is given showing that when nondifferentiable equality constraints are present, arbitrary feasible points can become stationary points of the barrier function $P(x; \varepsilon) = f(x) + (1/\varepsilon)\|g(x)_+, h(x)\|_\infty$ for sufficiently small values of ε. It is easy to see, using the example of [35], that the same result holds for $Z_q(x; \varepsilon)$. Hence, if one wants to employ $Z_q(x; \varepsilon)$ to find a solution of Problem (P) numerically, the differentiability of the equality constraints must be assumed (see the next section).

Remark 3. In this section we have assumed, for simplicity, that Assumptions A1 and A2 hold. However, since the results stated in this section are essentially local, it would not be difficult to show that they still hold if we just assume that Condition CQq holds in the points considered in the various theorems.

6. Globally Convergent Algorithm Models

In this section we propose some algorithm models which aim at solving the constrained Problem (P) by an unconstrained minimization algorithm. In view of the discussion made after Proposition 15, in this section we always suppose that h is continuously differentiable.

We assume that an unconstrained iterative minimization algorithm $\mathscr{U}$ with the following characteristics is available. First we suppose that, for every given value of the parameter ε and of the starting point $x_0 \in \mathscr{N}_\alpha$, algorithm $\mathscr{U}$ produces a sequence $\{x_k\}$ contained in $\mathscr{N}_\alpha$ and such that $Z_q(x_k; \varepsilon) \leq Z_q(x_0; \varepsilon)$. Recalling that $\mathscr{N}_\alpha$ is an open set, it is easy to see that the preceding requirement can be satisfied by any descent minimization algorithm: it is only necessary to ensure, by simple devices, that the trial points produced along the search direction remain in $\mathscr{N}_\alpha$.

The remaining requirements on algorithm $\mathscr{U}$ are standard. We suppose that, at each iteration, it is possible to associate to the current point x_k a nonnegative "stationarity measure" σ_k which can be 0 at (unconstrained) stationary points of $Z_q(x; \varepsilon)$ only and such that if $\{x_k\}$ is the sequence of points generated by algorithm $\mathscr{U}$ and for some subsequence K of indexes we have that $\{x_k\}_K \to \bar{x}$ and $\{\sigma_k\}_K \to 0$, then $\bar{x}$ is an (unconstrained) stationary point of $Z_q(x; \varepsilon)$. Finally, we assume that algorithm $\mathscr{U}$ either drives the penalty function value to $-\infty$ or the stationarity measure σ_k to 0.

The purpose of this section is to show that it is possible to use the *unconstrained* minimization algorithm $\mathscr{U}$, coupled with the penalty function studied in the previous sections, to solve the *constrained* Problem (P). We introduce a very general algorithm scheme and study its behavior when no assumptions are made on Problem (P). Then, we investigate how this behavior is improved when various, increasingly stronger, assumptions on Problem (P) are made. Finally, we also consider two variants of the basic algorithm that can be useful in particular cases.

To state the basic algorithm model more simply, we introduce the following notation:

$$\Delta_\mu^\nu := \{\delta \in \mathbb{R}_+ : \delta = \mu\nu^s, s = 0, 1, \ldots\} \cup \{0\},$$

$$\partial_\delta \varphi(x) := \operatorname{cl}\left[\bigcup_{y \in B(x;\, \delta) \cap \mathscr{N}_\alpha} \partial\varphi(y)\right].$$

Algorithm Model I.

Data. $x_0 \in \mathscr{N}_\alpha$, $\varepsilon_0 > 0$, $\mu > 0$, $\nu \in (0,1)$, $\theta_1 \in (0,1)$, $\theta_2 \in (\theta_1, 1)$.

Step 0. Set $r = 0$, $k = 0$.

Step 1. If $\|g_+(x_k), h(x_k)\|_q = 0$ go to Step 2, else go to Step 3.

Step 2. If $\sigma_k = 0$ stop, else go to Step 6.

Step 3. If $0 \in \partial\|g_+(x_k), h(x_k)\|_q$ stop.

Step 4. Compute

$$\delta_k = \max\{\delta \in \Delta_\mu^\nu : \delta \leq \|\mathrm{Nr}\, \partial_\delta Z_q(x_k; \varepsilon_k)\|_q\}.$$

Step 5. If $\delta_k \geq \nu\varepsilon_k\|g_+(x_k), h(x_k)\|_q$ go to Step 6, else go to Step 7.

Step 6. Compute x_{k+1} using algorithm $\mathscr{U}$, set $\varepsilon_{k+1} = \varepsilon_k$, $k = k+1$ and go to Step 1.

