SIAM J. OPTIM.
Vol. 25, No. 3, pp. 1388–1410
© 2015 Society for Industrial and Applied Mathematics

SMOOTHING SQP METHODS FOR SOLVING DEGENERATE NONSMOOTH CONSTRAINED OPTIMIZATION PROBLEMS WITH APPLICATIONS TO BILEVEL PROGRAMS*

MENGWEI XU†, JANE J. YE‡, AND LIWEI ZHANG§

Abstract. We consider a degenerate nonsmooth and nonconvex optimization problem for which the standard constraint qualification such as the generalized Mangasarian–Fromovitz constraint qualification (GMFCQ) may not hold. We use smoothing functions with the gradient consistency property to approximate the nonsmooth functions and introduce a smoothing sequential quadratic programming (SQP) algorithm under the l_∞ penalty framework. We show that any accumulation point of a selected subsequence of the iteration sequence generated by the smoothing SQP algorithm is a Clarke stationary point, provided that the sequence of multipliers and the sequence of penalty parameters are bounded. Furthermore, we propose a new condition called the weakly generalized Mangasarian–Fromovitz constraint qualification (WGMFCQ) that is weaker than the GMFCQ. We show that the extended version of the WGMFCQ guarantees the boundedness of the sequence of multipliers and the sequence of penalty parameters and thus guarantees the global convergence of the smoothing SQP algorithm. We demonstrate that the WGMFCQ can be satisfied by bilevel programs for which the GMFCQ never holds. Preliminary numerical experiments show that the algorithm is efficient for solving degenerate nonsmooth optimization problems such as the simple bilevel program.

Key words. nonsmooth optimization, constrained optimization, smoothing function, sequential quadratic programming algorithm, bilevel program, constraint qualification

AMS subject classifications. 65K10, 90C26, 90C30

DOI. 10.1137/140971580

1. Introduction.

In this paper, we consider the constrained optimization problem of the form

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

where the objective function and constraint functions $f, g_i\,(i = 1, \dots, p)$, $h_j\,(j = p+1, \dots, q) : \mathbb{R}^n \to \mathbb{R}$ are locally Lipschitz. In particular, our focus is on solving a degenerate problem for which the generalized Mangasarian–Fromovitz constraint qualification (GMFCQ) may not hold at a stationary point.

The sequential quadratic programming (SQP) method is one of the most effective methods for solving smooth constrained optimization problems. For the current iteration point x_k, the basic idea of the SQP method is to generate a descent direction

*Received by the editors June 4, 2014; accepted for publication (in revised form) May 8, 2015; published electronically July 14, 2015.
http://www.siam.org/journals/siopt/25-3/97158.html
†Department of Mathematics, School of Science, Tianjin University, Tianjin, 300072, China (xumengw@hotmail.com).
‡Department of Mathematics and Statistics, University of Victoria, Victoria V8W 2Y2, BC, Canada (janeye@uvic.ca). The research of this author was partially supported by NSERC.
§School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China (lwzhang@dlut.edu.cn). The research of this author was supported by the National Natural Science Foundation of China under project 91330206.

d_k by solving the following quadratic programming problem:

$$\begin{aligned}\min_{d}\ & \nabla f(x_k)^T d + \frac{1}{2} d^T W_k d \\ \text{s.t.}\ & g_i(x_k) + \nabla g_i(x_k)^T d \le 0,\ i = 1, \ldots, p, \\ & h_j(x_k) + \nabla h_j(x_k)^T d = 0,\ j = p+1, \ldots, q,\end{aligned}$$

where $\nabla f(x)$ denotes the gradient of function f at x and W_k is a symmetric positive definite matrix that approximates the Hessian matrix of the Lagrangian function. Then d_k is used to generate the next iteration point: $x_{k+1} := x_k + \alpha_k d_k$, where the stepsize α_k is chosen to yield a sufficient decrease of a suitable merit function. The SQP algorithm with $\alpha_k = 1$ was first studied by Wilson in [44], where the exact Hessian matrix of the Lagrangian function was used as W_k. Garcia-Palomares and Mangasarian [19] proposed to use an estimate to approximate the Hessian matrix. Han [21] proposed to update the matrix W_k by the Broyden–Fletcher–Goldfarb–Shanno (BFGS) formula. When the stepsize $\alpha_k = 1$, the convergence is only local. To obtain a global convergence, Han [22] proposed to use the classical l_1 penalty function as a merit function to determine the step size. While the l_1 penalty function is not differentiable, the authors of [36] suggested using the augmented Lagrangian function, which is a smooth function as a merit function. The inconsistency of the system of the linearized constraints is a serious limitation of the SQP method. Several techniques have been introduced to deal with the possible inconsistency. For example, Pantoja and Mayne [35] proposed to replace the standard SQP subproblem by the following penalized SQP subproblem:

$$\begin{aligned}\min_{d,\xi}\ & \nabla f(x_k)^T d + \frac{1}{2} d^T W_k d + r_k \xi \\ \text{s.t.}\ & g_i(x_k) + \nabla g_i(x_k)^T d \le \xi,\ i = 1, \ldots, p, \\ & -\xi \le h_j(x_k) + \nabla h_j(x_k)^T d \le \xi,\ j = p+1, \ldots, q, \\ & \xi \ge 0,\end{aligned}$$

where the penalty parameter $r_k > 0$. Unlike the standard SQP subproblem which may not have feasible solutions, the penalized SQP subproblem is always feasible. Other alternative methods for inconsistency of the SQP method are also presented [3, 17, 20, 29, 40, 41, 50]. For nonlinear programs which have some simple bound constraints on some of the variables, Heinkenschloss [23] proposed a projected SQP method which combines the ideas of the projected Newton methods and the SQP method.

Recently Curtis and Overton [12] pointed out that applying SQP methods directly to a general nonsmooth and nonconvex constrained optimization problem will fail in theory and in practice. They employed a process of gradient sampling (GS) method to make the search direction effective in nonsmooth regions and proved that the iteration points generated by the SQP-GS method converge globally to a stationary point of the penalty function with probability one. A smoothing method is a well-recognized technique for numerical solution of a nonsmooth optimization problem. Using a smoothing method, one replaces the nonsmooth function by a suitable smooth approximation, solves a sequence of smooth problems, and drives the approximation closer and closer to the original problem. The fundamental question is as follows: What property should a family of the smoothing functions have in order for the stationary points of the smoothing problems to approach a stationary point of the original problem? In most of the literature, a particular smoothing function is employed for the particular problem studied. It turns out that not all smooth approximations of the nonsmooth function can be used in the smoothing technique to obtain the desired result; an example for which the smoothing method fails to converge with almost all initial points was given by Kummer [26]. Zhang and Chen [49] (see also the recent survey on the subject by Chen [8]) identified the desired property as the gradient consistency property. Zhang and Chen [49] proposed a smoothing projected gradient algorithm for solving optimization problems with a convex set constraint by using a family of smoothing functions with the gradient consistency property to approximate the nonsmooth objective function. They proved that any accumulation point of the iteration sequence is a Clarke stationary point of the original nonsmooth optimization problem. Recently [27, 45] extended the result of [49] to a class of nonsmooth constrained optimization problems using the projected gradient method and the augmented Lagrangian method, respectively. Smoothing functions were proposed and the SQP method was used for the smooth problem in [18, 25] to solve the mathematical programs with complementarity constraints (MPCC) and in [28, 42] to solve the semi-infinite programming (SIP). In this paper we will combine the SQP method and the smoothing technique to design a smoothing SQP method for a class of general constrained optimization problems with smoothing functions satisfying the gradient consistency property.

For the SQP method under a penalty framework to converge globally, usually the set of the multipliers is required to be bounded (see, e.g., [2]). This amounts to saying that the MFCQ is required to hold. For the nonsmooth optimization problem, the corresponding MFCQ is referred to as the GMFCQ. Unfortunately, the GMFCQ is quite strong for certain classes of problems. For example, it is well known by now that the GMFCQ never holds for the bilevel program [46]. Another example of a nonsmooth optimization problem which does not satisfy the GMFCQ is a reformulation of an SIP [28]. In this paper we propose a new constraint qualification that is much weaker than the GMFCQ. We call it the weakly generalized Mangasarian–Fromovitz constraint qualification (WGMFCQ). WGMFCQ is not a constraint qualification in the classical sense. It is defined in terms of the smoothing functions and the sequence of iteration points generated by the smoothing algorithm. In our numerical experiment, WGMFCQ is very easy to satisfy for the bilevel programs.

Both the objective function and the constrained functions may be nonsmooth. We first use some smoothing functions approximating the nonsmooth functions and then consider the robust formulation which is proposed by Pantoja and Mayne. Under the EWGMFCQ, global convergence can be obtained.

The rest of the paper is organized as follows. In section 2, we present preliminaries which will be used in this paper and introduce the new constraint qualification WGMFCQ. In section 3, we consider the smoothing approximations of the original problem and propose the smoothing SQP method under an l_∞ penalty framework. Then we establish the global convergence for the algorithm. In section 4, we apply the smoothing SQP method to bilevel programs. The final section contains some concluding remarks.

We adopt the following standard notation in this paper. For any two vectors a and b in $\mathbb{R}^n$, we denote their inner product by $a^T b$. 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 the interior, the relative interior, the closure, the convex hull, and the distance from x to Ω by $\operatorname{int}\Omega$, $\operatorname{ri}\Omega$, $\operatorname{cl}\Omega$, $\operatorname{co}\Omega$, and $\operatorname{dist}(x, \Omega)$, respectively. 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. Preliminaries and the new constraint qualifications. In this section, we first present some background materials and results which will be used later. We then discuss the issue of constraint qualifications.

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

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

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

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

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

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

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 [10, 11].

For $\bar{x}$, a feasible solution of problem (P), we denote by $I(\bar{x}) := \{i = 1, \dots, p : g_i(\bar{x}) = 0\}$ the active set at $\bar{x}$. The following nonsmooth Fritz John–type multiplier rule holds by Clarke [10, Theorem 6.1.1]) and the nonsmooth calculus (see, e.g., [10]).

