JOURNAL OF INDUSTRIAL AND
MANAGEMENT OPTIMIZATION
Volume **15**, Number **3**, July **2019**

doi:10.3934/jimo.2018094

pp. **1241–1261**

A SMOOTHING AUGMENTED LAGRANGIAN METHOD FOR NONCONVEX, NONSMOOTH CONSTRAINED PROGRAMS AND ITS APPLICATIONS TO BILEVEL PROBLEMS

Qingsong Duan

School of Mathematical Sciences, Dalian University of Technology
Dalian 116024, China

Mengwei Xu*

School of Mathematics, Tianjin University
Tianjin, 300072, China

Yue Lu

School of Mathematical Sciences, Tianjin Normal University
Tianjin 300387, China

Liwei Zhang

School of Mathematical Sciences, Dalian University of Technology
Dalian 116024, China

(Communicated by Shunsuke Hayashi)

Abstract. In this paper, we consider a class of nonsmooth and nonconvex optimization problem with an abstract constraint. We propose an augmented Lagrangian method for solving the problem and construct global convergence under a weakly nonsmooth Mangasarian-Fromovitz constraint qualification. We show that any accumulation point of the iteration sequence generated by the algorithm is a feasible point which satisfies the first order necessary optimality condition provided that the penalty parameters are bounded and the upper bound of the augmented Lagrangian functions along the approximated solution sequence exists. Numerical experiments show that the algorithm is efficient for obtaining stationary points of general nonsmooth and nonconvex optimization problems, including the bilevel program which will never satisfy the nonsmooth Mangasarian-Fromovitz constraint qualification.

2010 *Mathematics Subject Classification.* 65K10, 90C26.

Key words and phrases. Nonsmooth constrained problem, smoothing function, augmented Lagrangian method, constraint qualification, bilevel program.

The second author is supported by NSFC grant 11601376. The third author is supported by NSFC grant 11601389, the Doctoral Foundation of Tianjin Normal University grant 52XB1513 and and 2017- Outstanding Young Innovation Team Cultivation Program of Tianjin Normal University grant 135202TD1703. The fourth author is supported by NSFC grant 11571059 and 11731013.

* Corresponding author: Mengwei Xu.

1. **Introduction.** In this paper, we consider a nonsmooth constrained optimization problem:

$$\begin{aligned}(\mathrm{P})\ \ \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \le 0,\ i = 1, \cdots, p, \\ & h_j(x) = 0,\ j = p+1, \cdots, q, \\ & x \in \Omega,\end{aligned}$$

where $\Omega \subset \mathbb{R}^n$ is a closed convex set and $f, g_i, i = 1, \cdots, p, h_j, j = p+1, \cdots, q : \mathbb{R}^n \to \mathbb{R}$ are Lipschitz continuous functions.

There are various methods to deal with nonsmooth programs. In this paper, we take our attention on the smoothing technique. The authors [40] introduced a function sequence converges continuously to the Lipschitz continuous functions.

Definition 1.1. Let $g : \mathbb{R}^n \to R$ be a locally Lipschitz function. Assume that, for a given $\rho > 0$, $g_\rho : \mathbb{R}^n \to R$ is a continuously differentiable function. We say that $\{g_\rho : \rho > 0\}$ is a family of smoothing functions of g if $\lim\limits_{z\to x,\ \rho\uparrow\infty} g_\rho(z) = g(x)$ for any fixed $x \in \mathbb{R}^n$.

From numerical point of view, we need the following property to obtain a stationary point for certain smoothing method.

Definition 1.2. [9, 16]Let $g : \mathbb{R}^n \to R$ be a locally Lipschitz continuous function. We say that a family of smoothing functions $\{g_\rho : \rho > 0\}$ of g satisfies the gradient consistency property if $\limsup\limits_{z\to x,\, \rho\uparrow\infty} \nabla g_\rho(z)$ is nonempty and $\limsup\limits_{z\to x,\, \rho\uparrow\infty} \nabla g_\rho(z) \subseteq \partial g(x)$ holds for any $x \in \mathbb{R}^n$, where $\limsup\limits_{z\to x,\, \rho\uparrow\infty} \nabla g_\rho(z)$ denotes the set of all outer limits of $\nabla g_\rho(z)$ when $z \to x$ and $\rho \to \infty$, namely,

$$\limsup_{z\to x,\, \rho\uparrow\infty} \nabla g_\rho(z) := \Big\{ \lim_{k\to\infty} \nabla g_{\rho_k}(z_k) : z_k \to x, \rho_k \uparrow \infty \Big\}.$$

∂g denotes the Clarke generalized gradient of g and ∇g denotes the gradient of g; see Section 2.

To generate a family of smoothing functions with the gradient consistency property for any locally Lipschitz function, one can use the integral-convolution with bounded supports. Rockafellar and Wets [40, Example 7.19 and Theorem 9.67] showed that for any locally Lipschitz function g,

$$g_\rho(x) := \int_{\mathbb{R}^n} g(x-y)\phi_\rho(y)dy = \int_{\mathbb{R}^n} g(y)\phi_\rho(x-y)dy \tag{1}$$

is a family of smoothing functions of g with the gradient consistency property, where $\phi_\rho : \mathbb{R}^n \to \mathbb{R}_+$ is a sequence of bounded, measurable functions with $\int_{\mathbb{R}^n} \phi_\rho(x)dx = 1$ such that the sets $B_\rho = \{x : \phi_\rho(x) > 0\}$ form a bounded sequence converging to $\{0\}$ as $\rho \uparrow \infty$. Note that for $g_\rho(x)$ defined by (1), the inclusion

$$\partial g(x) \subseteq co \limsup_{z\to x,\, \rho\uparrow\infty} \nabla g_\rho(z)$$

always holds by [40, Theorem 9.61 and Corollary 8.47 (b)]. Thus the definition of gradient consistency in Definition 1.2 is equivalent to saying that

$$\partial g(x) = co \limsup_{z\to x,\, \rho\uparrow\infty} \nabla g_\rho(z),$$

which is showed in [10, 12].

In practice, many Lipschitz functions can be considered as a composition of a smooth function with a plus function $(t)_+ := \max\{0, t\}$. By using the integral convolution with density functions, Chen and Mangasarian [15] constructed smoothing approximations to the plus function. Let $\phi : \mathbb{R} \to \mathbb{R}_+$ be a piecewise continuous density function with $\phi(s) = \phi(-s)$ and $\displaystyle\int_{-\infty}^{+\infty} |s|\phi(s)ds < \infty$. Then $\psi_\mu(t) := \displaystyle\int_{-\infty}^{+\infty} (t - \mu s)_+ \phi(s)ds$, with $\mu \downarrow 0$ is a family of smoothing functions of the plus function with the gradient consistency property. Different density functions derive different smoothing approximations. Choosing $\phi_1(s) := \frac{2}{(s^2+4)^{\frac{3}{2}}}$ results in the so-called the CHKS (Chen-Harker-Kanzow-Smale) smoothing function of $(t)_+$ ([13, 26, 42]):

$$\psi_\mu^1(t) := \frac{1}{2}\left(t + \sqrt{t^2 + 4\mu^2}\right).$$

Choosing $\phi_2(s) := \begin{cases} 0, & \text{if } |s| > \frac{1}{2} \\ 1, & \text{if } |s| \le \frac{1}{2}, \end{cases}$ results in the so-called uniform smoothing function of $(t)_+$:

$$\psi_\mu^2(t) := \begin{cases} (t)_+, & \text{if } |t| > \frac{\mu}{2} \\ \frac{1}{2\mu}(t + \frac{\mu}{2})^2, & \text{if } |t| \le \frac{\mu}{2}. \end{cases}$$

Since $|t| = (t)_+ + (-t)_+$, approximating $(t)_+$ by $\psi_\mu^2(t)$ and $(-t)_+$ by $\psi_\mu^2(-t)$ respectively results in the following smoothing function of $|t|$ which is used frequently:

$$\psi_\mu^3(t) := \begin{cases} |t|, & \text{if } |t| > \frac{\mu}{2} \\ \frac{t^2}{\mu} + \frac{\mu}{4}, & \text{if } |t| \le \frac{\mu}{2}. \end{cases}$$

There also exists many other smoothing functions which are not generated by the integral-convolution [10, 11, 12, 15, 33].

The augmented Lagrangian method [19] is known as a method of multipliers for solving constrained optimization problems, which is also basis for some high quality software such as ALGENCAN [1] and LANCELOT [27]. The augmented Lagrangian method has been studied for decades, the initial researches can be found in [6, 25, 35, 37, 38]. As an exact penalty method, the convergence result of the augmented Lagrangian method requires the boundedness of the penalty parameters. By assuming the LICQ or Mangasarian-Fromovitz constraint qualification (MFCQ) holds at all feasible and infeasible accumulation points of iteration sequence would derive boundedness parameters [20, 44]. Many researchers also studied the boundedness of the parameters under the so called Constant Positive Linear Dependence (CPLD) and its relaxed vision RCPLD [2, 3, 4, 8]. Recently, Curtis et al. [21] proposed an adaptive augmented Lagrangian method for smooth constrained problems which greatly improved the overall performance of the algorithm. However, the adaptive method may converges to an infeasible point.

To overcome the infeasibility problem of the augmented Lagrangian method, Lu and Zhang [29] proposed an augmented Lagrangian method for solving a nonlinear program where the objective function is a sum of a smooth term and a nonsmooth convex term and established global convergence to a feasible stationary point under MFCQ. The method differs from the classical augmented Lagrangian method in that: (i) the values of augmented Lagrangian function along the solutions generated by the algorithm are bounded from above, (ii) the magnitude of penalty parameters outgrows that of Lagrange multipliers. If there is a feasible point known at the beginning of the algorithm, the upper bound would be obtained easily. Chen et al. [14] applied such method to a non-Lipschitz nonconvex programming. While for many problems, such as the bilevel programming problems, it is not easy to get a feasible point and thus the upper bound may difficult to obtain.

Since the usual constraint qualifications are still too strong to hold for many problems, such as the mathematical program with equilibrium constraints (MPEC) and bilevel programs, recently Xu, Ye and Zhang [45] proposed a weaker version of the GMFCQ, which called weakly generalized Mangasarian Fromovitz constraint qualification (WGMFCQ). The WGMFCQ is based on the smoothing functions and the sequence of iteration points generated by the smoothing algorithm. It was shown in [45, 46] that although the GMFCQ will never hold for the bilevel program, the weaker version of the GMFCQ may hold for bilevel programs.