Step 7. Choose $\varepsilon_{k+1} \in [\theta_1\varepsilon_k, \theta_2\varepsilon_k]$, set $x_{k+1} = x_k$; set (for future reference purposes only): $w_r = x_k$, $\tilde{\delta}_r = \delta_k$, $\tilde{\varepsilon}_r = \varepsilon_k$, and $r = r+1$, set $k = k+1$ and go to Step 1.

Remark 4. Note that Algorithm Model I produces the sequences $\{x_k\}$ and $\{\varepsilon_k\}$. At each iteration, either $x_{k+1} \neq x_k$ or $\varepsilon_{k+1} < \varepsilon_k$, but never both. The first case $(x_{k+1} \neq x_k)$ occurs when the penalty parameter is not updated and a new point is generated using algorithm $\mathscr{U}$. The second case $(\varepsilon_{k+1} < \varepsilon_k)$, instead, occurs when the algorithm model detects the need to reduce the penalty parameter; in this case we do not change the current point x_k.

The algorithm model produces also the sequences $\{w_r\}$, $\{\tilde{\varepsilon}_r\}$, and $\{\tilde{\delta}_r\}$. These sequences are introduced only to facilitate the exposition of some proofs in what follows. These sequences keep track of the values of the corresponding quantities at the steps where the penalty parameter is updated.

Proposition 16. *Suppose that h is continuously differentiable in an open set containing $\mathscr{F}$ and let $\{x_k\}$ be the sequence produced by Algorithm Model* I.

(1) *If the sequence $\{x_k\}$ is finite with the last point $\bar{x}$, then $\bar{x}$ is either a KKT point or an infeasible stationary point of Problem* (P).

(2) *If $\{\varepsilon_k\} \downarrow 0$ (and hence the sequence $\{w_r\}$ is infinite and $\{\tilde{\varepsilon}_r\} \downarrow 0$), then every limit point $\bar{x}$ of the sequence $\{w_r\}$, is such that*

$$0 \in \partial_{\mathscr{F}^c}\|g_+(\bar{x}), h(\bar{x})\|_q.$$

(3) *If $\varepsilon_k = \bar{\varepsilon}$ for every k sufficiently large, then every limit point $\bar{x}$ of $\{x_k\}$ is a KKT point for Problem* (P).

Proof. (1) This point easily follows by the stopping criteria of Steps 2 and 3, recalling that if we arrive at Step 3 the current point is not feasible, and using Proposition 15.

(2) Suppose the contrary. Then a subsequence of $\{w_r\}$ (which we relabel $\{w_r\}$) exists which converges to a point $\bar{x}$ where $0 \notin \partial|_{\mathscr{F}^c}\|g_+(\bar{x}), h(\bar{x})\|_q$. By Step 5 we also have a sequence $\{\tilde{\delta}_r\}$ of nonnegative numbers such that

$$\tilde{\delta}_r < \nu\tilde{\varepsilon}_r\|g_+(w_r), h(w_r)\|_q. \tag{20}$$

Since $\{\tilde{\varepsilon}_r\} \downarrow 0$ and $\{w_r\}$ converges, (20) implies

$$\tilde{\delta}_r \to 0. \tag{21}$$

By the definition of $\tilde{\delta}_r$ we have

$$\frac{\tilde{\delta}_r}{\nu} \geq \|\mathrm{Nr}\, \partial_{\tilde{\delta}_r/\nu} Z_q(w_r; \tilde{\varepsilon}_r)\|_q,$$

that in turn implies, taking into account (21),

$$\|\mathrm{Nr}\, \partial_{\tilde{\delta}_r/\nu} Z_q(w_r; \tilde{\varepsilon}_r)\|_q \to 0. \tag{22}$$

Now, if $\bar{x} \notin \mathcal{F}$, we have,

$$\partial|_{\mathcal{F}^c} \|g_+(\bar{x}), h(\bar{x})\|_q = \partial \|g_+(\bar{x}), h(\bar{x})\|_q. \tag{23}$$

However, we also have, recalling (10),

$$\begin{aligned} \tilde{\varepsilon}_r \partial_{\tilde{\delta}_r/\nu} Z_q(w_r; \tilde{\varepsilon}_r) &\subseteq \tilde{\varepsilon}_r \partial_{\tilde{\delta}_r/\nu} f(w_r) + \partial_{\tilde{\delta}_r/\nu} \|\hat{g}_+(w_r), \hat{h}(w_r)\|_q \\ &\subseteq \tilde{\varepsilon}_r \partial_{\tilde{\delta}_r/\nu} f(w_r) + \left(\frac{\alpha}{b(w_r)^2} \right) \partial_{\tilde{\delta}_r/\nu} \|g_+(w_r), h(w_r)\|_q, \end{aligned} \tag{24}$$

that, by the boundedness of the subdifferential mapping, by the fact that $\alpha/b(w_r)^2$ is bounded away from zero in a neighborhood of $\bar{x}$, and by (22), implies

$$\|\mathrm{Nr}\, \partial_{\tilde{\delta}_r/\nu} \|g_+(w_r), h(w_r)\|_q\|_q \to 0. \tag{25}$$

