Some Exact Penalty Results for Nonlinear Programs and Their Applications to Mathematical Programs with Equilibrium Constraints

Gui-Hua Lin* and Masao Fukushima†

May 29, 2001

Abstract. Recently, some exact penalty results for nonlinear programs, especially for mathematical programs with equilibrium constraints were proved by Luo, Pang, and Ralph. In this paper, we will show that those results remain valid under some other mild conditions. Especially, one of these conditions, called strong convexity with order σ, is discussed in detail.

Key Words. mathematical program with equilibrium constraints, nonlinear complementarity problem, subanalytic set/function, Hölder continuity, strong convexity, error bound.

1 Introduction

A mathematical program with equilibrium constraints (MPEC) is a constrained optimization problem in which the essential constraints are defined by a parametric variational inequality or complementarity system. This problem plays an important role in many fields such as engineering design, economic equilibrium, and multilevel game, see [10].

Because of the presence of variational inequality or complementarity constraints, MPEC has such an intrinsic feature that its constraints fail to satisfy a standard constraint qualification and hence it is very difficult to handle. At present, a popular approach is to reformulate an MPEC as a standard nonlinear problem. In this respect, penalty functions have provided a powerful approach, both as a theoretical tool and as a computational vehicle. Recently, based on the study of subanalytic optimization problems and with the help of the theory of error bounds, some exact penalty results for nonlinear programs and MPECs were proved by Luo, Pang, and Ralph in [10]. However, the subanalytic property and error bounds are somewhat abstract and difficult to verify in practice. In this paper, we will show that those theorems remain valid under some other mild conditions. In particular, some of our results use a property

*Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan. E-mail Address: ghlin@amp.i.kyoto-u.ac.jp.

†Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan. E-mail Address: fuku@amp.i.kyoto-u.ac.jp.

called strong convexity with order σ, which is a generalization of the ordinary strong convexity [8].

The following notations and definitions will be used throughout this paper. We denote the projection operator from R^n onto its subset W by Π_W. For a real scalar u, $(u)_+$ denotes $\max(0, u)$. In addition, $B(0, c)$ stands for the closed ball $\{x \in R^n : ||x|| \leq c\}$, and $||\cdot||$ and $||\cdot||_1$ denote the norms defined by

$$||x|| = (\sum_{i=1}^{n} x_i^2)^{1/2} \quad \text{and} \quad ||x||_1 = \sum_{i=1}^{n} |x_i|$$

for $x \in R^n$, respectively.

Definition 1.1 ([10]) Let $f : R^n \to R$ and $X \subseteq R^n$. X is said to be subanalytic, if for any $u \in R^n$, there exist a neighborhood U of u and a bounded set $Z \subseteq R^{n+p}$ with some nonnegative integer p such that

(1) for any $v \in R^{n+p}$, there exist a neighborhood V of v and a finite family $\{Z_{ij} : 1 \leq i \leq l, 1 \leq j \leq q\}$ of sets $Z_{ij} = \{z \in V : f_{ij}(z) = 0\}$ or $\{z \in V : f_{ij}(z) < 0\}$ defined for some real analytic functions f_{ij} on V such that

$$Z \cap V = \bigcup_{i=1}^{l} \bigcap_{j=1}^{q} Z_{ij};$$

(2) $X \cap U = \{x \in R^n : (x, y) \in Z \text{ for some } y \in R^p\}$.
If the graph of f is subanalytic, we call f subanalytic.

The class of subanalytic functions is broader than the class of analytic functions, although it is somewhat abstract. For more details, we refer the reader to [1,5,9,10].

Definition 1.2 ([12]) Let $0 < p \leq 1$ be a constant and $G : R^n \to R^m$ be a mapping. G is said to be Hölder continuous with order p on $X \subseteq R^n$, if there exists a constant $L > 0$ such that

$$||G(x) - G(y)|| \leq L||x - y||^p, \quad \forall x, y \in X. \tag{1.1}$$

This concept is a generalization of the Lipschitz continuity when $0 < p \leq 1$. For example, the function

$$G(x) = \sqrt{||x||}, \quad \forall x \in R^n$$

is Hölder continuous with order $p = \frac{1}{2}$ on R^n and not Lipschitz continuous on R^n. So, for $0 < p \neq p' \leq 1$, Hölder continuous functions with order p and those with order p' constitute different classes of functions. But for $p > 1$, the condition (1.1) implies that all directional derivatives of G at any interior point are zero and so G is quite trivial.

