J Glob Optim (2015) 62:675–694
DOI 10.1007/s10898-014-0242-7

Smoothing augmented Lagrangian method for nonsmooth constrained optimization problems

Mengwei Xu · Jane J. Ye · Liwei Zhang

Received: 9 March 2014 / Accepted: 27 September 2014 / Published online: 8 October 2014

Abstract In this paper, we propose a smoothing augmented Lagrangian method for finding a stationary point of a nonsmooth and nonconvex optimization problem. We show that any accumulation point of the iteration sequence generated by the algorithm is a stationary point provided that the penalty parameters are bounded. Furthermore, we show that a weak version of the generalized Mangasarian Fromovitz constraint qualification (GMFCQ) at the accumulation point is a sufficient condition for the boundedness of the penalty parameters. Since the weak GMFCQ may be strictly weaker than the GMFCQ, our algorithm is applicable for an optimization problem for which the GMFCQ does not hold. Numerical experiments show that the algorithm is efficient for finding stationary points of general nonsmooth and nonconvex optimization problems, including the bilevel program which will never satisfy the GMFCQ.

Keywords Nonsmooth optimization · Constrained optimization · Smoothing function · Augmented Lagrangian method · Constraint qualification · Bilevel program

Mathematics Subject Classification 65K10 · 90C26

The research of Jane J. Ye was partially supported by NSERC and the research of Liwei Zhang was supported by the National Natural Science Foundation of China under Projects Nos. 11071029, 91330206 and 91130007.

M. Xu
School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China
e-mail: xumengw@hotmail.com

J. J. Ye (✉)
Department of Mathematics and Statistics, University of Victoria, Victoria, BC V8W 2Y2, Canada
e-mail: janeye@uvic.ca

L. Zhang
Dalian University of Technology, Dalian 116024, China
e-mail: lwzhang@dlut.edu.cn

1 Introduction

In this paper, we consider a constrained optimization problem with inequality and equality constraints:

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

Unless otherwise specified, we assume that the objective function and constraint functions $f, g_i(i = 1, \ldots, p), h_j(j = p+1, \ldots, q) : \mathbb{R}^n \to \mathbb{R}$ are Lipschitz around the point of interest.

The quadratic penalty method for problem (P) when all functions are smooth is an inexact penalty method which locates stationary points of a sequence of smooth penalized problems:

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

and takes $c \uparrow \infty$ to find the stationary point of the original constrained problem. However, when c is large, the penalized problems may become ill-conditioned and very difficult to solve. The augmented Lagrangian method (see e.g. [21]), which is also known as the method of multipliers, reduces the possibility of ill-conditioning by adding the estimates of Lagrange multipliers into the penalty function. This method is also the basis for some high quality software such as ALGENCAN [1] and LANCELOT [35]. The augmented Lagrangian method was first proposed by Hestenes [29] and Powell [41] for equality constrained problems. Bertsekas [6] and Rockafellar [43,44] extended the method to inequality constrained convex optimization problems and to nonconvex optimization problems respectively. The Powell–Hestenes–Rockafellar (PHR) augmented Lagrangian function [29,41,44] (see [8] for a comparison with other augmented Lagrangian functions) takes the form:

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

which is a sum of the standard Lagrangian function and the quadratic penalty function. Even when the functions f, g and h are twice continuously differentiable, the PHR augmented Lagrangian function is not twice continuously differentiable. To guarantee twice continuous differentiability, an exponential-type augmented Lagrangian function was proposed in [34,40,48]. The augmented Lagrangian functions were also used as merit functions for the sequential quadratic programming (SQP) methods [10,11,13,26]. Other related researches for augmented Lagrangian methods may be found in [21,27,28,30].

The boundedness of the penalty parameters is a basic requirement for the convergence result in most exact penalty methods for constrained optimization, including the augmented Lagrangian method. Rockafellar [43,44] showed that the augmented Lagrangian method is an exact penalty method. However, to ensure the boundedness of the penalty parameters, it has been common to make assumptions that the Linear Independence Constraint Qualification (LICQ) holds at all feasible and infeasible accumulation points of the iteration sequence (see e.g. [22]). In fact, the MFCQ holding at all feasible and infeasible accumulation points of the iteration sequence is sufficient to ensure the boundedness of the penalty parameters (see e.g. [50]). The MFCQ is a strong condition since many problems such as the mathematical program with equilibrium constraints (MPEC), will never satisfy the MFCQ. Recently some papers [2,3,9] studied the boundedness of the penalty parameters for the augmented Lagrangian algorithm under the Constant Positive Linear Dependence (CPLD) constraint qualification, the MFCQ and the LICQ. The CPLD condition was proposed by Qi and Wei [42] and proved to be a constraint qualification by Andreani et al. [4]. Although the CPLD condition is weaker than the MFCQ, it is still not reasonable to expect the CPLD condition to hold for general MPECs [31].

In the case where all functions except one constraint function are smooth, Xu and Ye [50] used a family of smoothing functions with gradient consistency properties to approximate the nonsmooth function, applied the augmented Lagrangian method to the smoothing problems and updated the smoothing parameter. The boundedness of the penalty parameters and hence the global convergence of the algorithm was shown under the assumptions that the nonsmooth version of the MFCQ called the generalized MFCQ (GMFCQ) holds at all feasible and infeasible accumulation points. Unfortunately GMFCQ never holds for bilevel programs (see [53]) which was our main motivation to study an augmented Lagrange algorithm for nonsmooth and nonconvex problems. It was observed that under the calmness condition [50], the sequence of multipliers is very likely to be bounded and hence the smoothing augmented Lagrangian algorithm is efficient for searching a stationary point of bilevel program. However up to now, there is no proof that the calmness condition would guarantee the boundedness of the penalty parameters. To cope with this difficulty, recently Xu et al. [51] proposed a weaker version of the GMFCQ called the weak GMFCQ (WGMFCQ). It was shown in [51] 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 smoothing augmented Lagrange algorithm to the general nonsmooth and nonconvex problem (P). We show that if either the exact penalty sequence is bounded or if all feasible or infeasible accumulation points of the iteration sequence generated by the algorithm satisfy the WGMFCQ, then any accumulation point is a stationary point of the original problem (P). We apply the smoothing augmented Lagrangian method to the bilevel program and verify that either the exact penalty sequence is bounded or the WGMFCQ holds for all bilevel programs in this paper.

The rest of the paper is organized as follows. In Sect. 2, we present a summary of constraint qualifications that will guarantee the global convergence of the algorithm. In Sect. 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. In Sect. 4, we report 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^Tb 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 matrix $A \in \mathbb{R}^{n\times m}$, A^T denotes its transpose. In addition, we let $\mathbf{N}$ be the set of nonnegative integers and $\exp[z]$ be the exponential function.

2 Constraint qualifications

The focus of this section is on constraint qualifications. Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz continuous near $\bar{x}$. We denote the Clarke generalized gradient of φ at $\bar{x}$ by $\partial\varphi(\bar{x})$. Definition of the Clarke generalized gradient and its properties can be found in [19,20].

From now on for $\bar{x}$, a feasible solution of problem (P), we denote by $I(\bar{x}) := \{i = 1, \ldots, p : g_i(\bar{x}) = 0\}$ the active set at $\bar{x}$.

Definition 2.1 (*Stationary point*) We call a feasible point $\bar{x}$ of problem (P) a stationary point if there exists a (normal) multiplier $\lambda \in \mathbb{R}^q$ such that

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

The above stationary condition is a Karush–Kuhn–Tucker (KKT) type necessary optimality condition. It holds at a locally optimal solution only under certain constraint qualifications. The Fritz John type necessary optimality condition, however, holds without constraint qualification but has a multiplier attached to the objective function. When the multiplier corresponding to the objective function in the Fritz John type necessary optimality condition vanishes, a multiplier is called an abnormal multiplier (see Clarke [19, p. 235]). The Fritz John type necessary optimality condition is equivalent to saying that at an optimal solution, either there exists a normal multiplier or there is a nonzero abnormal multiplier.Hence the nonexistence of a nonzero abnormal multiplier would imply the existence of a normal multiplier. Therefore the following commonly used constraint qualification, under which any locally optimal solution of problem (P) is a stationary point, follows naturally from the Fritz John type necessary optimality condition [19, Theorem 6.1.1].