In this paper, we extend the concept of WGMFCQ to the case where abstract constraint is involved. For the general nonsmooth and nonconvex problem (P), we propose a smoothing augmented Lagrange algorithm and show that any accumulation point is a stationary point of the original problem (P) under the extension version of WGMFCQ. Note that we do not assume that the WGMFCQ holds at all feasible and infeasible accumulation points as in [20, 44, 46]. Finally, we prove that if the upper bound of the augmented Lagrangian functions along the approximated solution sequence exists, any accumulation point will be a feasible point and the WGMFCQ guarantees a stationary point. We also apply the smoothing augmented Lagrangian method to the bilevel program.

The rest of the paper is organized as follows. In Section 2, we present a summary of the constraint qualifications and extend the WGMFCQ to the case where abstract constraint involved. In Section 3, we propose a smoothing augmented Lagrangian algorithm for locating a stationary point of a general nonsmooth and nonconvex optimization problem (P) and establish a convergence result for the algorithm. Furthermore, we prove that if a feasible point for problem (P) is known, then any accumulation point will be a feasible point and the WGMFCQ guarantees a stationary point. In Section 4, we report the results of our numerical experiments for some general nonsmooth and nonconvex constrained optimization problems as well as some bilevel programs.

We adopt the following standard notation in this paper. For any two vectors a and b in $\mathbb{R}^n$, we denote by $a^T b$ their inner product. Given a function $G : \mathbb{R}^n \to \mathbb{R}^m$, we denote its Jacobian by $\nabla G(z) \in \mathbb{R}^{m\times n}$ and, if $m = 1$, the gradient $\nabla G(z) \in \mathbb{R}^n$ is considered as a column vector. For a set $\Omega \subseteq \mathbb{R}^n$, we denote by aff Ω,int Ω, ri Ω, co Ω, and dist(x, Ω) the affine hull, interior, relative interior, the convex hull, and the distance from x to Ω respectively. For a matrix $A \in \mathbb{R}^{n\times m}$, A^T denotes its transpose. Let $\exp[z]$ be the exponential function.

2. Background and constraint qualifications.

In this section, we present some background materials on variational analysis. Detailed discussions on these subjects can be found in [17, 18, 32, 40]. Then we discuss the issue of constraint qualifications and extend the WGMFCQ to the case where abstract constraint involved.

Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz continuous near $\bar{x}$. The directional derivative of φ at $\bar{x}$ in direction d is defined by

$$\varphi'(\bar{x}; d) \quad := \quad \lim_{t\downarrow 0} \frac{\varphi(\bar{x} + td) - \varphi(\bar{x})}{t}.$$

The Clarke generalized directional derivative of φ at $\bar{x}$ in direction d is defined by

$$\varphi^\circ(\bar{x}; d) := \limsup_{x\to\bar{x},\ t\downarrow 0} \frac{\varphi(x+td)-\varphi(x)}{t}.$$

The Clarke generalized gradient of φ at $\bar{x}$ is a convex and compact subset of $\mathbb{R}^n$ defined by

$$\partial\varphi(\bar{x}) := \{\xi \in \mathbb{R}^n : \xi^T d \le \varphi^\circ(\bar{x}; d),\quad \forall d \in \mathbb{R}^n\}.$$

Note that when φ is convex, the Clarke generalized gradient coincides with the subdifferential in the sense of convex analysis, i.e.,

$$\partial\varphi(\bar{x}) = \{\xi \in \mathbb{R}^n : \xi^T(x-\bar{x}) \le \varphi(x) - \varphi(\bar{x}),\quad \forall x \in \mathbb{R}^n\}$$

and, when φ is continuously differentiable at $\bar{x}$, we have $\partial\varphi(\bar{x}) = \{\nabla\varphi(\bar{x})\}$. Detailed discussions of the Clarke generalized gradient and its properties can be found in [17, 18].

For a nonempty closed set $\Omega \subseteq \mathbb{R}^n$ and a point $\bar{x} \in \Omega$, the Clarke tangent cone [17, 18] of Ω at $\bar{x}$ is given by

$$\mathcal{T}_\Omega(\bar{x}) := \{d \in \mathbb{R}^n : \mathrm{dist}^o(\bar{x}, \Omega) = 0\},$$

and the Clarke normal cone [17, 18] of Ω at $\bar{x}$ is given by

$$\mathcal{N}_\Omega(\bar{x}) := \{\zeta \in \mathbb{R}^n : \zeta^T d \le 0,\ \forall d \in \mathcal{T}_\Omega(\bar{x})\}$$

respectively.

Lemma 2.1 (Moreau's decomposition theorem)**.** [40] *Let $\mathcal{K} \in \mathbb{R}^n$ be a closed convex cone and $\mathcal{K}^*$ its polar cone; that is, the closed convex cone defined by $\mathcal{K}^* = \{a \in \mathbb{R}^n | \langle a, b\rangle \le 0, \forall b \in \mathcal{K}\}$. Then for $x, y, z \in \mathbb{R}^n$ the following statements are equivalent:*

$$\begin{aligned} &z = x + y,\ x \in \mathcal{K},\ y \in \mathcal{K}^* \text{ and } \langle x, y\rangle = 0,\\ \iff\ & x = P_{\mathcal{K}}(z) \text{ and } y = P_{\mathcal{K}^*}(z), \end{aligned}$$

where $P_C(\cdot)$ denotes the projection to set C.

From Theorem 6.28 [40], we have

$$\mathcal{N}(z;\Omega)^* = \mathcal{T}(z;\Omega).$$

Definition 2.2. [Stationary point] We call a feasible point $\bar{x}$ of problem (P) a (Clarke) stationary point if there exists $\lambda \in \mathbb{R}^q$ such that

$$\begin{aligned} &0 \in \partial f(\bar{x}) + \sum_{i=1}^{p} \lambda_i \partial g_i(\bar{x}) + \sum_{j=p+1}^{q} \lambda_j \partial h_j(\bar{x}) + \mathcal{N}_\Omega(\bar{x}),\\ &\lambda_i \ge 0,\ \lambda_i g_i(\bar{x}) = 0,\ i = 1, \cdots, p. \end{aligned}$$

Following from the Fritz John type necessary optimality condition [17, Theorem 6.1.1],

Definition 2.3. [NNAMCQ] We say that the no nonzero abnormal multiplier constraint qualification (NNAMCQ) holds at a feasible point $\bar{x}$ of problem (P) if

$$\begin{aligned} &0 \in \sum_{i\in I(\bar{x})} \lambda_i \partial g_i(\bar{x}) + \sum_{j=p+1}^{q} \lambda_j \partial h_j(\bar{x}) + \mathcal{N}_\Omega(\bar{x}) \text{ and } \lambda_i \ge 0,\ i \in I(\bar{x}) \Longrightarrow \lambda_i = 0,\\ &\lambda_j = 0. \end{aligned}$$

where $I(\bar{x}) := \{i = 1, \cdots, p : g_i(\bar{x}) = 0\}$ the active set at $\bar{x}$.

The NNAMCQ is equal to the GMFCQ.

Definition 2.4. [GMFCQ] A feasible point $\bar{x}$ is said to satisfy the generalized Mangasarian-Fromovitz constraint qualification (GMFCQ) for problem (P) if
(i) $v_{p+1}, \cdots, v_q$ are linearly independent, where $v_j \in \partial h_j(\bar{x})$, $j = p+1, \cdots, q$,
(ii) there exists a direction $d \in \text{int } \mathcal{T}_\Omega(\bar{x})$ such that

$$\begin{aligned} &v_i^T d < 0, \quad \forall v_i \in \partial g_i(\bar{x}),\ i \in I(\bar{x}), \\ &v_j^T d = 0, \quad \forall v_j \in \partial h_j(\bar{x}),\ j = p+1, \cdots, q. \end{aligned}$$

However, NNAMCQ and GMFCQ may be too strong to hold for many problems, such as the bilevel programs.

Using the smoothing technique, one can approximate the Lipschitz functions $f(x)$, $g_i(x)$, $i = 1, \cdots, p$ and $h_j(x)$, $j = p+1, \cdots, q$ by families of smoothing functions $\{f_\rho(x) : \rho > 0\}$, $\{g_\rho^i(x) : \rho > 0\}$, $i = 1, \cdots, p$ and $\{h_\rho^j(x) : \rho > 0\}$, $j = p+1, \cdots, q$ which satisfy the gradient consistency property. Based on the sequence of iteration points generated by certain algorithm, [45] defined two new constraint qualifications for problem (P) without the abstract constraint. In this paper, we extend the conditions to the case where the abstract constraint is involved:

Definition 2.5. [WNNAMCQ] Let $\{x_k\}$ be a sequence of iteration points for problem (P) and $\rho_k \uparrow \infty$ as $k \to \infty$. Suppose that $\bar{x}$ is a feasible accumulation point of the sequence $\{x_k\}$. We say that the weakly no nonzero abnormal multiplier constraint qualification (WNNAMCQ) based on the smoothing functions $\{g_\rho^i(x) : \rho > 0\}$, $i = 1, \cdots, p$, $\{h_\rho^j(x) : \rho > 0\}$, $j = p+1, \cdots, q$ holds at $\bar{x}$ provided that $\lambda_i \geq 0$, $i \in I(\bar{x})$ and

$$0 \in \sum_{i \in I(\bar{x})} \lambda_i v_i + \sum_{j=p+1}^{q} \lambda_j v_j + \mathcal{N}_\Omega(\bar{x}) \tag{2}$$

implies that $\lambda_i = 0, \lambda_j = 0$ for any $K_0 \subseteq K \subseteq \mathbf{N}$ such that $\lim\limits_{k \to \infty, k \in K} x_k = \bar{x}$ and

$$\begin{aligned} v_i &= \lim_{k \to \infty, k \in K_0} \nabla g_{\rho_k}^i(x_k),\ i \in I(\bar{x}), \\ v_j &= \lim_{k \to \infty, k \in K_0} \nabla h_{\rho_k}^j(x_k),\ j = p+1, \cdots, q. \end{aligned}$$

Definition 2.6. [WGMFCQ] Let $\{x_k\}$ be a sequence of iteration points for problem (P) and $\rho_k \uparrow \infty$ as $k \to \infty$. Let $\bar{x}$ be a feasible accumulation point of the sequence $\{x_k\}$. We say that the weakly generalized Mangasarian Fromovitz constraint qualification (WGMFCQ) based on the smoothing functions $\{g_\rho^i(x) : \rho > 0\}$, $i = 1, \cdots, p$, $\{h_\rho^j(x) : \rho > 0\}$, $j = p+1, \cdots, q$ holds at $\bar{x}$ provided the following conditions hold. For any $K_0 \subseteq K \subseteq \mathbf{N}$ such that $\lim\limits_{k \to \infty, k \in K} x_k = \bar{x}$ and any

$$\begin{aligned} v_i &= \lim_{k \to \infty, k \in K_0} \nabla g_{\rho_k}^i(x_k),\ i \in I(\bar{x}) \\ v_j &= \lim_{k \to \infty, k \in K_0} \nabla h_{\rho_k}^j(x_k),\ j = p+1, \cdots, q, \end{aligned}$$

(i) $v_{p+1}, \cdots, v_q$ are linearly independent;
(ii) there exists a direction $d \in \text{int } \mathcal{T}_\Omega(\bar{x})$ such that

$$v_i^T d < 0, \text{ for all } i \in I(\bar{x}), \tag{3}$$

$$v_j^T d = 0, \text{ for all } j = p+1, \cdots, q.. \tag{4}$$

We now show the equivalence between the WGMFCQ and WNNAMCQ.

