Exact Penalty Principle*

Jane J. Ye†

March 15, 2011

Abstract

Exact penalty approach aims at replacing a constrained optimization problem by an equivalent unconstrained optimization problem. Most of results in the literature of exact penalization are mainly concerned with finding conditions under which a solution of the constrained optimization problem is a solution of an unconstrained penalized optimization problem and the reverse property is rarely studied. In this paper we study the reverse property. We give conditions under which the original constrained (single and/or multiobjective) optimization problem and the unconstrained exact penalized problem are exactly equivalent. The main conditions to ensure the exact penalty principle for optimization problems include the global and local error bound conditions. By using variational analysis, these conditions may be characterized by using generalized differentiation.

Key words: Exact penalty, constrained optimization, multiobjective optimization, error bounds, variational analysis

MSC2010 subject classification: 90C30, 90C29

*The research of this paper is partially supported by NSERC.

†Department of Mathematics and Statistics, University of Victoria, P.O. Box 3045 STN CSC, Victoria, BC, V8W 3R4 Canada (janeye@Uvic.CA).

1 Introduction

The exact penalization approach toward constrained (single-objective) optimization problems

$$(P) \qquad \min\{f(x)|g(x) \leq 0, h(x) = 0, x \in S\}$$

where $f : X \to R, g : X \to R^m, h : X \to R^n$ and $S \subset X$ goes back to Eremin [6] and Zangwill [21]. It aims at replacing the above constrained optimization problem by an equivalent unconstrained optimization problem

$$(P_\rho) \qquad \min_{x \in S} f(x) + \rho\psi(x)$$

for some $\rho > 0$, where ψ is a scalar-valued function satisfying

$$\psi(x) \geq 0 \qquad \text{if } x \in S,$$

$$\psi(x) = 0 \qquad \text{if and only if } g(x) \leq 0, h(x) = 0, x \in S$$

and the function ψ is usually referred to as an exact penalty function. Most of the literature of exact penalty functions is mainly concerned with conditions that ensure that a local (global) minimum of the problem (P) is a local (global) minimum of the penalized problem (P_ρ) for all sufficiently large but finite values of the penalty parameter ρ. However, the main motivation for the use of penalty methods is that of solving the original constrained problem by employing some unconstrained minimization algorithm to solve the penalized problem. Hence the reverse properties which ensure that the local (global) minimizers of the penalized problem are local (global) solutions of the original problem are very important. In this paper we aim at studying this property.

Let $d_C(y) := \inf\{\|y - c\| : c \in C\}$ denote the distance of y to a set C. Then $d_C(y)$ is a nonsmooth Lipschitz continuous function. In his seminal result [2, Proposition 2.4.3], Clarke shows that the distance function is always an exact penalty function without any extra condition. Precisely the following exact penalty principle is always true.

Theorem 1.1 (Clarke's exact penalty principle) *Let S be a subset of a normed space X and $f : X \to R$ be Lipschitz of rank L_f on S. Let x belong to a set $C \subset S$ and suppose that f attains a minimum over C at x. Then for any $L \geq L_f$, the function $g(y) = f(y) + Ld_C(y)$ attains a minimum over S at x. Conversely suppose that C is closed, then for any $L > L_f$, any other point minimizing g over S must also minimize the function f over C.*

One of the purposes of this paper is to extend Clarke's exact penalty principle to the case where f is vector-valued. Our result in the special case where the objective function f is scalar-valued proved the following improved Clarke Exact Penalty Principle which is a corollary of Theorem 3.1.

Theorem 1.2 (Improved Clarke's exact penalty principle) *Let X be a normed space, $C \subset S \subset X$ and $f : X \to R$ be Lipschitz of rank L_f on S. Then for $L > L_f$, f attains a minimum over C at x if and only if the function $g(y) = f(y) + Ld_C(y)$ attains a minimum over S at x.*

Unfortunately when *local optimal solutions* are considered, the reverse statement of Clarke's exact penalty principle does not hold without additional conditions. In [17], Scholtes and Stöhr gave some conditions which ensure the reverse statements hold for the distance function and the error bound function. In this paper we extend these results to the vector optimization case.

Under the assumption that S is compact and the local (global) optimal solutions of the problem (P) lies in the interior of the set S, Di Pillo and Grippo [4, 5] showed that the extended Mangasarian Fromovitz constraint qualification (EMFCQ) can be used to insure that the local (global) minimizers of the penalized problem with penalty function $\psi(x) := \rho(\|h(x)\|_1 + \|g(x)_+\|_1)$, where $\|\cdot\|_1$ denote the L^1 norm, for some large enough $\rho > 0$ coincide with the local (global) minimizers of the original problem (P). In this paper we show that the results of Di Pillo and L. Grippo hold under much weaker conditions. In particular we do not require that the optimal solutions lie in the interior of the set S and we show the results hold under only the existence of local error bounds. For any point $\bar{x}$ in the feasible region of (P), the condition that a local error bound holds at $\bar{x}$ is equivalent to the condition that the perturbed feasble region of the problem (P) as a set-valued mapping defined by

$$C(p, q) := \{x \in S : g(x) + p \leq 0, h(x) = q\} \tag{1}$$

is calm at $(0, 0, \bar{x})$ and hence will hold if the set-valued mapping $C(p, q)$ is pseudo-Lipschitz continuous at $(0, 0, \bar{x})$. Consequently, the pseudo-Lipschitz continuity can be characterized by using Mordukhovich criteria for pseudo-Lipschitz continuity [13].

Throughout the paper we use standard notation. Unless otherwise stated, all spaces considered are normed space whose norms are always denoted by $\|\cdot\|$. For any Banach space X we consider its dual space X^* equipped with the weak-star topology w^*, where $\langle \cdot, \cdot \rangle$ means the canonical pairing. For a set-valued mapping $\Phi : X \rightrightarrows Y$, we denote its graph by $gph\Phi := \{(x,y) : y \in \Phi(x)\}$. We denote by $\bar{B}$ the closed unit ball centered at the origin, $B(\bar{x},\delta)$ the open ball centered at $\bar{x}$ with radius $\delta > 0$ and $\bar{B}(\bar{x},\delta)$ the closed ball centered at $\bar{x}$ with radius $\delta > 0$.

2 Preliminaries

2.1 Preliminary results for variational analysis

We now recall some of the concepts in variational analysis that will be used in this paper. For more detailed discussion, the reader is referred to [3, 11, 12, 16].

Let X, Y be Banach spaces. For set-valued mapping $\Phi : X \rightrightarrows Y$, we denote by $\limsup_{x \to \bar{x}} \Phi(x)$ the sequential *Painlevé–Kuratowski upper limit* with respect to the norm topology in X and the weak-star topology in X^*, i.e.,

$$\begin{aligned}\limsup_{x \to \bar{x}} \Phi(x) \;\; := \;\; & \{x^* \in X^* : \exists \text{ sequences } x_k \to \bar{x}, x_k^* \stackrel{w^*}{\to} x^* \\ & \text{with } x_k^* \in \Phi(x_k) \;\; \forall k = 1, 2, \ldots\}.\end{aligned}$$

Definition 2.1 (Normal cones) *Let X be an Asplund space and $S \subset X$. Given $x \in clS$ where clS denotes the closure of S, the cone*

$$N_S^F(x) := \left\{ x^* \in X^* : \limsup_{u \xrightarrow{S} x} \frac{\langle x^*, u - x\rangle}{\|u - x\|} \leq 0 \right\}$$

is called the Fréchet normal cone (the regular normal cone or the prenormal cone) to S at x. Let $\bar{x} \in clS$. The nonempty cone

$$N_S^L(\bar{x}) := \limsup_{x \to \bar{x}} N_S^F(x)$$

is called the limiting (the basic or Mordukhovich) normal cone to S at $\bar{x}$.

Definition 2.2 (Subdifferentials) *Let X be an Asplund space and S be a subset of X. Let $\varphi : S \to R$ be Lipschitz around $\bar{x} \in S$. The set*

$$\partial^F \varphi(\bar{x}) := \left\{ x^* \in X^* : (x^*, -1) \in N^F_{epi\varphi}(\bar{x}, \varphi(\bar{x})) \right\}$$

is called the Fréchet (or the regular) subdifferential of φ at $\bar{x}$ and the set

$$\partial^L \varphi(\bar{x}) := \left\{ x^* \in X^* : (x^*, -1) \in N^L_{epi\varphi}(\bar{x}, \varphi(\bar{x})) \right\}$$

is called the limiting (the basic or Mordukhovich) subdifferential of φ at $\bar{x}$, where $epi\varphi := \{(x,r) : \varphi(x) \leq r\}$ denotes the epigraph of φ.