THEOREM 2.1 (Fritz John multiplier rule). *Let $\bar{x}$ be a local optimal solution of problem* (P). *Then there exist $r \ge 0$, $\lambda_i \ge 0$ $(i \in I(\bar{x}))$, $\lambda_j \in \mathbb{R}$ $(j = p+1, \dots, q)$ not all zero such that*

$$0 \in r\partial f(\bar{x}) + \sum_{i \in I(\bar{x})} \lambda_i \partial g_i(\bar{x}) + \sum_{j=p+1}^{q} \lambda_j \partial h_j(\bar{x}). \tag{2.1}$$

There are two possible cases in the Fritz John multiplier rule: $r > 0$ or $r = 0$. Let $\bar{x}$ be a feasible solution of problem (P). If the Fritz John condition (2.1) holds with $r > 0$, then we call $\bar{x}$ a (Clarke) stationary point of (P). According to Clarke [10], any multiplier $\lambda \in \mathbb{R}^q$ with $\lambda_i \ge 0$, $i = 1, \dots, p$, satisfying the Fritz John condition (2.1) with $r = 0$ is an abnormal multiplier. From the Fritz John multiplier rule, it is easy to see that if there is no nonzero abnormal multiplier, then any local optimal solution $\bar{x}$ must be a stationary point. Hence it is natural to define the following constraint qualification.

DEFINITION 2.1 (NNAMCQ). *We say that the no nonzero abnormal multiplier constraint qualification (NNAMCQ) holds at a feasible point $\bar{x}$ of problem* (P) *if*

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

It is easy to see that NNAMCQ amounts to saying that any collection of 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. NNAMCQ is equivalent to the generalized MFCQ which was first introduced by Hiriart-Urruty [24].

Definition 2.2 (GMFCQ). *A feasible point $\bar{x}$ is said to satisfy the generalized Mangasarian–Fromovitz constraint qualification (GMFCQ) for problem* (P) *if for any given collection of vectors $\{v_i, i \in I(\bar{x}), v_{p+1}, \ldots, v_q\}$, where $v_i \in \partial g_i(\bar{x})$ $(i \in I(\bar{x}))$, $v_j \in \partial h_j(\bar{x})$ $(j = p+1, \ldots, q)$, the following two conditions hold:*

(i) *$v_{p+1}, \ldots, v_q$ are linearly independent.*

(ii) *There exists a direction d such that*

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

In order to accommodate infeasible accumulation points in the numerical algorithm, we now extend the NNAMCQ and the GMFCQ to allow infeasible points. Note that when $\bar{x}$ is feasible, ENNAMCQ and EGMFCQ (see Definitions 2.3 and 2.4) reduce to NNAMCQ and GMFCQ, respectively.

Definition 2.3 (ENNAMCQ). *We say that the extended no nonzero abnormal multiplier constraint qualification (ENNAMCQ) holds at $\bar{x} \in \mathbb{R}^n$ if*

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

imply that $\lambda_i = 0$, $\lambda_j = 0$ for all $i = 1, \ldots, p$, $j = p+1, \ldots, q$.

Definition 2.4 (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 for any given collection of vectors $\{v_i, v_j : i = 1, \ldots, p, \ j = p+1, \ldots, q\}$, where $v_i \in \partial g_i(\bar{x})$, $v_j \in \partial h_j(\bar{x})$, the following two conditions hold:*

(i) *$v_{p+1}, \ldots, v_q$ are linearly independent.*

(ii) *There exists a direction d such that*

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

Note that under the extra assumption that the functions g_i are directionally differentiable, the EGMFCQ coincides with the conditions (B4) and (B5) in [25].

Since the set of the Clarke generalized gradient can be large, the ENNAMCQ and the EGMFCQ may be too strong for some problems to hold. In what follows, we propose two conditions that are much weaker than the ENNAMCQ and the EGMFCQ, respectively. For this purpose, we first recall the definition of smoothing functions.

Definition 2.5. *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 $x \in \mathbb{R}^n$.*

Such sequence $g_\rho(\cdot)$ converges continuously to $g(\cdot)$ as defined in [38].

DEFINITION 2.6 (see [4, 9]). *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* $\limsup_{z\to x,\,\rho\uparrow\infty} \nabla g_\rho(z)$ *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*

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

Note that according to [38, Theorem 9.61 and Corollary 8.47(b)], for a locally Lipschitz function g and its smoothing family $\{g_\rho : \rho > 0\}$, one always has the inclusion

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

Thus our definition of gradient consistency is equivalent to saying that

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

which is the definition used in [5, 8].

It is natural to ask whether one can always find a family of smoothing functions with the gradient consistency property for a locally Lipschitz function. The answer is yes. Rockafellar and Wets [38, Example 7.19 and Theorem 9.67] show that for any locally Lipschitz function g, one can construct 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 $\rho \uparrow \infty$. Although one can always generate a family of smoothing functions with the gradient consistency property by integral convolution with bounded supports, there are many other smoothing functions which are not generated by the integral convolution with bounded supports [5, 6, 7, 8, 32].

Using the smoothing technique, we approximate the locally Lipschitz functions $f(x)$, $g_i(x)$, $i = 1, \dots, p$, and $h_j(x)$, $j = p+1, \dots, q$, by families of smoothing functions $\{f_\rho(x) : \rho > 0\}$, $\{g^i_\rho(x) : \rho > 0\}$, $i = 1, \dots, p$, and $\{h^j_\rho(x) : \rho > 0\}$, $j = p+1, \dots, q$. We also assume that these families of smoothing functions satisfy the gradient consistency property. We use certain algorithms to solve the smooth problem and drive the smoothing parameter ρ to infinity. Based on the sequence of iteration points of the algorithm, we now define the new conditions.

DEFINITION 2.7 (WNNAMCQ). *Let $\{x_k\}$ be a sequence of iteration points for problem* (P)*, and let $\rho_k \uparrow \infty$ 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 $\{g^i_\rho(x) : \rho > 0\}$, $i = 1, \dots, p$, $\{h^j_\rho(x) : \rho > 0\}$, $j = p+1, \dots, q$, holds at $\bar{x}$, provided that for any $K_0 \subset K \subset \mathbf{N}$ such that* $\lim_{k\to\infty, k\in K} x_k = \bar{x}$ *and any collection of vectors $\{v_i(i \in I(\bar{x})), v_j(j = p+1, \dots, q)\}$, where*

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

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

DEFINITION 2.8 (WGMFCQ). *Let $\{x_k\}$ be a sequence of iteration points for problem* (P)*, and let $\rho_k \uparrow \infty$ 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 $\{g_\rho^i(x) : \rho > 0\}$, $i = 1, \ldots, p$, $\{h_\rho^j(x) : \rho > 0\}$, $j = p+1, \ldots, q$, holds at $\bar{x}$, provided that the following conditions hold. For any $K_0 \subset K \subset \mathbf{N}$ such that $\lim_{k\to\infty, k\in K} x_k = \bar{x}$ and any collection of vectors $\{v_i(i \in I(\bar{x})), v_j(j = p+1, \ldots, q)\}$, where $v_i = \lim_{k\to\infty, k\in K_0} \nabla g_{\rho_k}^i(x_k),\ i \in I(\bar{x}),\ v_j = \lim_{k\to\infty, k\in K_0} \nabla h_{\rho_k}^j(x_k),\ j = p+1, \ldots, q,$*

(i) *$v_{p+1}, \ldots, v_q$ are linearly independent;*

(ii) *there exists a direction d such that*

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

We now extend the WNNAMCQ and the WGMFCQ to accommodate infeasible points.

DEFINITION 2.9 (EWNNAMCQ). *Let $\{x_k\}$ be a sequence of iteration points for problem* (P)*, and let $\rho_k \uparrow \infty$ as $k \to \infty$. Let $\bar{x}$ be an 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 $\{g_\rho^i(x) : \rho > 0\}$, $i = 1, \ldots, p$, $\{h_\rho^j(x) : \rho > 0\}$, $j = p+1, \ldots, q$, holds at $\bar{x}$, provided that the following condition holds. For any $K_0 \subset K \subset \mathbf{N}$ such that $\lim_{k\to\infty, k\in K} x_k = \bar{x}$ and any*

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

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

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

imply that $\lambda_i = 0$, $\lambda_j = 0$, $i = 1, \ldots, p$, $j = p+1, \ldots, q$.

DEFINITION 2.10 (EWGMFCQ). *Let $\{x_k\}$ be a sequence of iteration points for problem* (P)*, and let $\rho_k \uparrow \infty$ as $k \to \infty$. Let $\bar{x}$ be an accumulation point of the sequence $\{x_k\}$. We say that the extended weakly generalized Mangasarian–Fromovitz constraint qualification (EWGMFCQ) based on the smoothing functions $\{g_\rho^i(x) : \rho > 0\}$, $i = 1, \ldots, p$, $\{h_\rho^j(x) : \rho > 0\}$, $j = p+1, \ldots, q$, holds at $\bar{x}$, provided that the following conditions hold. For any $K_0 \subset K \subset \mathbf{N}$ such that $\lim_{k\to\infty, k\in K} x_k = \bar{x}$ and any collection of vectors $\{v_i, v_j : i = 1, \ldots, p, j = p+1, \ldots, q\}$, where*

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

(i) *$v_{p+1}, \ldots, v_q$ are linearly independent;*

(ii) *there exists a direction* d *such that*

$$g_i(\bar{x}) + v_i^T d < 0 \quad \forall i = 1, \ldots, p, \tag{2.4}$$

$$h_j(\bar{x}) + v_j^T d = 0 \quad \forall j = p+1, \ldots, q. \tag{2.5}$$

Due to the gradient consistency property, it is easy to see that, in general, the EWNNAMCQ and the EWGMFCQ are weaker than the ENNAMCQ and the EGMFCQ, respectively. We finish this section with an equivalence between the EWGMFCQ and EWNNAMCQ.

THEOREM 2.2. *The following implication always holds:*

$$\text{EWGMFCQ} \Longleftrightarrow \text{EWNNAMCQ}.$$

Proof. We first show that EWGMFCQ implies EWNNAMCQ. To the contrary we suppose that EWGMFCQ holds, but EWNNAMCQ does not hold, which means that there exist scalars $\lambda_i \in \mathbb{R}$, $i = 1, \ldots, q$, not all zero such that conditions (2.2)–(2.3) hold. Suppose that d is the direction that satisfies condition (ii) of EWGMFCQ. Due to the linear independence of $v_{p+1}, \ldots, v_q$ (condition (i) of EWGMFCQ), the scalars λ_i, $i = 1, \ldots, p$, cannot all be equal to zero. Multiplying both sides of condition (2.2) by d, it follows from conditions (2.4) and (2.5) that

$$\begin{aligned} 0 &= \sum_{i=1}^{p} \lambda_i v_i^T d + \sum_{j=p+1}^{q} \lambda_j v_j^T d \\ &< -\sum_{i=1}^{p} \lambda_i g_i(\bar{x}) - \sum_{j=p+1}^{q} \lambda_j h_j(\bar{x}) \le 0, \end{aligned}$$

which is a contradiction. Therefore, EWNNAMCQ holds.