However, then, taking into account the upper semicontinuity of the subdifferential mapping and the definition of $\partial_{\tilde{\delta}_r/\nu}\|g_+(w_r), h(w_r)\|_q$, we have that (21), (23), and (25) contradict $0 \notin \partial|_{\mathcal{F}^c}\|g_+(\bar{x}), h(\bar{x})\|_q$. Hence $\bar{x} \in \mathcal{F}$ must hold. By the Lipschitzianity of $\|g_+(x), h(x)\|_q$ we have that a constant L and a neighborhood Ω of $\bar{x}$ exist such that

$$\|g_+(x), h(x)\|_q = 0 \quad \Rightarrow \quad \|w - x\|_q \geq \frac{\|g_+(w), h(w)\|_q}{L}, \qquad \forall w, x \in \Omega. \tag{26}$$

Subsequencing if necessary, two possibilities now arise: for all $w_r \in \Omega$, either $\tilde{\delta}_r/\nu < \|g_+(w_r), h(w_r)\|_q/L$ or $\tilde{\delta}_r/\nu \geq \|g_+(w_r), h(w_r)\|_q/L$.

If $\tilde{\delta}_r/\nu \geq \|g_+(w_r); h(w_r)\|_q/L$, we have a contradiction to the rule of Step 5. Hence, we only need consider the case $\tilde{\delta}_r/\nu < \|g_+(w_r), h(w_r)\|_q/L$. It is easy to see that in this case (21) and (26) imply that, for r sufficiently large,

$$B\left(w_r, \frac{\tilde{\delta}_r}{\nu}\right) \cap \mathcal{F} = \emptyset. \tag{27}$$

It is immediate to see, now, that (25) is still valid in the case $\bar{x} \in \mathcal{F}$. Hence, since, for r sufficiently large, $B(w_r, \tilde{\delta}_r/\nu) \cap \partial\mathcal{N}_\alpha = \emptyset$, the closure operation in the definition of $\partial_{\tilde{\delta}_r/\nu}\|g_+(w_r), h(w_r)\|_q$ is superfluous, and, recalling also (27), we can conclude that

$$\mathrm{Nr}\, \partial_{\tilde{\delta}_r/\nu} \|g_+(w_r), h(w_r)\|_q = \zeta_r,$$

with

$$\zeta_r \in \partial \|g_+(y_r), h(y_r)\|_q, \qquad y_r \in B\left(w_r, \frac{\tilde{\delta}_r}{\nu}\right), \quad y_r \notin \mathcal{F}.$$

Then, taking into account the definition of generalized gradient relative to a set, (21), and (25) we have

$$\{\zeta_i^r\} \to 0 \in \partial|_{\mathcal{F}^c}\|g_+(\bar{x}), h(\bar{x})\|_q,$$

which again contradicts $0 \notin \partial_{\mathcal{F}^c}\|g_+(\bar{x}), h(\bar{x})\|_q$,

(3) If a limit point $\bar{x}$ of the sequence $\{x_k\}$ exists, then, by continuity, $\{Z_q(x_k, \bar{\varepsilon})\} \nrightarrow -\infty$. Hence, by the assumptions on the unconstrained algorithm $\mathcal{U}$, we have $\{\sigma_k\} \to 0$, so that $\bar{x}$ is a stationary point for $Z_q(x; \bar{\varepsilon})$. This implies $\{\delta_k\} \to 0$, so that, by the rule of Step 5, we also have $\|g_+(x_k), h(x_k)\|_q \to 0$, that, in turn, implies the assertion by Proposition 15. □

By Propositions 8 and 16 we immediately obtain the following theorem which gives a fairly detailed picture of the behavior of the algorithm model.

Theorem 17. *Suppose that h is continuously differentiable in an open set containing $\mathcal{F}$ and let $\{x_k\}$ be the sequence produced by Algorithm Model* I.

1. *If the sequence $\{x_k\}$ is finite with the last point $\bar{x}$, then $\bar{x}$ is either a KKT point or an infeasible stationary point of Problem* (P).
2. *If $\{\varepsilon_k\} \downarrow 0$ (and hence the sequence $\{w_r\}$ is infinite and $\{\tilde{\varepsilon}_r\} \downarrow 0$), then every limit point $\bar{x}$ of the sequence $\{w_r\}$ is either an FJ point or an infeasible stationary point of Problem* (P).
3. *If $\varepsilon_k = \bar{\varepsilon}$ for every k sufficiently large, then every limit point $\bar{x}$ of $\{x_k\}$ is a KKT point for Problem* (P).