Theorem 2.7. *Let $\bar{x} \in \Omega$. Then the following implication always holds:*

$$\text{WGMFCQ} \Longrightarrow \text{WNNAMCQ},$$

and the reverse implication holds provided $\operatorname{int} \mathcal{T}_\Omega(\bar{\mathrm{x}}) \neq \emptyset$.

Proof. We first show that WGMFCQ implies WNNAMCQ. To the contrary we suppose that WGMFCQ holds but WNNAMCQ does not hold which means that there exist scalars $\lambda_i \geq 0$, $i \in I(\bar{x})$, $\lambda_j \geq 0, j = p+1, \cdots, q$ not all zero such that for any $K_0 \subseteq K \subseteq \mathbf{N}$ such that $\lim\limits_{k\to\infty, k\in K} x_k = \bar{x}$ and $\eta \in \mathcal{N}_\Omega(\bar{x})$,

$$\begin{aligned} v_i &= \lim_{k\to\infty, k\in K_0} \nabla g^i_{\rho_k}(x_k),\ i \in I(\bar{x}), \\ v_j &= \lim_{k\to\infty, k\in K_0} \nabla h^j_{\rho_k}(x_k),\ j = p+1, \cdots, q, \end{aligned}$$

such that

$$0 = \sum_{i\in I(\bar{x})} \lambda_i v_i + \sum_{j=p+1}^{q} \lambda_j v_j + \eta. \tag{5}$$

Suppose that $d \in \operatorname{int} \mathcal{T}_\Omega(\bar{x})$ is the direction that satisfies the condition (ii) of WGMFCQ. When λ_i, $i \in I(\bar{x})$ not all zero, multiplying both sides of condition (5) by d, it follows from conditions (3) and (4) that

$$0 = \sum_{i\in I(\bar{x})} \lambda_i v_i^T d + \sum_{j=p+1}^{q} \lambda_j v_j^T d < 0,$$

which is a contradiction.

When $\lambda_i = 0$, for all $i \in I(\bar{x})$, multiplying both sides of condition (5) with d, it follows from condition (4) that

$$\sum_{j=1}^{m} \lambda_j v_j^T d = 0.$$

Since $d \in \operatorname{int} \mathcal{T}_\Omega(\bar{x})$, there exists $\varepsilon > 0$ such that

$$d + \varepsilon\mathbb{B} \in \mathcal{T}_\Omega(\bar{x}),$$

where $\mathbb{B}$ denotes the unit ball in $\mathbb{R}^n$. Then multiplying both sides of condition (5) with $d + d'$, $\forall d' \in \varepsilon\mathbb{B}$, we have

$$0 \leq \sum_{j=p+1}^{q} \lambda_j v_j^T (d + d') = \sum_{j=1}^{m} \lambda_j v_j^T d'.$$

Since d' is an arbitrary vector in $\varepsilon\mathbb{B}$, we must have

$$\sum_{j=1}^{m} \lambda_j v_j = 0,$$

which contradicts with the linearly independence of $v_1, \cdots, v_m$. Therefore, WNNAMCQ holds.

We now prove the reverse implication. Assume the WNNAMCQ holds. WNNAMCQ implies (i) of WGMFCQ. If both (i) and (ii) of WGMFCQ hold, we are done. Suppose that the condition (ii) of WGMFCQ does not hold; that is, there exists a subsequence $K_0 \subset K \subset N$ and $v_1, \cdots, v_q$ with $\lim\limits_{k\to\infty, k\in K} x_k = \bar{x}$ and

$$\begin{aligned} v_i &= \lim_{k\to\infty, k\in K_0} \nabla g^i_{\rho_k}(x_k),\ i = 1, \cdots, p, \\ v_j &= \lim_{k\to\infty, k\in K_0} \nabla h^j_{\rho_k}(x_k),\ j = p+1, \cdots, q, \end{aligned}$$

such that for all direction $d \in \operatorname{int} \mathcal{T}_\Omega(\bar{x})$, (3) or (4) fails to hold. Let A be the matrix with $v_i, i \in I(\bar{x}), v_{p+1}, \ldots, v_q$ are columns and

$$\begin{aligned} &S_1 := \{z : \exists d \in \operatorname{int} \mathcal{T}_\Omega(\bar{\mathrm{x}}) \text{ such that } \mathrm{z} = \mathrm{A}^\mathrm{T}\mathrm{d}\}, \\ &S_2 := \{z : z_i < 0,\ i \in I(\bar{x}),\ z_j = 0,\ j = p+1, \cdots, q\}. \end{aligned}$$

Since the convex sets S_1 and $\mathrm{cl}S_2$ are nonempty and ri S_1 and ri $\mathrm{cl}S_2$ have no point in common by the violation of the condition (ii) of EWGMFCQ, there exists a hyperplane separating S_1 and $\mathrm{cl}S_2$ properly from [36, Theorem 11.3]. Since S_1 is a subspace and thus a cone, from [36, Theorem 11.7], there exists a hyperplane separating S_1 and $\mathrm{cl}S_2$ properly and passes through the origin. By the separation theorem (see e.g. [36, Theorem 11.1]), there exists a vector y such that

$$\begin{aligned} &\inf\{y^T z : z \in S_1\} \geq 0 \geq \sup\{y^T z : z \in \mathrm{cl}S_2\}, \\ &\sup\{y^T z : z \in S_1\} > \inf\{y^T z : z \in \mathrm{cl}S_2\}. \end{aligned} \tag{6}$$

From (6), we know that $y \neq 0$. Therefore, there exists $y \in \mathbb{R}^q$, $y \neq 0$ such that $y^T z \geq 0, \forall z \in S_1$ and $y^T z \leq 0, \forall z \in \mathrm{cl}S_2$.

(a) We first consider the inequality $y^T z \leq 0, \forall z \in \mathrm{cl}S_2$. By taking $z^0 \in \mathrm{cl}S_2$ such that $z^0_j = 0, j = p+1, \ldots, q$ and $z^0_i \to -\infty, i \in I(\bar{x})$, we conclude that

$$y_i \geq 0, \quad i \in I(\bar{x}). \tag{7}$$

(b) We now consider the inequality $y^T z \geq 0, \forall z \in S_1$. Select an arbitrary nonzero $d \in \operatorname{int} \mathcal{T}_\Omega(\bar{x})$. Then $z^1 := A^T d \in S_1$, and hence

$$\sum_{i\in I(\bar{x})} y_i v_i^T d + \sum_{j=p+1}^{q} y_j v_j^T d = y^T z^1 \geq 0.$$

That is,

$$-\Big(\sum_{i\in I(\bar{x})} y_i v_i + \sum_{j=p+1}^{q} y_j v_j\Big) \in \mathcal{N}_\Omega(\bar{x}). \tag{8}$$

Therefore, if there exists a nonzero vector y such that $y^T z \geq 0, \forall z \in S_1$ and $y^T z \leq 0$, $\forall z \in \mathrm{cl}S_2$, the vector should also satisfies conditions (7)-(8). While from the WNNAMCQ, conditions (7)-(8) imply that $y = 0$, which is a contradiction. Thus the condition (ii) of WGMFCQ must hold. The proof is therefore complete. □

The WNNAMCQ and the WGMFCQ can be extended to infeasible points.

Definition 2.8. [EWNNAMCQ] Let $\{x_k\}$ be a sequence of iteration points for problem (P) and $\rho_k \uparrow \infty$ as $k \to \infty$. Let $\bar{x} \in \Omega$ be a accumulation point of the sequence $\{x_k\}$. We say that the extended weakly no nonzero abnormal multiplier constraint qualification (EWNNAMCQ) based on the smoothing functions $\{g^i_\rho(x) :$

$\rho > 0\}$, $i = 1, \cdots, p$, $\{h_\rho^j(x) : \rho > 0\}$, $j = p+1, \cdots, q$ holds at $\bar{x}$ provided that

$$0 \in \sum_{i=1}^{p} \lambda_i v_i + \sum_{j=p+1}^{q} \lambda_j v_j + \mathcal{N}_\Omega(\bar{x}) \text{ and } \lambda_i \geq 0,\ i = 1, \cdots, p,$$

$$\sum_{i=1}^{p} \lambda_i g_i(\bar{x}) + \sum_{j=p+1}^{q} \lambda_j h_j(\bar{x}) \geq 0.$$

implies that $\lambda_i = 0, \lambda_j = 0$ for any $K_0 \subseteq K \subseteq \mathbf{N}$ such that $\lim_{k\to\infty, k\in K} x_k = \bar{x}$ and

$$\begin{aligned} v_i &= \lim_{k\to\infty, k\in K_0} \nabla g_{\rho_k}^i(x_k),\ i = 1, \cdots, p, \\ v_j &= \lim_{k\to\infty, k\in K_0} \nabla h_{\rho_k}^j(x_k),\ j = p+1, \cdots, q. \end{aligned}$$

Definition 2.9. [EWGMFCQ] Let $\{x_k\}$ be a sequence of iteration points for problem (P) and $\rho_k \uparrow \infty$ as $k \to \infty$. Let $\bar{x} \in \Omega$ be a accumulation point of the sequence $\{x_k\}$. We say that the extended weakly generalized Mangasarian Fromovitz constraint qualification (EWGMFCQ) based on the smoothing functions $\{g_\rho^i(x) : \rho > 0\}$, $i = 1, \cdots, p$, $\{h_\rho^j(x) : \rho > 0\}$, $j = p+1, \cdots, q$ holds at $\bar{x}$ provided that the following conditions hold. For any $K_0 \subseteq K \subseteq \mathbf{N}$ such that $\lim_{k\to\infty, k\in K} x_k = \bar{x}$ and any

$$\begin{aligned} v_i &= \lim_{k\to\infty, k\in K_0} \nabla g_{\rho_k}^i(x_k),\ i = 1, \cdots, p, \\ v_j &= \lim_{k\to\infty, k\in K_0} \nabla h_{\rho_k}^i(x_k),\ j = p+1, \cdots, q, \end{aligned}$$

(i) $v_{p+1}, \cdots, v_q$ are linearly independent;
(ii) there exists a nonzero direction $d \in \text{int } \mathcal{T}_\Omega(\bar{x})$ nonzero such that

$$\begin{aligned} &g_i(\bar{x}) + v_i^T d < 0, \text{ for all } i = 1, \cdots, p, \\ &h_j(\bar{x}) + v_j^T d = 0, \text{ for all } j = p+1, \cdots, q. \end{aligned}$$

The equivalence between the EWGMFCQ and EWNNAMCQ is an extension of Theorem 2.1 and [45, Theorem 2.2].