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

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

such that for all directions d, (2.4) or (2.5) fails to hold. Let $A := [v_1, \ldots, v_q]$ be the matrix where $v_1, \ldots, v_q$ are columns and

$$\begin{aligned} S_1 &:= \{z : \exists d \text{ such that } z = A^T d\}, \\ S_2 &:= \{z : z_i < -g_i(\bar{x}),\ i = 1, \ldots, p,\ z_j = -h_j(\bar{x}),\ j = p+1, \ldots, q\}. \end{aligned}$$

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 condition (ii) of EWGMFCQ, from [37, Theorem 11.3], there exists a hyperplane separating S_1 and $\mathrm{cl}S_2$ properly. Since S_1 is a subspace and thus a cone, from [37, Theorem 11.7], there exists a hyperplane separating S_1 and $\mathrm{cl}S_2$ properly and passing through the origin. By the separation theorem (see, e.g., [37, Theorem 11.1]), there exists a vector y such that

$$\begin{aligned} &\inf\{y^T z : z \in S_1\} \ge 0 \ge \sup\{y^T z : z \in \mathrm{cl}S_2\}, \\ &\sup\{y^T z : z \in S_1\} > \inf\{y^T z : z \in \mathrm{cl}S_2\}. \end{aligned} \tag{2.6}$$

From (2.6), we know that $y \neq 0$. Therefore, there exists $y \in \mathbb{R}^q$, $y \neq 0$, such that $y^T z \geq 0$ for all $z \in S_1$ and $y^T z \leq 0$ for all $z \in \mathrm{cl}S_2$.

(a) We first consider the inequality $y^T z \leq 0$ for all $z \in \mathrm{cl}S_2$. By taking $z^0 \in \mathrm{cl}S_2$ such that z_j^0, $j = p+1, \ldots, q$, are constants and $z_i^0 \to -\infty$, $i \in \{1, \ldots, p\}$, we conclude that

$$y_i \geq 0, \quad i = 1, \ldots, p. \tag{2.7}$$

Choosing $z^2 \in \mathrm{cl}S_2$ with $z_i^2 = -g_i(\bar{x}),\ i = 1, \ldots, p,\ z_j^2 = -h_j(\bar{x}),\ j = p+1, \ldots, q$, we have

$$\sum_{i=1}^{p} y_i g_i(\bar{x}) + \sum_{j=p+1}^{q} y_j h_j(\bar{x}) = -y^T z^2 \geq 0. \tag{2.8}$$

(b) We now consider the inequality $y^T z \geq 0$ for all $z \in S_1$. Select an arbitrary d. Then $z^1 := A^T d \in S_1$, $z' := -z^1 = A^T(-d) \in S_1$, and hence

$$\sum_{i=1}^{p} y_i v_i^T d + \sum_{j=p+1}^{q} y_j v_j^T d = y^T z^1 \geq 0,$$

$$\sum_{i=1}^{p} y_i v_i^T (-d) + \sum_{j=p+1}^{q} y_j v_j^T (-d) = y^T z' \geq 0.$$

That is,

$$\sum_{i=1}^{p} y_i v_i + \sum_{j=p+1}^{q} y_j v_j = 0. \tag{2.9}$$

Therefore, if there exists a nonzero vector y such that $y^T z \geq 0$ for all $z \in S_1$ and $y^T z \leq 0$ for all $z \in \mathrm{cl}S_2$, the vector should also satisfy conditions (2.7)–(2.9). However, from the EWNNAMCQ, conditions (2.7)–(2.9) imply that $y = 0$, which is a contradiction. Thus condition (ii) of EWGMFCQ must hold. The proof is therefore complete. $\square$

In the case when there is only one inequality constraint and no equality constraints in problem (P), the EWNNAMCQ and EWGMFCQ at $\bar{x}$ reduce to the following condition: There is no $K_0 \subset K \subset \mathbf{N}$ such that $\lim_{k\to\infty, k\in K} x_k = \bar{x}$ and $\lim_{k\to\infty, k\in K_0} \nabla g_{\rho_k}^1(x_k) \neq 0$. This condition is slightly weaker than a similar condition [28, Assumption (B4)] which requires that there is no $K_0 \subset \mathbf{N}$ such that $\lim_{k\to\infty, k\in K_0} \nabla g_{\rho_k}^1(x_k) \neq 0$.

3. Smoothing SQP method.

In this section we design the smoothing SQP algorithm and prove its convergence.

Suppose that $\{g_\rho^i(x) : \rho > 0\}$ and $\{h_\rho^j(x) : \rho > 0\}$ are families of smoothing functions for g_i, h_j, respectively. Let x_k be the current iterate, and let (W_k, r_k, ρ_k) be current updates of the positive definite matrix, the penalty parameter, and the smoothing parameter, respectively. We will try to find a descent direction of a smoothing merit function by using the smoothing SQP subprogram. In order to overcome the inconsistency of the smoothing SQP subprograms, following Pantoja and Mayne [35], we solve the penalized smoothing SQP subprogram:

$$(\mathrm{QP})_k \qquad \min_{d\in\mathbb{R}^n, \xi\in\mathbb{R}} \ \nabla f_{\rho_k}(x_k)^T d + \frac{1}{2} d^T W_k d + r_k \xi$$

$$\begin{aligned}\text{s.t.}\quad & g^i_{\rho_k}(x_k)+\nabla g^i_{\rho_k}(x_k)^T d\le \xi,\ i=1,\dots,p,\\ & h^j_{\rho_k}(x_k)+\nabla h^j_{\rho_k}(x_k)^T d\le \xi,\ j=p+1,\dots,q,\\ & -h^j_{\rho_k}(x_k)-\nabla h^j_{\rho_k}(x_k)^T d\le \xi,\ j=p+1,\dots,q,\\ & \xi\ge 0.\end{aligned}$$

If (d_k,ξ_k) is a solution of $(\mathrm{QP})_k$, then its Karush–Kuhn–Tucker (KKT) condition can be written as

$$0=\nabla f_{\rho_k}(x_k)+W_kd_k+\sum_{i=1}^{p}\lambda^g_{i,k}\nabla g^i_{\rho_k}(x_k)+\sum_{j=p+1}^{q}(\lambda^+_{j,k}-\lambda^-_{j,k})\nabla h^j_{\rho_k}(x_k), \tag{3.1}$$

$$0=r_k-\left(\sum_{i=1}^{p}\lambda^g_{i,k}+\sum_{j=p+1}^{q}(\lambda^+_{j,k}+\lambda^-_{j,k})+\lambda^\xi_k\right), \tag{3.2}$$

$$0\le\lambda^g_{i,k}\perp(g^i_{\rho_k}(x_k)+\nabla g^i_{\rho_k}(x_k)^Td_k-\xi_k)\le 0,\ i=1,\dots,p, \tag{3.3}$$

$$0\le\lambda^+_{j,k}\perp(h^j_{\rho_k}(x_k)+\nabla h^j_{\rho_k}(x_k)^Td_k-\xi_k)\le 0,\ j=p+1,\dots,q, \tag{3.4}$$

$$0\le\lambda^-_{j,k}\perp(-h^j_{\rho_k}(x_k)-\nabla h^j_{\rho_k}(x_k)^Td_k-\xi_k)\le 0,\ j=p+1,\dots,q, \tag{3.5}$$

$$0\le\lambda^\xi_k\perp-\xi_k\le 0, \tag{3.6}$$

where $\lambda_k=(\lambda^g_k,\lambda^+_k,\lambda^-_k,\lambda^\xi_k)$ is a corresponding Lagrange multiplier.

Let $\rho>0, r>0$. We define the smoothing merit function by

$$\theta_{\rho,r}(x):=f_\rho(x)+r\phi_\rho(x),$$

where $\phi_\rho(x):=\max\{0,g^i_\rho(x),\ i=1,\dots,p,\ |h^j_\rho(x)|,\ j=p+1,\dots,q\}$, and propose the following smoothing SQP algorithm.

ALGORITHM 3.1. *Let $\{\beta,\sigma_1\}$ be constants in $(0,1)$, and let $\{\sigma,\sigma',\hat\eta\}$ be constants in $(1,\infty)$. Choose an initial point x_0, an initial smoothing parameter $\rho_0>0$, an initial penalty parameter $r_0>0$, and an initial positive definite matrix $W_0\in\mathbb{R}^{n\times n}$, and set $k:=0$.*

1. *Solve $(\mathrm{QP})_k$ to obtain (d_k,ξ_k) with the corresponding Lagrange multiplier $\lambda_k=(\lambda^g_k,\lambda^+_k,\lambda^-_k,\lambda^\xi_k)$; go to step 2.*

2. *If $\xi_k=0$, set $r_{k+1}:=r_k$ and go to step 3. Otherwise, set $r_{k+1}:=\sigma' r_k$ and go to step 3.*

3. *Let $x_{k+1}:=x_k+\alpha_kd_k$, where $\alpha_k:=\beta^l$, $l\in\{0,1,2,\dots\}$ is the smallest nonnegative integer satisfying*

$$\theta_{\rho_k,r_k}(x_{k+1})-\theta_{\rho_k,r_k}(x_k)\le-\sigma_1\alpha_kd_kW_kd_k. \tag{3.7}$$

If

$$\|d_k\|\le\hat\eta\rho_k^{-1}, \tag{3.8}$$

set $\rho_{k+1}:=\sigma\rho_k$ and go to step 4. Otherwise, set $\rho_{k+1}:=\rho_k$ and go to step 1. In either case, update to a symmetric positive definite matrix W_{k+1} and $k=k+1$.

4. *If a stopping criterion holds, terminate. Otherwise, go to step 1.*

We now show the global convergence of the smoothing SQP algorithm. For this purpose, we need the following standard assumption.