The preceding theorem is quite interesting, especially because it is proved without imposing any regularity assumption on Problem (P). It is obvious, then, that it allows possible "pathological" behaviors. For example, it does not exclude that no limit points exist at all or that the penalty function value be driven to $-\infty$. It is of interest, then, to study under which assumptions these "pathological" behaviors can be excluded.

A first result in this direction is given in the following theorem, where, under Assumption A1, it is excluded that unbounded sequences be generated and so the existence of at least one limit point which is an extended stationary point for Problem (P) is guaranteed.

Theorem 18. *Suppose that h is continuously differentiable in an open set containing $\mathcal{F}$ and that Assumption* A1 *is satisfied. Let $\{x_k\}$ be the sequence produced by Algorithm Model* I. *Then the sequence $\{x_k\}$ is bounded and the sequences $\{x_k\}$ and $\{w_r\}$ admit at least a limit point.*

1. *If the sequence $\{x_k\}$ is finite with the last point $\bar{x}$, then $\bar{x}$ is either a KKT point or an infeasible stationary point of Problem* (P).

2. *If* $\{\varepsilon_k\} \downarrow 0$ *(and hence the sequence* $\{w_r\}$ *is infinite and also* $\{\tilde{\varepsilon}_r\} \downarrow 0$*), then every limit point* $\bar{x}$ *of the sequence* $\{w_r\}$ *is either an FJ point or an infeasible stationary point of Problem* (P).
3. *If* $\varepsilon_k = \bar{\varepsilon}$ *for every k sufficiently large, then every limit point* $\bar{x}$ *of* $\{x_k\}$ *is a KKT point for Problem* (P).

Proof. Since Assumption A1 is satisfied, the set $\mathcal{N}_\alpha$ is bounded; but by the assumptions made on the unconstrained algorithm $\mathcal{U}$ we also have that the sequence $\{x_k\}$ is contained in $\mathcal{N}_\alpha$, so that the thesis follows. □

We remark that two elements are essential to the preceding result: Assumption A1 and the particular structure of the barrier function $Z_q(x; \varepsilon)$. In fact, had we not employed the function Z_q, but, instead, a more usual penalty function, for example the function

$$P_q(x; \varepsilon) := f(x) + \frac{1}{\varepsilon}\|g_+(x), h(x)\|_q,$$

the result stated in the Theorem 18 would not hold. For instance, consider the following simple example:

$$\min x^3,$$
$$x = 0.$$

This is a very well-behaved problem, which satisfies any standard regularity assumption (included Assumptions A1 and A2) and has the obvious, unique solution $x = 0$. Nevertheless, if we consider the penalty function $P_q(x, \varepsilon)$, it is easily seen that, for any value of ε, the level sets of the penalty function are unbounded and that the value of the penalty function goes to $-\infty$ when $x \to -\infty$. Hence, if this penalty function is employed for solving the problem, unbounded sequences can be generated and the penalty parameter can be driven to 0, and there seem to be no simple way of avoiding these possibilities. Let us examine what happens, instead, if the barrier function $Z_q(x; \varepsilon)$ is employed. Suppose that we start the algorithm with a point x_0. It is sufficient to assume $\alpha = \|x_0\|_q + 1$, to ensure that $x_0 \in \mathcal{N}_\alpha$ so that the whole sequence $\{x_k\}$ is "trapped" in $\mathcal{N}_\alpha$ and at least a limit point exists (actually, we prove shortly that convergence occurs to the solution $x = 0$).

We now examine the other source of trouble: the possibility that extended stationary points other than the KKT points are generated. Thanks to the analysis carried out it is now easy to determine the weakest assumption that allows us to rule out the possibility of this event. This is done in the next theorem.

Theorem 19. *Suppose that h is continuously differentiable in an open set containing* $\mathcal{F}$, *that Assumptions* A1 *and* A2 *hold, and let* $\{x_k\}$ *be the sequence of points produced by Algorithm Model* I. *Then the sequence* $\{x_k\}$ *is bounded and* $\varepsilon_k = \bar{\varepsilon}$ *for every k sufficiently large. Furthermore:*

1. *If the sequence* $\{x_k\}$ *is finite with the last point* $\bar{x}$, *then* $\bar{x}$ *is a KKT point for Problem* (P).

2. *If the sequence* $\{x_k\}$ *is not finite, it admits a limit point, furthermore any limit point* $\bar{x}$ *of the sequence is a KKT point for Problem* (P).