Definition 2.3 (Normal compactness condition) *Let X be an Asplund space. A closed subset S of X is said to be normally compact around $\bar{x} \in S$ if there exist positive numbers γ, σ and a compact subset Ω of X such that*

$$N_S^L(\bar{x}) \subset K_\sigma(\Omega) := \left\{ x^* \in X^* : \sigma \|x^*\| \leq \max_{c \in \Omega} |\langle x^*, c \rangle| \right\} \forall x \in B(\bar{x}, \gamma) \cap S.$$

In proving the exact penalty principle of optimization problems, we need to discuss stability of some perturbed feasible regions. Hence we recall the following Lipschitz properties of a set-valued mapping. The definition of upper Lipschitz continuity was first studied by Robinson [14] and the concept of pseudo-Lipschitz continuity was introduced by Aubin in [1] and it is also referred to as "Lipschitz like" (see [11, 12]). Although the term "calmness" was coined in [16], the concept of the calmness was first introduced by Ye and Ye in [19] under the term "pseudo upper-Lipschitz continuity".

Definition 2.4 *Let $\Phi : X \rightrightarrows Y$ be a set-valued mapping. Φ is said to be (locally) upper-Lipschitz continuous at $\bar{x} \in X$ if there exist a constant $\mu \geq 0$ and a neighborhood U of $\bar{x}$ such that*

$$\Phi(x) \subset \Phi(\bar{x}) + \mu \|x - \bar{x}\| \bar{B}, \qquad \forall x \in U.$$

Φ is said to be pseudo-Lipschitz continuous (or Aubin continuous or Lipschitz like) at $(\bar{x}, \bar{y}) \in gph\Phi$ if there exist a constant $\mu \geq 0$, a neighborhood U of $\bar{x}$ and a neighborhood V of $\bar{y}$ such that

$$\Phi(x) \cap V \subset \Phi(x') + \mu \|x - x'\| \bar{B}, \qquad \forall x, x' \in U.$$

Φ is said to be calm (or pseudo upper-Lipschitz continuous [19]) at $(\bar{x}, \bar{y}) \in gph\Phi$ if there exist a constant $\mu \geq 0$, a neighborhood U of $\bar{x}$ and a neighborhood V of $\bar{y}$ such that

$$\Phi(x) \cap V \subset \Phi(\bar{x}) + \mu \|x - \bar{x}\| \bar{B}, \qquad \forall x \in U.$$

It is obvious from the definitions that both the local upper Lipschitz and the pseudo-Lipschitz continuity implies the calmness.

2.2 Preliminary results for vector optimization problems

Recall that a subset K of a real topological vector space W is a cone if $\lambda y \in K$ for all $y \in K$ and $\lambda \geq 0$, a convex cone is one for which $\lambda_1 y_1 + \lambda_2 y_2 \in K$ for all $\lambda_1, \lambda_2 \geq 0$ and a pointed cone is one for which $K \cap (-K) = \{0\}$ where 0 denotes the zero element in W. Let W be a normed space and K be a cone in W. We consider the preference relation for two vectors $x, y \in W$ introduced by cone K given as follows

$$x \prec y \Longleftrightarrow x - y \in K \setminus \{0\},$$
$$x \preceq y \Longleftrightarrow x - y \in K.$$

In particular, if $W = R^N$ and $K = R^N_- := \{z \in R^N : z \text{ has nonpositive components}\}$ then we have a preference in the Pareto sense and if $W = R^N$ and $K = intR^N_- \cup \{0\}$ where $intC$ denotes the interior of set C, then we have a preference in the weak Pareto sense.

Now consider the function $f : X \to W$ and assume that f is Lipschitz near $\bar{x}$ in the following sense.

Definition 2.5 *Let S be a subset of X. Suppose that $f : X \to W$ and K is a cone of W. We say that f is K-Lipschitz on S (of rank L_f) if there is a constant $L_f > 0$, an element $e \in (-K)$ with $\|e\| = 1$ such that*

$$f(x') \preceq f(x'') + L_f\|x' - x''\|e \quad \forall x', x'' \in S.$$

Let $\bar{x} \in X$. We say that f is K-Lipschitz near $\bar{x}$ if there is $U(\bar{x})$, a neighborhood of $\bar{x}$ such that f is K-Lipschitz on $U(\bar{x})$.

Note that the K-Lipschitz continuity is not the same as the usual Lipschitz continuity. The following property can be obtained easily from the definition.

Proposition 2.1 *Let $K_1 \subset K_2$. K_1-Lipschitz continuity implies K_2-Lipschitz continuity.*

When $W = R^n$, $K_1 = R^n_+$, the K_1-Lipschitz continuity is the same as the usual Lipschitz continuity. Therefore from the above proposition it is immediate that if $K_2 \supset R^n_+$ and f is Lipschitz in the usual sense then f is also K_2-Lipschitz continuous.

Definition 2.6 *$\bar{x}$ is said to be a global K-minimizer of f on C provided that there is no $x \in C$ such that*

$$f(x) \prec f(\bar{x}).$$

$\bar{x}$ is a local K-minimizer of f on C provided that it is a global K-minimizer of f on $C \cap \bar{B}(\bar{x}, \varepsilon)$ for some $\varepsilon > 0$.

When $W = R^N$ and $K = R^N_-$, the concept of a minimizer is the same as an efficient or a Pareto solution and when $K = intR^N_- \cup \{0\}$, it coincides with the concept of a weakly efficient or a weak Pareto solution.

Lemma 2.1 *(i) Let K be a cone and $e \in -K$. If $\alpha \leq \beta$ and $\alpha, \beta \in R$, then*

$$\alpha e \preceq \beta e;$$

(ii) Let K be a cone and $e \in -K$. If $\alpha < \beta$ and $\|e\| \neq 0$ then

$$\alpha e \prec \beta e.$$

(iii) If K is a convex and pointed cone, then $a \preceq b$ and $b \prec c$ or $a \prec b, b \preceq c$ implies that $a \prec c$.

(iv) If K is a convex cone then $a \preceq b$ and $b \preceq c$ imply $a \preceq c$.

Proof. (i) Since $\beta - \alpha \geq 0$ and $-e \in K$ and K is a cone, we have

$$\alpha e - \beta e = (\beta - \alpha)(-e) \in K$$

which implies that $\alpha e \preceq \beta e$.

(ii) The assertion follows from (i) and the assumptions that $\beta - \alpha > 0$ and $e \neq 0$.

(iii) Let $a \preceq b$ and $b \prec c$. Since $a - b \in K$ and $b - c \in K \setminus \{0\}$, we have by the convexity of K that

$$a - c = (a - b) + (b - c) \in K.$$

We now prove that $a \neq c$ by contradiction. If $a = c$, then $b - a = b - c \neq 0$ which implies that $b - a \in K \cap (-K)$ and $b - a \neq 0$. This contradicts the pointedness of the cone K. Hence $a = b$. The conclusion for $a \prec b, b \preceq c$ implying $a \prec c$ is similar.

■

The following result is an extension of [17, Lemma 2.5] to the case of infinite dimensional spaces and to the case where C is not necessarily closed.

Lemma 2.2 *Let C be a nonempty set and $\bar{x} \in C$. Then for any $\varepsilon > \delta > 0$ and any $y \in \bar{B}(\bar{x}, \frac{\varepsilon - \delta}{2})$,*

$$d_C(y) = d_{C \cap \bar{B}(\bar{x}, \varepsilon)}(y).$$

Moreover if C is a closed subset of a finite dimensional space, then δ can be chosen as 0 in the above statement.

Proof. One always have

$$d_C(y) \leq d_{C\cap \bar{B}(\bar{x},\varepsilon)}(y).$$

To prove the reverse inequality, let $\{y_i\}$ be a minimizing sequence for the distance function $d_C(y)$, i.e., $y_i \in C$ and

$$d_C(y) \leq \|y_i - y\| \leq d_C(y) + \frac{1}{i}.$$

Then

$$\begin{aligned}
\|y_i - \bar{x}\| &\leq \|y_i - y\| + \|y - \bar{x}\| \\
&= d_C(y) + \frac{1}{i} + \|y - \bar{x}\| \\
&\leq 2\|y - \bar{x}\| + \frac{1}{i} \qquad \text{since } \bar{x} \in C \\
&\leq \varepsilon \quad \forall y \in \bar{B}(\bar{x}, \frac{\varepsilon - \delta}{2}) \text{ and } \frac{1}{i} \leq \delta.
\end{aligned}$$

Therefore $y_i \in C \cap \bar{B}(\bar{x}, \varepsilon)$ for i large enough and hence

$$\|y - y_i\| \geq d_{C\cap \bar{B}(\bar{x},\varepsilon)}(y).$$

Taking the limit as $i \to \infty$, we have the reverse inequality.