Assumption 3.1. There exist two positive constants m and M, $m\le M$, such that for each k and each $d\in\mathbb{R}^n$,

$$m\|d\|^2\le d^TW_kd\le M\|d\|^2.$$

THEOREM 3.1. *Suppose that $\{(x_k, \rho_k, d_k, \xi_k, \lambda_k, r_k, W_k)\}$ is a sequence generated by Algorithm* 3.1. *Then for every k,*

$$\theta'_{\rho_k,r_k}(x_k, d_k) \leq -d_k W_k d_k \tag{3.9}$$

and d_k is a descent direction of function $\theta_{\rho_k,r_k}(x)$ at x_k, provided that Assumption 3.1 *holds. Furthermore, suppose that Algorithm* 3.1 *does not terminate within finite iterations. Suppose that the sequences $\{x_k\}$, $\{\lambda_k\}$, and $\{r_k\}$ are bounded. Then $\bar{K} := \{k : \|d_k\| \leq \hat{\eta}\rho_k^{-1}\}$ is an infinite set, and any accumulation point of sequence $\{x_k\}_{\bar{K}}$ is a stationary point of problem* (P).

Proof. Since (d_k, ξ_k) is a solution of $(\mathrm{QP})_k$, the KKT conditions (3.1)−(3.6) hold. The directional derivative of the function $x \to |h^j_{\rho_k}(x)|$ at x_k in direction d_k is

$$\begin{cases} -\nabla h^j_{\rho_k}(x_k)^T d_k & \text{if } h^j_{\rho_k}(x_k) < 0, \\ |\nabla h^j_{\rho_k}(x_k)^T d_k| & \text{if } h^j_{\rho_k}(x_k) = 0, \\ \nabla h^j_{\rho_k}(x_k)^T d_k & \text{if } h^j_{\rho_k}(x_k) > 0. \end{cases}$$

Denote the index sets

$$\begin{aligned} I_k &:= \{i = 1, \dots, p : g^i_{\rho_k}(x_k) = \phi_{\rho_k}(x_k)\}, \\ J_k^+ &:= \{j = p+1, \dots, q : h^j_{\rho_k}(x_k) = \phi_{\rho_k}(x_k)\}, \\ J_k^- &:= \{j = p+1, \dots, q : -h^j_{\rho_k}(x_k) = \phi_{\rho_k}(x_k)\}, \end{aligned}$$

and $\Gamma_k := I_k \cup J_k^+ \cup J_k^-$. Therefore the directional derivative of the function $x \to \phi_{\rho_k}(x)$ at x_k in direction d_k is

$$\begin{cases} 0 & \text{if } \phi_{\rho_k}(x_k) = 0 \text{ and } \Gamma_k = \emptyset, \\ \max\{0, \nabla g^i_{\rho_k}(x_k)^T d_k,\, i \in I_k,\ |\nabla h^j_{\rho_k}(x_k)^T d_k|,\, j \in J_k^+\} & \text{if } \phi_{\rho_k}(x_k) = 0 \text{ and } \Gamma_k \neq \emptyset, \\ \max\{\nabla g^i_{\rho_k}(x_k)^T d_k,\, i \in I_k,\ \nabla h^j_{\rho_k}(x_k)^T d_k,\, j \in J_k^+, & \\ \qquad -\nabla h^j_{\rho_k}(x_k)^T d_k,\, j \in J_k^-\} & \text{if } \phi_{\rho_k}(x_k) > 0. \end{cases}$$

From (3.3)–(3.5), we have

$$\begin{aligned} \nabla g^i_{\rho_k}(x_k)^T d_k &\leq \xi_k - g^i_{\rho_k}(x_k) = \xi_k - \phi_{\rho_k}(x_k), \quad i \in I_k, \\ \nabla h^j_{\rho_k}(x_k)^T d_k &\leq \xi_k - h^j_{\rho_k}(x_k) = \xi_k - \phi_{\rho_k}(x_k), \quad j \in J_k^+, \\ -\nabla h^j_{\rho_k}(x_k)^T d_k &\leq \xi_k + h^j_{\rho_k}(x_k) = \xi_k - \phi_{\rho_k}(x_k), \quad j \in J_k^-. \end{aligned}$$

Thus, $\phi'_{\rho_k}(x_k, d_k) \leq \xi_k - \phi_{\rho_k}(x_k)$. Therefore,

$$\begin{aligned} \theta'_{\rho_k,r_k}(x_k, d_k) &= \nabla f_{\rho_k}(x_k)^T d_k + r_k \phi'_{\rho_k}(x_k, d_k) \\ &\leq \nabla f_{\rho_k}(x_k)^T d_k + r_k \left(\xi_k - \phi_{\rho_k}(x_k)\right). \end{aligned}$$

From (3.2) and (3.6), we know that if $\xi_k > 0$, then

$$r_k = \left(\sum_{i=1}^{p} \lambda^g_{i,k} + \sum_{j=p+1}^{q} (\lambda^+_{j,k} + \lambda^-_{j,k}) \right),$$

which means

$$r_k \xi_k = \left(\sum_{i=1}^{p} \lambda^g_{i,k} + \sum_{j=p+1}^{q} (\lambda^+_{j,k} + \lambda^-_{j,k}) \right) \xi_k. \tag{3.10}$$

By taking conditions (3.1), (3.3)–(3.5), and (3.10) into account, we obtain that for each k,

$$\begin{aligned}
\theta'_{\rho_k,r_k}(x_k,d_k) &= \theta'_{\rho_k,r_k}(x_k,d_k) + \sum_{i=1}^{p} \lambda^g_{i,k}(g^i_{\rho_k}(x_k) + \nabla g^i_{\rho_k}(x_k)^T d_k - \xi_k) \\
&\quad + \sum_{j=p+1}^{q} \lambda^+_{j,k}(h^j_{\rho_k}(x_k) + \nabla h^j_{\rho_k}(x_k)^T d_k - \xi_k) + \sum_{j=p+1}^{q} \lambda^-_{j,k}(-h^j_{\rho_k}(x_k) - \nabla h^j_{\rho_k}(x_k)^T d_k - \xi_k) \\
&\le -d_k W_k d_k + \sum_{i=1}^{p} \lambda^g_{i,k}(g^i_{\rho_k}(x_k) - \xi_k) + \sum_{i=p+1}^{q} \lambda^+_{j,k}(h^j_{\rho_k}(x_k) - \xi_k) \\
&\quad + \sum_{i=p+1}^{q} \lambda^-_{j,k}(-h^j_{\rho_k}(x_k) - \xi_k) + r_k\,(\xi_k - \phi_{\rho_k}(x_k)) \\
&\le -d_k W_k d_k + r_k\,(\xi_k - \phi_{\rho_k}(x_k)) + \left(\sum_{i=1}^{p} \lambda^g_{i,k} + \sum_{i=p+1}^{q} \lambda^+_{j,k} + \sum_{i=p+1}^{q} \lambda^-_{j,k} \right) (\phi_{\rho_k}(x_k) - \xi_k) \\
&= -d_k W_k d_k - \left(r_k - \sum_{i=1}^{p} \lambda^g_{i,k} - \sum_{i=p+1}^{q} \lambda^+_{j,k} - \sum_{i=p+1}^{q} \lambda^-_{j,k} \right) \phi_{\rho_k}(x_k) \\
&\le -d_k W_k d_k.
\end{aligned}$$

Hence inequality (3.9) holds. Since W_k is assumed to be positive definite, it follows that d_k is a descent direction of function $\theta_{\rho_k,r_k}(x)$ at x_k for every k. Therefore, the algorithm is well defined.

We now suppose that Algorithm 3.1 does not terminate within finite iterations. We first prove that there always exists some d_k such that (3.8) holds; thus $\bar{K}$ is an infinite set.

To the contrary suppose that $\|d_k\| \ge c_0 > 0$ for each k. Then Assumption 3.1 together with condition (3.7) implies the existence of a positive constant c such that $\theta_{\rho_k,r_k}(x_{k+1}) \le \theta_{\rho_k,r_k}(x_k) - c$. Consequently, (3.8) fails. From the boundedness of $\{r_k\}$, we know that $\xi_k = 0$ when k is large. We can then assume that there exists a $\bar{k}$ large enough such that $\rho_k = \rho_{\bar{k}}$ and $r_k = \bar{r}$ for $k \ge \bar{k}$ by the updating rule of ρ_k and r_k.

Since the sequence $\{x_k\}$ is bounded, the sequence $\{\theta_{\rho_{\bar{k}},\bar{r}}(x_k)\}$ is bounded below. Moreover, $\theta_{\rho_k,r_k}(x_{k+1}) \le \theta_{\rho_k,r_k}(x_k) - c$, $c > 0$, which implies that the sequence $\{\theta_{\rho_{\bar{k}},\bar{r}}(x_k)\}$ is monotonously decreasing. Hence we have

$$\begin{aligned}
\sum_{k \ge \bar{k}} c &\le \sum_{k \ge \bar{k}} \left(\theta_{\rho_{\bar{k}},\bar{r}}(x_k) - \theta_{\rho_{\bar{k}},\bar{r}}(x_{k+1})\right) \\
&= \theta_{\rho_{\bar{k}},\bar{r}}(x_{\bar{k}}) - \lim_{k\to\infty} \theta_{\rho_{\bar{k}},\bar{r}}(x_k) \\
&< \infty,
\end{aligned}$$

which is a contradiction. Therefore $\bar{K}$ is an infinite set, which also implies that $\rho_k \uparrow \infty$ as $k \to \infty$.