Proof. The theorem easily follows from Proposition 16 and Theorem 18 taking into account that, by Assumption 2 and Proposition 8, case 2 cannot occur in Proposition 16 (and Theorem 18). □

Remark 5. Note that if we assume that Assumption A2 holds, the stopping rule of Step 3 can never be satisfied so that Step 3 itself can be suppressed.

Remark 6. It is interesting to observe that Theorem 19 guarantees the convergence toward KKT points of Problem (P), without the need to exclude the existence of FJ points (see Remark 2).

If a feasible point is known in advance, we can modify Algorithm Model I in a very simple way, to avoid any assumption on the constraints outside the feasible set (see also [18] for a similar result). All we need is that the following assumption holds:

Assumption A2 bis. *Condition* CQq *is satisfied for all* x *in* $\mathscr{F}$.

We note that Assumption A2 bis is weaker than Assumption A2, because $\mathscr{F}$ is strictly contained in $\bar{\mathscr{N}}_\alpha$.

Algorithm Model II.

Data. $x_0 \in \mathscr{F}$, $\varepsilon_0 > 0$, $\mu > 0$, $\nu \in (0,1)$, $\theta_1 \in (0,1)$, $\theta_2 \in (\theta_1, 1)$.

Step 0. Set $r = 0$, $k = 0$.

Step 1. If $\|g_+(x_k), h(x_k)\|_q = 0$ go to Step 2, else go to Step 3.

Step 2. If $\sigma_k = 0$ stop, else go to Step 5.

Step 3. Compute

$$\delta_k = \max\{\delta \in \Delta_\mu^\nu : \|\mathrm{Nr}\, \partial_\delta Z_q(x_k; \tilde{\varepsilon}_r)\|_q \geq \delta\}.$$

Step 4. If $\delta_k \geq \nu\varepsilon_r \|g_+(x_k), h(x_k)\|_q$ go to Step 5, else go to Step 6.

Step 5. Compute x_{k+1} using algorithm $\mathscr{U}$, set $\varepsilon_{k+1} = \varepsilon_k$, $k = k + 1$, and go to Step 1.

Step 6. Choose $\varepsilon_{k+1} \in [\theta_1\varepsilon_k, \theta_2\varepsilon_k]$; if $Z_q(x_k; \varepsilon_{k+1}) \leq Z_q(x_0; \varepsilon_{k+1}) = f(x_0)$, set $x_{k+1} = x_k$, otherwise set $x_{k+1} = x_0$; set (for future reference purposes only) $w_r = x_k$, $\tilde{\delta}_r = \delta_k$, $\tilde{\varepsilon}_r = \varepsilon_k$, and $r = r + 1$; set $k = k + 1$ and go to Step 1.

Remark 7. The main difference between Algorithm Models I and II is in what is done when ε is updated (Step 7 of Algorithm Model I and Step 6 of Algorithm Model II). In the first algorithm, when ε is updated, the current point is not changed ($x_{k+1} = x_k$). In Algorithm Model II, instead, when ε is updated, we perform a test: we compare the value of the barrier function in x_k (with the new value of the penalty parameter) to the value of the barrier function in x_0 which, since x_0 is feasible, does not depend on ε. If we find that the value of $Z_q(x_k; \varepsilon_{k+1})$ is not worse than the value of the barrier function in x_0 we proceed normally setting $x_{k+1} = x_k$. However, if we find that the value of $Z_q(x_k; \varepsilon_{k+1})$ is worse than the value of the barrier function in x_0, we backtrack and restart from a better point that we have at hand ($x_{k+1} = x_0$).

Theorem 20. *Suppose that h is continuously differentiable in an open set containing $\mathcal{F}$, and that Assumptions* A1 *and* A2 *bis hold. Let $\{x_k\}$ be the sequence of points produced by Algorithm Model* II. *Then the sequence $\{x_k\}$ is bounded. Furthermore*:

(1) *If the sequence $\{x_k\}$ is finite with the last point $\bar{x}$, then $\bar{x}$ is a KKT point for Problem* (P).
(2) *If the sequence $\{x_k\}$ is not finite, it admits a limit point, and any limit point $\bar{x}$ of the sequence is a KKT point for Problem* (P).

Proof. The fact that $\{x_k\}$ admits a limit point at least follows from Assumption A1 by reasonings that are, by now, standard.

(1) The assertion follows from Steps 2 and 3 of Algorithm Model II and from Proposition 15.