3. An augmented Lagrangian method for problem (P).

In this section, we use smoothing functions with gradient property to approximate the Lipschitz functions in problem (P), and then we introduce an augmented Lagrangian algorithm. We establish the convergence theorem of the algorithm under the WNNAMCQ.

We define the PHR augmented Lagrangian function as follows:

$$G_\rho^{\lambda,c}(x) := f_\rho(x) + \frac{1}{2c}\sum_{i=1}^{p}\left(\max\{0, \lambda_i + c g_\rho^i(x)\}^2 - \lambda_i^2\right) + \sum_{j=p+1}^{q}\left(\lambda_j h_\rho^j(x) + \frac{c}{2}(h_\rho^j(x))^2\right),$$

for each smoothing parameter $\rho > 0$ and thus we consider the following penalized problem:

$$(\mathrm{P}_\rho^{\lambda,c}) \ \min_{x\in\Omega} \qquad G_\rho^{\lambda,c}(x),$$

for each $\rho > 0, c > 0, \lambda \in \mathbb{R}^q$. In the algorithm, we denote a robust residual function measuring the infeasibility and the complementarity by

$$\sigma_\rho^\lambda(x) := \max\left\{|h_\rho^j(x)|, j = p+1, \cdots, q,\ |\min\{\lambda_i, -g_\rho^i(x)\}|, i = 1, \cdots, p\right\}.$$

Recently, Lu and Zhang [29] proposed a feasible augmented Lagrangian method where the values of the augmented Lagrangian functions along the approximated solution sequence are bounded above, and the magnitude of penalty parameters outgrows that of Lagrangian multipliers. Such procedures will derive feasible accumulation points.

Since $(\mathrm{P}_{\rho}^{\lambda,c})$ is a smooth optimization problem for any fixed $\rho > 0, c > 0, \lambda \in \mathbb{R}^q$, we first develop an augmented Lagrangian method to deal with the problem. Then we update the iteration by increasing the smoothing parameter ρ, updating the multiplier λ and increasing the penalty parameter c provided that $\sigma_{\rho}^{\lambda}(x)$ has an enough decrease.

Note that in the algorithm, we do not require the upper bound of augmented Lagrangian functions along the approximated solution sequence exists since many problems, such as the bilevel problems, the upper bound may not easy to obtain. We will show that any sequence of iteration points generated by the algorithm converges to some stationary point of problem (P) when ρ goes to infinity and the penalty parameter c is bounded. Furthermore, the EWNNAMCQ guarantees the boundedness of the sequence of the Lagrangian multiplier λ and any iteration sequence converge to a stationary point. Finally, we assume at least one feasible point of problem (P) exists, denoted by x^{feas}, and thus the upper bound of the augmented Lagrangian functions along the approximated solution sequence exists. In this case, any accumulation point will be a feasible point and the WNNAMCQ guarantees a stationary point.

We now describe the smoothing augmented Lagrangian algorithm as follows.

Algorithm 3.1. *Let τ be a constant in $(0,1)$ Let $\{\varepsilon_k\}$ be a positive sequence converging to 0 and σ be a constant in $(1,+\infty)$. Choose an initial point $x^0 = x^1 \in \Omega$, an initial smoothing parameter $\rho_0 > 0$, an initial penalty parameter $c_0 > 0$ and initial multipliers $\lambda_i^0 \geq 0,\ i = 1,\cdots,q$. Let $\lambda_i^1 = \max\{0, \lambda_i^0 + c_0 g_{\rho_0}^i(x^1)\},\ i = 1,\cdots,p; \lambda_j^1 = \lambda_j^0 + c_0 h_{\rho_0}^j(x^1),\ j = p+1,\cdots,q$. Set $k := 0$.*

1. *If $\operatorname{dist}(0, \nabla G_{\rho_k}^{\lambda^k,c_k}(x^{k+1}) + \mathcal{N}_\Omega(x^{k+1})) = 0$ and $\sigma_{\rho_k}^{\lambda^{k+1}}(x^{k+1}) = 0$, terminate. Otherwise, set $k = k+1$, and go to Step 2.*
2. *Compute an approximate solution $x^{k+1} \in \Omega$ for the subproblem $(P_\rho^{\lambda,c})$ with $\rho = \rho_k$, $\lambda = \lambda_k$, $c = c_k$ such that*

$$\operatorname{dist}(0, \nabla G_{\rho_k}^{\lambda^k,c_k}(x^{k+1}) + \mathcal{N}_\Omega(x^{k+1})) < \hat{\eta}\rho_k^{-1}. \tag{9}$$

   Set $\rho_{k+1} := \sigma\rho_k$, go to Step 3.
3. *Set*

$$\lambda_i^{k+1} = \max\{0, \lambda_i^k + c_k g_{\rho_k}^i(x^{k+1})\},\ i = 1,\cdots,p; \tag{10}$$

$$\lambda_j^{k+1} = \lambda_j^k + c_k h_{\rho_k}^j(x^{k+1}),\ j = p+1,\cdots,q. \tag{11}$$

   and go to Step 4.
4. *If*

$$\sigma_{\rho_k}^{\lambda^{k+1}}(x^{k+1}) < \varepsilon_k, \tag{12}$$

   go to Step 1. Otherwise, set

$$c_{k+1} := \max\{\sigma c_k, \|\lambda^{k+1}\|^{1+\tau}\},$$

   $k = k+1$ and go to Step 2.

The quantity $\mathrm{dist}(0, \nabla G_{\rho_k}^{\lambda^k, c_k}(x^{k+1}) + \mathcal{N}_\Omega(x^{k+1}))$ is generally convenient as a measure of how near x^{k+1} is to being a minimizer of $(P_{\rho_k}^{\lambda^k, c_k})$. To calculate a distance function to a normal cone seems not easy. However, there are some methods to make it easy.

(i) Since $\Omega \in \mathbb{R}^n$ is a closed and convex set, from the discussion in [39], we derive that

$$\mathrm{dist}(0, \nabla G_{\rho_k}^{\lambda^k, c_k}(x^{k+1}) + \mathcal{N}_\Omega(x^{k+1})) = \|P_{\mathcal{T}_\Omega(x^{k+1})}(-\nabla G_{\rho_k}^{\lambda^k, c_k}(x^{k+1}))\|. \tag{13}$$

Actually, setting $\mathcal{K} = \mathcal{N}_\Omega(x^{k+1})$ and $\mathcal{K}^* = \mathcal{T}_\Omega(x^{k+1})$. By observing the Lemma 2.1, we have that the following two equations are equivalent for $W^{k+1} \in \mathcal{N}_\Omega(x^{k+1}), T^{k+1} \in \mathcal{T}_\Omega(x^{k+1})$,

$$\begin{aligned}&(a) - \nabla G_{\rho_k}^{\lambda^k, c_k}(x^{k+1}) = W^{k+1} + T^{k+1},\ \langle W^{k+1}, T^{k+1}\rangle = 0,\\ &(b) W^{k+1} = P_{\mathcal{N}_\Omega(x^{k+1})}(-\nabla G_{\rho_k}^{\lambda^k, c_k}(x^{k+1})),\ T^{k+1} = P_{\mathcal{T}_\Omega(x^{k+1})}(-\nabla G_{\rho_k}^{\lambda^k, c_k}(x^{k+1})).\end{aligned}$$

Thus we can obtain

$$\mathrm{dist}(-\nabla G_{\rho_k}^{\lambda^k, c_k}(x^{k+1}), \mathcal{N}_\Omega(x^{k+1})) = \|T^{k+1}\| = \|P_{\mathcal{T}_\Omega(x^{k+1})}(-\nabla G_{\rho_k}^{\lambda^k, c_k}(x^{k+1}))\|,$$

which derives (13) holds. From the preliminaries in Section 2, the tangent cone is calculable. Thus $\|P_{\mathcal{T}_\Omega(x^{k+1})}(-\nabla G_{\rho_k}^{\lambda^k, c_k}(x^{k+1}))\| = 0$ could be used as the stopping criterion.

(ii) In some situations when the tangent cone is complicated to calculate, we could also use the following method.

Consider the stopping criterion $\mathrm{dist}(0, \nabla G_{\rho_k}^{\lambda^k, c_k}(x^{k+1}) + \mathcal{N}_\Omega(x^{k+1})) = 0$, which equals to $0 \in \nabla G_{\rho_k}^{\lambda^k, c_k}(x^{k+1}) + \mathcal{N}_\Omega(x^{k+1})$ and thus

$$x^{k+1} - \nabla G_{\rho_k}^{\lambda^k, c_k}(x^{k+1}) \in x^{k+1} + \mathcal{N}_\Omega(x^{k+1}).$$

From Proposition 6.17 [40], we have $P_C(x)^{-1} = x + \mathcal{N}_C(x)$ and thus there exists $y \in x + \mathcal{N}_C(x)$ such that $x = P_C(y)$. Therefore, $x^{k+1} = P_\Omega(x^{k+1} - \nabla G_{\rho_k}^{\lambda^k, c_k}(x^{k+1}))$ and the stopping criterion $\mathrm{dist}(0, \nabla G_{\rho_k}^{\lambda^k, c_k}(x^{k+1}) + \mathcal{N}_\Omega(x^{k+1})) = 0$ can be replaced by

$$\|x^{k+1} - P_\Omega(x^{k+1} - \nabla G_{\rho_k}^{\lambda^k, c_k}(x^{k+1}))\| = 0.$$

Many researchers assumed that certain constraint qualifications hold at all feasible and infeasible accumulation points such that the parameters are bounded. In this paper, we do not need such assumption since the magnitude of $\{c_k\}$ outgrows that of $\{\lambda^k\}$; see Step 4.