Definition 2.2 (*NNAMCQ*) We say that the no nonzero abnormal multiplier constraint qualification (NNAMCQ) holds at a feasible point $\bar{x}$ of problem (P) if

$$0 \in \sum_{i \in I(\bar{x})} \lambda_i \partial g_i(\bar{x}) + \sum_{j=p+1}^{q} \lambda_j \partial h_j(\bar{x}) \quad \text{and} \quad \lambda_i \geq 0, \quad i \in I(\bar{x}) \Longrightarrow \lambda_i = 0, \quad \lambda_j = 0. \tag{2.1}$$

When condition (2.1) holds, we say that the vectors

$$\{v_i, i \in I(\bar{x}), v_{p+1}, \ldots, v_q\}$$

where $v_i \in \partial g_i(\bar{x})(i \in I(\bar{x})), v_j \in \partial h_j(\bar{x})(j = p+1, \ldots, q)$ are positively linearly independent. Jourani [32] showed that the NNAMCQ is equivalent to the GMFCQ to be defined as follows.

Definition 2.3 (*GMFCQ*) A feasible point $\bar{x}$ is said to satisfy the generalized Mangasarian-Fromovitz constraint qualification (GMFCQ) for problem (P) if

(i) $v_{p+1}, \ldots, v_q$ are linearly independent, where $v_j \in \partial h_j(\bar{x}),\ j = p+1, \ldots, q$;
(ii) there exists a direction $d \in \mathbb{R}^n$ such that

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

Although the NNAMCQ and the GMFCQ are equivalent, it is some times easier to verify the NNAMCQ than the GMFCQ since verifying the NNAMCQ amounts to verifying the positive linear independence of some vectors. In particular when there is no inequality constraints or there are only two constraints, the positive linear independence is reduced to the linear independence. We will explain this point using the examples in Sect. 4.

In order to accommodate infeasible accumulation points in the numerical algorithm, we now extend the NNAMCQ and the GMFCQ to infeasible points. Note that when $\bar{x}$ is feasible, ENNAMCQ and EGMFCQ reduce to NNAMCQ and GMFCQ respectively.

Definition 2.4 (*ENNAMCQ*) We say that the extended no nonzero abnormal multiplier constraint qualification (ENNAMCQ) holds at $\bar{x} \in \mathbb{R}^n$ if

$$0 \in \sum_{i=1}^{p} \lambda_i \partial g_i(\bar{x}) + \sum_{j=p+1}^{q} \lambda_j \partial h_j(\bar{x}) \text{ and } \lambda_i \geq 0, \quad i = 1, \dots, p,$$
$$\sum_{i=1}^{p} \lambda_i g_i(\bar{x}) + \sum_{j=p+1}^{q} \lambda_j h_j(\bar{x}) \geq 0$$

imply that $\lambda_i = 0, \lambda_j = 0$.

Definition 2.5 (*EGMFCQ*) A point $\bar{x} \in \mathbb{R}^n$ is said to satisfy the extended generalized Mangasarian Fromovitz constraint qualification (EGMFCQ) for problem (P) if

(i) $v_{p+1}, \dots, v_q$ are linearly independent, where $v_j \in \partial h_j(\bar{x}),\ j = p+1, \dots, q$,
(ii) there exists a direction d such that

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

Since the ENNAMCQ and the EGMFCQ may be too strong for some problems to hold, in [51] we have proposed two weaker constraint qualifications. These two new conditions are defined for the nonsmooth problem (P) relatively with smoothing functions as to be defined next.

Definition 2.6 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 $\lim_{z \to x,\ \rho \uparrow \infty} g_\rho(z) = g(x)$ for any fixed $x \in \mathbb{R}^n$.

In order to guarantee the convergence to a stationary point, the smoothing function is required to have the following property.

Definition 2.7 [18] We say that a family of smoothing functions $\{g_\rho : \rho > 0\}$ satisfies the gradient consistency property if $\limsup_{z \to x,\ \rho \uparrow \infty} \nabla g_\rho(z)$ is nonempty and $\limsup_{z \to x,\ \rho \uparrow \infty} \nabla g_\rho(z) \subseteq \partial g(x)$ for any $x \in \mathbb{R}^n$, where $\limsup_{z \to x,\ \rho \uparrow \infty} \nabla g_\rho(z)$ denotes the set of all limiting points

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

Rockafellar and Wets [45, Example 7.19 and Theorem 9.67] showed that for any locally Lipschitz function g, one can always find a family of smoothing functions of g with the gradient consistency property by the integral convolution:

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

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 $\rho \uparrow \infty$. What is more, there are many other smoothing functions with the gradient consistency property which are not generated by the integral-convolution with bounded supports. The reader is referred to [12,14,16,17] for more details.

Using the smoothing technique, one approximates the locally Lipschitz functions $f(x), g_i(x), i = 1, \ldots, p$ and $h_j(x), j = p+1, \ldots, q$ by families of smoothing functions $\{f_\rho(x) : \rho > 0\}, \{g^i_\rho(x) : \rho > 0\}, i = 1, \ldots, p$ and $\{h^j_\rho(x) : \rho > 0\}, j = p+1, \ldots, q$ which satisfy the gradient consistency property. Based on the sequence of iteration points generated by the smoothing SQP algorithm, in [51] we defined the new conditions as follows:

Definition 2.8 (*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 $\{g^i_\rho(x) : \rho > 0\}, i = 1, \ldots, p, \{h^j_\rho(x) : \rho > 0\}, j = p+1, \ldots, q$ holds at $\bar{x}$ provided that

$$0 = \sum_{i \in I(\bar{x})} \lambda_i v_i + \sum_{j=p+1}^{q} \lambda_j v_j \quad \text{and} \quad \lambda_i \geq 0, \quad i \in I(\bar{x}) \Longrightarrow \quad \lambda_i = 0, \lambda_j = 0,$$

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

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

Definition 2.9 (*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 $\{g^i_\rho(x) : \rho > 0\}, i = 1, \ldots, p, \{h^j_\rho(x) : \rho > 0\}, j = p+1, \ldots, q$ holds at $\bar{x}$ provided 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

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

(i) $v_{p+1}, \ldots, v_q$ are linearly independent;
(ii) there exists a direction d such that

$$v_i^T d < 0, \quad \text{for all } i \in I(\bar{x}),$$
$$v_j^T d = 0, \quad \text{for all } j = p+1, \ldots, q.$$

The WNNAMCQ and the WGMFCQ can be extended to infeasible points [51].

Definition 2.10 (*EWNNAMCQ*) Let $\{x_k\}$ be a sequence of iteration points for problem (P) and $\rho_k \uparrow \infty$ as $k \to \infty$. Let $\bar{x}$ 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 $\{g^i_\rho(x) : \rho > 0\}, i = 1, \ldots, p, \{h^j_\rho(x) : \rho > 0\}, j = p+1, \ldots, q$ holds at $\bar{x}$ provided that

$$0 = \sum_{i=1}^{p} \lambda_i v_i + \sum_{j=p+1}^{q} \lambda_j v_j \quad \text{and} \quad \lambda_i \geq 0, \quad i = 1, \ldots, p, \tag{2.2}$$

$$\sum_{i=1}^{p} \lambda_i g_i(\bar{x}) + \sum_{j=p+1}^{q} \lambda_j h_j(\bar{x}) \geq 0. \tag{2.3}$$

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

$$v_i = \lim_{k\to\infty, k\in K_0} \nabla g^i_{\rho_k}(x_k), \quad i = 1, \ldots, p,$$
$$v_j = \lim_{k\to\infty, k\in K_0} \nabla h^j_{\rho_k}(x_k), \quad j = p+1, \ldots, q.$$

Definition 2.11 (*EWGMFCQ*) Let $\{x_k\}$ be a sequence of iteration points for problem (P) and $\rho_k \uparrow \infty$ as $k \to \infty$. Let $\bar{x}$ 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 $\{g^i_\rho(x) : \rho > 0\}, i = 1, \ldots, p, \{h^j_\rho(x) : \rho > 0\}, j = p+1, \ldots, q$ 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

$$v_i = \lim_{k\to\infty, k\in K_0} \nabla g^i_{\rho_k}(x_k), \quad i = 1, \ldots, p,$$
$$v_j = \lim_{k\to\infty, k\in K_0} \nabla h^i_{\rho_k}(x_k), \quad j = p+1, \ldots, q,$$

(i) $v_{p+1}, \ldots, v_q$ are linearly independent;
(ii) there exists a nonzero direction $d \in \mathbb{R}^n$ such that

$$g_i(\bar{x}) + v_i^T d < 0, \quad \text{for all} \quad i = 1, \ldots, p, \tag{2.4}$$
$$h_j(\bar{x}) + v_j^T d = 0, \quad \text{for all} \quad j = p+1, \ldots, q. \tag{2.5}$$

It was showed in [51] that the EWGMFCQ is equivalent to the EWNNAMCQ.