(2) We first prove by contradiction that the sequence $\{w_r\}$ is finite. Suppose the contrary, then we have, recalling Step 6 and the fact that, by assumption, $\mathcal{U}$ is a descent algorithm,

$$Z_q(w_r; \tilde{\varepsilon}_r) \leq Z_q(x_0; \tilde{\varepsilon}_r) = f(x_0).$$

Since $\{\tilde{\varepsilon}_r\} \downarrow 0$, this implies

$$\limsup_{r \to \infty} \|\hat{g}_+(w_r), \hat{h}(w_r)\|_q = 0. \tag{28}$$

Now suppose, without loss of generality, that $\{w_r\} \to \bar{w}$, then (28) implies $\bar{w} \in \mathcal{F}$. Then we can complete the proof of the finiteness of the sequence $\{w_r\}$ along the same lines adopted in point (2) of Proposition 16, since in that part of the proof the condition $0 \notin \partial|_{\mathcal{F}^c} \|g_+(x), h(x)\|_q$ for $x \in \bar{\mathcal{N}}_\alpha \setminus \mathcal{F}$ is not employed. The assertion of the theorem can then be proved as in point (3) of Proposition 16. □

Remark 8. Theorem 20 makes clear the role of the regularity condition expressed by Assumption A2 out of the feasible set. We need it, essentially, to guarantee that we can algorithmically find a feasible point (and hence that a feasible point at least exists). If we know a feasible point, globally convergent algorithms can be constructed without the need of any assumption outside the feasible set.

It should be clear that Algorithm Models I and II are, employing the terminology of [35], "conceptual" algorithms, for some of the calculations required cannot be carried out in a finite number of steps. Nevertheless, they settle a framework within which to build and evaluate implementable algorithms that, obviously, must be tailored to the particular algorithm $\mathscr{U}$ employed. We remark, however, that most existing algorithms for the minimization of Z will allow us to compute "suitable" approximations to the quantities needed in the algorithm models described so far. In any case we observe that, in a very important particular case, Algorithm Models I and II can be made fully implementable. In fact, most of the difficulties are due to the necessity to make up for the lack of continuity of the generalized gradient. If we can somehow ensure "sufficient smoothness," we should expect simpler algorithms. To this end, suppose that f, g, and h are continuously differentiable, $q \in (1, \infty)$, and consider the following implementable simplified version of Algorithm Model I:

Algorithm Model III.

Data. $x_0 \in \mathscr{N}_\alpha$, $\varepsilon_0 > 0$, $\rho \in (0, 1)$.

Step 0. Set $k = 0$, $r = 0$.

Step 1. If $\|g_+(x_k), h(x_k)\|_q = 0$ go to Step 2, else go to Step 3.

Step 2. If $\sigma_k = 0$ stop, else go to Step 4.

Step 3. If $\varepsilon_k \|\nabla f(x_k)\|_q > \rho \big\| \nabla \|\hat{g}_+(x_k), \hat{h}(x_k)\|_q \big\|_q$, set

$$\varepsilon_k = \rho \min \left\{ \frac{\big\| \nabla \|\hat{g}_+(x_k), \hat{h}(x_k)\|_q \big\|_q}{\|\nabla f(x_k)\|_q}, \varepsilon_k \right\},$$

set (for future reference purposes only) $\tilde{\varepsilon}_r = \varepsilon_k$, $w_r = x_k$, and $r = r + 1$, set $x_{k+1} = x_k$, $k = k + 1$ and go to Step 1.

Step 4. Compute x_{k+1} using algorithm $\mathscr{U}$, set $k = k + 1$ and go to Step 1.

The following theorem holds:

Theorem 21. *Suppose that f, g, and h are continuously differentiable, $q \in (1, \infty)$, and that Assumptions* A1 *and* A2 *hold. Let $\{x_k\}$ be the sequence of points produced by Algorithm Model* III. *Then the sequence $\{x_k\}$ admits a limit point at least. Furthermore*:

(1) *If the sequence $\{x_k\}$ is finite with the last point $\bar{x}$, then $\bar{x}$ is a KKT point for Problem* (P).

(2) *If the sequence $\{x_k\}$ is not finite, any limit point $\bar{x}$ of the sequence is a KKT point for Problem* (P).

Proof. The fact that the sequence $\{x_k\}$ admits a limit point at least follows from Assumption A1 by arguments which are, by now, standard.

(1) The assertion follows from Steps 1 and 2 of the algorithm model and from Proposition 15.