We now establish the convergence of Algorithm 3.1.

Theorem 3.1. *Suppose the Algorithm 3.1 does not terminate within finite iterations. Let x^* be an accumulation point of the sequence $\{x^k\}$ generated by Algorithm 3.1. If $\{c_k\}$ is bounded, then x^* is a stationary point of problem* (P).

Proof. Without loss of generality, assume that $\lim_{k\to\infty} x^k = x^*$. From Step 4, the boundedness of $\{c_k\}$ is equivalent to saying that condition (12) holds for sufficiently large k and thus $\lim_{k\to\infty} \sigma_{\rho_k}^{\lambda^{k+1}}(x^{k+1}) = 0$, i.e., $|h_{\rho_k}^j(x^{k+1})| = 0$, $j = p+1, \cdots, q$ and $g_{\rho_k}^i(x^{k+1}) \le 0$, $i = 1, \cdots, p$ for sufficiently large k from the definition of $\sigma_\rho^\lambda(\cdot)$. Thus $\{\lambda^k\}$ is bounded from the updating rule (10) – (11).

Since $\{g_\rho^i : \rho > 0\}$, $i = 1, \cdots, p$, $\{h_\rho^j : \rho > 0\}$, $j = p+1, \cdots, q$ are families of smoothing functions of g_i, $i = 1, \cdots, p$, h_j, $j = p+1, \cdots, q$, taking limits as $k \to \infty$, we have $h_j(x^*) = 0,\ j = p+1, \cdots, q,\ g_i(x^*) \le 0,\ i = 1, \cdots, p$ and thus x^* is a feasible point. Let

$$\mu_{\rho,i}^{\lambda,c}(x) := \max\{0, \lambda_i + c g_\rho^i(x)\},\ i = 1, \cdots, p,$$
$$\mu_{\rho,j}^{\lambda,c}(x) := \lambda_j + c h_\rho^j(x),\ j = p+1, \cdots, q.$$

By calculation, we have

$$\begin{aligned} \nabla G_{\rho_k}^{\lambda^k, c_k}(x^{k+1}) &= \nabla f_{\rho_k}(x^{k+1}) + \sum_{i=1}^{p} \mu_{\rho_k,i}^{\lambda^k,c_k}(x^{k+1}) \nabla g_{\rho_k}^i(x^{k+1}) \\ &\quad + \sum_{j=p+1}^{q} \mu_{\rho_k,j}^{\lambda^k,c_k}(x^{k+1}) \nabla h_{\rho_k}^j(x^{k+1}) \\ &= \nabla f_{\rho_k}(x^{k+1}) + \sum_{i=1}^{p} \lambda_i^{k+1} \nabla g_{\rho_k}^i(x^{k+1}) + \sum_{j=p+1}^{q} \lambda_j^{k+1} \nabla h_{\rho_k}^j(x^{k+1}). \end{aligned} \tag{14}$$

Since $\{\lambda^k\}$ is bounded, we assume there exists a subsequence $K_0 \subseteq \mathbf{N}$ and λ^* such that $\lambda^* := \lim_{k\to\infty,\, k\in K_0} \lambda^k$.

By the gradient consistency property of $f_\rho(\cdot)$, $g_\rho^i(\cdot)$, $i = 1, \cdots, p$ and $h_\rho^j(\cdot)$, $j = p+1, \cdots, q$, there exists a subsequence $\bar{K}_0 \subseteq K_0$ such that

$$\lim_{k\to\infty,\, k\in\bar{K}_0} \nabla f_{\rho_k}(x^{k+1}) \in \partial f(x^*),$$
$$\lim_{k\to\infty,\, k\in\bar{K}_0} \nabla g_{\rho_k}^i(x^{k+1}) \in \partial g_i(x^*),\ i = 1, \cdots, p,$$
$$\lim_{k\to\infty,\, k\in\bar{K}_0} \nabla h_{\rho_k}^j(x^{k+1}) \in \partial h_j(x^*),\ j = p+1, \cdots, q.$$

From condition (9), we know that there exists $\xi^{k+1} \in \mathbb{R}^n$ such that

$$\xi^{k+1} \in \nabla G_{\rho_k}^{\lambda^k, c_k}(x^{k+1}) + \mathcal{N}_\Omega(x^{k+1}), \tag{15}$$

where $\|\xi^{k+1}\| \le \hat{\eta}\rho_k^{-1}$ and thus $\xi_k \to 0$ follows by $\rho_k \to \infty$ as $k \to \infty$. Taking limits in (15) as $k \to \infty$, $k \in \bar{K}_0$, by the gradient consistency properties, we have

$$0 \in \partial f(x^*) + \sum_{i=1}^{p} \lambda_i^* \partial g_i(x^*) + \sum_{j=p+1}^{q} \lambda_j^* \partial h_j(x^*) + \mathcal{N}_\Omega(x^*). \tag{16}$$

It follows from (10) that $\lambda_i^* \ge 0, i = 1, \cdots, p$. We now show that the complementary slackness condition holds. If $g_i(x^*) < 0$ for certain $i \in \{1, \ldots, p\}$, we have $g_{\rho_k}^i(x^{k+1}) < 0$ for sufficiently large k since $\{g_\rho^i : \rho > 0\}$ are families of smoothing functions of g_i, $i = 1, \cdots, p$. Then $\lambda_i^* = \lim_{k\to\infty,\, k\in\bar{K}_0} \lambda_i^{k+1} = 0$ since $\sigma_{\rho_k}^{\lambda^{k+1}}(x^{k+1}) = 0$ for sufficiently large k. Therefore x^* is a stationary point of problem (P) and the proof of the theorem is complete. □

Theorem 3.2. *Suppose that the Algorithm 3.1 does not terminate within finite iterations and $\{x_k, \rho_k, \lambda_k, c_k\}$ is a sequence generated by Algorithm 3.1. If* EWNNAM CQ *holds at any accumulation point x^* for the problem (P), then the parameter sequence $\{\lambda^k\}$ is bounded and thus x^* is a stationary point of problem* (P).

Proof. From Theorem 3.1, if c_k is bounded, the conclusions hold automatically. We consider the case where $\{c_k\}$ is unbounded. Since x^* is an arbitrary accumulation point, we assume there exist $K \subseteq \mathbf{N}$ and x^* such that $\lim\limits_{k\to\infty, k\in K} x_k = x^*$, and a set $\tilde{K} \subseteq K$ such that condition (12) fails for every $k \in \tilde{K}$ sufficiently large. From the update rule, we have $c_{k+1} := \max\{\sigma c_k, \|\lambda^{k+1}\|^{1+\tau}\}$ for all large $k \in \tilde{K}$, thus

$$0 \leq \frac{\|\lambda^k\|}{c_k} \leq (c_k)^{-\frac{\tau}{1+\tau}} \to 0, \text{ as } k \to \infty. \tag{17}$$

Without loss of generality, we assume there exists a subset $\tilde{K}_0 \subseteq \tilde{K}$ such that

$$\begin{aligned}
&\lim_{k\to\infty,\, k\in\tilde{K}_0} f_{\rho_k}(x^{k+1}) = f(x^*),\\
&\lim_{k\to\infty,\, k\in\tilde{K}_0} g^i_{\rho_k}(x^{k+1}) = g_i(x^*),\ i = 1, \cdots, p,\\
&\lim_{k\to\infty,\, k\in\tilde{K}_0} h^j_{\rho_k}(x^{k+1}) = h_j(x^*),\ j = p+1, \cdots, q.
\end{aligned}$$

By the gradient consistency property of $f_\rho(\cdot)$, $g^i_\rho(\cdot)$, $i = 1, \cdots, p$ and $h^j_\rho(\cdot)$, $j = p+1, \cdots, q$, there exists a subsequence $\bar{K} \subseteq \tilde{K}_0$ such that

$$\begin{aligned}
v &:= \lim_{k\to\infty,\, k\in\bar{K}} \nabla f_{\rho_k}(x^{k+1}) \in \partial f(x^*),\\
v_i &:= \lim_{k\to\infty,\, k\in\bar{K}} \nabla g^i_{\rho_k}(x^{k+1}) \in \partial g_i(x^*),\ i = 1, \cdots, p,\\
v_j &:= \lim_{k\to\infty,\, k\in\bar{K}} \nabla h^j_{\rho_k}(x^{k+1}) \in \partial h_j(x^*),\ j = p+1, \cdots, q.
\end{aligned}$$

We now assume for a contradiction that λ^k is unbounded. By the definition of $\mu^{\lambda,c}_\rho(\cdot)$, we have $\|\lambda^{k+1}\| = \|\mu^{\lambda^k,c_k}_{\rho_k}(x^{k+1})\| \to \infty$. There exists a subsequence $\bar{K}_0 \subseteq \bar{K}$ and nonzero vector $\mu \in \mathbb{R}^q$ nonzero such that

$$\lim_{k\to\infty, k\in\bar{K}_0} \frac{\mu^{\lambda^k,c_k}_{\rho_k}(x^{k+1})}{\|\mu^{\lambda^k,c_k}_{\rho_k}(x^{k+1})\|} = \mu.$$

It follows from the definition of $\mu(\cdot)$ that $\mu_i \geq 0,\ i = 1, \cdots, p$.

Similarly as in Theorem 3.1, (15) holds. Dividing by $\|\mu^{\lambda^k,c_k}_{\rho_k}(x^{k+1})\|$ in both sides of (15) and letting $k \to \infty$ in $\bar{K}_0$, we have

$$0 \in \sum_{i=1}^{p} \mu_i v_i + \sum_{j=p+1}^{q} \mu_j v_j + \mathcal{N}_\Omega(x^*). \tag{18}$$

If $g_i(x^*) < 0$ for certain $i \in \{1, \ldots, p\}$, we have $g^i_{\rho_k}(x^{k+1}) < 0$ for sufficiently large k, respectively, since $\{g^i_\rho : \rho > 0\}$ are families of smoothing functions of g_i, $i = 1, \cdots, p$. Thus $\mu_i = \lim\limits_{k\to\infty,\, k\in\bar{K}_0} \mu^{\lambda^k,c_k}_{\rho_k,i}(x^{k+1}) = 0$ by the definitions $\mu^{\lambda,c}_\rho(\cdot)$ and the unboundedness of $\{c_k\}$. Consequently we have $\mu_i = 0$ if $g_i(x^*) < 0$, for $i \in \{1, \cdots, p\}$.