3 Smoothing augmented Lagrangian method

In this section, we propose a smoothing augmented Lagrangian algorithm and show its convergence.

For each smoothing parameter $\rho > 0$, we use the PHR augmented Lagrangian function to define the smoothing augmented Lagrangian function as follows:

$$G^{\lambda,c}_\rho(x) := f_\rho(x) + \frac{1}{2c} \sum_{i=1}^{p} \left(\max\{0, \lambda_i + c g^i_\rho(x)\}^2 - \lambda_i^2 \right)$$
$$+ \sum_{j=p+1}^{q} \left(\lambda_j h^j_\rho(x) + \frac{c}{2} (h^j_\rho(x))^2 \right).$$

For each $\rho > 0, c > 0, \lambda \in \mathbb{R}^q$, we consider the following penalized problem:

$$(\mathrm{P}^{\lambda,c}_\rho) \qquad \min_x \quad G^{\lambda,c}_\rho(x).$$

In the algorithm, we denote the residual function measuring the infeasibility and the complementarity by

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

Since $(\mathrm{P}_\rho^{\lambda,c})$ is a smooth unconstrained optimization problem for each fixed $\rho > 0, c > 0, \lambda \in \mathbb{R}^q$, we suggest to use a gradient descent method to find a stationary point of the problem. Then we increase the smoothing parameter ρ, update the multiplier λ and increase the penalty parameter c provided that the residual $\sigma_\rho^\lambda(x)$ has sufficiently decreased.

We will show that any sequence of the 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 penalty parameters.

We propose the following smoothing augmented Lagrangian algorithm.

Algorithm 3.1 *Let $\{\beta, \sigma_1\}$ be constants in $(0, 1)$ and $\{\hat{\eta}, \sigma\}$ be constants in $(1, \infty)$. Let $\{\varepsilon_k\}$ be a positive sequence converging to 0 and σ_k be a sequence approaching $+\infty$ with $\sigma_k\varepsilon_k \to \infty$ as $k \to \infty$. Choose an initial point $x^0 = x^1 \in \mathbb{R}^n$, an initial smoothing parameter $\rho_0 > 0$, an initial penalty parameter $c_0 > 0$ and an initial multiplier $\lambda^0 \in \mathbb{R}^q$. Let constants $\lambda_{min} < 0$ and $\lambda_{max} > 0$ and take $\bar{\lambda}^0$ as the Euclidean projection of λ^0 onto $\bigotimes_{i=1}^p[0, \lambda_{max}] \times \bigotimes_{i=p+1}^q[\lambda_{min}, \lambda_{max}]$. Set $k := 0$.*

Let $\lambda_i^1 = \max\{0, \bar{\lambda}_i^0 + c_0 g_{\rho_0}^i(x^1)\},\ i = 1,\ldots,p;\ \lambda_j^1 = \bar{\lambda}_j^0 + c_0 h_{\rho_0}^j(x^1),\ j = p+1,\ldots,q$ and take $\bar{\lambda}^1$ as the Euclidean projection of λ^1 onto $\bigotimes_{i=1}^p[0, \lambda_{max}] \times \bigotimes_{i=p+1}^q[\lambda_{min}, \lambda_{max}]$.

1. *If $\nabla G_{\rho_k}^{\bar{\lambda}^k,c_k}(x^{k+1}) = 0$ and $\sigma_{\rho_k}^{\lambda^{k+1}}(x^{k+1}) = 0$, terminate. Otherwise, set $k = k+1$, and go to Step 2.*
2. *Compute $d_k := -\nabla G_{\rho_k}^{\bar{\lambda}^k,c_k}(x^k)$. Let $x^{k+1} := x^k + \alpha_k d_k$, where $\alpha_k := \beta^{l_k}, l_k \in \{0, 1, 2\ldots\}$ is the smallest number satisfying*
$$G_{\rho_k}^{\bar{\lambda}^k,c_k}(x^{k+1}) - G_{\rho_k}^{\bar{\lambda}^k,c_k}(x^k) \le -\alpha_k\sigma_1\|\nabla G_{\rho_k}^{\bar{\lambda}^k,c_k}(x^k)\|^2. \tag{3.1}$$
If
$$\|\nabla G_{\rho_k}^{\bar{\lambda}^k,c_k}(x^{k+1})\| < \hat{\eta}\rho_k^{-1}, \tag{3.2}$$
set $\rho_{k+1} := \sigma\rho_k$ and go to Step 3. Otherwise, set $k = k+1$ and repeat Step 2.
3. *Set*
$$\lambda_i^{k+1} = \max\{0, \bar{\lambda}_i^k + c_k g_{\rho_k}^i(x^{k+1})\}, \quad i = 1,\ldots,p; \tag{3.3}$$
$$\lambda_j^{k+1} = \bar{\lambda}_j^k + c_k h_{\rho_k}^j(x^{k+1}), \quad j = p+1,\ldots,q. \tag{3.4}$$
Take $\bar{\lambda}^{k+1}$ as the Euclidean projection of λ^{k+1} onto $\bigotimes_{i=1}^p[0, \lambda_{max}] \times \bigotimes_{i=p+1}^q$ $[\lambda_{min}, \lambda_{max}]$, and go to Step 4.
4. *If*
$$\sigma_{\rho_k}^{\lambda^{k+1}}(x^{k+1}) < \varepsilon_k, \tag{3.5}$$
go to Step 1. Otherwise, set $c_{k+1} := \sigma_{k+1} + c_k$, $k = k+1$ and go to Step 2.

To prove the convergence theorems we need the following lemma.

Lemma 3.1 *Suppose that Algorithm* 3.1 does not terminate within finite iterations. Let $\{x^k\}$ be a sequence generated by Algorithm 3.1 which has an accumulation point. Then there exists an infinite subset $K \subseteq \mathbf{N}$ such that condition (3.2) holds for each $k \in K$ and hence $\lim_{k\to\infty} \rho_k = +\infty$.

Proof Assume for a contradiction that for any large k, the condition (3.2) fails and thus there exist $\bar{k}$, $\bar{\rho}$, $\bar{\lambda}$ and $\bar{c}$ such that when $k \geq \bar{k}$, $\rho_k = \bar{\rho}$, $\bar{\lambda}_k = \bar{\lambda}$ and $c_k = \bar{c}$.