Moreover if X is a finite dimensional space and C is closed, then the projection exists, that is $d_C(y) = \|y - \bar{x}\|$ for some $\bar{x} \in C$. Therefore from the above proof it is obvious that δ can be chosen as 0 in this case.

■

3 Exact penalization for distance function

In this section we discuss global and local exact penalization for distance functions. We first extend Clarke's exact penalty principle to the vector optimization case.

Theorem 3.1 (Global exact penalization for distance function) *Let X, W be normed spaces, $S \subset X$, $C \subset S$ and $K \subset W$ be a convex and pointed cone. Let $f : S \to W$ be K-Lispchitz on S of rank L_f. Let e be the element in $-K$ given by the K-Lipschitz continuity of f.*

(i) Assume that $K \setminus \{0\}$ is an open set. Then any global K-minimizer of f on C is a global K-minimizer of the exact penalty function $f(x) + L_f d_C(x)e$ on S.

(ii) Assume that either C is closed or $K\setminus\{0\}$ is an open set. Then for any $L > L_f$, $\bar{x}$ is a global K-minimizer of f on C if and only if it is a global K-minimizer of the exact penalty function $f(x) + Ld_C(x)e$ on S.

Proof. By the K-Lipschitz continuity of f, there is a constant $L_f > 0$, an element $e \in (-K)$ with $\|e\| = 1$, such that

$$f(x^*) \preceq f(x) + L_f\|x - x^*\|e \quad \forall x^*, x \in S. \tag{2}$$

We prove (i) by contradiction. Suppose that $\bar{x}$ is a K-global minimizer of f on C but not a global K-minimizer for $f(x) + L_f d_C(x)e$ on S. Then there exists $x \in S$ such that

$$f(x) + L_f d_C(x)e \prec f(\bar{x}). \tag{3}$$

Since $K \setminus \{0\}$ is open, (3) implies the existence of a small enough $\varepsilon > 0$ such that

$$f(x) + L_f d_C(x)e \prec f(\bar{x}) - L_f\varepsilon e. \tag{4}$$

By definition of the distance function, there exists $x^*_\varepsilon \in C$ such that $\|x - x^*_\varepsilon\| \leq d_C(x) + \varepsilon$. Therefore we have

$$\begin{aligned} f(x^*_\varepsilon) &\preceq f(x) + L_f\|x - x^*_\varepsilon\|e, \quad \text{by (2)} \\ &\preceq f(x) + L_f(d_C(x) + \varepsilon)e, \quad \text{by Lemma 2.1 (i)} \\ &\prec f(\bar{x}) \quad \text{by (4)} \end{aligned}$$

which implies by Lemma 2.1 (iii) that $f(x^*_\varepsilon) \prec f(\bar{x})$. This contradicts the fact that $\bar{x}$ minimizes f on C and hence the conclusion of (i) holds.

We now prove (ii). Suppose that $\bar{x}$ is a global minimizer of f on C but not a global minimizer for $f(x) + Ld_C(x)e$ on S and $L > L_f$. Then there exists $x \in S$ such that

$$f(x) + Ld_C(x)e \prec f(\bar{x}). \tag{5}$$

Observe that x can not lie in the set C since otherwise $\bar{x}$ would not be a global minimizer of f on C. Suppose that C is closed. Then $x \notin C$ would imply $d_C(x) > 0$. Therefore since $\frac{L}{L_f} > 1$, one can pick $x^* \in C$ such that $\|x - x^*\| < \frac{L}{L_f} d_C(x)$. Therefore one has

$$\begin{aligned} f(x^*) &\preceq f(x) + L_f\|x - x^*\|e \quad \text{by (2)} \\ &\prec f(x) + Ld_C(x)e \quad \text{by Lemma 2.1 (ii)} \\ &\prec f(\bar{x}) \quad \text{by (5)} \end{aligned}$$

which implies by Lemma 2.1 (iii) that $f(x^*) \prec f(\bar{x})$. This contradicts the fact that $\bar{x}$ minimizes f on C and hence the necessity in (ii) under the assumption that C is closed is proved. In the case where C is not closed but $K \setminus \{0\}$ is open, (5) implies the existence of a small enough $\varepsilon > 0$ such that

$$f(x) + Ld_C(x)e \prec f(\bar{x}) - L\varepsilon e. \tag{6}$$

Pick $x^*_\varepsilon \in C$ such that $\|x - x^*_\varepsilon\| \leq d_C(x) + \varepsilon$. Then

$$\begin{aligned} f(x^*_\varepsilon) &\preceq f(x) + L_f\|x - x^*_\varepsilon\|e \quad \text{by (2)} \\ &\preceq f(x) + L_f(d_C(x) + \varepsilon)e \quad \text{by Lemma 2.1 (i)} \\ &\prec f(x) + L(d_C(x) + \varepsilon)e \quad \text{by Lemma 2.1 (ii)} \\ &\prec f(\bar{x}) \quad \text{by (6)} \end{aligned}$$

which implies by Lemma 2.1 (iii) that $f(x^*_\varepsilon) \prec f(\bar{x})$. This contradicts the fact that $\bar{x}$ minimizes f on C and hence the necessity in (ii) holds.

We now prove the sufficiency in (ii) by contradiction. Let $\bar{x}$ is a minimizer of $f(x) + Ld_C(x)e$ on S with $\bar{x} \in C$ but not a minimizer of f on C and $L > L_f$. Then there is $x \in C$ such that

$$f(x) \prec f(\bar{x}).$$

Since $x \in C$ implies that $d_C(x) = 0$, the above relationship implies that

$$f(x) + Ld_C(x)e \prec f(\bar{x}) + Ld_C(\bar{x})e$$

contradicting the fact that $\bar{x}$ is a minimizer of $f(x) + Ld_C(x)e$ on S. Now it remains to prove that it is not possible to have $\bar{x} \notin C$ and $\bar{x}$ be a minimizer of $f(x) + Ld_C(x)e$ on S. In the case where C is closed, since $\bar{x} \notin C$ and $\frac{L}{L_f} > 1$, one can pick $x^* \in C$ such that $\|\bar{x} - x^*\| < \frac{L}{L_f} d_C(\bar{x})$. Then

$$\begin{aligned} f(x^*) &\preceq f(\bar{x}) + L_f\|x^* - \bar{x}\|e \quad \text{by K-Lipschitz continuity of } f \\ &\prec f(\bar{x}) + Ld_C(\bar{x})e \quad \text{by Lemma 1.1 (i) .} \end{aligned}$$

Hence by Lemma 1.1 (iii) the above implies that

$$f(x^*) + Ld_C(x^*)e \prec f(\bar{x}) + Ld_C(\bar{x})e.$$

This contradicts the fact that $\bar{x}$ is a minimizer of $f(x) + Ld_C(x)e$ on S. Therefore $\bar{x}$ must lie in C. Now suppose that C is not closed but $K \setminus \{0\}$ is open. Let $\varepsilon > 0$ and

$x_\varepsilon^* \in C$ be such that $\|\bar{x} - x_\varepsilon^*\| \leq d_C(\bar{x}) + \varepsilon$. Then

$$\begin{aligned} f(x_\varepsilon^*) &\preceq f(\bar{x}) + L_f\|x_\varepsilon^* - \bar{x}\|e \quad \text{by K-Lipschitz continuity of } f \\ &\preceq f(\bar{x}) + L_f(d_C(\bar{x}) + \varepsilon)e \quad \text{by Lemma 1.1 (i)} \\ &\prec f(\bar{x}) + L(d_C(\bar{x}) + \varepsilon)e \quad \text{by Lemma 1.1 (ii).} \end{aligned}$$

Hence by Lemma 1.1 (iii) the above implies that

$$f(x_\varepsilon^*) + Ld_C(x_\varepsilon^*)e \prec f(\bar{x}) + Ld_C(\bar{x})e + L\varepsilon e.$$

Since $\varepsilon > 0$ is arbitrary and $K \setminus \{0\}$ is open, this contradicts the fact that $\bar{x}$ is a minimizer of $f(x) + Ld_C(x)e$ on S. Therefore $\bar{x}$ must lie in C. ■

Remark 3.1 *In the case of the usual single objective optimization problem, $K = (-\infty, 0]$ and $K \setminus \{0\}$ is open. Our theorem has recovered and improved Clarke's exact penalty principle for global minimium in that C is not required to be a closed set.*

The following result is an extension of [17, Corollary 2.6] to the case of infinite dimensional spaces and to the case where C is not necessarily closed. Note that although the assumption $C \cap \bar{B}(\bar{x}, \varepsilon) \neq \emptyset$ is not explicitly stated in [17, Corollary 2.6], it is clear that it is needed from the proof of the result.

