JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 118, No. 1, pp. 67–80, July 2003 (© 2003)

Some Exact Penalty Results for Nonlinear Programs and Mathematical Programs with Equilibrium Constraints[1]

G. H. Lin[2] and M. Fukushima[3]

Communicated by Z. Q. Luo

Abstract. Recently, some exact penalty results for nonlinear programs and mathematical programs with equilibrium constraints were proved by Luo, Pang, and Ralph (Ref. 1). In this paper, we show that those results remain valid under some other mild conditions. One of these conditions, called strong convexity with order σ, is discussed in detail.

Key Words. Mathematical programs with equilibrium constraints, nonlinear complementarity problems, subanalytic sets, subanalytic functions, Hölder continuity, strong convexity, error bounds.

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 Ref. 1.

Because of the presence of variational inequality or complementarity constraints, MPEC has such an intrinsic feature that its feasible region is nonconvex or nonsmooth in general; hence, it is very difficult to handle. At

[1]This work was supported in part by a Scientific Research Grant-in-Aid from the Ministry of Education, Science, Sports, and Culture of Japan. The authors are grateful to an anonymous referee for critical comments.

[2]PhD Student, Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto, Japan and Department of Applied Mathematics, Dalian University of Technology, Dalian, China.

[3]Professor, Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto, Japan.

present, a popular approach is to reformulate an MPEC as a standard nonlinear program. In this respect, penalty functions have provided a powerful approach, both as theoretical tool and as 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 (Ref. 1). In this paper, we show that those results remain valid under some other mild conditions. Instead of the subanalytic property and error bounds, which are somewhat abstract and difficult to verify in practice, some of our results use a property called strong convexity with order σ, which is a generalization of the ordinary strong convexity (Ref. 2).

The following notations and definitions will be used throughout this paper. For $x \in R^n$, $\|\cdot\|$ and $\|\cdot\|_1$ denote the norms defined by

$$\|x\| = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}, \qquad \|x\|_1 = \sum_{i=1}^{n} |x_i|.$$

For a nonempty closed set $W \subseteq R^n$, we denote

$$\mathrm{dist}(x, W) = \min_{z \in W} \|x - z\|,$$

$$\Pi_W(x) = \{z \in W: \|x - z\| = \mathrm{dist}(x, W)\}.$$

In addition, $B(0, c)$ stands for the closed ball $\{x \in R^n: \|x\| \leq c\}$ and R^n_+ denotes the nonnegative orthant in R^n. For a real scalar u, we denote

$$(u)_+ = \max\{0, u\}.$$

Definition 1.1. See Ref. 1. A set $X \subseteq R^n$ 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:

(a) 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

$$Z_{ij} = \{z \in V: f_{ij}(z) = 0\} \quad \text{or} \quad \{z \in V: f_{ij}(z) < 0\},$$

defined for some real analytic functions f_{ij} on V, such that

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

(b) $X \cap U = \{x \in R^n: (x, y) \in Z, \text{ for some } y \in R^p\}$.

A function f: $R^n \to R$ is said to be subanalytic if its graph is subanalytic.

The class of subanalytic functions is broader than the class of analytic functions and is employed by many papers, although it is somewhat abstract. For more details, we refer the reader to Refs. 1 and 3–5.

Definition 1.2. See Ref. 6. Let $0 < p \leq 1$ be a constant, and let $G: R^n \to R^m$ be a mapping. We say that G is Hölder continuous with order p on $X \subseteq R^n$ if there exists a constant L such that

$$\|G(x) - G(y)\| \leq L\|x - y\|^p, \qquad \forall x, y \in X. \tag{1}$$

This concept is a generalization of the Lipschitz continuity, which is, by definition, Hölder continuity with order $p = 1$. Note that Hölder continuity makes sense only when $0 < p \leq 1$. In fact, when $p > 1$, condition (1) implies that all directional derivatives of G at any interior point are zero and so G is quite trivial. In addition, for $0 < p \neq p' \leq 1$, Hölder continuous functions with order p and those with order p' constitute different classes of functions. For example, the function

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

is Hölder continuous with order $p = 1/2$ on R^n and is not Lipschitz continuous on R^n.