We first show that there exists an infinite subset $K_1 \subseteq \mathbf{N}$ such that

$$\lim_{k\to\infty, k\in K_1} \|\nabla G_{\bar{\rho}}^{\bar{\lambda},\bar{c}}(x^k)\| = 0. \tag{3.6}$$

To the contrary, suppose that there exists $\varepsilon > 0$ such that for sufficiently large $k > \bar{k}$,

$$\|\nabla G_{\bar{\rho}}^{\bar{\lambda},\bar{c}}(x^k)\| > \varepsilon.$$

From (3.1), we have that for all $k > \bar{k}$

$$G_{\bar{\rho}}^{\bar{\lambda},\bar{c}}(x^{k+1}) - G_{\bar{\rho}}^{\bar{\lambda},\bar{c}}(x^k) \leq -\alpha_k \sigma_1 \|\nabla G_{\bar{\rho}}^{\bar{\lambda},\bar{c}}(x^k)\|^2 < -\alpha_k \sigma_1 \varepsilon^2. \tag{3.7}$$

Since the Armijo line search in Step 2 only requires a small number of iterations, α_k will never approach to 0. It follows from (3.7) that $G_{\bar{\rho}}^{\bar{\lambda},\bar{c}}(x^k) \to -\infty$ as $k \to \infty$. But this is impossible since $G_{\bar{\rho}}^{\bar{\lambda},\bar{c}}(x)$ is continuous and $\{x^k\}$ has an accumulation point x^*. Therefore, there exists an infinite subset $K := \{k-1 : k \in K_1\}$ such that the condition (3.2) holds for each $k \in K$. By the updating rule in Algorithm 3.1 we must have $\lim_{k\to\infty} \rho_k = +\infty$. □

Theorem 3.1 *Suppose 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 the updating rule of c_k in the algorithm, the boundedness of $\{c_k\}$ is equivalent to saying that condition (3.5) holds for sufficiently large k and thus $\lim_{k\to\infty} \sigma_{\rho_k}^{\lambda^{k+1}}(x^{k+1}) = 0$. It follows from the definition of $\sigma_\rho^\lambda(\cdot)$ that $|h_{\rho_k}^j(x^{k+1})| < \varepsilon_k$, $j = p+1, \ldots, q$ and $g_{\rho_k}^i(x^{k+1}) < \varepsilon_k$, $i = 1, \ldots, p$ for sufficiently large k. Thus $\{\lambda^k\}$ is bounded from the updating rule (3.3) and (3.4).

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

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

By calculation, we have

$$\nabla G_{\rho_k}^{\bar{\lambda}^k, c_k}(x^{k+1}) = \nabla f_{\rho_k}(x^{k+1}) + \sum_{i=1}^{p} \mu_{\rho_k,i}^{\bar{\lambda}^k, c_k}\left(x^{k+1}\right) \nabla g_{\rho_k}^i\left(x^{k+1}\right)$$
$$+ \sum_{j=p+1}^{q} \mu_{\rho_k,j}^{\bar{\lambda}^k, c_k}\left(x^{k+1}\right) \nabla h_{\rho_k}^j\left(x^{k+1}\right). \tag{3.8}$$

From the definition of $\mu_\rho^{\lambda,c}(\cdot)$ and the updating rule (3.3)–(3.4), we have $\mu_{\rho_k,i}^{\bar\lambda^k,c_k}(x^{k+1}) = \lambda_i^{k+1}$, $i = 1,\ldots,p$ and $\mu_{\rho_k,j}^{\bar\lambda^k,c_k}(x^{k+1}) = \lambda_j^{k+1}$, $j = p+1,\ldots,1$. Note that, by the boundedness of $\{\lambda^k\}$, there is a subsequence $K_0 \subseteq \mathbf{N}$ such that $\{\lambda^k\}_{K_0}$ is convergent. Let $\lambda^* := \lim_{k\to\infty,\, k\in K_0} \lambda^k$.

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

$$\lim_{k\to\infty,\, k\in\bar K_0} \nabla f_{\rho_k}(x_k) \in \partial f(x^*),$$
$$\lim_{k\to\infty,\, k\in\bar K_0} \nabla g_{\rho_k}^i(x_k) \in \partial g_i(x^*), \quad i = 1,\ldots,p,$$
$$\lim_{k\to\infty,\, k\in\bar K_0} \nabla h_{\rho_k}^j(x_k) \in \partial h_j(x^*), \quad j = p+1,\ldots,q.$$

From Lemma 3.1 and condition (3.2), we know that $\lim_{k\to\infty} \nabla G_{\rho_k}^{\bar\lambda^k,c_k}(x^{k+1}) = 0$. Taking limits in (3.8) as $k\to\infty$, $k\in\bar K_0$, by the gradient consistency properties, we have

$$0 \in \partial f(x^*) + \sum_{i=1}^{p} \lambda_i^* \partial g_i(x^*) + \sum_{j=p+1}^{q} \lambda_j^* \partial h_j(x^*). \tag{3.9}$$

The feasibility of x^* follows from taking the limits in $\lim_{k\to\infty} \sigma_{\rho_k}^{\lambda^{k+1}}(x^{k+1}) = 0$. It follows from (3.3) that $\lambda_i^* \geq 0, i = 1,\ldots,p$. 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, $i = 1,\ldots,p$. Then $\lambda_i^* = \lim_{k\to\infty,\, k\in\bar K_0} \lambda_i^{k+1} = 0$ since $\lim_{k\to\infty} \sigma_{\rho_k}^{\lambda^{k+1}}(x^{k+1}) = 0$. Therefore x^* is a stationary point of problem (P) and the proof of the theorem is complete. □

Theorem 3.2 *Suppose 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 for every $K \subseteq \mathbf{N}$ and x^* such that $\lim_{k\to\infty, k\in K} x_k = x^*$, the EWNNAMCQ holds at x^*, then $\{c_k\}$ is bounded.

Proof Assume for a contradiction that $\{c_k\}$ is unbounded. Since x^* is an arbitrary accumulation point, we assume there exists an infinite index set $\tilde K \subseteq K$ such that condition (3.5) fails for every $k \in \tilde K$ sufficiently large. Then for sufficiently large $k \in \tilde K$, at least one of the following conditions hold:

(a) there exists an index $i_1 \in \{1,\ldots,p\}$ such that $g_{\rho_k}^{i_1}(x^{k+1}) \geq \varepsilon_k$,
(b) there exists an index $j_1 \in \{p+1,\ldots,q\}$ such that $|h_{\rho_k}^{j_1}(x^{k+1})| \geq \varepsilon_k$.

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

$$v_0 := \lim_{k\to\infty,\, k\in\bar K} \nabla f_{\rho_k}(x_k) \in \partial f(x^*),$$
$$v_i := \lim_{k\to\infty,\, k\in\bar K} \nabla g_{\rho_k}^i(x_k) \in \partial g_i(x^*), \quad i = 1,\ldots,p,$$
$$v_j := \lim_{k\to\infty,\, k\in\bar K} \nabla h_{\rho_k}^j(x_k) \in \partial h_j(x^*), \quad j = p+1,\ldots,q.$$

From the updating rule of c_k, we have $c_k\varepsilon_k \to +\infty$ as $k\to\infty$. Thus by the definition of $\mu_\rho^{\lambda,c}(\cdot)$, under either case (a) or case (b), we have $\|\mu_{\rho_k}^{\bar\lambda^k,c_k}(x^{k+1})\| \to \infty$. There exists a subsequence $\bar K_0 \subseteq \bar K$ and $\mu \in \mathbb{R}^q$ nonzero such that

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

It follows from the definition of $\mu_\rho^{\lambda,c}(\cdot)$ that $\mu_i \geq 0,\ i = 1, \ldots, p$.