Theorem 3.2 (Local exact penalization for distance functions) *Let X, W be normed spaces, $C \subset S \subset X$ and $K \subset W$ be a convex and pointed cone. Let $\bar{x} \in S$. Suppose that there exists a positive constant ε such that $f : S \to W$ is K-Lipschitz on $\bar{B}(\bar{x}, \varepsilon)$ of rank L_f. Let e be the element in $-K$ given by the Lipschitz continuity of f.*

(i) Assume that $K \setminus \{0\}$ is an open set. Let $\bar{x}$ be a local minimizer of f on C. Then for any $L \geq L_f$ it is a local minimizer of the exact penalty function: $f(x) + Ld_C(x)e$ on S. Assume that C is closed then for any $L > L_f$, if $\bar{x}$ is a local minimizer of f on C then it is a local minimizer of the exact penalty function: $f(x) + Ld_C(x)e$ on S.

(ii) Assume that either C is closed or $K \setminus \{0\}$ is an open set and $L > L_f$. Suppose that $\bar{x}$ is a minimizer of the exact penalty function $f(x) + Ld_C(x)e$ on $S \cap \bar{B}(\bar{x}, \varepsilon)$ and $C \cap \bar{B}(\bar{x}, \varepsilon) \neq \emptyset$. Then $\bar{x}$ is a minimizer of the function f on $C \cap \bar{B}(\bar{x}, \varepsilon)$.

Proof. Proof of (i) under the assumption that $K \setminus \{0\}$ is an open set and $L \geq L_f$: Suppose that $\bar{x}$ is a local minimizer of f on C but not a local minimizer of the exact penalty function: $f(x) + Ld_C(x)e$ on S. Then there exists a closed ball $\bar{B}(\bar{x}, \varepsilon)$ such that $\bar{x}$ is a global minimizer of f over $C \cap \bar{B}(\bar{x}, \varepsilon)$ and f is K-Lipschitz of rank L_f on $\bar{B}(\bar{x}, \varepsilon)$. Hence Theorem 3.1 (i) implies that $\bar{x}$ is a global minimizer of $f(x) + Ld_{C \cap \bar{B}(\bar{x},\varepsilon)}(x)e$ on $\bar{B}(\bar{x}, \varepsilon) \cap S$. By Lemma 2.2, $\bar{x}$ is a global minimizer of $f(x) + Ld_C(x)e$ on $\bar{B}(\bar{x}, \varepsilon/3) \cap S$. Therefore $\bar{x}$ is a local minimizer of $f + Ld_C(x)e$ on S.

Using Theorem 3.1 (ii) instead of Theorem 3.1 (i) in the above proof one can prove (i) under the assumption that C is closed and $L > L_f$.

Proof of (ii): Let $\bar{x}$ be a minimizer of $f(x) + Ld_C(x)e$ on $S \cap \bar{B}(\bar{x}, \varepsilon)$ with $\bar{x} \in C$ but not a local minimizer of f on $C \cap \bar{B}(\bar{x}, \varepsilon)$. Then there is $x \in C \cap \bar{B}(\bar{x}, \varepsilon)$ such that

$$f(x) \prec f(\bar{x})$$

which implies that

$$f(x) + Ld_C(x)e \prec f(\bar{x}) + Ld_C(\bar{x})e$$

contradicting that $\bar{x}$ is a minimizer of $f(x) + Ld_C(x)e$ on $S \cap \bar{B}(\bar{x}, \varepsilon)$. Now it remains to prove that it is not possible to have $\bar{x} \notin C$, $C \cap \bar{B}(\bar{x}, \varepsilon) \neq \emptyset$ for some $\varepsilon > 0$ and $\bar{x}$ is a minimizer of $f(x) + Ld_C(x)e$ on $S \cap \bar{B}(\bar{x}, \varepsilon)$. By assumption f is K-Lipschitz on $\bar{B}(\bar{x}, \varepsilon)$ of rank L_f and $C \cap \bar{B}(\bar{x}, \varepsilon) \neq \emptyset$. In the case where C is closed, since $\bar{x} \notin C$ and $\frac{L}{L_f} > 1$, one can pick $x^* \in C \cap \bar{B}(\bar{x}, \varepsilon)$ be such that $\|\bar{x} - x^*\| < \frac{L}{L_f} d_C(\bar{x})$. Then

$$\begin{aligned} f(x^*) &\preceq f(\bar{x}) + L_f \|x^* - \bar{x}\| e \quad \text{by K-Lipschitz continuity of } f \\ &\prec f(\bar{x}) + Ld_C(\bar{x})e \quad \text{by Lemma 1.1 (i) .} \end{aligned}$$

Hence by Lemma 1.1 (iii) the above implies that

$$f(x^*) + Ld_C(x^*)e \prec f(\bar{x}) + Ld_C(\bar{x})e.$$

This contradicts the fact that $\bar{x}$ is a minimizer of $f(x) + Ld_C(x)e$ on $S \cap \bar{B}(\bar{x}, \varepsilon)$. Therefore $\bar{x}$ must lie in C. For the case where $K \setminus \{0\}$ is open. Let $\delta > 0$ be small enough and $x^*_\delta \in C$ be such that $\|\bar{x} - x^*_\delta\| \leq d_C(\bar{x}) + \delta$ and x^*_δ is included in the ball $\bar{B}(\bar{x}, \varepsilon)$. Then we have

$$\begin{aligned} f(x^*_\delta) &\preceq f(\bar{x}) + L_f \|x^*_\delta - \bar{x}\| e \\ &\preceq f(\bar{x}) + L_f (d_C(\bar{x}) + \delta) e \\ &\prec f(\bar{x}) + L(d_C(\bar{x}) + \delta) e. \end{aligned}$$

Hence the above implies that

$$f(x_\delta^*) + Ld_C(x_\delta^*)e \prec f(\bar{x}) + Ld_C(\bar{x})e + L\delta e.$$

Since $\delta > 0$ is arbitrary and $K \setminus \{0\}$ is an open set, this contradicts the fact that $\bar{x}$ is a minimizer of $f(x) + Ld_C(x)e$ on $S \cap \bar{B}(\bar{x}, \varepsilon)$. Therefore $\bar{x}$ must lie in C.

■

4 Exact penalization for merit function

Although the distance function is always an exact penalty function, it is difficult to evaluate since it is usually an implicit nonsmooth function of the data in the original problem. One is therefore interested in finding exact penalty functions which are easier to evaluate. Let X be a normed space and $C \subset S \subset X$. According to [20], we call a function $\psi(y) : S \to R$ a merit function provided that

$$\psi(y) \geq 0 \quad \forall y \in S \text{ and } \psi(y) = 0 \text{ if and only if } y \in C.$$

Although the distance function $d_C(y)$ is obviously a merit function, one can usually find some merit functions that are more trackable than the distance function. It is obvious that a merit function provides the following equivalent formulation of the problem of minimizing f over C:

$$\min f(x) \quad s.t. \quad \psi(x) = 0, \quad x \in S.$$

Definition 4.1 (Global error bound function) *We say that a merit function $\psi : S \to R$ is a global error bound function if*

1. *$\psi(x) \geq d_C(x)$ for every $x \in S$,*

2. *$\psi(x) = d_C(x)$ for every $x \in C$.*

By definition, an error bound function is a majorant of the distance function. Therefore the following exact penalty result follows from applying Theorem 3.1.