Definition 1.3 Let $G : R^n \to R^m$, $X \subseteq R^n$, and $S = \{x \in X : G(x) = 0\}$ be a closed subset of X. We say that $||G||$ provides an error bound of order σ on S, if there exist positive constants σ and β such that

$$||G(x)|| \geq \beta||x - \Pi_S(x)||^\sigma, \quad \forall x \in X. \tag{1.2}$$

The condition (1.2) may not be easy to verify in practice. In this paper, we show that if X is convex and G is strongly convex with order σ on X as defined below, then (1.2) holds. We will further discuss this property in Section 4.

Definition 1.4 A function $f : R^n \to R$ is said to be strongly convex with order $\sigma > 0$ on a convex set $X \subseteq R^n$, if there exists a constant $c > 0$ such that

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - ct(1-t)\|x - y\|^\sigma \tag{1.3}$$

holds for all $x, y \in X$ and $t \in [0, 1]$.

When $\sigma = 2$, this property reduces to the strong convexity in the ordinary sense [8]. But if $\sigma \neq 2$, they are different. For example, we can see from the results given in Section 4 that the function $f(x) = x^4$ is strongly convex with order 4 and not strongly convex (with order 2) on $K = (0, +\infty)$.

2 Penalty Results for Nonlinear Programs

Consider the following nonlinear program:

$$\begin{array}{ll} \min & \theta(x) \\ \text{s.t.} & x \in X \\ & g(x) \leq 0,\ h(x) = 0, \end{array} \tag{2.1}$$

where $\theta : R^n \to R, g : R^n \to R^m$, and $h : R^n \to R^l$ are all continuous functions and $X \subseteq R^n$ is a nonempty closed set. Let W denote the feasible region of (2.1) and

$$r(x) = \sum_{i=1}^{m} (g_i(x))_+ + \sum_{j=1}^{l} |h_j(x)|$$

be the residual for the constraints in (2.1) at $x \in X$. Then the function $r : R^n \to R$ may be used as a penalty function for problem (2.1). Let $r|_X$ denote the restriction of r on X. The following theorem is shown in [10]:

Theorem 2.1 *Let $X \subseteq R^n$ be a compact subanalytic set, θ be Lipschitz continuous on X, and g_i, h_j be continuous subanalytic for all i, j. Suppose problem (2.1) is feasible. Then there exist positive constants α^* and γ^* such that for $\alpha \geq \alpha^*$ and $\gamma \geq \gamma^*$, problem (2.1) is equivalent to*

$$\begin{array}{ll} \min & \theta(x) + \alpha r(x)^{1/\gamma} \\ \text{s.t.} & x \in X \end{array} \tag{2.2}$$

in the sense that x^* solves (2.1) if and only if it solves (2.2).

In particular, the following result is useful both in [10] and in our paper:

Theorem 2.2 (Lojasiewicz inequality, [1]) *Let $\phi, \psi : S \to R$ be continuous subanalytic and $S \subseteq R^n$ be compact subanalytic. If $\phi^{-1}(0) \subseteq \psi^{-1}(0)$, then there exist constants $\rho > 0$ and $N^* > 0$ such that*

$$\rho|\psi(x)|^{N^*} \leq |\phi(x)|, \quad \forall x \in S. \tag{2.3}$$

Now we give our penalty results for problem (2.1).

Theorem 2.3 *Suppose that X, g, h are the same as in Theorem 2.1, problem (2.1) is feasible, and the function $(\theta - \theta \circ \Pi_W)$ is continuous subanalytic. Then the conclusion of Theorem 2.1 is also valid.*

Proof. Noticing that both r and $(\theta - \theta \circ \Pi_W)$ are continuous subanalytic and

$$(r|_X)^{-1}(0) = W \subseteq (\theta - \theta \circ \Pi_W)^{-1}(0),$$

we have from Theorem 2.2 that there exist constants $\rho > 0$ and $N^* > 0$ such that

$$\rho|\theta(x) - \theta(\Pi_W(x))|^{N^*} \leq r(x), \quad \forall x \in X.$$

Let

$$\mu = \max\{1, \max_{x \in X} r(x)\}, \quad \alpha^* > (\mu/\rho)^{1/N^*}, \quad \gamma^* = N^*$$

and

$$\alpha \geq \alpha^*, \quad \gamma \geq \gamma^*.$$

(a) Let $\bar{x}$ solve problem (2.1). Then for any $x \in X$,

$$\begin{aligned}
\theta(x) + \alpha r(x)^{1/\gamma} &= \theta(z) + (\theta(x) - \theta(z)) + \alpha r(x)^{1/\gamma^*} r(x)^{1/\gamma - 1/\gamma^*} \\
&\geq \theta(\bar{x}) - |\theta(x) - \theta(z)| + \alpha \rho^{1/\gamma^*} |\theta(x) - \theta(z)| \mu^{1/\gamma - 1/\gamma^*} \\
&\geq \theta(\bar{x}) + (\alpha \rho^{1/\gamma^*} \mu^{-1/\gamma^*} - 1)|\theta(x) - \theta(z)| \\
&\geq \theta(\bar{x}) \\
&= \theta(\bar{x}) + \alpha r(\bar{x})^{1/\gamma},
\end{aligned}$$

where $z \in \Pi_W(x)$. Therefore, $\bar{x}$ is a global optimal solution of problem (2.2).