Suppose there exist $K \subseteq \bar{K}$ and $\bar{x}$ such that $\lim_{k\to\infty, k\in K} x_k = \bar{x}$. Since the sequence $\{\lambda_k\}$ is bounded, without loss of generality, assume there exists a subsequence $K_1 \subset K$ such that $(\lambda^g_k, \lambda^+_k, \lambda^-_k, \lambda^\xi_k) \to (\bar{\lambda}^g, \bar{\lambda}^+, \bar{\lambda}^-, \bar{\lambda}^\xi)$ as $k \to \infty$, $k \in K_1$ and

$\bar{\lambda} \geq 0$. By the gradient consistency property of $f_\rho(\cdot)$, $g^i_\rho(\cdot)$, $i = 1, \ldots, p$, and $h^j_\rho(\cdot)$, $j = p+1, \ldots, q$, there exists a subsequence $\tilde{K}_1 \subset K_1$ such that

$$\begin{aligned}
&\lim_{k\to\infty,\, k\in\tilde{K}_1} \nabla f_{\rho_k}(x_k) \in \partial f(\bar{x}),\\
&\lim_{k\to\infty,\, k\in\tilde{K}_1} \nabla g^i_{\rho_k}(x_k) \in \partial g_i(\bar{x}),\ i = 1, \ldots, p,\\
&\lim_{k\to\infty,\, k\in\tilde{K}_1} \nabla h^j_{\rho_k}(x_k) \in \partial h_j(\bar{x}),\ j = p+1, \ldots, q.
\end{aligned}$$

Taking limits in (3.1) and (3.4)–(3.6) as $k \to \infty$, $k \in \tilde{K}_1$, by the gradient consistency properties and $\xi_k \to 0$, it is easy to see that $\bar{x}$ is a stationary point of problem (P). The proof of the theorem is complete. $\square$

In the rest of this section, we give a sufficient condition for the boundedness of sequences $\{r_k\}$ and $\{\lambda_k\}$. We first give the following result on error bounds.

LEMMA 3.1. *For each $k \in \mathbf{N}$, $j = 1, \ldots, l$, let $F^j_k, F^j : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable. Assume that for each $j = 1, \ldots, l$, $\{F^j_k(\cdot)\}$ and $\{\nabla F^j_k(\cdot)\}$ converge pointwise to $F^j(\cdot)$ and $\nabla F^j(\cdot)$, respectively, as k goes to infinity. Let $\hat{d}$ be the point such that $F^j(\hat{d}) = 0, j = 1, \ldots, l$. Suppose that there exist $\kappa > 0$ and $\delta > 0$ such that for all $\mu_j \in [-1, 1]$, $j = 1, \ldots, l$, not all zero and all $d \in \hat{d} + \delta B$ it holds that*

$$\left\| \sum_{j=1}^{l} \mu_j \nabla F^j(d) \right\| > \frac{1}{\kappa}.$$

Then for sufficiently large k,

$$\operatorname{dist}(\hat{d}, S_k) \leq \kappa \sum_{j=1}^{l} |F^j_k(\hat{d})|, \tag{3.11}$$

where $S_k := \{d \in \mathbb{R}^n : F^j_k(d) = 0, j = 1, \ldots, l\}$.

Proof. Denote $F(d) := \sum_{j=1}^{l} |F^j(d)|$, $F_k(d) := \sum_{j=1}^{l} |F^j_k(d)|$. If $\hat{d} \in S_k$, then (3.11) holds trivially. Now suppose that $\hat{d} \notin S_k$. Since $F_k(\hat{d}) \to F(\hat{d})$ as $k \to \infty$, there exists a $\bar{k} \in \mathbf{N}$ such that $F_k(\hat{d}) < \kappa^{-1}\delta$ when $k \geq \bar{k}$. Let $\varepsilon := F_k(\hat{d})$. Then $\varepsilon\kappa < \delta$. Take $\lambda \in (\varepsilon\kappa, \delta)$. Then by Ekeland's variational principle [38, Proposition 1.43], there exists an ω such that $\|\omega - \hat{d}\| \leq \lambda$, $F_k(\omega) \leq F_k(\hat{d})$, and the function $\varphi(d) := F_k(d) + \frac{\varepsilon}{\lambda}\|d - \omega\|$ attains minimum at ω. Hence by the nonsmooth calculus of the Clarke generalized gradient, we have

$$0 \in \partial F_k(w) + \frac{\varepsilon}{\lambda} B,$$

where B denotes the closed unit ball of $\mathbb{R}^n$. Thus $\|v_k\| \leq \frac{\varepsilon}{\lambda} < \frac{1}{\kappa}$ for all $v_k \in \partial F_k(\omega)$, for $k \geq \bar{k}$. We now show that $F_k(w) = 0$ by contradiction. Suppose that $F_k(w) \neq 0$. Then there exists at least one j such that $F^j_k(w) \neq 0$. For such a j, $\partial|F^j_k(w)| = \{\pm\nabla F^j_k(w)\}$. Therefore there exist $\mu^k_j \in [-1, 1]$, $j = 1, \ldots, l$, not all zero such that $v_k = \sum_{j=1}^{l} \mu^k_j \nabla F^j_k(\omega)$. We assume that there exist a subsequence $K \subset \mathbf{N}$ and $\mu_j \in [-1, 1]$, $j = 1, \ldots, l$, not all zero such that for every $k \in K$, $F_k(w) \neq 0$, $\lim_{k\to\infty, k\in K} \mu^k_j = \mu_j$, $j = 1, \ldots, l$. Since $\{\nabla F^j_k(w)\}_k$ converge to $\nabla F^j(w)$, we have $v := \lim_{k\to\infty, k\in K} v_k = \sum_{j=1}^{l} \mu_j \nabla F^j(\omega)$ and $\|v\| \leq \frac{1}{\kappa}$, which is a contradiction. The contradiction shows that we must have $F_k(w) = 0$ and hence $w \in S_k$. Therefore we have

$$\mathrm{dist}(\hat{d}, S_k) \le \|\hat{d} - \omega\| \le \lambda.$$

Since this is true for every $\lambda \in (\varepsilon\kappa, \delta)$, we have that for all $k \ge \bar{k}$,

$$\mathrm{dist}(\hat{d}, S_k) \le \varepsilon\kappa = \kappa|F_k(\hat{d})|. \qquad \square$$

THEOREM 3.2. *Assume that Assumption* 3.1 *holds. Suppose that Algorithm* 3.1 *does not terminate within finite iterations and* $\{(x_k, \rho_k, d_k, \xi_k, \lambda_k, r_k)\}$ *is a sequence generated by Algorithm* 3.1. *If EWGMFCQ holds (or, equivalently, EWNNAMCQ holds) at any accumulation point* $\bar{x}$, *then the following two statements are true:*

(a) $\{d_k\}$ *and* $\{\xi_k\}$ *are bounded.*

(b) $\{r_k\}$ *and* $\{\lambda_k\}$ *are bounded. Furthermore, when* k *is large enough,* $\xi_k = 0$.

Proof. (a) Assume that there exists a subset $K \subseteq \mathbf{N}$ such that $\lim_{k\to\infty, k\in K} x_k = \bar{x}$. To the contrary, suppose that $\{d_k\}_K$ is unbounded. Then there exists a subset $K_0 \subseteq K$ such that $\lim_{k\to\infty,k\in K_0} \|d_k\| = \infty$ and $\lim_{k\to\infty,k\in K_0} x_k = \bar{x}$. By the gradient consistency property, without loss of generality we may assume that

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

By EWGMFCQ, $v_{p+1}, \dots, v_q$ are linearly independent and there exists $\hat{d}$ such that

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

Since the vectors $\{\lim_{k\to\infty,\, k\in K_0} \nabla h^j_{\rho_k}(x_k) : j = p+1, \dots, q\}$ are linearly independent, it is easy to see that for sufficiently large $k \in K_0$, the vectors $\{\nabla h^j_{\rho_k}(x_k),\ j = p+1, \dots, q\}$ are also linearly independent. Denote

$$F^j(d) := h_j(\bar{x}) + v_j^T d,\ \ j = p+1, \dots, q,$$
$$F^j_k(d) := h^j_{\rho_k}(x_k) + \nabla h^j_{\rho_k}(x_k)^T d,\ \ j = p+1, \dots, q.$$

Then $F^j(\hat{d}) = 0$, $j = p+1, \dots, q$. Since $v_{p+1}, \dots, v_q$ are linearly independent, there is κ such that $0 < \frac{1}{\kappa} < \min\{\| \sum_{j=p+1}^q \mu_j v_j\| : \mu_j \in [-1, 1]$ not all equal to zero$\}$. By Lemma 3.1, for sufficiently large k,

$$\mathrm{dist}(\hat{d}, S_k) \le \kappa \sum_{j=p+1}^{q} |F^j_k(\hat{d})|, \tag{3.12}$$

where $S_k := \{d \in \mathbb{R}^n : F^j_k(d) = 0,\ j = p+1, \dots, q\}$. Since S_k is closed, there exists $\hat{d}_k \in S_k$ such that $\|\hat{d} - \hat{d}_k\| = \mathrm{dist}(\hat{d}, S_k)$. Moreover, by virtue of (3.12), the fact that $\lim_{k\to\infty,k\in K_0} F^j_k(\hat{d}) = F^j(\hat{d}) = 0$ for all $j = p+1, \dots, q$ implies that $\|\hat{d} - \hat{d}_k\| \to 0$ as $k \to \infty, k \in K_0$. Hence for sufficiently large k, we have

$$h^j_{\rho_k}(x_k) + \nabla h^j_{\rho_k}(x_k)^T \hat{d}_k = 0,\ \ j = p+1, \dots, q, \tag{3.13}$$
$$g^i_{\rho_k}(x_k) + \nabla g^i_{\rho_k}(x_k)^T \hat{d}_k < 0,\ \ i = 1, \dots, p. \tag{3.14}$$

Conditions (3.13)–(3.14) imply that $(\hat{d}_k, 0)$ is a feasible solution for $(\mathrm{QP})_k$. Since

(d_k, ξ_k) is an optimal solution to problem $(\mathrm{QP})_k$, we have that for any $k \geq \bar{k}$, $k \in K_0$,

$$\begin{aligned}\nabla f_{\rho_k}(x_k)^T d_k + \frac{1}{2} d_k^T W_k d_k &\leq \nabla f_{\rho_k}(x_k)^T d_k + \frac{1}{2} d_k^T W_k d_k + r_k \xi_k \\ &\leq \nabla f_{\rho_k}(x_k)^T \hat{d}_k + \frac{1}{2} \hat{d}_k^T W_k \hat{d}_k. \end{aligned} \tag{3.15}$$

Since $\nabla f_{\rho_k}(x_k)^T \hat{d}_k + \frac{1}{2} \hat{d}_k^T W_k \hat{d}_k$ is bounded, it follows that $\{d_k\}_K$ is bounded from Assumption 3.1. Since (d_k, ξ_k) are feasible for problem $(\mathrm{QP})_k$, by the definition of the smoothing function and the gradient consistency property, it is easy to see that if $\{d_k\}_K$ is bounded, then $\{\xi_k\}_K$ is also bounded. Since K and $\bar{x}$ are an arbitrary subset and an arbitrary accumulation point, $\{d_k\}$ and $\{\xi_k\}$ are bounded for the whole sequence.