Theorem 4.1 (Global exact penalty for merit function) *Let X, W be normed spaces, $C \subset S \subset X$ and $K \subset W$ be a convex and pointed cone. Assume that $f : S \to W$ is K-Lispchitz on S and e is the element in $-K$ given by the K-Lipschitz continuity of f. Suppose that $\psi : S \to R$ is a global error bound function.*

(i) Assume that $K \setminus \{0\}$ is an open set. Then any global K-minimizer of f on C is a global K-minimizer of the exact penalty function $f(x) + L_f\psi(x)e$ on S.

(ii) Assume that either C is closed or $K \setminus \{0\}$ is an open set. Then $\bar{x}$ is a global minimizer of f on C if and only if it is a global minimizer of the exact penalty function $f(x) + L\psi(x)e$ on S for any $L > L_f$.

Proof. Proof of (i): Suppose that $\bar{x}$ is a global K-minimizer of f on C and to the contrary suppose that $\bar{x}$ is not a global K-minimizer of the exact penalty function: $f(x) + L_f\psi(x)e$ on S. Then by definition of the K optimality, there is $x \in S$ such that

$$f(x) + L_f\psi(x)e \quad \prec f(\bar{x}) + L_f\psi(\bar{x})e.$$

But since $d_C(x) \leq \psi(x)$ for every $x \in S$ and $\psi(\bar{x}) = 0$, the above relationship implies that

$$f(x) + L_f d_C(x)e \prec f(\bar{x}) + L_f d_C(\bar{x})e.$$

But this is a contradiction since by Theorem 3.1, $\bar{x}$ is a global K-minimizer of $f(x) + L_f d_C(x)e$ on S.

Proof of (ii): Let $L > L_f$ and $\bar{x}$ be a K-minimizer of $f(x) + L\psi(x)e$ on S with $\bar{x} \in C$ but not a K-minimizer of f on C. Then there is $x \in C$ such that

$$f(x) \prec f(\bar{x})$$

which implies that

$$f(x) + L\psi(x)e \prec f(\bar{x}) + L\psi(\bar{x})e$$

contradicting that $\bar{x}$ is a minimizer of $f(x) + L\psi(x)e$ on S. Now it remains to prove that it is not possible to have $\bar{x} \notin C$ and $\bar{x}$ is a minimizer of $f(x) + L\psi(x)e$ on S. In the case where C is closed, since $\bar{x} \notin C$ and $\frac{L}{L_f} > 1$, one can pick $x^* \in C$ such that $\|\bar{x} - x^*\| < \frac{L}{L_f} d_C(\bar{x})$. Then

$$\begin{aligned} f(x^*) &\preceq f(\bar{x}) + L_f\|x^* - \bar{x}\|e \quad \text{by K-Lipschitz continuity of } f \\ &\prec f(\bar{x}) + L d_C(\bar{x})e \quad \text{by Lemma 1.1 (i) .} \end{aligned}$$

which implies that

$$f(x^*) + L\psi(x^*)e \prec f(\bar{x}) + L d_C(\bar{x})e \preceq f(\bar{x}) + L\psi(\bar{x})e.$$

This contradicts the fact that $\bar{x}$ is a K-minimizer of $f(x) + L\psi(x)e$ on S. Therefore $\bar{x}$ must lie in C. For the case where $K \setminus \{0\}$ is open. Let $\delta > 0$ be small enough and $x^*_\delta \in C$ be such that $\|\bar{x} - x^*_\delta\| \leq d_C(\bar{x}) + \delta$. Then we have

$$\begin{aligned} f(x^*_\delta) &\preceq f(\bar{x}) + L_f\|x^*_\delta - \bar{x}\|e \\ &\preceq f(\bar{x}) + L_f(d_C(\bar{x}) + \delta)e \\ &\prec f(\bar{x}) + L(d_C(\bar{x}) + \delta)e. \end{aligned}$$

Hence the above implies that

$$f(x^*_\delta) + L\psi(x^*_\delta)e \prec f(\bar{x}) + Ld_C(\bar{x})e + L\delta e \preceq f(\bar{x}) + L\psi(\bar{x})e + L\delta e.$$

Since $\delta > 0$ is arbitrary and $K \setminus \{0\}$ is an open set, this contradicts the fact that $\bar{x}$ is a minimizer of $f(x) + L\psi(x)e$ on S. Therefore $\bar{x}$ must lie in C.

■

We can also give the local version of the above results as follows.

Definition 4.2 (Local error bound function) *Let $\bar{x} \in S$ and $C \subset S \subset X$. We say that a merit function $\psi : S \to R$ is a local error bound function on $\bar{B}(\bar{x}, \varepsilon)$ with $\varepsilon > 0$ if*

1. *$\psi(x) \geq d_C(x)$ for every $x \in S \cap \bar{B}(\bar{x}, \varepsilon)$,*
2. *$\psi(x) = d_C(x)$ for every $x \in C$.*

Theorem 4.2 (Local exact penalty for merit function) *Let X, W be normed spaces and $C \subset S \subset X$ and $K \subset W$ be a convex and pointed cone. Let $\bar{x} \in S$. Suppose that one can find a positive constant $\varepsilon > 0$ such that $f : S \to W$ is K-Lipschitz on $\bar{B}(\bar{x}, \varepsilon)$ of rank L_f and $\psi : S \to R$ is an error bound function on $\bar{B}(\bar{x}, \varepsilon)$. Let e be the element in $-K$ given by the Lipschitz continuity of f. Then the following statements hold.*

(i) *Assume that $K \setminus \{0\}$ is an open set. For any $L \geq L_f$ if $\bar{x}$ is a local minimizer of f on $C \subset S$ then it is a local minimizer of the exact penalty function: $f(x) + L\psi(x)e$ on S. Conversely assume that C is closed then for any $L > L_f$. If $\bar{x}$ is a local minimizer of f on $C \subset S$ then it is a local minimizer of the exact penalty function: $f(x) + L\psi(x)e$ on S.*

(ii) Assume that either C is closed or $K \setminus \{0\}$ is an open set and $L > L_f$. Suppose that $\bar{x}$ is a minimizer of the exact penalty function $f(x)+L\psi(x)e$ on $S\cap\bar{B}(\bar{x},\varepsilon)$ and $C \cap \bar{B}(\bar{x},\varepsilon) \neq \emptyset$. Then $\bar{x}$ is a minimizer of the function f on $C \cap \bar{B}(\bar{x},\varepsilon)$.

Proof. Proof of (i): Suppose that $\bar{x}$ is a local minimizer of f on C. By the assumption, one can find ε small enough so that $\bar{x}$ is a minimizer of f on $C\cap\bar{B}(\bar{x},\varepsilon)$, f is Lipschitz continuous on $\bar{B}(\bar{x},\varepsilon)$ and ψ is an error bound function on $\bar{B}(\bar{x},\varepsilon)$. Suppose that $\bar{x}$ is not a minimizer of the exact penalty function: $f(x) + L\psi(x)e$ on $S \cap \bar{B}(\bar{x},\varepsilon)$. Then there is $x \in S \cap \bar{B}(\bar{x},\varepsilon)$ such that

$$f(x) + L\psi(x)e \quad \prec f(\bar{x}) + L\psi(\bar{x})e = f(\bar{x}).$$

But since $d_C(x) \leq \psi(x)$ for every $x \in S \cap \bar{B}(\bar{x},\varepsilon)$, the above implies

$$f(x) + Ld_C(x)e \prec f(\bar{x}) + Ld_C(\bar{x})e.$$

But this is a contradiction since by Theorem 3.2, $\bar{x}$ is a local minimizer of $f(x) + Ld_C(x)e$ on S.