Definition 1.3. 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

$$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y) - ct(1 - t)\|x - y\|^\sigma, \tag{2}$$

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

When $\sigma = 2$, this property reduces to the strong convexity in the ordinary sense (Ref. 2). 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 is not strongly convex with order 2 on R.

2. Penalty Results for Nonlinear Programs

Consider the following nonlinear program:

$$\min \quad \theta(x), \tag{3a}$$

$$\text{s.t.} \quad x \in X, \tag{3b}$$

$$g(x) \leq 0, \qquad h(x) = 0, \tag{3c}$$

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 (3) and let

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

be the residual for the constraints in (3) at $x \in X$. Then, the function r may be used as a penalty function for problem (3). The following theorem is shown in Ref. 1.

Theorem 2.1. Let $X \subseteq R^n$ be a compact subanalytic set, let θ be Lipschitz continuous on X, and let g_i, h_j be continuous subanalytic. Suppose that problem (3) is feasible. Then, there exist positive constants α^* and γ^* such that, for $\alpha \geq \alpha^*$ and $\gamma \geq \gamma^*$, problem (3) is equivalent to

$$\min \quad \theta(x) + \alpha r(x)^{1/\gamma}, \tag{4a}$$

$$\text{s.t.} \quad x \in X, \tag{4b}$$

in the sense that x^* solves (3) if and only if it solves (4).

In particular, the following result is useful in both Ref. 1 and our paper.

Theorem 2.2. Lojasiewicz Inequality (Ref. 3). Let ϕ, $\psi: S \to R$ be continuous subanalytic, and let $S \subseteq R^n$ be compact subanalytic. If $\phi^{-1}(0) \subseteq \psi^{-1}(0)$, then there exist constants $\rho > 0$ and $N^* > 0$ such that

$$\rho |\psi(x)|^{N^*} \leq |\phi(x)|, \qquad \forall x \in S. \tag{5}$$

Now, we give our penalty results for problem (3). First of all, we define a new function. Suppose that problem (3) is feasible, i.e., $W \neq \emptyset$. Then, we can take a vector $d \in W$ and define a function θ_d on X by

$$\theta_d(x) = \theta(x_d), \qquad \forall x \in X,$$

where

$$x_d = (1 - t_d)x + t_d d,$$

with t_d the smallest number $t \in [0, 1]$ such that

$$(1 - t)x + td \in W.$$

Theorem 2.3. Suppose that X, g, h are the same as in Theorem 2.1, that problem (3) is feasible, and that the function $\theta - \theta_d$ is continuous subanalytic for some $d \in W$. Then, the conclusion of Theorem 2.1 remains valid.

Proof. Let $r|_X$ denote the restriction of r on X. Noticing that both r and $\theta - \theta_d$ are continuous subanalytic and that

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

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

$$\rho|\theta(x) - \theta(x_d)|^{N^*} \leq r(x), \qquad \forall x \in X. \tag{6}$$

Let

$$\mu = \max\left\{1, \max_{x \in X} r(x)\right\},$$

$$\alpha^* > (\mu/\rho)^{1/N^*}, \qquad \gamma^* = N^*,$$

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

(a) Assume that $\bar{x}$ solves problem (3). Then, for any $x \in X$, we have

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

Therefore, $\bar{x}$ is a global optimal solution of problem (4).

(b) If $\bar{x}$ solves (4), we can claim that $\bar{x}$ is an optimal solution of problem (3). In fact, since W is compact and since problem (3) is feasible, it has an optimal solution, denoted by $\tilde{x}$. In a way similar to (a), we have

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

This implies $\theta(\bar{x}) = \theta(\bar{x}_d)$ and then

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

where the first inequality holds because $\tilde{x}$ solves (3) and $\bar{x}_d$ is feasible to (3). Hence, we have

$$r(\bar{x}) = 0 \quad \text{and} \quad \theta(\bar{x}) = \theta(\tilde{x}).$$

The former implies $\bar{x} \in W$, and so $\bar{x}$ is an optimal solution to problem (3). This completes the proof. □

The new condition given in Theorem 2.3 may be satisfied by choosing d appropriately even if W is not convex and θ is not Lipschitz continuous, as the next example shows.