(b) If $\bar{x}$ solves (2.2), we can claim that $\bar{x}$ is an optimal solution of problem (2.1). In fact, since W is compact and (2.1) is feasible, it has an optimal solution, which we denote $\tilde{x}$. In a similar way to (a), we have

$$\begin{aligned}
\theta(\tilde{x}) &= \theta(\tilde{x}) + \alpha r(\tilde{x})^{1/\gamma} \\
&\geq \theta(\bar{x}) + \alpha r(\bar{x})^{1/\gamma} \\
&\geq \theta(\tilde{x}) + (\alpha \rho^{1/\gamma^*} \mu^{-1/\gamma^*} - 1)|\theta(\bar{x}) - \theta(\bar{z})| \\
&\geq \theta(\tilde{x}),
\end{aligned}$$

where $\bar{z} \in \Pi_W(\bar{x})$. This implies $\theta(\bar{x}) = \theta(\bar{z})$ and then

$$\theta(\bar{x}) \geq \theta(\tilde{x}) \geq \theta(\bar{x}) + \alpha r(\bar{x})^{1/\gamma},$$

where the first inequality holds because $\tilde{x}$ solves (2.1), $\bar{x}$ is feasible to (2.1), and $\theta(\bar{x}) = \theta(\bar{z})$. Hence, we have $r(\bar{x}) = 0$ and $\theta(\bar{x}) = \theta(\tilde{x})$. The former implies $\bar{x} \in W$ and so $\bar{x}$ is an optimal solution to problem (2.1). This completes the proof. □

Theorem 2.4 *Assume that X, g, and h are the same as in Theorem 2.3, θ is Hölder continuous with order $p > 0$ on X, and problem (2.1) is feasible. Then the conclusion of Theorem 2.1 holds.*

Proof. Let $\psi(x) = \text{dist}(x, W)$. Then ψ is continuous subanalytic and

$$(r|_X)^{-1}(0) = \psi^{-1}(0).$$

We have from Theorem 2.2 that there exist constants $\rho > 0$ and $N^* > 0$ such that

$$\rho\psi(x)^{N^*} \leq r(x), \quad \forall x \in X. \tag{2.4}$$

Since θ is Hölder continuous with order p on X, there exists some $L > 0$ such that

$$|\theta(x) - \theta(y)| \leq L||x - y||^p, \quad \forall x, y \in X. \tag{2.5}$$

Let

$$\mu = \max\{1, \max_{x \in X} r(x)\}, \ \alpha^* > L(\mu/\rho)^{p/N^*}, \ \gamma^* = N^*/p$$

and

$$\alpha \geq \alpha^*, \quad \gamma \geq \gamma^*.$$

(a) Suppose that $\bar{x}$ solves problem (2.1). Then for any $x \in X$,

$$\begin{aligned}
\theta(x) + \alpha r(x)^{1/\gamma} &= \theta(z) + (\theta(x) - \theta(z)) + \alpha r(x)^{1/\gamma^*} r(x)^{1/\gamma - 1/\gamma^*} \\
&\geq \theta(\bar{x}) - L||x - z||^p + \alpha\rho^{1/\gamma^*}||x - z||^p \mu^{1/\gamma - 1/\gamma^*} \\
&\geq \theta(\bar{x}) + (\alpha\rho^{1/\gamma^*}\mu^{-1/\gamma^*} - L)||x - z||^p \\
&\geq \theta(\bar{x}) \\
&= \theta(\bar{x}) + \alpha r(\bar{x})^{1/\gamma},
\end{aligned}$$

where $z \in \Pi_W(x)$ and the first inequality follows from (2.5), (2.4), and the fact that $\psi(x) = ||x - z||$. Therefore, $\bar{x}$ is a global optimal solution of problem (2.2).