Proof of (ii): Let $\bar{x}$ be a minimizer of $f(x) + L\psi(x)e$ on $S \cap \bar{B}(\bar{x},\varepsilon)$ with $\bar{x} \in C$ but not a local minimizer of f on $C \cap \bar{B}(\bar{x},\varepsilon)$. Then there is $x \in C \cap \bar{B}(\bar{x},\varepsilon)$ such that

$$f(x) \prec f(\bar{x})$$

which implies that

$$f(x) + L\psi(x)e \prec f(\bar{x}) + L\psi(\bar{x})e$$

contradicting that $\bar{x}$ is a minimizer of $f(x) + L\psi(x)e$ on $S \cap \bar{B}(\bar{x},\varepsilon)$. Now it remains to prove that it is not possible to have $\bar{x} \notin C$, $C \cap \bar{B}(\bar{x},\varepsilon) \neq \emptyset$ for some $\varepsilon > 0$ and $\bar{x}$ is a minimizer of $f(x) + L\psi(x)e$ on $S \cap \bar{B}(\bar{x},\varepsilon)$. By assumption f is K-Lipschitz on $\bar{B}(\bar{x},\varepsilon)$ rank L_f and $C \cap \bar{B}(\bar{x},\varepsilon) \neq \emptyset$. In the case where C is closed, since $\bar{x} \notin C$ and $\frac{L}{L_f} > 1$, one can pick $x^* \in C \cap \bar{B}(\bar{x},\varepsilon)$ such that $\|\bar{x} - x^*\| < \frac{L}{L_f} d_C(\bar{x})$. Then

$$\begin{aligned} f(x^*) &\preceq f(\bar{x}) + L_f\|x^* - \bar{x}\|e \quad \text{by K-Lipschitz continuity of } f \\ &\prec f(\bar{x}) + Ld_C(\bar{x})e \quad \text{by Lemma 1.1 (i) .} \end{aligned}$$

which implies that

$$f(x^*) + L\psi(x^*)e \prec f(\bar{x}) + Ld_C(\bar{x})e \preceq f(\bar{x}) + L\psi(\bar{x})e.$$

This contradicts the fact that $\bar{x}$ is a minimizer of $f(x) + L\psi(x)e$ on $S \cap \bar{B}(\bar{x}, \varepsilon)$. Therefore $\bar{x}$ must lie in C. For the case where $K \setminus \{0\}$ is open. Let $\delta > 0$ be small enough and $x_\delta^* \in C$ be such that $\|\bar{x} - x_\delta^*\| \leq d_C(\bar{x}) + \delta$ and x_δ^* is included in the ball $\bar{B}(\bar{x}, \varepsilon)$. Then we have

$$\begin{aligned} f(x_\delta^*) &\preceq f(\bar{x}) + L_f\|x_\delta^* - \bar{x}\|e \\ &\preceq f(\bar{x}) + L_f(d_C(\bar{x}) + \delta)e \\ &\prec f(\bar{x}) + L(d_C(\bar{x}) + \delta)e. \end{aligned}$$

Hence the above implies that

$$f(x_\delta^*) + L\psi(x_\delta^*)e \prec f(\bar{x}) + Ld_C(\bar{x})e + L\delta e \preceq f(\bar{x}) + L\psi(\bar{x})e + L\delta e.$$

Since $\delta > 0$ is arbitrary and $K \setminus \{0\}$ is an open set, this contradicts the fact that $\bar{x}$ is a minimizer of $f(x) + L\psi(x)e$ on $S \cap \bar{B}(\bar{x}, \varepsilon)$. Therefore $\bar{x}$ must lie in C.

■

5 Exact penalty principle for optimization problems

We first consider the following multiobjective optimization problem with linear inequality, equality and a polyhedral convex abstract constraint:

$$\begin{array}{llll} (\widetilde{P}) & \text{“min”} & f(x) & \\ & \text{s.t.} & Ax = b, & \\ & & \langle x_i^*, x\rangle \leq c_i, & i = 1, \cdots, n, \\ & & x \in S & \end{array}$$

where X, Y are Banach spaces, S is a polyhedral convex set of X, W is a normed space, $f : X \to W$, $A : X \to Y$ is a linear continuous mapping such that the range of A is closed and $x_i^* \in X^*$, $c_i \in R$ and “min” means K-optimality where K is a convex and pointed cone in W. For the case where X, Y are finite dimensional space, Hoffman [8] showed that a linear inequality system must have a global error bound. Ioffe [9] extended Hoffman's celebrated result to allow the spaces X, Y to be Banach spaces and hence we have the following global exact penalty result.

Theorem 5.1 *Suppose that f is $K-$Lipschitz on S of rank L_f and e is the element in $-K$ given by the Lipschitz continuity of f .*

(i) Assume that $K \setminus \{0\}$ is an open set. Then there exists a constant $\mu > 0$ such that for all $L \geq L_f$, any global K-minimizer of $(\widetilde{P})$ is a global K-minimizer of the penalized problem:

$$(\widetilde{P}_{L\mu}) \qquad \min \quad f(x) + L\mu(\|Ax - b\| + \sum_{i=1}^{n}(\langle x_i^*, x\rangle - c_i)_+)e$$
$$\text{s.t.} \quad x \in S,$$

where $a_+ := \max\{0, a\}$ for a real number a.

(ii) Assume that either S is closed or $K \setminus \{0\}$ is an open set. Let $L > L_f$ and suppose that $\bar{x}$ is a global K-minimizer of $(\widetilde{P}_{L\mu})$. Then $\bar{x}$ is a global K-minimizer of $(\widetilde{P})$.

Proof. By Ioffe [9, Theorem 3], there exists a constant $\mu > 0$ such that

$$d_C(x) \leq \mu(\|Ax - b\| + \sum_{i=1}^{n}(\langle x_i^*, x\rangle - c_i)_+) \qquad x \in S.$$

Since $\psi(x) := \mu(\|Ax - b\| + \sum_{i=1}^{n}(\langle x_i^*, x\rangle - c_i)_+) = 0$ for any x in C, the feasible region of $(\widetilde{P})$, the above implies that $\psi(x)$ is an error bound function. The results follow from applying Theorem 4.1.

■

We now consider the exact penalty results for the following multiobjective optimization problem with linear inequality, equality and an abstract constraint:

$$(P) \qquad \text{“}\min\text{”} \quad f(x)$$
$$\text{s.t.} \quad g(x) \leq 0,$$
$$h(x) = 0,$$
$$x \in S$$

where $f : X \to W, g : X \to R^m, h : X \to R^n$, $S \subset X$ and “min” means K-optimality where K is a convex and pointed cone in W. We denote the feasible region of problem (P) by C.

Theorem 5.2 *Let X be a finite dimensional space, the mappings g, h be affine and the set S be a union of finitely many polyhedral convex sets. Suppose one can find a* *positive number $\varepsilon > 0$ such that f is K-Lipschitz on $\bar{B}(\bar{x}, \varepsilon)$ of rank L_f. Let e be the element in $-K$ given by the Lipschitz continuity of f. Then the following statements are true.*

(i) Assume that $K \setminus \{0\}$ is an open set . Let $L \geq L_f$. If $\bar{x}$ is a local K- minimizer of (P) then it is a local K-minimizer of the exact penalized problem:

$$(P_{L\mu}) \qquad \text{“min”} \quad f(x) + L\mu(\|h(x)\|_1 + \|g(x)_+\|_1)e$$
$$\text{s.t.} \quad x \in S.$$

(ii) Assume that either S is closed or $K \setminus \{0\}$ is an open set and $L > L_f$. Suppose that $\bar{x}$ is a local K-minimizer of $(P_{L\mu})$ on $B(\bar{x}, \varepsilon)$ and $C \cap B(\bar{x}, \varepsilon) \neq \emptyset$. Then $\bar{x}$ is a local minimizer of the problem (P) .

Proof. By Robinson [15], under the assumption of the theorem, the perturbed feasible region as a set-valued mapping $C(p, q)$ defined as in (1) is upper Lipschitz, i.e., there exists $\mu \geq 0$ and U, a neighborhood of $(0, 0)$ such that

$$C(p, q) \subset C(0, 0) + \mu\|(p, q)\|\bar{B} \qquad \forall(p, q) \in U.$$

It is easy to see that the upper Lipschitz continuity of the set valued mapping $C(p, q)$ implies the existence of a local error bound i.e., there exist constants $\mu_1 > 0, \varepsilon > 0$ such that

$$d_C(x) \leq \mu_1(\|h(x)\|_1 + \|g(x)_+\|_1) \qquad x \in S \cap B(\bar{x}, \varepsilon).$$

Since $\psi(x) := L\mu_1(\|h(x)\|_1 + \|g(x)_+\|_1) = 0$ for $x \in C$, the feasible region of (P), $\psi(x)$ is a local error bound function on $B(\bar{x}, \varepsilon)$. The results follow from applying Theorem 4.2.

■

In our global exact penalty results Theorems 4.1 and 5.1, we require the existence of a global error bound and in our local exact penalty results Theorems 4.2 and 5.2 we require the existence of a local error bound and the Lipschitz continuity of the objective function on a ball which intersects the feasible region. The global error bound condition, however, is quite strong for nonlinear functions and the Lipschitz continuity of the objective function on a ball which intersects the feasible region is restrictive. In the following result we show that these conditions can be replaced by the local error bound condition when the set S is compact in the case where the problem is a single-objective optimization problem. Our results extend the results of Di Pillo and Grippo [4, 5] to a more general setting. In particular we do not require that the local (global) optimal solutions to lie in the interior of the set S.