For each $j = p+1, \cdots, q$ such that $h_j(x^*) \neq 0$. We only consider when $h_j(x^*) < 0$ which also implies that $h^j_{\rho_k}(x^{k+1}) < 0$ for sufficiently large k. From the definition of $\mu^{\lambda,c}_\rho(\cdot)$, for sufficiently large $k \in \bar{K}_0$, $j = p+1, \cdots, q$, we have

$$\begin{aligned}
\mu^{\lambda^k,c_k}_{\rho_k,j}(x^{k+1}) h^j_{\rho_k}(x^{k+1}) &= (\lambda^k_j + c_k h^j_{\rho_k}(x^{k+1})) h^j_{\rho_k}(x^{k+1})\\
&= \lambda^k_j h^j_{\rho_k}(x^{k+1}) + c_k (h^j_{\rho_k}(x^{k+1}))^2 > 0.
\end{aligned} \tag{19}$$

Otherwise, we assume for a contrary there exists a subsequence $K_1 \subseteq \bar{K}_0$ such that for $k \in K_1$,

$$\lambda_j^k h_{\rho_k}^j(x^{k+1}) + c_k (h_{\rho_k}^j(x^{k+1}))^2 \le 0.$$

Dividing c_k in both sides of the above inequality and letting $k \to \infty, k \in K_1$, from (17), $\frac{\|\lambda^k\|}{c_k} \to 0$ and thus $h_j^2(x^*) \le 0$, which contradicts with $h_j(x^*) < 0$. Therefore, (19) holds for sufficiently large $k \in \bar{K}_0$.

From the definition of $\mu(\cdot)$,

$$\mu_j = \lim_{k\to\infty, k\in\bar{K}_0} \frac{\mu_{\rho_k,j}^{\lambda^k,c_k}(x^{k+1})}{\|\mu_{\rho_k}^{\lambda^k,c_k}(x^{k+1})\|}$$

for $j = p+1, \cdots, q$. Thus the limit of $\frac{\mu_{\rho_k,j}^{\lambda^k,c_k}(x^{k+1})}{\|\mu_{\rho_k}^{\lambda^k,c_k}(x^{k+1})\|} h_{\rho_k}^j(x^{k+1})$ exists for $k \in \bar{K}_0$ and followed by (19),

$$\mu_j h_j(x^*) = \lim_{k\to\infty, k\in\bar{K}_0} \frac{\mu_{\rho_k,j}^{\lambda^k,c_k}(x^{k+1})}{\|\mu_{\rho_k}^{\lambda^k,c_k}(x^{k+1})\|} h_{\rho_k}^j(x^{k+1}) \ge 0. \tag{20}$$

Similarly, $\mu_j h_j(x^*) \ge 0$ if $h_j(x^*) > 0$. Therefore,

$$\sum_{i=1}^{p} \mu_i g_i(x^*) + \sum_{j=p+1}^{q} \mu_j h_j(x^*) \ge 0. \tag{21}$$

Conditions (18) and (21) contradict with the EWNNAMCQ assumption. Thus λ^k is bounded. From the update rule of λ^k, x^* should be a feasible point since the unboundedness of c_k. Similarly with the proof of Theorem 3.1, x^* is a stationary point of problem (P). □

We now discuss the situation if a feasible point of problem (P) is known, denotes by $x^{feas} \in \Omega$, then the values of the augmented Lagrangian functions along the approximated solution sequence are bounded above, i.e., for any k,

$$G_{\rho_k}^{\lambda_k,c_k}(x^{k+1}) \le \Upsilon, \tag{22}$$

for a constant $\Upsilon > f(x^{\mathrm{feas}}) + 1$.

In this case, choosing a proper initial point is important. For any k, we denote the initial point for solving the subproblem $(\mathrm{P}_\rho^{\lambda,c})$ with $\rho = \rho_k$, $\lambda = \lambda_k$, $c = c_k$ as follows:

$$x_{\mathrm{int}}^{k+1} = \begin{cases} x^{\mathrm{feas}}, & \text{if } G_{\rho_k}^{\lambda_k,c_k}(x^k) > \frac{\Upsilon}{2}, \\ x^k, & \text{otherwise,} \end{cases}$$

where x^k is an approximate stationary point of the kth subproblem $(\mathrm{P}_\rho^{\lambda,c})$ satisfying condition (22).

Since x^{feas} is a feasible point of problem (P), $G_{\rho_k}^{\lambda^k,c_k}(x^{\mathrm{feas}})$ would approximate to $f(x^{feas})$ when k is sufficiently large. The above inequality together with the choice of x_{int}^k implies that $G_{\rho_k}^{\lambda^k,c_k}(x_{\mathrm{int}}^{k+1}) \le \Upsilon$. Additionally, the objective values at all subsequent iterates generated by the non-monotone gradient method are bounded above by the one at the initial point. That is

$$G_{\rho_k}^{\lambda^k,c_k}(x^{k+1}) \le G_{\rho_k}^{\lambda^k,c_k}(x_{\mathrm{int}}^{k+1}) \le \Upsilon,$$

and thus the condition (22) holds at x^{k+1}.

Theorem 3.3. *Suppose that the Algorithm 3.1 does not terminate within finite iterations and $\{x_k, \rho_k, \lambda_k, c_k\}$ is a sequence generated by Algorithm 3.1. Assume a feasible point of problem (P) is known. If* WNNAMCQ *holds at any accumulation point x^* for the problem (P), then*
(i) x^* is a feasible point of (P)*.*
(ii) The parameter sequence $\{\lambda^k\}$ is bounded and thus x^* is a stationary point of problem (P)*.*

Proof. Assume that there exists $K \subseteq \mathbf{N}$ and x^* such that $\lim\limits_{k\to\infty, k\in K} x_k = x^*$. Similarly as in Theorem 3.2, we consider the case where $\{c_k\}$ is unbounded and for all large $k \in K$, (17) holds.

(i) Without loss of generality, we assume there exists a subset $\tilde{K}_0 \subseteq K$ such that

$$\begin{aligned}
&\lim_{k\to\infty,\, k\in\tilde{K}_0} f_{\rho_k}(x_k) = f(x^*),\\
&\lim_{k\to\infty,\, k\in\tilde{K}_0} g^i_{\rho_k}(x_k) = g_i(x^*),\ \ i = 1, \cdots, p,\\
&\lim_{k\to\infty,\, k\in\tilde{K}_0} h^j_{\rho_k}(x_k) = h_j(x^*),\ \ j = p+1, \cdots, q.
\end{aligned}$$

From the condition (22), we have

$$\begin{aligned}
&f_{\rho_k}(x^{k+1}) + \frac{1}{2c_k}\sum_{i=1}^{p}\big(\max\{0, \lambda_i^k + c_k g^i_{\rho_k}(x^{k+1})\}^2 - (\lambda_i^k)^2\big)\\
&+ \sum_{j=p+1}^{q} \lambda_j^k h^j_{\rho_k}(x^{k+1}) + \frac{c_k}{2}(h^j_{\rho_k}(x^{k+1}))^2 \leq \Upsilon.
\end{aligned}$$

It follows that

$$\begin{aligned}
&\frac{1}{2}\sum_{i=1}^{p}\left(\max\{0, \frac{\lambda_i^k}{c_k} + g^i_{\rho_k}(x^{k+1})\}^2 - (\frac{\lambda_i^k}{c_k})^2\right)\\
&+ \sum_{j=p+1}^{q}\left(\frac{\lambda_j^k}{c_k} h^j_{\rho_k}(x^{k+1}) + \frac{1}{2}(h^j_{\rho_k}(x^{k+1}))^2\right) \leq \frac{1}{c_k}(\Upsilon - f_{\rho_k}(x^{k+1})).
\end{aligned}$$

Taking limits on both sides as $k \to \infty, k \in \tilde{K}_0$ and using (17), we have

$$\frac{1}{2}\sum_{i=1}^{p}\max\{0, g_i(x^*)\}^2 + \frac{1}{2}\sum_{j=p+1}^{q} h_j^2(x^*) = 0$$

from the definition of smoothing function. Therefore x^* is a feasible point of problem (P).

(ii) The second conclusion follows from Theorem 3.2. $\square$

4. **Applications and numerical examples.**

In this section, we first test our algorithm on two general nonsmooth and nonconvex constrained optimization problems. Then we apply the algorithm to the bilevel programs.

In numerical practise, it is impossible to obtain an exact '0', thus we select some small enough $\epsilon > 0$, $\epsilon_1 > 0$ and terminate the algorithm when

$$\|P_{\mathcal{T}_\Omega(x^{k+1})}(-\nabla G^{\lambda^k, c_k}_{\rho_k}(x^{k+1}))\| < \epsilon \ \text{ and } \ \sigma^{\lambda^{k+1}}_{\rho_k}(x^{k+1}) < \epsilon_1.$$

The former terminate condition is derived by (13).

4.1. **Illustrative examples for general problems.** In this subsection, we illustrate the Algorithm 3.1 by two general nonsmooth and nonconvex constrained optimization problems.

Example 4.1. [22, Example 5.1] Consider the nonsmooth constrained optimization program of minimizing a nonsmooth Rosenbrock function subject to an inequality constraint on a weighted maximum value of the variables:

$$\begin{array}{ll} \min & f(x,y) := 8|x^2 - y| + (1-x)^2 \\ \text{s.t.} & g(x,y) := \max\{\sqrt{2}x, 2y\} - 1 \le 0. \end{array}$$

The unique optimal solution of the problem is $(\bar{x}, \bar{y}) = (\frac{\sqrt{2}}{2}, \frac{1}{2})$.

From easily calculation,the NNAMCQ is satisfied at every point in $\mathbb{R}^2$ and a feasible point $(x^{feas}, y^{feas}) = (0.5, 0.3)$ is known. Our convergent theorem guarantees that any accumulation point of the iteration sequence must be a stationary point from Theorem 3.3.