(b) Let $\bar{x}$ solve problem (2.2). Since W is compact and (2.1) is feasible, it has an optimal solution $\tilde{x}$. It follows from (a) that

$$\begin{aligned}
\theta(\tilde{x}) &= \theta(\tilde{x}) + \alpha r(\tilde{x})^{1/\gamma} \\
&\geq \theta(\bar{x}) + \alpha r(\bar{x})^{1/\gamma} \\
&\geq \theta(\tilde{x}) + (\alpha\rho^{1/\gamma^*}\mu^{-1/\gamma^*} - L)||\bar{x} - \bar{z}||^p \\
&\geq \theta(\tilde{x}),
\end{aligned}$$

where $\bar{z} \in \Pi_W(\bar{x})$. This implies $\bar{x} = \bar{z}$ and $\theta(\bar{x}) = \theta(\tilde{x})$, i.e., $\bar{x}$ is an optimal solution to problem (2.1). □

For the special case of (2.1) in which $l = 0$, namely,

$$\begin{aligned}
\min \quad & \theta(x) \\
\text{s.t.} \quad & x \in X \\
& g(x) \leq 0,
\end{aligned} \tag{2.6}$$

we have the following penalty results which will be applied to the mathematical program with a nonlinear complementarity system in the next section. In the rest of this section, φ denotes the function defined by

$$\varphi(x) = \max_{1 \leq i \leq m} g_i(x).$$

Theorem 2.5 *Let X be a closed subset of R^n, θ be Hölder continuous with order p and Hölder constant L on X, and g_i be continuous on X for each $i = 1, \cdots, m$. Assume that (1.2) holds for $G = \varphi_+$ with the corresponding constants σ and β on*

$$S = \{x \in X : G(x) = 0\} = \{x \in X : \varphi(x) \leq 0\}$$

and suppose that problem (2.6) is feasible. Then problem (2.6) has the same solution set as the problem

$$\begin{array}{ll} \min & \theta(x) + \alpha r(x)^{N^*} \\ \text{s.t.} & x \in X, \end{array} \tag{2.7}$$

where

$$r(x) = \sum_{i=1}^{m} (g_i(x))_+, \ N^* = \frac{p}{\sigma}, \ \alpha > L\beta^{-N^*}.$$

Proof. Note that S is just the feasible set of (2.6). Then for any $x \in X$, we have

$$r(x) \geq \beta ||x - z||^{\sigma}, \tag{2.8}$$

where $z \in \Pi_S(x)$. In fact, we have for each $x \notin S$ that

$$\varphi(x) > 0 \quad \text{and} \quad \varphi(z) = 0. \tag{2.9}$$

It follows from (2.9) and (1.2) with $G = \varphi_+$ on S that

$$r(x) = \sum_{i=1}^{m} (g_i(x))_+ \geq (\varphi(x))_+ = |G(x)| \geq \beta ||x - z||^{\sigma}$$

and hence (2.8) holds for any $x \in X$.

(a) If $\bar{x}$ solves problem (2.6), then for any $x \in X$, we have from (2.8) and the Hölder continuity of θ that

$$\begin{aligned} \theta(x) + \alpha r(x)^{N^*} &= \theta(z) + (\theta(x) - \theta(z)) + \alpha r(x)^{p/\sigma} \\ &\geq \theta(\bar{x}) + (\alpha\beta^{p/\sigma} - L)||x - z||^p \\ &\geq \theta(\bar{x}) \\ &= \theta(\bar{x}) + \alpha r(\bar{x})^{N^*}, \end{aligned}$$

where $z \in \Pi_S(x)$. Therefore, $\bar{x}$ is a global optimal solution of problem (2.7).

(b) Let $\bar{x} \in X$ be a solution of problem (2.7). Then for any $x \in S$,

$$\theta(\bar{x}) + \alpha r(\bar{x})^{N^*} \leq \theta(x). \tag{2.10}$$

Let $t = \inf_{x \in S} \theta(x)$. Then for any $\varepsilon > 0$, we can find an $x_\varepsilon \in S$ such that $\theta(x_\varepsilon) \leq t + \varepsilon$. By (2.8), (2.10), and the Hölder continuity of θ, we have

$$\begin{aligned} t + \varepsilon &\geq \theta(x_\varepsilon) \\ &\geq \theta(\bar{x}) + \alpha r(\bar{x})^{N^*} \\ &= \theta(\bar{z}) + (\theta(\bar{x}) - \theta(\bar{z})) + \alpha r(\bar{x})^{N^*} \\ &\geq t + (\alpha\beta^{p/\sigma} - L)||\bar{x} - \bar{z}||^p, \end{aligned}$$

where $\bar{z} \in \Pi_S(\bar{x})$. Therefore,

$$||\bar{x} - \bar{z}||^p \leq (\alpha\beta^{p/\sigma} - L)^{-1}\varepsilon$$

for any $\varepsilon > 0$ and so $\bar{x} = \bar{z} \in S$. Therefore, (2.10) becomes

$$\theta(\bar{x}) \leq \theta(x), \quad \forall x \in S,$$

i.e., $\bar{x}$ solves problem (2.6). This completes the proof. □

The condition (1.2) is difficult to verify in general. The proof of the following theorem indicates that (1.2) holds for $G = \varphi_+$ when X is convex and φ is strongly convex with order σ on X.