Theorem 5.3 *Suppose that X is a Banach space and S is compact subset of X. $f : X \to R$ is Lipschitz on S, $W = R_-$ and the mappings g, h are continuous. Suppose that the local error bound holds for all $\bar{x} \in C$, i.e., for each $\bar{x} \in C$, there exist positive numbers $\mu_{\bar{x}}$ and $\varepsilon_{\bar{x}}$ such that the following local error bound holds:*

$$d_C(x) \quad \leq \quad \mu_{\bar{x}}(\|h(x)\|_1 + \|g(x)_+\|_1) \qquad \forall x \in S \cap B(\bar{x}, \varepsilon_{\bar{x}}),$$

or equivalently the perturbed set-valued mapping $C(p, q)$ defined by (1) is calm around $(0, 0, \bar{x})$ for all $\bar{x} \in C$. Let S_{opt} denote the set of optimal solutions of problem (P). Then there exists a constant $\hat{\rho}$ such that S_{opt} coincides with S^{ρ}_{opt}, the set of global minimizers of the penalized problem:

$$\begin{array}{llll} (P_\rho) & \quad \min & \quad f(x) + \rho(\|h(x)\|_1 + \|g(x)_+\|_1) \\ & \quad \text{s.t.} & \quad x \in S, \end{array}$$

for all $\rho > \hat{\rho}$.

Proof. Since S is compact both sets S_{opt} and S^{ρ}_{opt} are not empty. We first prove that there exists some $\hat{\rho} > 0$ such that $S^{\rho}_{opt} \subset S_{opt}$ for $\rho > \hat{\rho}$. Assume for a contradiction that this is not true. Then for any $\rho_k > 0$, there exists $x_k \in S^{\rho_k}_{opt}$ but $x_k \notin S_{opt}$. Let $\rho_k \to \infty$. The compactness of S ensures that we can draw a subsequence if necessary such that $x_k \to x^* \in S$. Let $\bar{x} \in S_{opt}$. Since $x_k \in S^{\rho_k}_{opt}$, we have

$$f(x_k) + \rho_k(\|h(x_k)\|_1 + \|g(x_k)_+\|_1) \leq f(\bar{x})$$

which implies that

$$\rho_k(\|h(x_k)\|_1 + \|g(x_k)_+\|_1) \leq f(\bar{x}) - f(x_k)$$

Taking limits on both sides of the formula above by the continuity of the functions f, g, h we obtain

$$\|h(x^*)\|_1 + \|g(x^*)_+\|_1 = 0, \qquad 0 \leq f(\bar{x}) - f(x^*).$$

This means that x^* is an optimal solution of (P). Under the local error bound assumption, there exists a positive constant $\hat{\rho}$ such that x^* is a local minimizer of

$f(x) + \rho(\|h(x)\|_1 + \|g(x)_+\|_1)$ on S for all $\rho > \hat{\rho}$. Since $x_k \to x^*$ and $x_k \in S_{opt}^{\rho_k}$, we may find a neighborhood of x^*, denoted by $U(x^*)$, such that both x_k and x^* are minima of $f(x) + \rho_k(\|h(x)\|_1 + \|g(x)_+\|_1)$ over the set $S \cap U(x^*)$ for all $\rho_k \geq \hat{\rho}$. Consequently

$$\begin{aligned} f(x_k) + \rho_k(\|h(x_k)\|_1 + \|g(x_k)_+\|_1) &= f(x^*) + \rho_k(\|h(x^*)\|_1 + \|g(x^*)_+\|_1) \\ &= f(x^*) + \frac{\rho_k + \hat{\rho}}{2}(\|h(x^*)\|_1 + \|g(x^*)_+\|_1) \\ &\leq f(x_k) + \frac{\rho_k + \hat{\rho}}{2}(\|h(x_k)\|_1 + \|g(x_k)_+\|_1) \end{aligned}$$

Since $\rho_k > \hat{\rho}$, the above inequality implies that $\|h(x_k)\|_1 + \|g(x_k)_+\|_1 = 0$ and hence $x_k \in S_{opt}$ which contradicts the fact that $x_k \notin S_{opt}$. This shows that $S_{opt}^{\rho} \subseteq S_{opt}$ for all $\rho > \hat{\rho}$.

We are now ready to show that for any $\rho \geq \hat{\rho}$, $S_{opt}^{\rho} = S_{opt}$. Let $\tilde{x} \in S_{opt}$ and $x_\rho \in S_{opt}^{\rho}$. Then for any $\rho > \hat{\rho}$, since $S_{opt}^{\rho} \subset S_{opt}$, we have $x_\rho \in S_{opt}$. Therefore

$$f(x_\rho) + \rho(\|h(x_\rho)\|_1 + \|g(x_\rho)_+\|_1) = f(\tilde{x}) + \rho(\|h(\tilde{x})\|_1 + \|g(\tilde{x})_+\|_1),$$

which means that $\tilde{x}$ is also in S_{opt}^{ρ}. The proof is complete. ■

Corollary 5.1 *Suppose that X is a finite dimensional space, $f : X \to R$ is Lipschitz on S, $W = R_-$ and h, g are affine. Furthermore suppose that S is compact and it is a union of finite number of polyhedral convex sets. Then there exists a constant $\hat{\rho}$ such that $S_{opt} = S_{opt}^{\rho}$ for all $\rho > \hat{\rho}$.*

Proof. As shown in the proof of Theorem 5.2, by Robinson [15] the local error bound holds for each $\bar{x} \in C$. The results follow from applying Theorem 5.3.

■

Definition 5.1 (NNAMCQ) *Let X be an Asplund space, $S \subset X$ and $x \in S$. Assume that g, h are Lipschitz near x and S is a closed set. We say that the no nonzero abnormal multiplier constraint qualification (NNAMCQ) holds at x if there is no nonzero vector $(\lambda^g, \lambda^h) \in R^m_+ \times R^n$ such that*

$$\begin{aligned} &0 \in \partial^L[\langle \lambda^g, g\rangle + \langle \lambda^h, h\rangle](x) + N_S^L(x) \\ &\lambda_i^g g_i(x) = 0 \quad \forall i \in I(x) \end{aligned}$$

where $I(x) := \{i : g_i(x) = 0\}$ is the index set of active inequality constraints at x.

Corollary 5.2 *Suppose that X is an Asplund space and S is a compact subset of X which is normally compact around each point of C (which holds automatically if X is a finite dimensional space). Assume that $f : X \to R$ is Lipschitz on S, $W = R_-$ and h, g are locally Lipschitz at each point of S. Furthermore suppose that NNAMCQ holds at each point of C. Then there exists a constant $\hat{\rho}$ such that $S_{opt} = S_{opt}^{\rho}$ for all $\rho > \hat{\rho}$.*

Proof. By Mordukhovich and Shao [13, Corollary 6.2], the condition NNAMCQ at $\bar{x} \in C$ implies that the perturbed set-valued mapping $C(p, q)$ defined by (1) is pseudo-Lipschitzian around $(0, 0, \bar{x})$, i.e., there exist $\mu \geq 0$, V a neighborhood of $\bar{x}$ and U, a neighborhood of $(0, 0)$ such that

$$C(p', q') \cap V \subset C(p, q) + \mu \|(p', q') - (p, q)\| \bar{B} \qquad \forall (p', q'), (p, q) \in U.$$

It is easy to see that the pseudo-Lipschitz continuity of the set valued mapping $C(p, q)$ implies the existence of a local error bound i.e., there exist constants $\mu_1 > 0, \varepsilon > 0$ such that

$$d_C(x) \leq \mu_1 (\|h(x)\|_1 + \|g(x)_+\|_1) \qquad x \in S \cap B(\bar{x}, \varepsilon).$$

Since $\psi(x) := L\mu_1(\|h(x)\|_1 + \|g(x)_+\|_1) = 0$ for $x \in C$, $\psi(x)$ is a local error bound function on $B(\bar{x}, \varepsilon)$. The results follow from applying Theorem 5.3.