Rewrite the objective function and the constraint function as

$$\begin{aligned} F(x,y) &= 8\left((x^2-y)_+ + (-x^2+y)_+\right) + (1-x)^2 \\ g(x,y) &= \sqrt{2}x + (2y - \sqrt{2}x)_+ - 1. \end{aligned}$$

We use the following functions to approximate the Lipchitz functions:

$$\begin{aligned} F_\rho(x,y) &:= 8\sqrt{(x^2-y)^2 + \rho^{-1}} + (1-x)^2, \\ g_\rho(x,y) &:= \frac{1}{2}\left(\sqrt{2}x + 2y + \sqrt{(2y-\sqrt{2}x)^2 + \rho^{-1}}\right) - 1. \end{aligned}$$

In our test, we choose the initial point $(x_0, y_0) = (0.5, 0.3)$ and the parameters $\rho_0 = 100$, $c_0 = 100$, $\hat{\eta} = 10^3$, $\sigma = 10, \Gamma = 2, \tau = 0.5$, $\lambda = 100$ and $\epsilon = 10^{-5}$, $\epsilon_1 = 10^{-6}$. The stopping criteria

$$\operatorname{dist}(0, \nabla G_{\rho_k}^{\lambda^k, c_k}(x^{k+1}) + \mathcal{N}_\Omega(x^{k+1})) < \epsilon \quad \text{and} \quad \sigma_{\rho_k}^{\lambda^{k+1}}(x^{k+1}) < \epsilon_1$$

hold with $(x^{k+1}, y^{k+1}) = (0.70708, 0.49996)$, which is a good approximation of the true optimal solution.

Example 4.2. [7, Example 5.1] Consider the nonsmooth constrained optimization program of minimizing a nonsmooth Rosenbrock function subject to one nonsmooth inequality constraint and one linear equality constraint:

$$\begin{array}{ll} \min & f(x,y) := 8|x^2 - y| + (1-x)^2 \\ \text{s.t.} & g(x,y) := x^2 + |y| - 4 \le 0, \\ & h(x,y) := x - \sqrt{2}y = 0. \end{array}$$

The unique optimal solution of the problem is $(\bar{x}, \bar{y}) = (\frac{\sqrt{2}}{2}, \frac{1}{2})$.

From easily calculation,the NNAMCQ is satisfied at every point in $\mathbb{R}^2$ and a feasible point $(x^{feas}, y^{feas}) = (0.5, \frac{\sqrt{2}}{4})$ is known. Our convergent theorem guarantees that any accumulation point of the iteration sequence must be a stationary point from Theorem 3.3.

We use the following functions to approximate $F(x,y)$ and $g(x,y)$:

$$\begin{aligned} F_\rho(x,y) &:= 8\psi_\rho^3(x^2 - y) + (1-x)^2, \\ g_\rho(x,y) &:= x^2 + \psi_\rho^3(y) - 4. \end{aligned}$$

In our test, we choose the initial point $(x_0, y_0) = (0.8, 0.6)$, and the parameters $\rho_0 = 20$, $c_0 = 100$, $\hat{\eta} = 5 * 10^3$, $\sigma = 10, \Gamma = 4, \tau = 0.5$, $\lambda = 100$ and $\epsilon = 10^{-3}$, $\epsilon_1 = 10^{-6}$. The stopping criteria

$$\mathrm{dist}(0, \nabla G_{\rho_k}^{\lambda^k, c_k}(x^{k+1}) + \mathcal{N}_\Omega(x^{k+1})) < \epsilon \quad \text{and} \quad \sigma_{\rho_k}^{\lambda^{k+1}}(x^{k+1}) < \epsilon_1$$

hold with $(x^{k+1}, y^{k+1}) = (0.70710, 0.5000)$, which is a good approximation of the true optimal solution.

4.2. **Applications to the bilevel program.** In this subsection, we consider the simple bilevel program

$$\begin{aligned} \text{(SBP)} \quad \min \quad & F(x,y) \\ \text{s.t.} \quad & g_i(x,y) \leq 0, i = 1, \cdots, l, \\ & x \in X, y \in S(x), \end{aligned}$$

where $S(x)$ denotes the set of solutions of the lower level program

$$(\mathrm{P}_x) \ \min_{y \in Y} \ f(x,y),$$

where $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ are compact subsets, $f, F, g_i, i = 1, \cdots, l : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ are continuously differentiable functions and f is twice continuously differentiable in variable y. When the lower level constraint set Y is depend on the variable x, (SBP) turns to a general bilevel program. Applications and recent developments of general bilevel programs can be found in [5, 23, 24, 41, 43].

A general practice to solve the bilevel program is to replace the lower level program by its first order conditions and thus the problem reduces to a MPEC problem. While when the lower level problem is not convex on variable y, Mirrlees [30] pointed out that the true optimal solution may not be found by such approach.

By defining $V(x) := \inf_{y \in Y} f(x,y)$ as the value function of the lower level program, (SBP) equivalents to a single level problem [34, 47, 48]:

$$\begin{aligned} \text{(VP)} \quad \min \quad & F(x,y) \\ \text{s.t.} \quad & f(x,y) - V(x) = 0, \\ & g_i(x,y) \leq 0, i = 1, \cdots, l, \\ & x \in X, y \in Y. \end{aligned} \tag{23}$$

However, the optimal solution may not be a stationary point of (VP). To overcome such difficulty, Ye and Zhu [49] proposed to consider a combined program with both the first order condition and the value function constraint involved. By assuming that every optimal solution of the lower level problem is an interior point of set Y, the combined program takes the form:

$$\begin{aligned} \text{(CP)} \quad \min \quad & F(x,y) \\ \text{s.t.} \quad & f(x,y) - V(x) \leq 0, \\ & g_i(x,y) \leq 0, i = 1, \cdots, l, \\ & \nabla_y f(x,y) = 0, \\ & x \in X, y \in Y. \end{aligned}$$

It is already shown that the value function $V(x)$ is usually nonsmooth, even when the function $f(x,y)$ is smooth, Lin, Xu and Ye [28] proved that the integral entropy function:

$$\begin{aligned}\gamma_\rho(x) &:= -\rho^{-1}\ln\left(\int_Y \exp[-\rho f(x,y)]dy\right)\\ &= V(x)-\rho^{-1}\ln\left(\int_Y \exp[-\rho(f(x,y)-V(x))]dy\right)\end{aligned}$$

is a smoothing function of $V(x)$ and satisfies the gradient consistency property. Based on the smoothing method, many algorithms were proposed to solve the bilevel program [28, 44, 45, 46].

In the rest of this subsection, we apply Algorithm 3.1 to (CP), and verify that the EWNNAMCQ holds and thus the solution of the algorithm is a stationary point of the bilevel problem. Furthermore, we also compare the Algorithm 3.1 with some existing algorithms to solve the bilevel program. Since the method in [28] is an approximated method, we only consider the smoothing sequential quadratic programming (SQP) algorithm in [45] and the smoothing augmented Lagrangian method (SAL) in [46].

Example 4.3 (Mirrlees' problem)**.** [30] Consider Mirrlees' problem

$$\begin{aligned}\min \quad & F(x,y):=(x-2)^2+(y-1)^2\\ \text{s.t.} \quad & x\in X:=[-1,1], y\in S(x),\end{aligned}$$

where $S(x)$ is the solution set of the lower level program

$$\begin{aligned}\min \quad & f(x,y):=-x\exp[-(y+1)^2]-\exp[-(y-1)^2]\\ \text{s.t.} \quad & y\in Y:=[-1,1].\end{aligned}$$

It was shown in [30] that the unique optimal solution is $(\bar{x},\bar{y})$ with $\bar{x}=1$, $\bar{y}\approx 0.9575$.

In our test, we choose the initial point $(x_0,y_0)=(0.7,0.5)$ and the parameters $\rho_0=100$, $c_0=100$, $\hat{\eta}=10^3$, $\sigma=10, \tau=0.5$, $\lambda=(100,100)$ and $\epsilon=10^{-3}$, $\epsilon_1=10^{-5}$. The stopping criteria hold with $(x^{k+1},y^{k+1})\approx(1,0.957504)$. Since $\mathcal{T}_{X\times Y}(x^{k+1},y^{k+1})=\left\{(d_1,d_2)^T: d_1\le 0, d_2\in\mathbb{R}\right\}$. From calculation,

$$\nabla G_{\rho_k}^{\lambda^k,c_k}(x^{k+1},y^{k+1}))=10^{-3}\times(-0.5038,0.8099)^T,$$

thus

$$\begin{aligned}\|P_{\mathcal{T}_{X\times Y}(x^{k+1},y^{k+1})}(-\nabla G_{\rho_k}^{\lambda^k,c_k}(x^{k+1},y^{k+1}))\| &= \|10^{-3}\times(0,0.8099)^T\|\\ &= 8.099\times 10^{-4}<\epsilon.\end{aligned}$$

It seems that the sequence converges to $(\bar{x},\bar{y})$.

Since

$$\begin{aligned}&\nabla f(x^{k+1},y^{k+1})-(\nabla\gamma_{\rho_k}(x^{k+1}),0)=(0.0188,0)^T,\\ &\nabla(\nabla_y f)(x^{k+1},y^{k+1})=(0.08484,1.7004)^T,\end{aligned}$$

it is easy to see that the vectors $\nabla f(\bar{x},\bar{y})-(\lim_{k\to\infty}\nabla\gamma_{\rho_k}(x^{k+1}),0)$ and $\nabla(\nabla_y f)(\bar{x},\bar{y})$ are linearly independent. Thus the EWNNAMCQ holds at $(\bar{x},\bar{y})$ and our algorithm guarantees that $(\bar{x},\bar{y})$ is a stationary point of (CP) from Theorem 3.2.

We now compare the Algorithm 3.1 with the SQP algorithm and the SAL algorithm. Let (x^*,y^*) be the point generated by certain algorithm which is approximated to the accumulation point. The results are reported in Table 1, in which

$d(x^*, y^*)$ means the distance between (x^*, y^*) and the optimal point $(\bar{x}, \bar{y})$ defined by

$$d(x^*, y^*) \approx |x^* - 1| + |y^* - 0.9575| .$$

TABLE 1. Mirrlees' problem

	(x^*, y^*)	$d(x^*, y^*)$
Algorithm 3.1	(1,0.957504)	5.73e-006
SQP algorithm	(1.000002,0.957598)	9.79e-005
SAL algorithm	(1.000905,0.957459)	9.06e-004

Example 4.4. [31, Example 3.20] The bilevel program

$$\begin{aligned} \min \quad & F(x,y) := (x-0.25)^2 + y^2 \\ \text{s.t.} \quad & y \in S(x) := \underset{y\in[-1,1]}{\operatorname{argmin}} f(x,y) := \tfrac{1}{3}y^3 - x^2 y, \\ & x, y \in [-1,1] \end{aligned}$$

has the optimal solution point $(\bar{x}, \bar{y}) = (\frac{1}{2}, \frac{1}{2})$ with an objective value of $\frac{5}{16}$.