Theorem 2.6 *Assume that $X \subseteq R^n$ is a closed convex set, θ is Hölder continuous on X, and φ is strongly convex with order σ and the corresponding constant c on X. Then the conclusion of Theorem 2.5 is also valid with $\beta = \frac{c}{2}$.*

Proof. It is enough to prove that (2.8) holds for any $x \in X$. In fact, assume that $\varphi(x) > 0$ and $\varphi(z) = 0$, where $z \in \Pi_S(x)$ with $S = \{x \in X : \varphi(x) \leq 0\}$. Since φ is strongly convex with order σ and constant c on X, it follows from (1.3) that

$$\varphi\Big(\frac{x+z}{2}\Big) \leq \frac{1}{2}\varphi(x) - \frac{c}{4}||x - z||^\sigma.$$

Note that $\varphi(\frac{x+z}{2}) > 0$. (Otherwise, since $\frac{x+z}{2} \in X$, this will contradict $z \in \Pi_S(x)$.) In consequence,

$$\frac{c}{2}||x - z||^\sigma \leq \varphi(x) = (\varphi(x))_+ \leq r(x),$$

i.e., (2.8) holds with $\beta = \frac{c}{2}$. This completes the proof. □

Note that it is easy to verify that if each g_i is strongly convex with order σ, then the function φ is also strongly convex with order σ.

In addition, the set X need not be compact in the last two theorems. If X is compact, as in Theorems 2.3 and 2.4, the exponent of the penalty term can be chosen elastically. This result is stated in the following theorem, whose proof is omitted.

Theorem 2.7 *Assume that $X \subseteq R^n$ is compact and the other conditions are the same as in Theorems 2.5 or 2.6 (if φ is strongly convex, X is assumed to be convex). Let*

$$\gamma \geq \frac{\sigma}{p} \quad \textit{and} \quad \alpha > L\beta^{-p/\sigma}.$$

Then problem (2.6) has the same solution set as the problem

$$\begin{array}{ll} \min & \theta(x) + \alpha r(x)^{1/\gamma} \\ \text{s.t.} & x \in X. \end{array}$$

3 Penalty Results for MPECs

Consider the following mathematical program with equilibrium constraints (MPEC):

$$\begin{array}{ll} \min & f(x,y) \\ \text{s.t.} & (x,y) \in Z \\ & y \text{ solves VI}(F(x,\cdot), C(x)), \end{array} \tag{3.1}$$

where $f : R^{n+m} \to R, F : R^{n+m} \to R^m, Z \subseteq R^{n+m}$, and $C : R^n \to 2^{R^m}$ is defined by a continuously differentiable function $g : R^{n+m} \to R^l$ as

$$C(x) = \{y \in R^m : g(x,y) \leq 0\},$$

and VI$(F(x,\cdot), C(x))$ denotes the variational inequality problem defined by the pair $(F(x,\cdot), C(x))$, i.e., y solves VI$(F(x,\cdot), C(x))$ if and only if $y \in C(x)$ and

$$(v-y)^T F(x,y) \geq 0$$

holds for any $v \in C(x)$.

Let $\mathcal{F}$ denote the feasible region of problem (3.1) which is assumed to be nonempty. If F is continuous, $g_i(x,\cdot)$ is convex for all $x \in X = \{x \in R^n : (x,y) \in Z$ for some $y \in R^m\}$, and $\nabla_y g_i(x,y)$ exists and is continuous at every (x,y) in an open set containing $\mathcal{F}$ for each $i = 1, \cdots, l$, Z is compact, and the constraint qualification called SBCQ [10] holds on $\mathcal{F}$, then problem (3.1) is equivalent to the following mathematical program for some $\delta > 0$ [10, Theorem 1.3.5]:

$$\begin{array}{ll} \min & f(x,y) \\ \text{s.t.} & (x,y,\lambda) \in Z \times (B(0,\delta) \cap R^l_+) \\ & F(x,y) + \sum_{i=1}^{l} \lambda_i \nabla_y g_i(x,y) = 0 \\ & g(x,y) \leq 0, \lambda^T g(x,y) = 0, \end{array} \tag{3.2}$$

where $B(0,\delta)$ is the sphere in R^l with center at the origin and the radius δ.