Similarly as in Theorem 3.1, $\lim_{k\to\infty} \nabla G_{\rho_k}^{\bar{\lambda}^k, c_k}(x^{k+1}) = 0$. Dividing by $\|\mu_{\rho_k}^{\bar{\lambda}^k, c_k}(x^{k+1})\|$ in both sides of (3.8) and letting $k \to \infty$ in $\bar{K}_0$, we have

$$0 = \sum_{i=1}^{p} \mu_i v_i + \sum_{j=p+1}^{q} \mu_j v_j. \tag{3.10}$$

We now show that

$$\sum_{i=1}^{p} \mu_i g_i(x^*) + \sum_{j=p+1}^{q} \mu_j h_j(x^*) \geq 0 \tag{3.11}$$

and consequently conditions (3.10) and (3.11) contradict with the assumption that the EWNNAMCQ holds. We first show that $\mu_i g_i(x^*) \geq 0$. If $g_i(x^*) < 0$ for certain $i \in \{1, \ldots, p\}$, $g_{\rho_k}^i(x^{k+1}) < 0$ for sufficiently large k since $\{g_\rho^i : \rho > 0\}$ are families of smoothing functions of $g_i, i = 1, \ldots, p$. Thus $\mu_i = \lim_{k\to\infty,\, k\in\bar{K}_0} \mu_{\rho_k, i}^{\bar{\lambda}^k, c_k}(x^{k+1}) = 0$ by the definitions $\mu_\rho^{\lambda,c}(\cdot)$ and the unboundedness of $\{c_k\}$. Consequently we have $\mu_i = 0$ if $g_i(x^*) < 0$, for $i \in \{1, \ldots, p\}$. Next we show that $\mu_j h_j(x^*) \geq 0$. Since $c_k \to \infty$ as $k \to \infty$, we have for sufficiently large $k \in \bar{K}_0$,

$$\bar{\lambda}_j^k h_{\rho_k}^j(x^{k+1}) + c_k (h_{\rho_k}^j(x^{k+1}))^2 > 0.$$

Thus

$$\begin{aligned} \mu_j h_j(x^*) &= \lim_{k\to\infty, k\in\bar{K}_0} \frac{\mu_{\rho_k, j}^{\bar{\lambda}^k, c_k}(x^{k+1})}{\|\mu_{\rho_k}^{\bar{\lambda}^k, c_k}(x^{k+1})\|} h_{\rho_k}^j(x^{k+1}) \\ &= \lim_{k\to\infty, k\in\bar{K}_0} \frac{\bar{\lambda}_j^k h_{\rho_k}^j(x^{k+1}) + c_k (h_{\rho_k}^j(x^{k+1}))^2}{\|\mu_{\rho_k}^{\bar{\lambda}^k, c_k}(x^{k+1})\|} \geq 0 \end{aligned}$$

and hence (3.11) holds. The contradiction shows that $\{c_k\}$ is bounded. □

The following corollary follows immediately from Theorems 3.1 and 3.2.

Corollary 3.1 *Suppose that Algorithm 3.1 does not terminate within finite iterations. If the* EWNNAMCQ *holds at any accumulation point of the sequence* $\{x^k\}$ *generated by Algorithm 3.1, then any accumulation point is a stationary point of problem* (P).

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.

Before presenting numerical examples, we first give some remarks on choosing the parameters in the algorithm. In the Armijo line search, a small σ_1 gives the step-size too much flexibility while increasing σ_1 makes the search for step-size costly. $\sigma_1 = 0.8$ is a good selection from our experience.

Since we would like to increase ρ_k fast in the beginning of the algorithm and then let ρ_k grow slowly to infinity, $\hat{\eta}$ should be a large constant. In practice taking different initial parameter ρ_0 usually leads to similar results. According to our experience, if a relatively large initial parameter ρ_0 is chosen, then a smaller σ should be taken to let ρ_k goes to infinity slowly. We suggest to choose a relatively small ρ_0 to guarantee a faster convergence rate.

λ_{max} and λ_{min} are upper and lower bounds for the projected multiplier $\bar{\lambda}$ and can be chosen arbitrarily. In the algorithm, we select $\lambda_{max} = 10^4$ and $\lambda_{min} = -10^4$. There are different ways for choosing $\{\varepsilon_k\}$ and $\{\sigma_k\}$. In our experiments, unless otherwise specified, we select the sequences as $\varepsilon_k := \frac{\sigma'}{\sqrt{k}}$, $\sigma_k := k^2$ for each k,where $\sigma' > 0$ is a small constant.

In numerical practise, it is impossible to obtain an exact '0', thus we select some small enough $\epsilon > 0$, $\epsilon_1 > 0$ and change the update rule of c_k to the case when

$$\sigma_{\rho_k}^{\lambda^{k+1}}(x^{k+1}) < \max\{\varepsilon_k, \epsilon_1\} \tag{4.1}$$

and terminate the algorithm when

$$\nabla G_{\rho_k}^{\bar{\lambda}^k, c_k}(x^{k+1}) < \epsilon \quad \text{and} \quad \sigma_{\rho_k}^{\lambda^{k+1}}(x^{k+1}) < \epsilon_1.$$

4.1 Illustrative examples for general problems

In this subsection, we illustrate Algorithm 3.1 by two general nonsmooth and nonconvex constrained optimization problems.

The first task in designing a smoothing method is to find a family of smoothing functions with the gradient consistency property for nonsmooth functions involved. Many nonsmooth functions can be considered as a composition of a smooth function with a plus function $(t)_+ := \max\{0, t\}$. Chen and Mangasarian [16] constructed smooth approximations to the plus function by using the integral convolution with density functions as follows. Let $\phi : \mathbb{R} \to \mathbb{R}_+$ be a piecewise continuous density function with $\phi(s) = \phi(-s)$ and $\int_{-\infty}^{+\infty} |s|\phi(s)ds < \infty$. Then $\psi_\mu(t) := \int_{-\infty}^{+\infty}(t - \mu s)_+\phi(s)ds$, with $\mu \downarrow 0$ is a family of smoothing functions of the plus function with the gradient consistency property. Choosing $\phi(s) := \frac{2}{(s^2+4)^{\frac{3}{2}}}$ results in the so-called the Chen–Harker–Kanzow–Smale (CHKS) smoothing function of $(t)_+$ ([15,33,47]):

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

Choosing $\phi(s) := \begin{cases} 0, & \text{if } |s| > \frac{1}{2} \\ 1, & \text{if } |s| \le \frac{1}{2}, \end{cases}$ 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:

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

Example 4.1 [23, 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:

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

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

Since the Clarke generalized gradient of the constraint function is

$$\partial g(x,y) = \begin{cases} (\sqrt{2}, 0), & \text{if } \sqrt{2}x > 2y \\ co\{(\sqrt{2}, 0), (0, 2)\}, & \text{if } \sqrt{2}x = 2y \\ (0, 2), & \text{if } \sqrt{2}x < 2y, \end{cases}$$

where *co* denotes the convex hull, we have $(0,0) \notin \partial g(x,y), \forall (x,y)$ which implies that the ENNAMCQ is satisfied at every point in $\mathbb{R}^2$. Our convergent theorem guarantees that any accumulation point of the iteration sequence must be a stationary point.

Rewrite the objective function and the constraint function as

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

Since $(x^2 - y)_+$ is a composition of the smooth function $x^2 - y$ with the plus function and $\nabla(x^2 - y)$ has full rank, by [12, Theorem 4.6], $\psi_\mu^1(x^2 - y)$ is a smoothing function of $(x^2 - y)_+$ with the gradient consistency property. Similarly $\psi_\mu^1(-x^2 + y)$ and $\psi_\mu^1(2y - \sqrt{2}x)$ are smoothing functions of $(-x^2 + y)_+$ and $(2y - \sqrt{2}x)_+$ with the gradient consistency property respectively. Taking $\rho = \frac{1}{4\mu^2}$, it follows that

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

form the family of smoothing functions for $f(x,y)$ and $g(x,y)$ with the gradient consistency property.

In our test, we choose the initial point $(x^0, y^0) = (0.5, 0.3)$ and the parameters $\beta = 0.8$, $\sigma_1 = 0.7$, $\rho_0 = 100$, $c_0 = 100$, $\hat{\eta} = 10^3$, $\sigma = 10, \sigma' = 10^{-3}, \lambda^0 = 100$ and $\epsilon = 10^{-5}, \epsilon_1 = 10^{-6}$. The stopping criteria

$$\nabla G_{\rho_k}^{\bar{\lambda}^k, c_k}(x^{k+1}) < \epsilon \quad \text{and} \quad \sigma_{\rho_k}^{\lambda^{k+1}}(x^{k+1}) < \epsilon_1$$

hold with $(x^{k+1}, y^{k+1}) = (0.7071, 0.500)$, 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:

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

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

Note that $\nabla h(x, y) = (1, -\sqrt{2})$ and

$$\partial g(x, y) = \begin{cases} (2x, 1), & \text{if } y > 0 \\ co\{(2x, 1), (2x, -1)\}, & \text{if } y = 0 \\ (2x, -1), & \text{if } y < 0. \end{cases}$$

For all (x, y) in a sufficiently small neighbourhood of $(\bar{x}, \bar{y})$, $\partial g(x, y) = \{(2x, 1)\}$. For each (x, y) in a sufficiently small neighbourhood of $(\bar{x}, \bar{y})$, the vectors $\{(2x, 1), (1, -\sqrt{2})\}$ are linearly independent, thus the ENNAMCQ holds. Our convergent theorem guarantees that any accumulation point of the iteration sequence which is in a sufficiently small neighbourhood of $(\bar{x}, \bar{y})$ must be a stationary point.