(b) To the contrary, suppose that $\{\lambda_k\}$ is unbounded. Then there exists a subset $K_1 \subseteq K$ such that $\lim_{k\to\infty, k\in K_1} \|\lambda_k\| = \infty$ and $\xi_k > 0$ for $k \in K_1$ sufficiently large. By the gradient consistency property, without loss of generality we may assume that

$$\begin{aligned} v_i &= \lim_{k\to\infty, k\in K_1} \nabla g^i_{\rho_k}(x_k),\ i = 1, \dots, p, \\ v_j &= \lim_{k\to\infty, k\in K_1} \nabla h^j_{\rho_k}(x_k),\ j = p+1, \dots, q, \end{aligned}$$

and $\lim_{k\to\infty, k\in K_1} \frac{\lambda_k}{\|\lambda_k\|} = \bar{\lambda}$ for some nonzero vector $\bar{\lambda} = (\bar{\lambda}^g, \bar{\lambda}^+, \bar{\lambda}^-, \bar{\lambda}^\xi) \geq 0$. Dividing by $\|\lambda_k\|$ in both sides of (3.1) and letting $k \to \infty$, $k \in K_1$, we have

$$0 = \sum_{i=1}^p \bar{\lambda}^g_i v_i + \sum_{j=p+1}^q (\bar{\lambda}^+_j - \bar{\lambda}^-_j) v_j. \tag{3.16}$$

Letting $k \to \infty$, $k \in K_1$, in conditions (3.3)–(3.6) and assuming that $(\bar{d}, \bar{\xi})$ is the limiting point of $\{(d_k, \xi_k)\}_{K_1}$, we have

$$\begin{aligned} &0 \leq \bar{\lambda}^g_i \perp (g_i(\bar{x}) + v_i^T \bar{d} - \bar{\xi}) \leq 0,\ i = 1, \dots, p, \\ &0 \leq \bar{\lambda}^+_j \perp (h_j(\bar{x}) + v_j^T \bar{d} - \bar{\xi}) \leq 0,\ j = p+1, \dots, q, \\ &0 \leq \bar{\lambda}^-_j \perp (-h_j(\bar{x}) - v_j^T \bar{d} - \bar{\xi}) \leq 0,\ j = p+1, \dots, q, \\ &0 \leq \bar{\lambda}^{\bar{\xi}} \perp -\bar{\xi} \leq 0. \end{aligned}$$

Multiplying both sides of (3.16) by $\bar{d}$, since

$$\begin{aligned} &\bar{\lambda}^g_i (g_i(\bar{x}) + v_i^T \bar{d} - \bar{\xi}) = 0,\ i = 1, \dots, p, \\ &\bar{\lambda}^+_j (h_j(\bar{x}) + v_j^T \bar{d} - \bar{\xi}) = 0,\ j = p+1, \dots, q, \\ &\bar{\lambda}^-_j (-h_j(\bar{x}) - v_j^T \bar{d} - \bar{\xi}) = 0,\ j = p+1, \dots, q, \end{aligned}$$

we have

$$\begin{aligned} 0 &= \sum_{i=1}^p \bar{\lambda}^g_i v_i^T \bar{d} + \sum_{j=p+1}^q (\bar{\lambda}^+_j - \bar{\lambda}^-_j) v_j^T \bar{d} \\ &= \sum_{i=1}^p \bar{\lambda}^g_i (\bar{\xi} - g_i(\bar{x})) + \sum_{j=p+1}^q \bar{\lambda}^+_j (\bar{\xi} - h_j(\bar{x})) + \sum_{j=p+1}^q \bar{\lambda}^-_j (\bar{\xi} + h_j(\bar{x})). \end{aligned}$$

Thus,

$$\sum_{i=1}^{p} \bar{\lambda}_i^g g_i(\bar{x}) + \sum_{j=p+1}^{q} (\bar{\lambda}_j^+ - \bar{\lambda}_j^-) h_j(\bar{x}) = \sum_{i=1}^{p} \bar{\lambda}_i^g \bar{\xi} + \sum_{j=p+1}^{q} (\bar{\lambda}_j^+ + \bar{\lambda}_j^-) \bar{\xi} \geq 0. \tag{3.17}$$

From EWGMFCQ (equivalently, EWNNAMCQ), condition (3.17) together with condition (3.16) implies that $\bar{\lambda}_i^g = 0$, $i = 1, \ldots, p$, and $\bar{\lambda}_j^+ - \bar{\lambda}_j^- = 0$, $j = p+1, \ldots, q$.

Consider the case where $\bar{\lambda}_i^g = 0$, $i = 1, \ldots, p$, and there exists an index $j \in \{p+1, \ldots, q\}$ such that $\bar{\lambda}_j^+ = \bar{\lambda}_j^- > 0$. Then for sufficiently large $k \in K_1$, $\lambda_{j,k}^+ > 0$ and $\lambda_{j,k}^- > 0$. From the complementary condition (3.4)–(3.5), we must have $\xi_k = 0$ for sufficiently large $k \in K_1$, which is a contradiction.

Otherwise, consider the case where $\bar{\lambda}_i^g = 0$, $i = 1, \ldots, p$, and $\bar{\lambda}_j^+ = \bar{\lambda}_j^- = 0$, $j = p+1, \ldots, q$. Then since $\bar{\lambda}$ is a nonzero vector, we must have $\bar{\lambda}^\xi > 0$, which implies that $\lambda_k^\xi > 0$ for sufficiently large $k \in K_1$. From the complementarity condition (3.6), $\xi_k = 0$ for sufficiently large $k \in K_1$, which is a contradiction.

The contradiction shows that $\{\lambda_k\}$ must be bounded. By the relationship between $\{\lambda_k\}$ and $\{r_k\}$ given in (3.2), the boundedness of $\{\lambda_k\}$ implies the boundedness of $\{r_k\}$. Furthermore, from the updating rule of the algorithm, the boundedness of the sequences $\{\lambda_k\}$ and $\{r_k\}$ implies that when k is large enough, $\xi_k = 0$. We complete the proof. $\square$

The following corollary follows immediately from Theorems 3.1 and 3.2.

COROLLARY 3.2. *Let Assumption* 3.1 *hold, and suppose that Algorithm* 3.1 *does not terminate within finite iterations. Suppose that the sequence* $\{x_k\}$ *is bounded. Assume that EWGMFCQ (or, equivalently, EWNNAMCQ) holds at any accumulation point of sequence* $\{x_k\}$*; then* $\bar{K} := \{k : \|d_k\| \leq \hat{\eta}\rho_k^{-1}\}$ *is an infinite set, and any accumulation point of sequence* $\{x_k\}_{\bar{K}}$ *is a stationary point of problem* (P).

In the case where the objective function is smooth and there is only one inequality constraint and no equality constraints in problem (P), Corollary 3.2 extends [28, Theorem 4.3] to allow the general smoothing function instead of the specific smoothing function.

4. Applications to bilevel programs. The purpose of this section is to apply the smoothing SQP algorithm to the bilevel program. We illustrate how we can apply our algorithm to solve the bilevel program, and we demonstrate through some numerical examples that although the GMFCQ never holds for bilevel programs, the WGMFCQ may be satisfied easily.

In this section we consider the simple bilevel program

$$\text{(SBP)} \qquad \begin{aligned} \min\ & F(x,y) \\ \text{s.t.}\ & y \in S(x), \end{aligned}$$

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 $F, f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ are continuously differentiable and twice continuously differentiable, respectively, and Y is a compact subset of $\mathbb{R}^m$. Our smoothing SQP algorithm can easily handle any extra upper level constraint, but we omit it for simplicity. For a general bilevel program, the lower level constraint may depend on the upper level variables. By "simple," we mean that the lower level constraint Y is independent of x. Although (SBP) is a simple case of the general bilevel program, it has many applications such as the principal-agent problem [30] in economics. We refer the reader to [1, 15, 16, 39, 43] for applications of general bilevel programs.

When the lower level program is a convex program in variable y, the first order approach to solving a bilevel program is to replace the lower level program by its KKT conditions. In the case where f is not convex in variable y, Mirrlees [30] showed that this approach may not be valid in the sense that the true optimal solution for the bilevel problem may not even be a stationary point of the reformulated problem by the first order approach.

For a numerical purpose, Outrata [34] proposed to reformulate a bilevel program as a nonsmooth single level optimization problem by replacing the lower level program by its value function constraint, which in our simple case is

$$\begin{aligned}
\text{(VP)} \qquad & \min \quad F(x,y) \\
& \text{s.t.} \quad f(x,y) - V(x) = 0, \qquad (4.1)\\
& \qquad\quad x \in \mathbb{R}^n,\ y \in Y,
\end{aligned}$$

where $V(x) := \min_{y \in Y} f(x,y)$ is the value function of the lower level problem. By Danskin's theorem (see [11, page 99] or [14]), the value function is Lipschitz continuous but not necessarily differentiable, and hence problem (VP) is a nonsmooth optimization problem with Lipschitz continuous problem data. Ye and Zhu [46] pointed out that the usual constraint qualifications such as the GMFCQ never hold for problem (VP). Ye and Zhu [46, 47] derived the first order necessary optimality condition for the general bilevel program under the so-called partial calmness condition under which the difficult constraint (4.1) is moved to the objective function with a penalty.

Based on the value function approach, Lin, Xu, and Ye [27] recently proposed to approximate the value function by its integral entropy function, i.e.,

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

and developed a smoothing projected gradient algorithm to solve problem (VP) when problem (SBP) is partially calm and to solve an approximate bilevel problem $(\text{VP})_\varepsilon$, where the constraint (4.1) is replaced by $f(x,y) - V(x) \le \varepsilon$ for small $\varepsilon > 0$ when (SBP) is not partially calm.

Unfortunately, the partial calmness condition is rather strong, and hence a local optimal solution of a bilevel program may not be a stationary point of (VP). Ye and Zhu [48] proposed to study the following combined program by adding the first order condition of the lower level problem into the problem (VP). Although the partial calmness condition is a very strong condition for (VP), it is likely to hold for the combined problem under some reasonable conditions [48].