Roughly speaking, the SBCQ means that for any $(x,y) \in \mathcal{F}$, problem (3.2) is feasible and for a bounded subset of $\mathcal{F}$, the corresponding set of Lagrange multipliers is also bounded.

Comparing (3.2) with (2.1) and applying Theorems 2.3 and 2.4, we obtain the following result directly:

Theorem 3.1 *Let $F, g_i, \nabla_y g_i$ be continuous subanalytic and Z be compact subanalytic. Let f be Hölder continuous with order p on Z or $(f - f \circ \Pi_W)$ be continuous subanalytic, where $W = \{(x,y) \in R^{n+m} : (x,y,\lambda)$ satisfies the constraints of (3.2) for some $\lambda\}$. Furthermore, assume that each $g_i(x,\cdot)$ is convex for all $x \in X$ and the SBCQ holds on $\mathcal{F}$. Then there exist constants $\delta > 0, \alpha^* > 0$, and $\gamma^* > 0$ such that for any $\alpha \geq \alpha^*$ and $\gamma \geq \gamma^*$, problem (3.1) is equivalent to the problem*

$$\begin{array}{ll} \min & f(x,y) + \alpha r(x,y,\lambda)^{1/\gamma} \\ \text{s.t.} & (x,y,\lambda) \in Z \times (B(0,\delta) \cap R^l_+), \end{array} \tag{3.3}$$

where

$$r(x,y,\lambda) = ||F(x,y) + \sum_{i=1}^{l} \lambda_i \nabla_y g_i(x,y)||_1 + \sum_{i=1}^{l} ((g_i(x,y))_+ + \lambda_i |g_i(x,y)|),$$

in the sense that (x^*, y^*) *solves (3.1) if and only if* (x^*, y^*, λ^*) *solves (3.3) for some* $\lambda^* \in R^l_+$.

Now we consider a special class of MPECs:

$$\begin{array}{ll} \min & f(x,y) \\ \text{s.t.} & (x,y) \in Z \\ & y \geq 0,\ F(x,y) \geq 0 \\ & y^T F(x,y) = 0, \end{array} \tag{3.4}$$

i.e., mathematical programs with nonlinear complementarity constraints. Let

$$\psi(x,y) = \min\{\min_{1 \leq i \leq m} F_i(x,y), -y^T F(x,y)\}.$$

In a way similar to Theorem 2.5, we can show the following theorem.

Theorem 3.2 *Assume that* Z *is a convex closed subset of* R^{n+m} *and* f *is Hölder continuous with order* p *and Hölder constant* L *on* $Z_1 = Z \cap (R^n \times R^m_+)$. *Assume that for each* $i = 1, \cdots, m$, F_i *is continuous on* Z_1 *and (1.2) holds for* $G = (-\psi)_+$ *with the corresponding constants* σ *and* β *on the set*

$$S = \{(x,y) \in Z_1 : G(x,y) = 0\} = \{(x,y) \in Z_1 : \psi(x,y) \geq 0\}.$$

Suppose problem (3.4) is feasible. Then problem (3.4) has the same solution set as the problem

$$\begin{array}{ll} \min & f(x,y) + \alpha r(x,y)^{N^*} \\ \text{s.t.} & (x,y) \in Z_1, \end{array} \tag{3.5}$$

where

$$r(x,y) = \sum_{i=1}^{m} (-F_i(x,y))_+ + |y^T F(x,y)|$$

and

$$N^* = \frac{p}{\sigma}, \quad \alpha > L\beta^{-N^*}.$$

Proof. Since

$$\begin{aligned} S &= \{(x,y) \in Z_1 : \psi(x,y) \geq 0\} \\ &= \{(x,y) \in Z_1 : F(x,y) \geq 0,\ y^T F(x,y) = 0\}, \end{aligned}$$

then for any $(x,y) \in Z_1$,

$$r(x,y) \geq \beta ||(x,y) - (\bar{x}, \bar{y})||^{\sigma}, \tag{3.6}$$

where $(\bar{x}, \bar{y}) \in \Pi_S(x, y)$. In fact, if $(x, y) \in S$, (3.6) holds clearly. If $(x, y) \notin S$, then $(\bar{x}, \bar{y})$ belongs to the boundary of S and so

$$G(\bar{x}, \bar{y}) = (-\psi(\bar{x}, \bar{y}))_+ = 0.$$

From (1.2), we have

$$r(x, y) \geq (-\psi(x, y))_+ = |G(x, y)| \geq \beta ||(x, y) - (\bar{x}, \bar{y})||^\sigma.$$

Therefore, (3.6) holds for any $(x, y) \in Z_1$.