(2) We first prove by contradiction that the sequence $\{w_r\}$ is finite. Suppose the contrary, then we have sequences $\{w_r\}$ and $\{\tilde{\varepsilon}_r\}$ such that

$$\{\tilde{\varepsilon}_r\} \downarrow 0, \qquad w_r \in \mathscr{N}_\alpha,$$

$$\tilde{\varepsilon}_r \|\nabla f(w_r)\|_q > \rho \Big\| \nabla \|\hat{g}_+(w_r), \hat{h}(w_r)\|_q \Big\|_q. \tag{29}$$

We can suppose, without loss of generality, that $\{w_r\} \to \overline{w} \in \bar{\mathscr{N}}_\alpha$. By Step 1 we also have $w_r \notin \mathscr{F}$, so that (recall that $q \in (1, \infty)$) $\nabla \|\hat{g}_+(\cdot), \hat{h}(\cdot)\|_q$ is continuously differentiable around w_r for all r and

$$\nabla \|\hat{g}_+(w_r), \hat{h}(w_r)\|_q = \frac{\alpha}{b(w_r)^2} \nabla \|g_+(w_r), h(w_r)\|_q.$$

However, then, by (29) and by the boundedness of $\nabla f(w_r)$, we get

$$\nabla \|g_+(w_r), h(w_r)\|_q \to 0,$$

that, in turn, contradicts Assumption A2. Hence, the sequence $\{w_r\}$ is finite.

Let $\bar{r}$ be the largest index r produced by the algorithm, and indicate by $\bar{x}$ any limit point of the sequence $\{x_k\}$. Suppose by contradiction that $\bar{x} \notin \mathscr{F}$. Since by assumption $\bar{x}$ is a stationary point for $Z_q(x; \varepsilon_{\bar{r}})$ we have

$$0 = \nabla f(\bar{x}) + \frac{1}{\varepsilon_{\bar{r}}} \nabla \|\hat{g}_+(\bar{x}), \hat{h}(\bar{x})\|_q. \tag{30}$$

Since $\bar{x} \notin \mathscr{F}$, $\|\hat{g}_+(\bar{x}), \hat{h}(\bar{x})\|_q \neq 0$, and hence (30) and Assumption A2 imply $\nabla \|\hat{g}_+(\bar{x}), \hat{h}(\bar{x})\|_q \neq 0$, so that

$$\varepsilon_{\bar{r}} \|\nabla f(\bar{x})\|_q > \rho \Big\| \nabla \|\hat{g}_+(\bar{x}), \hat{h}(\bar{x})\|_q \Big\|_q. \tag{31}$$

However, then, by continuity, we have that, for some k large enough, the test of Step 3 is passed and the penalty parameter is reduced. This contradicts the definition of $\bar{r}$ and hence $\bar{x} \in \mathscr{F}$. The last assertion of the theorem then follows by Proposition 15, taking into account that, by the assumptions made on $\mathscr{U}$, $\bar{x}$ is an unconstrained stationary point of $Z_q(x; \varepsilon_{\bar{r}})$. □

Obviously a similar modification could be envisaged for Algorithm Model II, it is straightforward and we therefore omit it.

References

1. M. S. Bazaraa and J. J. Goode. Sufficient conditions for a globally exact penalty function without convexity. Mathematical Programming Study, 19:1–15, 1982.
2. D. P. Bertsekas. Necessary and sufficient conditions for a penalty method to be exact. Mathematical Programming, 9:87–99, 1975.
3. D. P. Bertsekas. Constrained Optimization and Lagrange Multipliers Methods. Academic Press, New York, 1982.