Recently Xu and Ye [45] proposed a smoothing augmented Lagrangian method to solve the combined problem with the assumption that each lower level solution lies in the interior of Y:

$$\begin{aligned}
\text{(CP)} \qquad & \min_{(x,y) \in \mathbb{R}^n \times Y} \quad F(x,y) \\
& \text{s.t.} \quad f(x,y) - V(x) \le 0, \qquad (4.2)\\
& \qquad\quad \nabla_y f(x,y) = 0. \qquad (4.3)
\end{aligned}$$

They showed that if the sequence of penalty parameters is bounded, then any accumulation point is a Clarke stationary point of (CP). They argued that since the problem (CP) is very likely to satisfy the partial calmness or the weak calmness condition (see [48]), the sequence of penalty parameters is likely to be bounded.

To simplify our discussion so that we can concentrate on the main idea, we make the following assumption.

Assumption 4.1. Every optimal solution of the lower level problem is an interior point of set Y.

Under Assumption 4.1, every optimal solution to the lower level constrained problem is a local minimizer to the objective function of the lower level problem, and hence the necessary optimality condition of the lower level problem is simply equal to $\nabla_y f(x,y) = 0$. For some practical problems, it may be possible to set the set Y large enough so that all optimal solutions of the lower level problem are contained in the interior of Y. For example, for the principal-agent problem in economics [30], a very important application of simple bilevel programs, the lower level constraint is an interval and the solution of the lower level problem can usually be estimated to lie in the interior of a certain bounded interval Y. If it is difficult to find a compact set Y that includes all optimal solutions of the lower level problem, but the set Y can be represented by some equality or inequality constraints, then one can use the KKT condition to replace the constraint (4.3) in the problem (CP). In this case the problem (CP) will become a nonsmooth mathematical program with equilibrium constraints. We will study this case in a separate paper.

Since problem (CP) is a nonconvex and nonsmooth optimization problem, in general the best we can do is to look for its Clarke stationary points. Since we assume that all lower level solutions lie in the interior of set Y, any local optimal solution of (CP) must be the Clarke stationary point of (CP) with the constraint $y \in Y$ removed. Hence the smoothing SQP method introduced in this paper can be used to find the stationary points of (CP).

Let $(\bar{x}, \bar{y})$ be a local optimal solution of (CP). Then by the Fritz John–type multiplier rule, there exist $r \geq 0, \lambda_1 \geq 0, \lambda_2 \in \mathbb{R}^m$ not all zero such that

$$0 \in r\nabla F(\bar{x}, \bar{y}) + \lambda_1(\nabla f(\bar{x}, \bar{y}) - \partial V(\bar{x}) \times \{0\}) + \nabla(\nabla_y f)(\bar{x}, \bar{y})^T \lambda_2. \tag{4.4}$$

In the case when r is positive, $(\bar{x}, \bar{y})$ is a stationary point of (CP). A sufficient condition for r to be positive is that in the Fritz John condition, $r = 0$ implies that λ_1, λ_2 are all equal to zero. Unfortunately we now show that r can always be taken as zero in the above Fritz John condition for problem (CP). Indeed, from the definition of $V(x)$, we always have $f(x, y) - V(x) \geq 0$ for any $y \in Y$. Hence any feasible point $(\bar{x}, \bar{y})$ of problem (CP) is always an optimal solution of the problem

$$\min_{(x,y) \in \mathbb{R}^n \times Y} f(x,y) - V(x) \quad \text{s.t. } \nabla_y f(x,y) = 0.$$

By the Fritz John–type multiplier rule, there exist $\lambda_1 \geq 0, \lambda_2 \in \mathbb{R}^m$ not all equal to zero such that

$$0 \in \lambda_1(\nabla f(\bar{x}, \bar{y}) - \partial V(\bar{x}) \times \{0\}) + \nabla(\nabla_y f)(\bar{x}, \bar{y})^T \lambda_2. \tag{4.5}$$

Observe that (4.5) is (4.4) with $r = 0$. Since (λ_1, λ_2) is nonzero, we have shown that the Fritz John condition (4.4) for problem (CP) holds with $r = 0$. In other words, NNAMCQ (or, equivalently, GMFCQ) for problem (CP) never holds.

It follows from [27, Theorems 5.1 and 5.5] that the integral entropy function $\gamma_\rho(x)$ is a smoothing function with the gradient consistency property for the value function $V(x)$. That is,

$$\lim_{z\to x,\,\rho\uparrow\infty} \gamma_\rho(z) = V(x) \quad \text{and} \quad \emptyset \neq \limsup_{z\to x,\,\rho\uparrow\infty} \nabla\gamma_\rho(z) \subseteq \partial V(x).$$

For a sequence of iteration points $\{(x^k, y^k)\}$, the set $\limsup_{k\to\infty} \nabla\gamma_{\rho_k}(x^k)$ may be strictly contained in $\partial V(x)$. Therefore while (4.5) holds for some $\lambda_1 \geq 0, \lambda_2 \in \mathbb{R}^m$ not all equal to zero, the following inclusion may hold only when $\lambda_1 = 0, \lambda_2 = 0$:

$$0 \in \lambda_1\Big(\nabla f(\bar{x},\bar{y}) - \limsup_{k\to\infty} \nabla\gamma_{\rho_k}(x^k) \times \{0\}\Big) + \nabla(\nabla_y f)(\bar{x},\bar{y})^T\lambda_2.$$

Then, consequently, the WNNAMCQ may hold. We illustrate this point by using some numerical examples. In these examples, since $y \in \mathbb{R}$, the problem (CP) has one inequality constraint $f(x,y) - V(x) \leq 0$ and one equality constraint $\nabla_y f(x,y) = 0$. Hence the WNNAMCQ

$$0 \in \lambda_1\Big(\nabla f(\bar{x},\bar{y}) - \limsup_{k\to\infty} \nabla\gamma_{\rho_k}(x^k) \times \{0\}\Big) + \lambda_2\nabla(\nabla_y f)(\bar{x},\bar{y}), \quad \lambda_1 \geq 0 \Longrightarrow \lambda_1 = \lambda_2 = 0,$$

amounts to saying that for $\lim_{k\to\infty}(x^k, y^k) = (\bar{x},\bar{y})$ and $v = \lim_{k\to\infty} \nabla\gamma_{\rho_k}(x^k)$, the vectors

$$\nabla f(\bar{x},\bar{y}) - (v, 0) \quad \text{and} \quad \nabla(\nabla_y f)(\bar{x},\bar{y})$$

are linearly independent.

In our numerical experiments, we use the so-called limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) approach proposed by Nocedal [33], which is a modification to the BFGS method for unconstrained optimization problems, to update the matrix W_k. Define $s_k := x_{k+1} - x_k$ and

$$\begin{aligned} y_k := \nabla f_{\rho_k}(x_{k+1}) - \nabla f_{\rho_k}(x_k) - \sum_{i=1}^{p} \lambda^g_{i,k}(\nabla g^i_{\rho_k}(x_{k+1}) - \nabla g^i_{\rho_k}(x_k)) \\ - \sum_{j=p+1}^{q} (\lambda^+_{j,k} - \lambda^-_{j,k})(\nabla h^j_{\rho_k}(x_{k+1}) - \nabla h^j_{\rho_k}(x_k)). \end{aligned}$$

We update W_{k+1} by

$$W_{k+1} = W_k - \frac{W_k s_k s_k^T W_k}{s_k^T W_k s_k} + \frac{y_k y_k^T}{s_k^T y_k}$$

if and only if

$$\|s_k\| \leq \gamma_s \epsilon, \quad \|y_k\| \leq \gamma_y \epsilon, \quad \text{and} \quad s_k^T y_k \geq \gamma_{sy}\epsilon^2$$

for given $(\gamma_s, \gamma_y, \gamma_{sy}) > 0$. Otherwise, we skip the update. As shown in [13], these restrictions guarantee the existence of $M \geq m > 0$ such that

$$m\|d\|^2 \leq d^T W_k d \leq M\|d\|^2.$$

In numerical practice, it is impossible to obtain an exact "0"; thus we select some small enough $\varepsilon > 0$, $\varepsilon' > 0$ and change the update rule of r_k and ρ_k to the case when $\xi_k < \varepsilon'$ and

$$\|d_k\| \leq \max\{\hat{\eta}\rho_k^{-1}, \varepsilon\},$$

respectively. Also the stopping criterion is considered as follows: For a given $\epsilon_1 > 0$, we terminate the algorithm at the kth iteration if

$$\|d_k\| < \varepsilon \quad \text{and} \quad \|\xi_k\| < \varepsilon'.$$

In the remainder of this section, we test the algorithm for some bilevel problems.

Example 4.1 (see [30]). Consider Mirrlees' problem. Note that the solution of Mirrlees' problem does not change if we add the constraint $y \in [-2, 2]$ into the problem:

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

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

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

It was shown in [30] that the unique optimal solution is $(\bar{x}, \bar{y})$ with $\bar{x} = 1$, $\bar{y} \approx 0.958$ being the positive solution of the equation

$$(1+y) = (1-y)\exp[4y].$$

In our test, we chose the initial point $(x_0, y_0) = (0.6, 0.3)$ and the parameters $\beta = 0.8$, $\sigma_1 = 10^{-6}$, $\rho_0 = 100$, $r_0 = 100$, $\hat{\eta} = 5 * 10^5$, $\sigma = 10$, $\sigma' = 10$, $\varepsilon = 10^{-7}$, and $\varepsilon' = 10^{-10}$. Since the stopping criteria hold, we terminate at the 16th iteration with $(x^k, y^k) = (1, 0.95759)$. It seems that the sequence converges to $(\bar{x}, \bar{y})$.

Since

$$\begin{aligned} &\nabla f(x^k, y^k) - (\nabla\gamma_{\rho_k}(x^k), 0) = (0.01784, 0.00015), \\ &\nabla(\nabla_y f)(x^k, y^k) = (0.084813, 1.70049), \end{aligned}$$

by virtue of the continuity of the gradients it is easy to see that the vectors

$$\nabla f(\bar{x}, \bar{y}) - \left(\lim_{k\to\infty} \nabla\gamma_{\rho_k}(x^k), 0\right) \quad \text{and} \quad \nabla(\nabla_y f)(\bar{x}, \bar{y})$$

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). Indeed, $(\bar{x}, \bar{y})$ is the unique global minimizer of Mirrlees' problem.