The rest of the proof is similar to Theorem 2.5. □

We also have the following result:

Theorem 3.3 *Assume that the conditions are the same as in Theorem 3.2 except that $(-\psi)$ is strongly convex with order σ and the corresponding constant c on Z. Then the conclusion of Theorem 3.2 is valid with $\beta = \frac{c}{2}$.*

4 Some Properties Related to Strong Convexity

For the strong convexity employed in Theorems 2.6–2.7 and 3.3, we have the following results:

Theorem 4.1 *If each $f_i, i = 1, \cdots, m$, is strongly convex with order σ on a convex set X, then $\sum_{i=1}^m t_i f_i$ and $\max_{1 \leq i \leq m} f_i$ are also strongly convex with order σ on X, where $t_i > 0$, $i = 1, \cdots, m$.*

Proof. Immediate from Definition 1.4. □

Theorem 4.2 *Suppose that $X \subseteq R^n$ is convex and $f : R^n \to R$ is continuously differentiable on an open set containing X. Then f is strongly convex with order σ on X if and only if there exists a constant $c > 0$ such that*

$$f(y) \geq f(x) + (y - x)^T \nabla f(x) + c\|x - y\|^\sigma, \quad \forall x, y \in X. \tag{4.1}$$

Proof. Assume that f is strongly convex with order σ on X and c is a constant that appears in (1.3). Then for any $x, y \in X$ and $t \in (0, 1)$, we have

$$\begin{aligned} f(y) - f(x) &\geq \frac{1}{t}(f(ty + (1 - t)x) - f(x)) + c(1 - t)\|x - y\|^\sigma \\ &= (y - x)^T \nabla f(x + \theta(y - x)) + c(1 - t)\|x - y\|^\sigma, \end{aligned}$$

for some $\theta \in (0, t)$. Letting $t \to 0$, we have (4.1) from the continuity of ∇f.

Conversely, suppose (4.1) holds for some $c > 0$. For any $x, y \in X$ and $t \in (0, 1)$, we have

$$f(x) - f(tx + (1 - t)y) \geq (1 - t)(x - y)^T \nabla f(tx + (1 - t)y) + c(1 - t)^\sigma \|x - y\|^\sigma$$

and

$$f(y) - f(tx + (1-t)y) \geq t(y-x)^T \nabla f(tx + (1-t)y) + ct^\sigma \|x-y\|^\sigma.$$

In consequence, we have

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(x) - ct(1-t)((1-t)^{\sigma-1} + t^{\sigma-1})||x-y||^\sigma. \quad (4.2)$$

If $0 < \sigma \leq 2$, then

$$(1-t)^{\sigma-1} + t^{\sigma-1} \geq (1-t) + t = 1.$$

If $\sigma > 2$, since the real function $\phi(t) = t^{\sigma-1}$ is convex on $(0,1)$, then

$$(1-t)^{\sigma-1} + t^{\sigma-1} \geq \left(\frac{1}{2}\right)^{\sigma-2}.$$

It follows from (4.2) that there exists some constant $c' > 0$ independent of x, y and t such that

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - c't(1-t)\|x-y\|^\sigma,$$

i.e., f is strongly convex with order σ on X. □

For a given concept of convexity, there usually exists some kind of monotonicity relevant to it, see [8] and the references therein. Now we define the strong monotonicity with order σ and discuss its relation to the strong convexity with order σ.

Definition 4.1 A mapping $G : R^n \to R^n$ is said to be strongly monotone with order σ on a convex set X if there exists a constant $\beta > 0$ such that

$$(y-x)^T(G(y) - G(x)) \geq \beta\|y-x\|^\sigma, \quad \forall x, y \in X. \quad (4.3)$$

Theorem 4.3 *Let $X \subseteq R^n$ be convex and $f : R^n \to R$ be continuously differentiable on an open set containing X. Then f is strongly convex with order σ on X if and only if ∇f is strongly monotone with order σ on X.*

Proof. Suppose that f is strongly convex with order σ on X. By Theorem 4.2, there exists a constant $c > 0$ such that (4.1) holds. Then for any $x, y \in X$, one has

$$f(y) - f(x) \geq (y-x)^T \nabla f(x) + c\|x-y\|^\sigma$$

and

$$f(x) - f(y) \geq (x-y)^T \nabla f(y) + c\|x-y\|^\sigma.$$

Therefore,

$$(y-x)^T(\nabla f(y) - \nabla f(x)) \geq 2c\|x-y\|^\sigma,$$

i.e., ∇f is strongly monotone with order σ on X with $\beta = 2c$.