In our test, we choose the initial point $(x_0, y_0) = (0.7, 0.2)$ and the parameters $\rho_0 = 100,\ c_0 = 100,\ \hat{\eta} = 10^3,\ \sigma = 10, \tau = 0.5,\ \lambda = (100, 100)$ and $\epsilon = 6 * 10^{-4}$, $\epsilon_1 = 5 \times 10^{-6}$. The stopping criteria hold with $(x^{k+1}, y^{k+1}) \approx (0.500003, 0.500003)$. It seems that the sequence converges to $(\bar{x}, \bar{y})$.

Since

$$\begin{aligned} &\nabla f(x^{k+1}, y^{k+1}) - (\nabla \gamma_{\rho_k}(x^{k+1}), 0) = (-0.0151, -0)^T, \\ &\nabla(\nabla_y f)(x^{k+1}, y^{k+1}) = (-1, 1)^T, \end{aligned}$$

it is easy to see that the vectors $\nabla f(\bar{x}, \bar{y}) - (\lim_{k\to\infty} \nabla \gamma_{\rho_k}(x^{k+1}), 0)$ and $\nabla(\nabla_y f)(\bar{x}, \bar{y})$ are linearly independent. Thus the EWNNAMCQ holds at $(\bar{x}, \bar{y})$ and our algorithm guarantees that $(\bar{x}, \bar{y})$ is a stationary point of (CP) from Theorem 3.2.

We now compare the Algorithm 3.1 with the SQP algorithm and the SAL algorithm. Let (x^*, y^*) be the point generated by certain algorithm which is approximated to the accumulation point. The results are reported in Table 2, in which $d(x^*, y^*)$ means the distance between (x^*, y^*) and the optimal point $(\bar{x}, \bar{y})$ defined by

$$d(x^*, y^*) := |x^* - \bar{x}| + |y^* - \bar{y}| .$$

TABLE 2. Example 4.4

	(x^*, y^*)	$d(x^*, y^*)$
Algorithm 3.1	(0.500003,0.500003)	4.08e-006
SQP algorithm	(0.499996,0.499996)	5.85e-006
SAL algorithm	(0.500000,0.499995)	2.89e-005

Examples 4.3 and 4.4 show that the Algorithm 3.1 succeeds in solving the bilevel problems. The algorithm derives a sequence of points converging to the solution. What is more, the accumulation point seems more accurate than the SQP algorithm in [45] and SAL algorithm in [46].

REFERENCES

[1] ALGENCAN, `http://www.ime.usp.br/$\sim$egbirgin/tango/`.

[2] R. Andreani, E. G. Birgin, J. M. Martínez and M. L. Schuverdt, On Augmented Lagrangian methods with general lower-level constraints, *SIAM J. Optim.*, **18** (2007), 1286–1309.

[3] R. Andreani, E. G. Birgin, J. M. Martínez and M. L. Schuverdt, Augmented Lagrangian methods under the constant positive linear dependence constraint qualification, *Math. Program., Ser. B*, **111** (2008), 5–32.

[4] R. Andreani, G. Haeser and M. L. Schuverdt, A relaxed constant positive linear dependence constraint qualification and applications, *Math. Program.*, **135** (2012), 255–273.

[5] J. F. Bard, *Practical Bilevel Optimization: Algorithms and Applications*, Kluwer Academic Publications, Dordrecht, Netherlands, 1998.

[6] D. P. Bertsekas, *Constrained Optimization and Lagrangian Multiplier Methods*, Academic Press, New York, 1982.

[7] W. Bian and X. Chen, Neural network for nonsmooth, nonconvex constrained minimization via smooth approximation, *IEEE Trans. Neural Netw. Learn. Syst.*, **25** (2014), 545–556.

[8] E. G. Birgin, D. Fernández and J. M. Martínez, The boundedness of penalty parameters in an Augmented Lagrangian method with lower level constraints, *Optim. Methods Soft.*, **27** (2012), 1001–1024.

[9] J. V. Burke and T. Hoheisel, Epi-convergent smoothing with applications to convex composite functions, *SIAM J. Optim.*, **23** (2013), 1457–1479.

[10] J. V. Burke, T. Hoheisel and C. Kanzow, Gradient consistency for integral-convolution smoothing functions, *Set-Valued Var. Anal.*, **21** (2013), 359–376.

[11] B. Chen and X. Chen, A global and local superlinear continuation-smoothing method for P_0 and R_0 NCP or monotone NCP, *SIAM J. Optim.*, **9** (1999), 624–645.

[12] X. Chen, Smoothing methods for nonsmooth, nonconvex minimization, *Math. Program.*, **134** (2012), 71–99.

[13] B. Chen and P. T. Harker, A non-interior-point continuation method for linear complementarity problems, *SIAM J. Matrix Anal. Appl.*, **14** (1993), 1168–1190.

[14] X. Chen, L. Guo, Z. Lu and J. J. Ye, An augmented Lagrangian method for non-Lipschitz nonconvex programming, *SIAM J. Numer. Anal.*, **55** (2017), 168–193.

[15] C. Chen and O. L. Mangasarian, A class of smoothing functions for nonlinear and mixed complementarity problems, *Math. Program.*, **71** (1995), 51–70.

[16] X. Chen, R. S. Womersley and J. J. Ye, Minimizing the condition number of a gram matrix, *SIAM J. Optim.*, **21** (2011), 127–148.

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

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

[19] A. R. Conn, N. I. M. Gould and Ph. L. Toint, A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bound, *SIAM J. Numer. Anal.*, **28** (1991), 545–572.

[20] A. R. Conn, N. I. M. Gould and Ph. L. Toint, *Trust Region Methods*, MPS/SIAM Series on Optimization, SIAM, Philadelphia, PA, 2000.

[21] F. E. Curtis, H. Jiang and D. P. Robinson, An adaptive augmented Lagrangian method for large-scale constrained optimization, *Math. Program.*, **152** (2015), 201–245.

[22] F. E. Curtis and M. L. Overton, A sequential quadratic programming algorithm for nonconvex, nonsmooth constrained optimization, *SIAM J. Optim.*, **22** (2012), 474–500.

[23] S. Dempe, *Foundations of Bilevel Programming*, Kluwer Academic Publishers, 2002.

[24] S. Dempe, Annotated bibliography on bilevel programming and mathematical programs with equilibrium constraints, *Optim.*, **52** (2003), 333–359.

[25] M. R. Hestenes, Multiplier and gradient methods, *J. Optim. Theory Appl.*, **4** (1969), 303–320.

[26] C. Kanzow, Some noninterior continuation methods for linear complementarity problems, *SIAM J. Matrix Anal. Appl.*, **17** (1996), 851–868.

[27] LANCELOT, `http://www.cse.scitech.ac.uk/nag/lancelot/lancelot.shtml`.

[28] G. H. Lin, M. Xu and J. J. Ye, On solving simple bilevel programs with a nonconvex lower level program, *Math. Program., series A*, **144** (2014), 277–305.

[29] Z. Lu and Y. Zhang, An augmented Lagrangian approach for sparse principal component analysis, *Math. Program. series A,* **135** (2012), 149–193.

[30] J. Mirrlees, The theory of moral hazard and unobservable behaviour: Part I, *Rev. Econ. Stud.*, **66** (1999), 3–22.

[31] A. Mitsos and P. Barton, *A Test Set for Bilevel Programs*, Technical Report, Massachusetts Institute of Technology, 2006.

[32] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation, Vol.1: Basic Theory, Vol.2: Applications*, Springer, 2006.

[33] Y. Nesterov, Smoothing minimization of nonsmooth functions, *Math. Program.*, **103** (2005), 127–152.

[34] J. V. Outrata, On the numerical solution of a class of Stackelberg problems, *Z. Oper. Res.*, **34** (1990), 255–277.

[35] M. J. D. Powell, A method for nonlinear constraints in minimization problems, in *Optimization*(eds. R. Fletcher), 283–298, London and New York, 1969. Academic Press.

[36] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, New Jersey, 1970.

[37] R. T. Rockfellar, A dual approach to solving nonlinear programming problems by unconstrained optimization, *Math. Program.*, **5** (1973), 354–373.

[38] R. T. Rockfellar, Augmented Lagrange multiplier functions and duality in nonconvex programming, *SIAM J. Con.*, **12** (1974), 268–285.

[39] R. T. Rockfellar, Monotone operators and the proximal point algorithm, *SIAM J. Con. Optim.*, **14** (1976), 877–898.

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

[41] K. Shimizu, Y. Ishizuka and J. F. Bard, *Nondifferentiable and Two-Level Mathematical Programming*, Kluwer Academic Publishers, Boston, 1997.

[42] S. Smale, Algorithms for solving equations, In: *Proceedings of the International Congress of Mathematicians*,Berkeley, CA., (1986), 172–195.

[43] L. N. Vicente and P. H. Calamai, Bilevel and multilevel programming: A bibliography review, *J. Global Optim.*, **5** (1994), 291–306.

[44] M. Xu and J. J. Ye, A smoothing augmented Lagrangian method for solving simple bilevel programs, *Compu. Optim. App.*, **59** (2014), 353–377.

[45] M. Xu, J. J. Ye and L. Zhang, Smoothing sequential quadratic programming method for solving nonconvex, nonsmooth constrained optimization problems, *SIAM J. Optim.*, **25** (2015), 1388–1410.

[46] M. Xu, J. J. Ye and L. Zhang, Smoothing augmented Lagrangian method for nonsmooth constrained optimization problems, *J. Glob. Optim.*, **62** (2015), 675–694.

[47] J. J. Ye and D. L. Zhu, Optimality conditions for bilevel programming problems, *Optim.*, **33** (1995), 9–27.

[48] J. J. Ye and D. L. Zhu, A note on optimality conditions for bilevel programming problems, *Optim.*, **39** (1997), 361–366.

[49] J. J. Ye and D. L. Zhu, New necessary optimality conditions for bilevel programs by combining MPEC and the value function approach, *SIAM J. Optim.*, **20** (2010), 1885–1905.

Received July 2017; revised October 2017.

E-mail address: duanqs@mail.dlut.edu.cn
E-mail address: xumengw@hotmail.com
E-mail address: jinjin403@sina.com
E-mail address: lwzhang@dlut.edu.cn