Since $|x^2 - y|$ is a composition of the smooth function $x^2 - y$ with the absolute value function and $\nabla(x^2 - y)$ has full rank, by [12, Theorem 4.6], $\psi^3_\mu(x^2 - y)$ is a smoothing function of $|x^2 - y|$ with the gradient consistency property. Similarly $\psi^3_\mu(y)$ is a smoothing function of $|y|$ with the gradient consistency property respectively. Taking $\mu = \frac{1}{\rho}$, it follows that

$$f_\rho(x, y) := 8\psi^3_\rho(x^2 - y) + (1 - x)^2,$$
$$g_\rho(x, y) := x^2 + \psi^3_\rho(y) - 4$$

form the family of smoothing functions for $f(x, y)$ and $g(x, y)$ with the gradient consistency property respectively.

In our test, we choose the initial point $(x^0, y^0) = (0.3, 0.2)$ and the parameters $\beta = 0.8$, $\sigma_1 = 0.7$, $\rho_0 = 100$, $c_0 = 100$, $\hat{\eta} = 5 \times 10^3$, $\sigma = 10$, $\sigma' = 10^{-3}$, $\lambda^0 = (100, 100)$ and $\epsilon = 5 \times 10^{-4}$, $\epsilon_1 = 10^{-5}$. The stopping criteria

$$\nabla G^{\bar{\lambda}^k, c_k}_{\rho_k}(x^{k+1}) < \epsilon \quad \text{and} \quad \sigma^{\lambda^{k+1}}_{\rho_k}(x^{k+1}) < \epsilon_1$$

hold with $(x^{k+1}, y^{k+1}) = (0.70711, 0.50001)$, 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

$$\begin{aligned} \text{(SBP)} \qquad \min \quad & F(x, y) \\ \text{s.t.} \quad & g_i(x, y) \le 0, \quad i = 1, \ldots, l, \\ & x \in \mathbb{R}^n, \ y \in S(x), \end{aligned}$$

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

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

where Y is a compact subset of $\mathbb{R}^m$ respectively, $f, F, g_i, i = 1, \ldots, l : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ are continuously differentiable functions and f is twice continuously differentiable in variable y.

The principal-agent problem [37] is one of the most important applications of the simple bilevel program. Applications and recent developments of general bilevel programs where the constraint set Y may depend on x can be found in [5,24,25,46,49].

When the lower level program is convex in variable y, it is common to replace the lower level program by its first order conditions. For a general SBP where the lower level problem may not be convex, Mirrlees [37] pointed out that the first order approach is not valid to solve (SBP) since the true optimal solution may not even be a stationary point of the problem reformulated by the first order approach.

By using the value function of the lower level program defined as $V(x) := \inf_{y\in Y} f(x, y)$, Ye and Zhu [53,54] studied the first order condition for the following equivalent formulation under the partial calmness condition:

$$\text{(VP)} \quad \begin{aligned} \min \quad & F(x, y) \\ \text{s.t.} \quad & f(x, y) - V(x) = 0, \\ & g_i(x, y) \le 0, \quad i = 1, \ldots, l, \\ & x \in \mathbb{R}^n, \quad y \in Y. \end{aligned} \tag{4.2}$$

However, Ye and Zhu [55] illustrated that the partial calmness condition is still too strong to hold for many bilevel problems (for example, the Mirrlees' problem) and proposed a new first order necessary optimality condition by considering the combined program with both the first order condition of the lower level problem and the value function constraint. For a general bilevel program, the lower level problem may have equality and/or inequality constraints and the first order condition is the KKT condition. If the lower level has inequality constraints, then the resulting combined program is a nonsmooth mathematical program with complementarity constraints and necessary conditions of Clarke, Mordukhovich and Strong (C, M, S) type have been studied in Ye and Zhu [55] and Ye [52]. For the simple bilevel program we consider in this paper the constraint of the lower level problem is a fixed set Y independent of x. If Y can be represented by some equality and/or inequality constraints, then one can use the KKT condition as the first order condition and the resulting combined program can be studied using the result of previous section. However to concentrate the main idea and simplify the exposition we assume that for any accumulation point $(\bar{x}, \bar{y})$, $\bar{y}$ lies in the interior of the set Y. Moreover this assumption is not too strong since in practice usually one has some idea on where the optimal solution lies and can enlarge the set Y so that the optimal solution lies in the interior of Y. Taking this into consideration we should try to find the stationary point of the following combined program:

$$\text{(CP)} \quad \begin{aligned} \min \quad & F(x, y) \\ \text{s.t.} \quad & f(x, y) - V(x) \le 0, \\ & g_i(x, y) \le 0, \quad i = 1, \ldots, l, \\ & 0 = \nabla_y f(x, y) \\ & x \in \mathbb{R}^n, \quad y \in Y. \end{aligned}$$

Since the value function $V(x)$ is usually nonsmooth, the problems (VP) and (CP) are both nonsmooth and nonconvex optimization problems. Note that for the stationary point whose y component lies in the interior of set Y, the constraint $y \in Y$ can be ignored and hence it is a problem of the type we study in this paper. To develop a numerical algorithm for problem (VP), Lin et al. [36] proposed to approximate the value function by its integral entropy function:

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

and proved that $\gamma_\rho(x)$ satisfies the gradient consistency property. The advantage of using the integral entropy function to approximate the value function is that we do not need to evaluate $V(x)$ or its Clarke generalized gradient. This means that we do not need to solve the lower level program during the iteration process of our smoothing algorithms.

Xu and Ye [50] proposed a projected smoothing augmented Lagrangian method to solve the combined program (CP). They showed that if the sequence of penalty parameters is bounded, then any accumulation point is a stationary point of the combined program. Although it was observed in [50] that the smoothing augmented Lagrangian method is efficient for the problem (CP), no sufficient conditions were given under which the sequence of penalty parameters is bounded. From the results in Sect. 3, we know that EWNNAMCQ is a sufficient condition under which the sequence of penalty parameters for problem (CP) is bounded. In the rest of this subsection, we apply Algorithm 3.1 to the combined program (CP), and verify that the WNNAMCQ holds for all bilevel programs presented here except one.

Example 4.3 (*Mirrlees' problem*) [37] Consider Mirrlees' problem

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

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

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

It was shown in [37] 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 $(x^0, y^0) = (0.5, 0.6)$ and the parameters $\beta = 0.8,\ \sigma_1 = 0.7,\ \rho_0 = 100,\ c_0 = 100,\ \hat{\eta} = 10^3,\ \sigma = 10, \sigma' = 10^{-3}, \lambda^0 = (100, 100)$ and $\epsilon = 10^{-5}, \epsilon_1 = 5 \times 10^{-4}$. The stopping criteria

$$\nabla G_{\rho_k}^{\bar{\lambda}^k, c_k}(x^{k+1}, y^{k+1}) < \epsilon \quad \text{and} \quad \sigma_{\rho_k}^{\lambda^{k+1}}(x^{k+1}, y^{k+1}) < \epsilon_1$$

hold with $(x^{k+1}, y^{k+1}) = (1.0004, 0.95749)$. It seems that the sequence converges to $(\bar{x}, \bar{y})$.

Since

$$\begin{aligned} & \nabla f(x^{k+1}, y^{k+1}) - (\nabla \gamma_{\rho_k}(x^{k+1}), 0) = (0.9755, 0), \\ & \nabla(\nabla_y f)(x^{k+1}, y^{k+1}) = (0.08484, 1.7002), \end{aligned}$$

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 WNNAMCQ holds at $(\bar{x}, \bar{y})$ and our algorithm guarantees that $(\bar{x}, \bar{y})$ is a stationary point of (CP).