Example 2.1. Consider the following problem:

$$\begin{aligned} \min \quad & \theta(x) = \sin^2(3\, x^{1/3}), \\ \text{s.t.} \quad & x \in [0, \pi^3], \\ & \cos(3x^{1/3}) \leq 0. \end{aligned}$$

Then, the feasible region is given by

$$W = [\pi^3/216, \pi^3/8] \cup [125\pi^3/216, \pi^3],$$

which is nonconvex. We note that θ is not Lipschitz continuous on $[0, \pi^3]$, which means that the conditions of Theorem 2.1 are not satisfied for this problem. However, we can show that the assumptions of Theorem 2.3 hold. In fact, for $d = \pi^3/10$, the function

$$\theta_d(x) = \begin{cases} 1, & x \in [0, \pi^3/216), \\ \sin^2(3x^{1/3}), & x \in [\pi^3/216, \pi^3/8], \\ 1, & x \in (\pi^3/8, 125\pi^3/216), \\ \sin^2(3x^{1/3}) & x \in [125\pi^3/216, \pi^3], \end{cases}$$

is continuous and piecewise smooth and so it is continuous subanalytic on $[0, \pi^3]$.

We consider next another kind of error bounds for problem (3), which is different from (6). We say that a function $u{:}X \to [0, \infty)$ provides an error bound of order $\nu > 0$ on W if there exists a positive constant β such that

$$u(x) \geq \beta(\operatorname{dist}(x, W))^{\nu}, \qquad \forall x \in X.$$

For more details of error bounds, we refer the reader to Refs. 6–7 and the references therein.

Theorem 2.4. Let X be a closed subset of R^n, let g and h be continuous on X, and let θ be Hölder continuous with order $p > 0$ and Hölder constant L on X. Assume that $r(x)$ provides an error bound of order $\nu > 0$ on W with corresponding constant β and suppose that problem (3) is feasible. Then, problem (3) has the same solution set as the problem

$$\min \quad \theta(x) + \alpha r(x)^{N^*}, \tag{7a}$$

$$\text{s.t.} \quad x \in X, \tag{7b}$$

where

$$N^* = p/\nu, \qquad \alpha > L\beta^{-N^*}.$$

Proof. By the assumption of the theorem, we have

$$r(x) \geq \beta[\mathrm{dist}(x, W)]^{\nu}, \qquad \forall x \in X. \tag{8}$$

(a) If $\bar{x}$ solves problem (3), then for any $x \in X$, we have from (8) and the Hölder continuity of θ that

$$\begin{aligned}\theta(x) + \alpha r(x)^{N^*} &= \theta(z) + (\theta(x) - \theta(z)) + \alpha r(x)^{p/\nu} \\ &\geq \theta(\bar{x}) + (\alpha\beta^{p/\nu} - L)[\mathrm{dist}(x, W)]^p \\ &\geq \theta(\bar{x}) \\ &= \theta(\bar{x}) + \alpha r(\bar{x})^{N^*},\end{aligned}$$

where $z \in \Pi_W(x)$. Therefore, $\bar{x}$ is a global optimal solution of problem (7).

(b) Let $\bar{x} \in X$ be a solution of problem (7). Then, for any $x \in W$,

$$\theta(\bar{x}) + \alpha r(\bar{x})^{N^*} \leq \theta(x) + \alpha r(x)^{N^*} = \theta(x). \tag{9}$$

Let

$$t = \inf_{x \in W} \theta(x).$$

Then, for any $\epsilon > 0$, we can find an $x_\epsilon \in W$ such that

$$\theta(x_\epsilon) \leq t + \epsilon.$$

By (8), (9), and the Hölder continuity of θ, we have

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

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

$$\|\bar{x} - \bar{z}\|^p \leq (\alpha\beta^{p/\nu} - L)^{-1}\epsilon,$$

for any $\epsilon > 0$ and so $\bar{x} = \bar{z} \in W$. Therefore, (9) becomes

$$\theta(\bar{x}) \leq \theta(x), \qquad \forall x \in W;$$

i.e., $\bar{x}$ solves problem (3). This completes the proof. □

The set X need not be compact and the functions g and h need not be subanalytic in the last theorem, in contrast with Theorems 2.1 and 2.3. If X is compact and g, h are subanalytic, as in Theorems 2.1 and 2.3, the exponent of the penalty term can be chosen elastically. This result is stated in the following theorem, whose proof is omitted.