■

Definition 5.2 (Extended MFCQ) *Let X be a finite dimensional space and $S \subset X$. Assume that g, h are continuously differentiable at $\bar{x}$ and S is a closed set. We say that the extended MFCQ (EMFCQ) holds at $\bar{x}$ if the vectors $\{\nabla h_j(\bar{x}), j = 1, \ldots, n\}$ are linearly independent and there exists a vector $d \in intT_S(\bar{x})$ such that*

$$g_i'(\bar{x}; d) < 0 \qquad i \in I_+(\bar{x})$$
$$h_j'(\bar{x}; d) = 0 \qquad j = 1, \ldots, n,$$

where $I_+(\bar{x}) := \{i : g_i(\bar{x}) \geq 0\}$, $\phi'(\bar{x}; d)$ denotes the directional derivative of ϕ at $\bar{x}$ in direction d and and $T_S(\bar{x})$ denotes the Clarke tangent cone to S at $\bar{x}$.

It is easy to see that for a feasible solution $\bar{x}$, EMFCQ is reduced to the usual MFCQ which is equivalent to the NNAMCQ (see e.g. [10]).

The next lemma is a slight generalization of [4, Lemma 3.1] which is a slight generalization of the "only if" part of [7, Theorem 2.2].

Lemma 5.1 *Let $\bar{x} \in X$, a finite dimensional space; let I be a given subset of $\{1, \ldots, m\}$; and assume that EMFCQ holds at $\bar{x}$. Then exists an open neighborhood $B(\bar{x}, \rho)$ and a bounded function $d : B(\bar{x}, \rho) \to R^n$ such that, for all $x \in B(\bar{x}; \rho)$, we have:*

$$d(x) \in intT_S(\bar{x}), \tag{7}$$

$$\nabla g_i(x)^T d(x) < -1 \quad i \in I_+(\bar{x}), \tag{8}$$

$$\nabla h_j(x)^T d(x) = \begin{cases} -1, & \text{if } h_j(x) > 0 \\ 0, & \text{if } h_j(x) = 0 \\ 1, & \text{if } h_j(x) < 0 \end{cases} \qquad \forall j = 1, \ldots, m, \tag{9}$$

where $\nabla f(x)$ denote the gradient of a function f at x.

Theorem 5.4 *Let X be a finite dimensional space. Suppose that $f : X \to R$ is Lipschitz on S of rank L_f, $W = R_-$ and g, h are continuously differentiable. Assume that S is compact and EMFCQ holds at each point of S. Then there exists a constant $\hat{\rho}$ such that for all $\rho \geq \hat{\rho}$, if x_ρ is a local minimum for the penalized problem:*

$$(P_\rho) \qquad \min \qquad f_\rho(x) := f(x) + \rho(\|h(x)\|_1 + \|g(x)_+\|_1)$$
$$x \in S$$

then x_ρ is also a local constrained minimizer of Problem (P).

Proof. We first show that x_ρ must lie in the feasible region C. Reasoning by contracdition, assume that the assertion is false. Then for any integer k, there exist an $\mu_k \geq k$ and a point $x_k \in S$ which is a local minimum of (P_k) such that $x_k \notin C$. Since S is compact, there exists a convergent subsequence [relabel it again $\{x_k\}$] such that

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

By assumption, we have that the hypotheses of Lemma 5.1 are statisfied at $\bar{x}$ for $I = I_+(\bar{x})$. Then by Lemma 5.1, there exists $B(\bar{x}, \rho(\bar{x}))$ and a bounded function $d(x) : B(\bar{x}, \rho(\bar{x})) \to R^n$ such that for all $x \in B(\bar{x}; \rho(\bar{x}))$, (7)-(9) hold.

By continuity, we can find $\sigma(\bar{x}) \leq \rho(\bar{x})$ such that $g_i(x) < 0$ for $x \in B(\bar{x}, \sigma(\bar{x}))$ and $i \notin I_+(\bar{x})$, whence $I_+(x) \subset I_+(\bar{x})$. Therefore we have that for $x \in B(\bar{x}, \sigma(\bar{x}))$,

$$\xi_i(x, d(x)) \leq 0, \quad i = 1, \ldots, n$$

$$\zeta_j(x, d(x)) \leq 0, \quad j = 1, \ldots, m.$$

Suppose that $x \in B(\bar{x}, \sigma(\bar{x}))$ be a point in S which is not in C. Then there must exist at least an index $i \in I_+(\bar{x})$ such that $g_i(x) > 0$ or an index j such that $|h_j(x)| > 0$. Moreover since

$$f'_\rho(x; d) = \nabla f(x)^T d + \rho(\sum_{i=1}^{m} \zeta_j(x, d) + \sum_{i=1}^{n} \xi_i(x, d)),$$

where

$$\zeta_j(x, d) = \begin{cases} \nabla h_j(x)^T d & \text{if } h_j(x) > 0 \\ |\nabla h_j(x)^T d| & \text{if } h_j(x) = 0 \\ -\nabla h_j(x)^T d & \text{if } h_j(x) < 0 \end{cases}$$

$$\xi_i(x, d) = \begin{cases} \nabla g_i(x)^T d & \text{if } g_i(x) > 0 \\ |(\nabla g_i(x)^T d| & \text{if } g_i(x) = 0 \\ 0 & \text{if } g_i(x) < 0 \end{cases}$$

by (7)-(9), we have

$$f'_\rho(x; d(x)) \leq \nabla f(x)^T d(x) - \rho.$$

This implies that, for a sufficient large value of k, we have

$$x_k \in B(\tilde{x}, \sigma(\tilde{x})) \quad \text{and } f'_{\rho_k}(x_k; d(x_k)) < 0, \quad d(x_k) \in intT_S(x_k)$$

which contradicts the assumption that x_k is a local optimal minimizer of $f_{\rho_k}(x)$ on S.

■

References

[1] J.P. Aubin, *Lipschitz behavior of solutions to convex minimization problems*, Math. Oper. Res. **9**(1994), pp. 87-111.

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

[3] F.H. Clarke, Yu. S. Ledyaev, R. J. Stern, and P.R. Wolenski, *Nonsmooth Analysis and Control Theory*, Springer, New York, 1998.

[4] G. Di Pillo and L. Grippo, *On the exactness of a class of nondifferentiable penalty functions*, J. Optim. Theo. Appl. **57**(1988), pp. 399-410.

[5] G. Di Pillo and L. Grippo, *Exact penalty functions in constrained optimization*, SIAM J. Contr. Optim. **27**(1989), pp. 1333-1360.

[6] I.I. Eremin, *Penalty method in convex programming*, Soviet Math. Dolk. **8**(1966), pp. 459-462.

[7] S.-P. Han and O. L. Mangasarian, *Exact penalty functions in nonlinear programming*, **17**(1979), pp. 251-269.

[8] A. J. Hoffman, *On approximate solutions of systems of linear inequalities*, J. Res. Nat. Bur. Stand., **49**(1952), pp. 263-265.

[9] A.D. Ioffe, *Regular points of Lipschitz functions*, Trans. AMS **251**(1979), pp. 61-69.

[10] A. Jourani, *Constraint qualifications and Lagrange multipliers in nondifferentiable programming problems*, J. Optim. Theo. Appl. **81**(1994), pp. 533-548.

[11] B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation, Vol.1: Basic Theory*, Springer, 2006.

[12] B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation, Vol.2: Applications*, Springer, 2006.

[13] B.S. Mordukhovich and Y. Shao, *Stability of set-valued mappings in infinite dimensions: point criteria and applications*, SIAM J. Control Optim. **32**(1997), pp. 285-314.

[14] S. M. Robinson, *Stability theory for systems of inequalities, part I: linear systems*, SIAM J. Numer. Anal., **12**(1975), pp. 754-769.

[15] S. M. Robinson, *Some continuity properties of polyhedral multifunctions*, Math. Program., **14**(1981), pp. 206-214.

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

[17] S. Scholtes and M. Stöhr, *Exact penalization of mathematical programs with equilibrium constraints*, SIAM J. Contr. Optim. **37**(1999), pp. 617-652.

[18] Z. Wu and J.J. Ye, *First-order and second-order conditions for error bounds*, SIAM J. Optim. **14**(2003), pp. 621-645.

[19] J.J. Ye and X.Y. Ye, *Necessary optimality conditions for optimization problems with variational inequality constraints*, Math. Oper. Res. **22**(1997), pp.977-997.

[20] J.J. Ye, D.L. Zhu and Q.J. Zhu, *Exact penalization and necessary optimality conditions for generalized bilevel programming problems*, SIAM J. Optim. **2**(1997), pp. 481-507.

[21] W.I. Zangwill, *Nonlinear programming via penalty functions*, Manage. Sci. **13**(1967), pp. 344-358.