4. J. M. Borwein. Stability and regular points of inequality systems. Journal of Optimization Theory and Applications, 48:9–52, 1986.
5. J. V. Burke. A sequential quadratic programming method for potentially infeasible mathematical programs. Journal of Mathematical Analysis and Applications, 139:319–351, 1989.
6. J. V. Burke, Calmness and exact penalization. SIAM Journal on Control and Optimization, 29:493–497, 1991.
7. J. V. Burke. An exact penalization viewpoint of constrained optimization. SIAM Journal on Control and Optimization, 29:968–998, 1991.
8. J. V. Burke. A robust trust region method for constrained nonlinear programming problems. SIAM Journal on Optimization, 2:325–347, 1992.
9. J. V. Burke and S.-P. Han. A robust sequential quadratic programming method. Mathematical Programming, 43:277–303, 1989.
10. C. Charalambous. A lower bound for the controlling parameters of the exact penalty functions. Mathematical Programming, 15:278–290, 1978.
11. C. Charalambous. On conditions for optimality of a class of nondifferentiable functions. Journal of Optimization Theory and Applications, 43:135–142, 1982.
12. F. H. Clarke. Optimization and Nonsmooth Analysis. Wiley, New York, 1983.
13. G. Di Pillo and F. Facchinei. Exact penalty functions for nondifferential programming problems. In F. H. Clarke, V. F. Demyanov, and F. Giannessi, editors, Nonsmooth Optimization and Related Topics, pp. 89–107, Plenum, New York, 1989.
14. G. Di Pillo and F. Facchinei. Regularity conditions and exact penalty functions in Lipschitz programming problems. In F. Giannessi, editor, Nonsmooth Optimization Methods and Applications, pp. 107–120. Gordon and Breach, New York, 1992.
15. G. Di Pillo and L. Grippo. An exact penalty method with global convergence properties for nonlinear programming problems. Mathematical Programming, 36:1–18, 1986.
16. G. Di Pillo and L. Grippo. On the exactness of a class of nondifferentiable penalty functions. Journal of Optimization Theory and Applications, 57:399–410, 1988.
17. G. Di Pillo and L. Grippo. Exact penalty functions in constrained optimization. SIAM Journal on Control and Optimization, 27:1333–1360, 1989.
18. G. Di Pillo, L. Grippo, and S. Lucidi. A smooth method for the minimax problem. Mathematical Programming, 60:187–214, 1993.
19. J. P. Evans, F. J. Gould, and J. W. Tolle. Exact penalty functions in nonlinear programming. Mathematical Programming, 4:72–97, 1973.
20. F. Facchinei. Exact penalty functions and Lagrange multipliers. Optimization, 22:579–606, 1991.
21. F. Facchinei. Refinements of necessary conditions for optimality in nonlinear programming. Journal of Optimization Theory and Applications, 73:65–74, 1992.
22. A. V. Fiacco and G. P. McCormick. Nonlinear Programming: Sequential Unconstrained Minimization Techniques. Wiley, New York, 1968.
23. R. Fletcher. Practical Methods of Optimization. Wiley, New York, 1987.
24. S. P. Han. A globally convergent method for nonlinear programming. Journal of Optimization Theory and Applications, 22:297–309, 1977.
25. S. P. Han and O. L. Mangasarian. Exact penalty functions in nonlinear programming. Mathematical Programming, 17:251–269, 1979.
26. J.-B. Hiriart-Urruty. Refinements of necessary optimality conditions in nondifferentiable programming, I. Applied Mathematics and Optimization, 5:63–82, 1979.
27. J.-B. Hiriart-Urruty. Refinements of necessary optimality conditions in nondifferentiable programming, II. Mathematical Programming Study, 19:120–139, 1982.
28. S. Howe. New conditions for the exactness of a simple penalty function. SIAM Journal on Control, 11:378–381, 1973.
29. A. D. Ioffe. Necessary and sufficient conditions for a local minimum. 1: A reduction theorem and first-order conditions. SIAM Journal on Control and Optimization, 17:245–250, 1979.
30. A. D. Ioffe. Necessary and sufficient conditions for a local minimum. 2: Conditions of Levitin–Miljutin–Osmolovskii type. SIAM Journal on Control and Optimization, 17:251–265, 1979.

31. A. D. Ioffe. Necessary and sufficient conditions for a local minimum. 3: Second-order conditions and augmented duality. SIAM Journal on Control and Optimization, 17:266–288, 1979.
32. V. H. Nguyen, J.-J. Strodiot, and R. Mifflin. On conditions to have bounded multipliers in locally Lipschitz programming. Mathematical Programming, 18:100–106, 1980.
33. T. Pietrzykowski. An exact potential method for constrained maxima. SIAM Journal on Numerical Analysis, 6:299–304, 1969.
34. T. Pietrzykowski. The potential method for conditional maxima in the locally compact metric spaces. Numerische Mathematik, 14:325–329, 1970.
35. E. Polak, D. Q. Mayne, and Y. Ward. On the extension of constrained optimization algorithms from differentiable to nondifferentiable problems. SIAM Journal on Control and Optimization, 21:179–203, 1983.
36. R. T. Rockafellar. Convex Analysis. Princeton University Press, Princeton, NJ, 1970.
37. E. Rosenberg. Exact penalty functions and stability in locally lipschitz programming. Mathematical Programming, 30:340–356, 1984.
38. J. Stoer. Principles of sequential quadratic programming methods for solving nonlinear programs. In K. Schittkowski, editor, Computational Mathematical Programming, NATO ASI Series, Vol. F15. Springer-Verlag, Berlin, 1985.
39. D. E. Ward. Exact penalties and sufficient conditions for optimality in nonsmooth optimization. Journal of Optimization Theory and Applications, 57:485–499, 1988.
40. D. E. Ward and J. M. Borwein. Nonsmooth calculus in finite dimensions. SIAM Journal on Control and Optimization, 25:1312–1340, 1987.
41. W. I. Zangwill. Non-linear programming via penalty functions. Management Science, 13:344–358, 1967.

Accepted 25 August 1993