Conversely, assume that (4.3) holds for some $\beta > 0$ and $F = \nabla f$. Set

$$t_i = \frac{i}{m+1}, \quad i = 0, 1, \cdots, m+1,$$

where m is a positive integer. By the mean-value theorem, there exist $\theta_i \in (t_i, t_{i+1}), 0 \leq i \leq m$, such that

$$f(x + t_{i+1}(y-x)) - f(x + t_i(y-x)) = (t_{i+1} - t_i)(y-x)^T \nabla f(x + \theta_i(y-x)).$$

Hence, it follows from (4.3) that

$$\begin{aligned} f(y) - f(x) &= \sum_{i=0}^{m} (f(x + t_{i+1}(y-x)) - f(x + t_i(y-x))) \\ &= \sum_{i=0}^{m} (t_{i+1} - t_i)(y-x)^T (\nabla f(x + \theta_i(y-x)) - \nabla f(x)) + (y-x)^T \nabla f(x) \\ &\geq \beta \|y-x\|^\sigma \sum_{i=0}^{m} \theta_i^{\sigma-1}(t_{i+1} - t_i) + (y-x)^T \nabla f(x). \end{aligned}$$

Letting $m \to +\infty$ and noticing that

$$\lim_{m \to +\infty} \sum_{i=0}^{m} \theta_i^{\sigma-1}(t_{i+1} - t_i) = \int_0^1 t^{\sigma-1} dt = \frac{1}{\sigma},$$

we have

$$f(y) - f(x) \geq \frac{\beta}{\sigma} \|y-x\|^\sigma + (y-x)^T \nabla f(x).$$

By Theorem 4.2, f is strongly convex with order σ on X. □

5 Conclusions

The theory of penalty functions for constrained nonlinear programs has a long history. Burke [2] provides a comprehensive review of the theory of exact penalization and gives a unified approach for deriving many known results in the literature. In this paper, we have shown some exact penalty results for nonlinear programs and MPECs. In particular, we proved Theorems 2.3–2.4 and Theorem 3.1, which are generalizations of Theorem 2.1.2 and Theorem 2.2.1 in [10], respectively, and obtained Theorems 2.5, 2.7, and 3.2 using the notion of error bound of order σ (Definition 1.3). Furthermore, replacing these conditions by a kind of convexity assumption, we proved Theorems 2.6 and 3.3. Although Theorems 2.5–2.7 are concerned with problem (2.6), these conclusions remain valid for problem (2.1) by defining the function φ appropriately.

References

[1] E. Bierstone and P.D. Milman, *Semianalytic and subanalytic sets*, Institut des Hautes Etudes Scientifiques, Publications Mathématiques, 67 (1988), 5–42.

[2] J.V. Burke, *An exact penalization viewpoint of constrained optimization*, SIAM Journal on Control and Optimization, 29 (1991), 968–998.

[3] J.V. Burke and M.C. Ferris, *Weak sharp minima in mathematical programming,* SIAM Journal on Control and Optimization, 31 (1993), 1340–1359.

[4] F.H. Clarke, Optimization and Nonsmooth Analysis, John Wiley & Sons , New York, 1983.

[5] J.P. Dedieu, *Penalty functions in subanalytic optimization,* Optimization, 26 (1992), 27–32.

[6] M. Fukushima, Z.Q. Luo, and J.S. Pang, *A globally convergent sequential quadratic programming algorithm for mathematical programs with linear complementarity constraints*, Computational Optimization and Applications, 10 (1998), 5–34.

[7] M. Fukushima and J.S. Pang, *Some feasibility issues in mathematical programs with equilibrium constraints*, SIAM Journal on Optimization, 8 (1998), 673–681.

[8] S. Karamardian and S. Schaible, *Seven kinds of monotone maps*, Journal of Optimization Theory and Applications, 66 (1990), 37–46.

[9] M.S. Lojasiewicz, *Ensembles semi-analytiques,* Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette, 1964.

[10] Z.Q. Luo, J.S. Pang, and D. Ralph, Mathematical Programs with Equilibrium Constraints, Cambridge University Press, Cambridge, United Kingdom, 1996.

[11] Z.Q. Luo, J.S. Pang, D. Ralph, and S.Q. Wu, *Exact penalization and stationary conditions of mathematical programs with equilibrium constraints* , Mathematical Programming, 75 (1996), 19–76.

[12] J.M. Ortega and W.C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York, 1970.

[13] J.S. Pang, *Error bounds in mathematical programming,* Mathematical Programming, 79 (1997), 299–332.