Example 4.2 (see [31, Example 3.14]). The bilevel program

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

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 chose the initial point $(x_0, y_0) = (0.3, 0.3)$ and the parameters $\beta = 0.9,\ \sigma_1 = 10^{-6},\ \rho_0 = 100,\ r_0 = 100,\ \hat{\eta} = 5000,\ \sigma = 10,\ \sigma' = 10,\ \varepsilon = 10^{-7}$, and $\varepsilon' = 10^{-10}$. Since the stopping criteria hold, we terminate at the 7th iteration with $(x^k, y^k) = (0.25, 0.5)$. It seems that the sequence converges to $(\bar{x},\bar{y})$.

Since

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

by virtue of the continuity of the gradients it is easy to see that the vectors

$$\nabla f(\bar{x},\bar{y}) - \Big(\lim_{k\to\infty}\nabla\gamma_{\rho_k}(x^k),0\Big) \quad\text{and}\quad \nabla(\nabla_y f)(\bar{x},\bar{y})$$

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). Indeed, $(\bar{x},\bar{y})$ is the unique global minimizer of the problem.

Example 4.3 (see [31, Example 3.20]). The bilevel program

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

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 chose the parameters $\beta = 0.9,\ \sigma_1 = 10^{-6},\ \rho_0 = 100,\ r_0 = 100,\ \hat{\eta} = 500,\ \sigma = 10,\ \sigma' = 10,\ \varepsilon = 10^{-7}$, and $\varepsilon' = 10^{-10}$. We chose the initial point $(x_0, y_0) = (0.3, 0.8)$. Since the stopping criteria hold, we terminate at the 8th iteration with $(x^k, y^k) = (0.4999996, 0.4999996)$. It seems that the sequence converges to $(\bar{x},\bar{y})$.

Since

$$\begin{aligned}&\nabla f(x^k,y^k) - (\nabla\gamma_{\rho_k}(x^k),0) = (-1.499898,0),\\&\nabla(\nabla_y f)(x^k,y^k) = (-1,1),\end{aligned}$$

by virtue of the continuity of the gradients it is easy to see that the vectors

$$\nabla f(\bar{x},\bar{y}_1) - \Big(\lim_{k\to\infty}\nabla\gamma_{\rho_k}(x^k),0\Big) \quad\text{and}\quad \nabla(\nabla_y f)(\bar{x},\bar{y}_1)$$

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). Indeed, $(\bar{x},\bar{y})$ is the unique global minimizer of the problem.

5. Conclusion. In this paper, we propose a smoothing SQP method for solving nonsmooth and nonconvex optimization problems with Lipschitz inequality and equality constraints. The algorithm is applicable even to degenerate constrained optimization problems which do not satisfy the GMFCQ, the standard constraint qualification for a local minimizer to satisfy the KKT conditions. Our main motivation comes from solving the bilevel program which is nonsmooth, nonconvex, and never satisfies the GMFCQ. In this paper, we have proposed the concept of the WGMFCQ (equivalently, WNNAMCQ), a weaker version of the GMFCQ, and have shown the global convergence of the smoothing SQP algorithm under the WGMFCQ. Moreover, we have demonstrated the applicability of the smoothing SQP algorithm for solving the combined program of a simple bilevel program with a nonconvex lower level problem. For smooth optimization problems, it is well known that the SQP methods converge very quickly when the iterates are close to the solution. The rapid local convergence of the SQP is due to the fact that the positive definite matrix W_k in the SQP subproblem is an approximation of the Hessian matrix of the Lagrangian function. For our nonsmooth problem, the Lagrangian function is only locally Lipschitz, and no classical Hessian matrix can be defined. However, it would be interesting to study the local behavior of the smoothing SQP algorithm by using the generalized second order subderivatives [38] of the Lagrangian function. This remains a topic of our future research.

Acknowledgments. The authors are grateful to the anonymous referees for their careful reading of the paper and helpful comments and to Shaoyan Guo for helpful discussions.

REFERENCES

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

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

[3] J.V. Burke and S.-P. Han, *A robust sequential quadratic programming method*, Math. Programming, 43 (1989), pp. 277–303.

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

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

[6] 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), pp. 624–645.

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

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

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

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

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

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

[13] F.E. Curtis and X. Quez, *An adaptive gradient sampling algorithm for nonsmooth optimization*, Optim. Methods Softw., 28 (2013), pp. 1302–1324.

[14] J.M. Danskin, *The Theory of Max-Min and Its Applications to Weapons Allocation Problems*, Springer, New York, 1967.

[15] S. Dempe, *Foundations of Bilevel Programming*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002.

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

[17] F. Facchinei, *Robust recursive quadratic programming algorithm model with global and superlinear convergence properties*, J. Optim. Theory Appl., 92 (1997), pp. 543–579.

[18] M. Fukushima and J.-S. Pang, *Some feasibility issues in mathematical programs with equilibrium constraints*, SIAM J. Optim., 8 (1998), pp. 673–681.

[19] U.M. Garcia-Palomares and O.L. Mangasarian, *Superlinearly convergent quasi-Newton methods for nonlinearly constrained optimization problems*, Math. Programming, 11 (1976), pp. 1–13.

[20] P.E. Gill and E. Wong, *Sequential quadratic programming methods*, in Mixed Integer Nonlinear Programming, IMA Vol. Math. Appl. 154, Springer-Verlag, Berlin, 2012, pp. 147–224.

[21] S.P. Han, *Superlinearly convergent variable metric algorithms for general nonlinear programming problems*, Math. Programming, 11 (1976), pp. 263–282.
[22] S.P. Han, *A globally convergent method for nonlinear programming*, J. Optim. Theory Appl., 22 (1977), pp. 297–309.
[23] M. Heinkenschloss, *Projected sequential quadratic programming methods*, SIAM J. Optim., 6 (1996), pp. 373–417.
[24] J.B. Hiriart-Urruty, *Refinements of necessary optimality conditions in nondifferentiable programming.* I, Appl. Math. Optim., 5 (1979), pp. 63–82.
[25] H. Jiang and D. Ralph, *Smooth SQP methods for mathematical programs with nonlinear complementarity constraints*, SIAM J. Optim, 10 (2000), pp. 779–808.
[26] B. Kummer, *Newton's method for nondifferentiable functions*, in Advances in Mathematical Optimization, Math. Res. 45, Akademie-Verlag, Berlin, 1988.
[27] G.-H. Lin, M. Xu, and J.J. Ye, *On solving simple bilevel programs with a nonconvex lower level program*, Math. Program. Ser. A, 144 (2014), pp. 277–305.
[28] C. Ling, L. Qi, G.L. Zhou, and S.Y. Wu, *Global convergence of a robust smoothing SQP method for semi-infinite programming*, J. Optim. Theory Appl., 129 (2006), pp. 147–164.
[29] X.-W. Liu and Y.-X. Yuan, *A robust algorithm for optimization with general equality and inequality constraints*, SIAM J. Sci. Comput., 22 (2000), pp. 517–534.
[30] J.A. Mirrlees, *The theory of moral hazard and unobservable behaviour: Part I*, Rev. Econ. Stud., 66 (1999), pp. 3–21.
[31] A. Mitsos and P.I. Barton, *A Test Set for Bilevel Programs*, Technical report, Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, 2006.
[32] Y. Nesterov, *Smoothing minimization of nonsmooth functions*, Math. Program., 103 (2005), pp. 127–152.
[33] J. Nocedal, *Updating quasi-Newton matrices with limited storage*, Math. Comp., 35 (1980), pp. 773–782.
[34] J.V. Outrata, *On the numerical solution of a class of Stackelberg problems*, Z. Oper. Res., 34 (1990), pp. 255–277.
[35] J.F.A. Pantoja and D.Q. Mayne, *Exact penalty function algorithm with simple updating of the penalty parameter*, J. Optim. Theory Appl., 69 (1991), pp. 441–467.
[36] M.J.D. Powell and Y. Yuan, *A recursive quadratic programming algorithm that uses differentiable exact penalty functions*, Math. Programming, 35 (1986), pp. 265–278.
[37] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
[38] R.T. Rockafellar and R.J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.
[39] K. Shimizu, Y. Ishizuka, and J.F. Bard, *Nondifferentiable and Two-Level Mathematical Programming*, Kluwer Academic Publishers, Boston, 1997.
[40] P. Spellucci, *A new technique for inconsistent QP problems in the SQP method*, Math. Methods Oper. Res., 47 (1998), pp. 355–400.
[41] K. Tone, *Revision of constraint approximations in the successive QP-method for nonlinear programming problems*. Math. Programming, 26 (1983), pp. 144–152.
[42] X. Tong, L.Q. Qi, G.L. Zhou, and S.Y. Wu, *A smoothing SQP method for nonlinear programs with stability constraints arising from power systems*, Comput. Optim. Appl., 51 (2012), pp. 175–197.
[43] L.N. Vicente and P.H. Calamai, *Bilevel and multilevel programming: A bibliography review*, J. Global Optim., 5 (1994), pp. 291–306.
[44] R.B. Wilson, *A Simplicial Algorithm for Concave Programming*, Ph.D. thesis, Graduate School of Business Administration, Harvard University, Cambridge, MA, 1963.
[45] M. Xu and J.J. Ye, *A smoothing augmented Lagrangian method for solving simple bilevel programs*, Comput. Optim. Appl., 59 (2014), pp. 353–377.
[46] J.J. Ye and D.L. Zhu, *Optimality conditions for bilevel programming problems*, Optimization, 33 (1995), pp. 9–27.
[47] J.J. Ye and D.L. Zhu, *A note on: "Optimality conditions for bilevel programming problems,"* Optimization, 39 (1997), pp. 361–366.
[48] J.J. Ye and D. Zhu, *New necessary optimality conditions for bilevel programs by combining the MPEC and value function approaches*, SIAM J. Optim., 20 (2010), pp. 1885–1905.
[49] C. Zhang and X. Chen, *Smoothing projected gradient method and its application to stochastic linear complementarity problems*, SIAM J. Optim., 20 (2009), pp. 627–649.
[50] J. Zhang and X. Zhang, *A robust SQP method for optimization with inequality constraints*, J. Comput. Math, 21 (2003), pp. 247–256.