Theorem 2.5. Assume that X, g, h are the same as in Theorem 2.1, that θ is Hölder continuous on X, and that problem (3) is feasible. Then, the conclusion of Theorem 2.1 remains true.

Now, we consider a special case of problem (3):

$$\min \quad \theta(x), \tag{10a}$$

$$\text{s.t.} \quad x \in X, \tag{10b}$$

$$g(x) \leq 0. \tag{10c}$$

We will show some new penalty results for problem (10) which will be applied in Section 3 to a mathematical program with a nonlinear complementarity system. In the rest of this section, we let φ denote the function defined by

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

In general, condition (8) is difficult to verify in practice. The proof of the following theorem indicates that it holds 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, that θ is Hölder continuous with order $p > 0$ and Hölder constant L on X, and that φ is strongly convex with order $\sigma > 0$ and corresponding constant c on X. Suppose that problem (10) is feasible. Then, problem (10) has the same solution set as problem (7) with

$$r(x) = \sum_{i=1}^{m} (g_i(x))_+, \qquad N^* = p/\sigma, \qquad \alpha > L(c/2)^{-N^*}.$$

Proof. By Theorem 2.4 and its proof, it is enough to prove that (8) holds with

$$\beta = c/2, \qquad \nu = \sigma,$$

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 (2) that

$$\varphi((x+z)/2) \leq (1/2)\varphi(x) - (c/4)\|x-z\|^\sigma.$$

Note that

$$\varphi((x+z)/2) > 0.$$

Otherwise, since $(x+z)/2 \in X$, this will contradict $z \in \Pi_S(x)$. In consequence,

$$(c/2)\|x-z\|^\sigma \leq \varphi(x) = (\varphi(x))_+ \leq r(x);$$

i.e., (8) holds with $\beta = c/2$ and $\nu = \sigma$. This completes the proof. □

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.

We have also the following result.

Theorem 2.7. Assume that $X \subseteq R^n$ is compact and convex and that the other conditions are the same as in Theorem 2.6. Let

$$\gamma \geq \sigma/p, \qquad \alpha > L(c/2)^{-p/\sigma}.$$

Then, problem (10) has the same solution set as the problem

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

3. Penalty Results for MPECs

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

$$\min \quad f(x, y), \tag{11a}$$

$$\text{s.t.} \quad (x, y) \in Z, \tag{11b}$$

$$y \text{ solves VI } (F(x, \cdot), C(x)), \tag{11c}$$

where f: $R^{n+m} \to R$, F: $R^{n+m} \to R^m$, $Z \subseteq R^{n+m}$, 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 $\mathrm{VI}(F(x,\cdot), C(x))$ denotes the variational inequality problem defined by the pair $(F(x,\cdot), C(x))$; i.e., y solves $\mathrm{VI}(F(x,\cdot), C(x))$ if and only if $y \in C(x)$ and

$$(v-y)^T F(x,y) \geq 0, \qquad \forall v \in C(x).$$

Let $\mathcal{F}$ denote the feasible region of problem (11), which is assumed to be nonempty. If F is continuous, $g_i(x,\cdot)$ is convex for all $x \in X$, where

$$X = \{x \in R^n: (x,y) \in Z, \text{ for some } y \in R^m\},$$

$\nabla_y g_i(x,y)$ exists and is continuous at every (x, y) in an open set containing $\mathcal{F}$ for each $i = 1, \ldots, l$, Z is compact, and the constraint qualification SBCQ (Ref. 1) holds on $\mathcal{F}$, then problem (11) is equivalent to the following mathematical program for some $\delta > 0$ (Ref. 1, Theorem 1.3.5):