Example 4.4 [39, Example 3.3]

$$\begin{aligned} \min \quad & F(x, y) := x^2 - y \\ \text{s.t.} \quad & y \in \operatorname*{argmin}_{y \in [0,3]} \{ f(x, y) := ((y - 1 - 0.1x)^2 - 0.5 - 0.5x)^2 \}. \end{aligned}$$

Mitsos et al. [39] found an approximate optimal solution for the problem to be $(\bar{x}, \bar{y}) = (0.2106, 1.799)$.

In our test, we choose the initial point $(x^0, y^0) = (0.3, 1.5)$ and the parameters $\beta = 0.8,\ \sigma_1 = 0.7,\ \rho_0 = 100,\ c_0 = 100,\ \hat{\eta} = 200,\ \sigma = 5, \sigma' = 10^{-2}, \lambda^0 = (100, 100)$ and $\epsilon = 5 \times 10^{-5}, \epsilon_1 = 5 \times 10^{-6}$. We select $\sigma_k := k$ for each k. The stopping criteria

Table 1 Example 4.4

k	(x^k, y^k)	$\nabla f(x^k, y^k) - (\nabla\gamma_{\rho_k}(x^k), 0)$	$\sigma_{\rho_k}^{\lambda^{k+1}}(x^{k+1}, y^{k+1})$	c_k
663	(0.2145255,1.800723)	$(-7.188 \times 10^{-5}, -1.32 \times 10^{-7})$	3.426965×10^{-6}	305
664	(0.2145255,1.800723)	$(-7.193 \times 10^{-5}, -7.557 \times 10^{-8})$	3.426966×10^{-6}	305
665	(0.2145255,1.800723)	$(-7.222 \times 10^{-5}, -9.728 \times 10^{-8})$	3.426966×10^{-6}	305
666	(0.2145255,1.800723)	$(-7.210 \times 10^{-5}, -7.155 \times 10^{-8})$	3.426966×10^{-6}	305
667	(0.2145255,1.800723)	$(-7.233 \times 10^{-5}, -8.46 \times 10^{-8})$	3.426967×10^{-6}	305
668	(0.2145255,1.800723)	$(-7.247 \times 10^{-5}, -7.339 \times 10^{-8})$	3.426967×10^{-6}	305
669	(0.2145255,1.800723)	$(-7.26 \times 10^{-5}, -7.739 \times 10^{-8})$	3.426967×10^{-6}	305
⋮	⋮	⋮	⋮	⋮

$$\nabla G_{\rho_k}^{\bar{\lambda}^k, c_k}(x^{k+1}, y^{k+1}) < \epsilon \quad \text{and} \quad \sigma_{\rho_k}^{\lambda^{k+1}}(x^{k+1}, y^{k+1}) < \epsilon_1$$

hold with $(x^{k+1}, y^{k+1}) = (0.2145255, 1.800723)$.

Table 1 reports results of some iterations. From the table, it seems that $(\bar{x}, \bar{y}) \approx (0.2145255, 1.800723)$ is a limit point of the iteration sequence (x^k, y^k). Since the sequence $\nabla f(x^{k+1}, y^{k+1}) - (\nabla\gamma_{\rho_k}(x^{k+1}), 0)$ tends to 0 as $k \to \infty$, the WNNAMCQ may not hold at $(\bar{x}, \bar{y})$. However we observe that $\sigma_{\rho_k}^{\lambda^{k+1}}(x^{k+1}, y^{k+1})$ also approaches 0. Therefore the feasibility and the complementarity follow the gradient consistency property and the continuity of functions. Moreover from the update rule (4.1), we do not update c_k when $\sigma_{\rho_k}^{\lambda^{k+1}}(x^{k+1}, y^{k+1})$ is sufficiently small and hence c_k is likely to be bounded. Therefore, the limit point should be a stationary point of the bilevel program from Theorem 3.1.

Example 4.5 [38, Example 3.14] The bilevel program

$$\begin{aligned} \min \quad & F(x, y) := (x - \frac{1}{4})^2 + y^2 \\ \text{s.t.} \quad & y \in \underset{y \in [-1,1]}{\operatorname{argmin}} \{ f(x, y) := \frac{y^3}{3} - xy \}. \end{aligned}$$

has the optimal solution point $(\bar{x}, \bar{y}) = (\frac{1}{4}, \frac{1}{2})$ with an objective value of $\frac{1}{4}$.

In our test, we choose the initial point $(x^0, y^0) = (0.3, 0.7)$ and the parameters $\beta = 0.8,\ \sigma_1 = 0.8,\ \rho_0 = 100,\ c_0 = 100,\ \hat{\eta} = 10^3,\ \sigma = 10, \sigma' = 10^{-3}, \lambda^0 = (100, 100)$ and $\epsilon = 10^{-4}, \epsilon_1 = 10^{-5}$. The stopping criteria

$$\nabla G_{\rho_k}^{\bar{\lambda}^k, c_k}(x^{k+1}, y^{k+1}) < \epsilon \quad \text{and} \quad \sigma_{\rho_k}^{\lambda^{k+1}}(x^{k+1}, y^{k+1}) < \epsilon_1$$

hold with $(x^{k+1}, y^{k+1}) = (0.2500, 0.49999)$. It seems that the sequence converges to $(\bar{x}, \bar{y})$.

Since

$$\begin{aligned} &\nabla f(x^{k+1}, y^{k+1}) - (\nabla\gamma_{\rho_k}(x^{k+1}), 0) = (-0.0991, -9.09 \times 10^{-6}), \\ &\nabla(\nabla_y f)(x^{k+1}, y^{k+1}) = (-1, 0.9998), \end{aligned}$$

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 WNNAMCQ holds at $(\bar{x}, \bar{y})$ and our algorithm guarantees that $(\bar{x}, \bar{y})$ is a stationary point of (CP).

Example 4.6 [38, Example 3.20] The bilevel program

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

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 $(x^0, y^0) = (0.3, 0.2)$ and the parameters $\beta = 0.8$, $\sigma_1 = 0.8$, $\rho_0 = 100$, $c_0 = 100$, $\hat{\eta} = 10^3$, $\sigma = 10, \sigma' = 10^{-3}, \lambda^0 = (100, 100)$ and $\epsilon = 10^{-5}, \epsilon_1 = 5 \times 10^{-5}$. The stopping criteria

$$\nabla G_{\rho_k}^{\bar{\lambda}^k, c_k}(x^{k+1}, y^{k+1}) < \epsilon \ \text{ and } \ \sigma_{\rho_k}^{\lambda^{k+1}}(x^{k+1}, y^{k+1}) < \epsilon_1$$

hold with $(x^{k+1}, y^{k+1}) = (0.5000, 0.49997)$. It seems that the sequence converges to $(\bar{x}, \bar{y})$.

Since

$$\begin{aligned} &\nabla f(x^{k+1}, y^{k+1}) - (\nabla \gamma_{\rho_k}(x^{k+1}), 0) = (-1.5, -2.8 \times 10^{-5}), \\ &\nabla(\nabla_y f)(x^{k+1}, y^{k+1}) = (-1, 0.9999), \end{aligned}$$

by continuity of the functions and the gradient consistency properties, it seems 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 WNNAMCQ holds at $(\bar{x}, \bar{y})$. Our algorithm guarantees that $(\bar{x}, \bar{y})$ is a stationary point of (CP).

Acknowledgments The authors would like to thank Xijun Liang for helping with the numerical experiments, Lei Guo for giving suggestions in an earlier version and the anonymous referees for their suggestions.