$$\min \quad f(x,y), \tag{12a}$$

$$\text{s.t.} \quad (x,y,\lambda) \in Z \times (B(0,\delta) \cap R^l_+), \tag{12b}$$

$$F(x,y) + \sum_{i=1}^{l} \lambda_i \nabla_y g_i(x,y) = 0, \tag{12c}$$

$$g(x,y) \leq 0, \qquad \lambda^T g(x,y) = 0. \tag{12d}$$

Roughly speaking, SBCQ means that, for any $(x,y) \in \mathcal{F}$, problem (12) is feasible and that, for a bounded subset of $\mathcal{F}$, the corresponding set of Lagrange multipliers is also bounded. Let

$$W = \{(x,y) \in R^{n+m}: (x,y,\lambda) \text{ satisfies the constraints of (12) for some } \lambda\}.$$

This set is nonempty if, under SBCQ, $\mathcal{F}$ is nonempty. We choose some $d \in W$ and define the function f_d in a way similar to the definition of θ_d in Section 2. Then, comparing (12) with (3) and applying Theorems 2.3 and 2.5, we obtain the following result directly.

Theorem 3.1. Let $F, g_i, \nabla_y g_i$ be continuous subanalytic, and let Z be compact subanalytic. Let f be Hölder continuous with order p on Z or let $f - f_d$ be continuous subanalytic for some $d \in W$. Furthermore, assume that each $g_i(x,\cdot)$ is convex for all $x \in X$ and that 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 (11) is equivalent to the problem

$$\min \quad f(x,y) + \alpha r(x,y,\lambda)^{1/\gamma}, \tag{13a}$$

$$\text{s.t.} \quad (x,y,\lambda) \in Z \times (B(0,\delta) \cap R^l_+), \tag{13b}$$

where

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

in the sense that (x^*, y^*) solves (11) if and only if (x^*, y^*, λ^*) solves (13) for some $\lambda^* \in R^l_+$.

Now, we consider a special class of MPECs:

$$\min \quad f(x, y), \tag{14a}$$

$$\text{s.t.} \quad (x, y) \in Z, \tag{14b}$$

$$y \geq 0, \qquad F(x, y) \geq 0, \tag{14c}$$

$$y^T F(x, y) = 0, \tag{14d}$$

i.e., mathematical programs with complementarity constraints. Let S denote the feasible region of problem (14); let

$$Z_1 = Z \cap (R^n \times R^m_+), \tag{15a}$$

$$r(x, y) = \sum_{i=1}^{m} (-F_i(x, y))_+ + |y^T F(x, y)|, \tag{15b}$$

$$\psi(x, y) = \min\left\{ \min_{1 \leq i \leq m} F_i(x, y), -y^T F(x, y) \right\}. \tag{15c}$$

In a way similar to Theorems 2.4 and 2.6, we can show the following results.

Theorem 3.2. Assume that Z is a closed subset of R^{n+m}, that f is Hölder continuous with order p and Hölder constant L on Z_1, and that F is continuous on Z_1. Assume that $r(x, y)$ defined by (15b) provides an error bound of order $\nu > 0$ with corresponding constant β on S and that problem (14) is feasible. Then, problem (14) has the same solution set as the problem

$$\min \quad f(x, y) + \alpha r(x, y)^{N^*}, \tag{16a}$$

$$\text{s.t.} \quad (x, y) \in Z_1, \tag{16b}$$

where

$$N^* = p/\nu, \qquad \alpha > L\beta^{-N^*}.$$

Theorem 3.3. Assume that F, f are the same as in Theorem 3.2 and that Z is closed and convex. Suppose that problem (14) is feasible. If the function $-\psi$ is strongly convex with order σ and corresponding constant c on Z, then problem (14) is equivalent to problem (16) with

$$N^* = p/\sigma, \qquad \alpha > L(c/2)^{-N^*},$$

in the sense that (x^*, y^*) solves (14) if and only if it solves (16).