References

1. ALGENCAN. http://www.ime.usp.br/~egbirgin/tango/
2. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: On augmented Lagrangian methods with general lower-level constraints. SIAM J. Optim. **18**, 1286–1309 (2007)
3. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. Math. Program. Ser. B **111**, 5–32 (2008)
4. Andreani, R., Martínez, J.M., Schuverdt, M.L.: On the relation between the constant positive linear dependence condition and quasinormality constraint qualification. J. Optim. Theory Appl. **125**, 473–483 (2005)
5. Bard, J.F.: Practical Bilevel Optimization: Algorithms and Applications. Kluwer, Boston (1998)
6. Bertsekas, D.P.: Constrained Optimization and Lagrangian Multiplier Methods. Academic Press, New York (1982)
7. Bian, W., Chen, X.: Neural network for nonsmooth, nonconvex constrained minimization via smooth approximation. IEEE Trans. Neural Netw. Learn. Syst. **25**, 545–556 (2014)
8. Birgin, E.G., Castillo, R., Martínez, J.M.: Numerical comparison of augmented Lagrangian algorithms for nonconvex problems. Comput. Optim. Appl. **23**, 101–125 (2002)
9. Birgin, E.G., Fernández, D., Martínez, J.M.: The boundedness of penalty parameters in an augmented Lagrangian method with lower level constraints. Optim. Methods Softw. **27**, 1001–1024 (2012)
10. Boggs, P.T., Kearsley, A.J., Tolle, J.W.: A global convergence analysis of an algorithm for large-scale nonlinear optimization problems. SIAM J. Optim. **9**, 833–862 (1999)
11. Boggs, P.T., Tolle, J.W.: Sequential Quadratic Programming, pp. 1–51. Cambridge University Press, Cambridge (1995)
12. Burke, J.V., Hoheisel, T., Kanzow, C.: Gradient consistency for integral-convolution smoothing functions. Set-Valued Var. Anal. **21**, 359–376 (2013)
13. Byrd, R.H., Tapia, R.A., Zhang, Y.: An SQP augmented Lagrangian BFGS algorithm for constrained optimization. SIAM J. Optim. **2**, 210–241 (1992)

14. Chen, B., Chen, X.: A global and local superlinear continuation-smoothing method for P_0 and R_0 NCP or monotone NCP. SIAM J. Optim. **9**, 624–645 (1999)
15. Chen, B., Harker, P.T.: A non-interior-point continuation method for linear complementarity problems. SIAM J. Matrix Anal. Appl. **14**, 1168–1190 (1993)
16. Chen, C., Mangasarian, O.L.: A class of smoothing functions for nonlinear and mixed complementarity problems. Math. Program. **71**, 51–70 (1995)
17. Chen, X.: Smoothing methods for nonsmooth, nonconvex minimization. Math. Program. **134**, 71–99 (2012)
18. Chen, X., Womersley, R.S., Ye, J.J.: Minimizing the condition number of a Gram matrix. SIAM J. Optim. **21**, 127–148 (2011)
19. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley-Interscience, New York (1983)
20. Clarke, F.H., Ledyaev, YuS, Stern, R.J., Wolenski, P.R.: Nonsmooth Analysis and Control Theory. Springer, New York (1998)
21. Conn, A.R., Gould, N.I.M., Toint, PhL: A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bound. SIAM J. Numer. Anal. **28**, 545–572 (1991)
22. Conn, A.R., Gould, N.I.M., Toint, PhL: Trust Region Methods, MPS/SIAM Series on Optimization. SIAM, Philadelphia (2000)
23. Curtis, F.E., Overton, M.L.: A sequential quadratic programming algorithm for nonconvex, nonsmooth constrained optimization. SIAM J. Optim. **22**, 474–500 (2012)
24. Dempe, S.: Foundations of Bilevel Programming. Kluwer, Dordrecht (2002)
25. Dempe, S.: Annotated bibliography on bilevel programming and mathematical programs with equilibrium constraints. Optimization **52**, 333–359 (2003)
26. Gill, P.E., Murray, W., Saunders, M.A.: SNOPT: an SQP algorithm for large-scale constrained optimization. SIAM J. Optim. **47**, 99–131 (2005)
27. Goldfarb, D., Polyak, R., Scheinberg, K., Yuzefovich, I.: A modified barrier-augmented Lagrangian method for constrained minimization. Comput. Optim. Appl. **14**, 55–74 (1999)
28. Griva, I., Polyak, R.: Primal-dual nonliner rescaling method with dynamic scaling parameter update. Math. Program. Ser. A **106**, 237–259 (2006)
29. Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. **4**, 303–320 (1969)
30. Huang, X.X., Yang, X.Q.: Further study on augmented Lagrangian duality theory. J. Glob. Optim. **31**, 193–210 (2005)
31. Izmailov, A.F., Solodov, M.V., Uskov, E.I.: Global convergence of augmented Lagrangian methods applied to optimization problems with degenerate constraints, including problems with complementarity constraints. SIAM J. Optim. **22**, 1579–1606 (2012)
32. Jourani, A.: Constraint qualifications and Lagrange multipliers in nondifferentiable programming problems. J. Optim. Theory Appl. **81**, 533–548 (1994)
33. Kanzow, C.: Some noninterior continuation methods for linear complementarity problems. SIAM J. Matrix Anal. Appl. **17**, 851–868 (1996)
34. Kort, B.W., Bertsekas, D.P.: A new penalty method for constrained minimization. In: Proceddings of the 1972 IEEE Conference on Decision and Control, New Orleans, pp. 162–166 (1972)
35. LANCELOT. http://www.cse.scitech.ac.uk/nag/lancelot/lancelot.shtml
36. Lin, G.H., Xu, M., Ye, J.J.: On solving simple bilevel programs with a nonconvex lower level program. Math. Program. Ser. A **144**, 277–305 (2014)
37. Mirrlees, J.: The theory of moral hazard and unobservable behaviour: part I. Rev. Econ. Stud. **66**, 3–22 (1999)
38. Mitsos, A., Barton, P.: A test set for bilevel programs. Technical Report, Massachusetts Institute of Technology (2006)
39. Mitsos, A., Lemonidis, P., Barton, P.: Global solution of bilevel programs with a nonconvex inner program. J. Glob. Optim. **42**, 475–513 (2008)
40. Nguyen, V.H., Strodiot, J.J.: On the convergence rate of a penalty function method of exponential type. J. Optim. Theory Appl. **27**, 495–508 (1979)
41. Powell, M.J.D.: A method for nonlinear constraints in minimization problems. In: Fletcher, R. (ed.) Optimization, pp. 283–298. Academic Press, London and New York (1969)
42. Qi, L., Wei, Z.: On the constant positive linear dependence condition and its application to SQP methods. SIAM J. Optim. **10**, 963–981 (2000)
43. Rockafellar, R.T.: A dual approach to solving nonlinear programming problems by unconstrained optimization. Math. Program. **5**, 354–373 (1973)
44. Rockafellar, R.T.: Augmented Lagrange multiplier functions and duality in nonconvex programming. SIAM J. Control **12**, 268–285 (1974)
45. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin (1998)

46. Shimizu, K., Ishizuka, Y., Bard, J.F.: Nondifferentiable and Two-Level Mathematical Programming. Kluwer, Boston (1997)
47. Smale, S.: Algorithms for solving equations. In: Proceedings of the International Congress of Mathematicians, Berkeley, CA, pp. 172–195 (1986)
48. Tseng, P., Bertsekas, D.P.: On the convergence of the exponential multiplier method for convex programming. Math. Program. **60**, 1–19 (1993)
49. Vicente, L.N., Calamai, P.H.: Bilevel and multilevel programming: a bibliography review. J. Glob. Optim. **5**, 291–306 (1994)
50. Xu, M., Ye, J.J.: A smoothing augmented Lagrangian method for solving simple bilevel programs. Comput. Optim. Appl. **59**, 353–377 (2013)
51. Xu, M., Ye, J.J., Zhang, L.: Smoothing SQP method for solving degenerate nonsmooth constrained optimization problems with applications to bilevel programs. Preprint, arXiv:1403.1636v1 [Math.OC] (2014)
52. Ye, J.J.: Necessary optimality conditions for multiobjective bilevel programs. Math. Oper. Res. **36**, 165–184 (2011)
53. Ye, J.J., Zhu, D.L.: Optimality conditions for bilevel programming problems. Optimization **33**, 9–27 (1995)
54. Ye, J.J., Zhu, D.L.: A note on optimality conditions for bilevel programming problems. Optimization **39**, 361–366 (1997)
55. Ye, J.J., Zhu, D.L.: New necessary optimality conditions for bilevel programs by combining MPEC and the value function approach. SIAM J. Optim. **20**, 1885–1905 (2010)