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, $i=1,\ldots,m$, is strongly convex with order σ on a convex set X, then $\sum_{i=1}^{m} t_i f_i$ and $\max_{1\le i\le m} f_i$ are also strongly convex with order σ on X, where $t_i>0$, $i=1,\ldots,m$.

Proof. It is immediate from Definition 1.3. □

Theorem 4.2. Suppose that $X\subseteq R^n$ is convex and that 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

$$f(y)\ge f(x)+(y-x)^T\nabla f(x)+c\|x-y\|^\sigma,\qquad \forall x,y\in X. \tag{17}$$

Proof. Assume that f is strongly convex with order σ on X and that c is a constant which appears in (2). Then, for any x, $y\in X$ and $t\in(0,1)$, we have

$$\begin{aligned}f(y)-f(x)&\ge(1/t)[f(ty+(1-t)x)-f(x)]+c(1-t)\|x-y\|^\sigma\\&=(y-x)^T\nabla f(x+\xi(y-x))+c(1-t)\|x-y\|^\sigma,\end{aligned}$$

for some $\xi\in(0,t)$. Letting $t\to 0$, we have (17) from the continuity of ∇f.

Conversely, suppose that (17) holds for some $c>0$. For any x, $y\in X$ and $t\in(0,1)$, we have

$$\begin{aligned}&f(x)-f(tx+(1-t)y)\\&\ge(1-t)(x-y)^T\nabla f(tx+(1-t)y)+c(1-t)^\sigma\|x-y\|^\sigma,\\&f(y)-f(tx+(1-t)y)\\&\ge t(y-x)^T\nabla f(tx+(1-t)y)+ct^\sigma\|x-y\|^\sigma.\end{aligned}$$

In consequence, we have

$$\begin{aligned}&f(tx+(1-t)\,y)\\&\le tf(x)+(1-t)f(x)-ct(1-t)((1-t)^{\sigma-1}+t^{\sigma-1})\|x-y\|^\sigma.\end{aligned} \tag{18}$$

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

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

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

$$(1-t)^{\sigma-1} + t^{\sigma-1} \geq (1/2)^{\sigma-2}.$$

It follows from (18) that there exists some constant $c' > 0$ independent of x, y, 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 exists usually some kind of monotonicity relevant to it; see Ref. 2 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: $R^n \rightarrow R^n$ is said to be strongly monotone with order σ on a convex set X if there exists a constant $\beta > 0$ such that

$$(y-x)^T(G(y) - G(x)) \geq \beta\|y-x\|^{\sigma}, \qquad \forall x, y \in X. \tag{19}$$

Theorem 4.3. Let $X \subseteq R^n$ be convex, and let f: $R^n \rightarrow R$ 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 (17) holds. Then, for any $x,y \in X$, one has

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

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

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 (19) holds for some $\beta > 0$ and $F = \nabla f$. Set

$$t_i = i/(m+1), \qquad i = 0, 1, \ldots, m+1,$$

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

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

Hence, it follows from (19) that

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

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

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

we have

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

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

References

1. Luo, Z. Q., Pang, J. S., and Ralph, D., *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge, United Kingdom, 1996.
2. Karamardian, S., and Schaible, S., *Seven Kinds of Monotone Maps*, Journal of Optimization Theory and Applications, Vol. 66, pp. 37–46, 1990.
3. Bierstone, E., and Milman, P. D., *Semianalytic and Subanalytic Sets*, Institut des Hautes Etudes Scientifiques, Publications Mathématiques, Vol. 67, pp. 5–42, 1988.
4. Dedieu, J. P., *Penalty Functions in Subanalytic Optimization*, Optimization, Vol. 26, pp. 27–32, 1992.
5. Lojasiewicz, M. S., *Ensembles Semi-Analytiques*, Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette, France, 1964.
6. Ortega, J. M., and Rheinboldt, W. C., *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, NY, 1970.
7. Pang, J. S., *Error Bounds in Mathematical Programming*, Mathematical Programming, Vol. 79, pp. 299–332, 1997.