CSIAM Trans. Appl. Math.
doi: 10.4208/csiam-am.2020-0043

Vol. **2**, No. 3, pp. 551-584
September 2021

Optimization with Least Constraint Violation

Yu-Hong Dai[1,2] and Liwei Zhang[3,*]

[1] *LSEC, ICMSEC, AMSS, Chinese Academy of Sciences, Beijing 100190, China.*
[2] *School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China.*
[3] *School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China.*

Received 6 November 2020; Accepted 25 March 2021

Abstract. Study about theory and algorithms for nonlinear programming usually assumes that the feasible region of the problem is nonempty. However, there are many important practical nonlinear programming problems whose feasible regions are not known to be nonempty or not, and optimizers of the objective function with the least constraint violation prefer to be found. A natural way for dealing with these problems is to extend the nonlinear programming problem as the one optimizing the objective function over the set of points with the least constraint violation. Firstly, the minimization problem with least constraint violation is proved to be an Lipschitz equality constrained optimization problem when the original problem is a convex nonlinear programming problem with possible inconsistent constraints, and it can be reformulated as an MPCC problem; And the minimization problem with least constraint violation is relaxed to an MPCC problem when the original problem is an nonlinear programming problem with possible inconsistent non-convex constraints. Secondly, for nonlinear programming problems with possible inconsistent constraints, it is proved that a local minimizer of the MPCC problem is an M-stationary point and an elegant necessary optimality condition, named as L-stationary condition, is established from the classical optimality theory of Lipschitz continuous optimization. Thirdly, properties of the penalty method for the minimization problem with the least constraint violation are developed and the proximal gradient method for the penalized problem is studied. Finally, the smoothing Fischer-Burmeister function method is constructed for solving the MPCC problem related to minimizing the objective function with the least constraint violation. It is demonstrated that, when the positive smoothing parameter approaches to zero, any point in the outer limit of the KKT-point mapping is an L-stationary point of the equivalent MPCC problem.

AMS subject classifications: 90C30

Key words: Nonlinear programming, least constraint violation, L-stationary point, MPCC, smoothing function.

*Corresponding author. *Email addresses:* `dyh@lsec.cc.ac.cn` (Y.-H. Dai), `lwzhang@dlut.edu.cn` (L. Zhang)

1 Introduction

For studying an ordinary nonlinear optimization problem, a basic assumption is that feasible region of the optimization problem is nonempty. Many important theoretical issues are well studied for an optimization problem under this assumption. For example, optimality theory and sensitivity analysis are two main theoretical topics. Optimality theory consists of necessary optimality conditions and sufficient optimality conditions. Sensitivity analysis studies continuity properties of the optimal value and the solution mapping when the optimization is perturbed. For nonlinear programming, for a local minimizer, the first-order necessary optimality conditions and the second-order optimality conditions can be developed under certain constraint qualifications, and the second-order sufficient optimality conditions imply the second-order growth condition, see for instance the famous textbook [13]. For nonlinear programming, a series of stability results were obtained by Robinson, see [14–16]. Bonnans and Shapiro [3] established the optimality theory and the stability theory for general optimization problems, including problems whose decision variables are infinite dimensional, nonlinear semidefinite programming problems and other conic optimization problems.

However, when the feasible set is empty or the constraints are inconsistent, infeasibility detection is an important issue for algorithmic design. Many numerical algorithms have been proposed to find infeasible stationary points; namely, stationary points for minimizing certain infeasibility measure. Byrd, Curtis and Nocedal [4] presented a set of conditions to guarantee the superlinear convergence of their SQP algorithm to an infeasible stationary point. Burke, Curtis and Wang [5] considered the general program with equality and inequality constraints, and proved that their SQP method has strong global convergence and rapid convergence to the KKT point, and has superlinear/quadratic convergence to an infeasible stationary point. Recently, Dai, Liu and Sun [7] proposed a primal-dual interior-point method, which can be superlinearly or quadratically convergent to the Karush-Kuhn-Tucker point if the original problem is feasible, and can be superlinearly or quadratically convergent to the infeasible stationary point when the problem is infeasible.

These algorithms can find a stationary point of the infeasibility measure, which have nothing to do with the objective function of the problem. In practice, there are many important problems that we need to find minimizers of the objective function over the points with the least constraint violation. A natural way to deal with such problems is to extend the constrained optimization problem as the one that optimizes the objective function over the set of points with least constraint violation. When the feasible region is nonempty, the set of points with least constraint violation coincides with the feasible region of the constrained optimization problem and hence the extended constrained optimization problem coincides with the original problem.

Now we give a formal definition of infeasibility measure of an nonlinear programming problem. Suppose that the nonlinear programming problem is of the following form

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & h(x)=0, \\ & g(x)\geq 0, \end{aligned} \tag{1.1}$$

where $f:\Re^n\to\Re$, $h:\Re^n\to\Re^q$ and $g:\Re^n\to\Re^p$. Let Φ denotes the feasible set

$$\Phi=\{x\in\Re^n: h(x)=0,\ g(x)\geq 0\}.$$

Definition 1.1. *A function $\theta:\Re^n\to\Re$ is said to be an infeasibility measure of the constraints $(h(x),g(x))\in\{0_q\}\times\Re^p_+$ if there exists an increasing continuous function $\varrho:\Re_+\to\Re_+$ with $\varrho(0)=0$ such that*

$$\theta(x)=\varrho\left(\operatorname{dist}\left((h(x),g(x)),\{0_q\}\times\Re^p_+\right)\right),$$

where

$$\operatorname{dist}\left((h(x),g(x)),\{0_q\}\times\Re^p_+\right)=\inf\left\{\|(h(x),y-g(x))\|: y\in\Re^p_+\right\}$$

is the distance from $((h(x),g(x))$ to $\{0_q\}\times\Re^p_+)$ under a norm $\|\cdot\|$ of $\Re^q\times\Re^p$.

Obviously $\theta(x)$ depends on the function $\varrho(\cdot)$ and the norm $\|\cdot\|$. In this paper, we will adopt $\varrho(t)=t^2/2$ and $\|\cdot\|$ as the standard Euclidean norm.

Under the infeasibility measure defined above, we introduce the mathematical model of minimizing the objective $f(x)$ over the set of points with least infeasibility measure.

Definition 1.2. *For an infeasibility measure $\theta(x)$ of the constraint $(h(x),g(x))\in\{0_q\}\times\Re^p_+$, the mathematical model of minimizing the objective $f(x)$ over the set of points with least constraint violation associated with θ, is defined by*

$$\left\{\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x\in \operatorname{Argmin}_z \theta(z), \end{array}\right. \tag{1.2}$$

where Argmin *denotes the solution set of an optimization problem.*

Obviously, if the feasible region Φ is nonempty, then $\min_z\theta(z)=0$, $\operatorname{Argmin}_z\theta(z)=\Phi$, and Problem (1.2) is just the original problem (1.1). Thus Problem (1.2) can be regarded as an extension of the original problem (1.1). As the minimization of the constraint violation is considered absolutely prior to the optimization of the objective function, we call the optimum (optimizer) of Problem (1.2) as the *optimum (optimizer) of Problem* (1.1) *with least constraint violation*, or simply *constrained optimum (optimizer) of Problem* (1.1).

We do not identify the notion of $\operatorname{Argmin}_z\theta(z)$ in Definition 1.2. If $\theta(z)$ is convex, it is obvious that the solution set is just the set of global minimizers. However, if $\theta(z)$ is non-convex (this happens when Φ is not a convex set if feasible), $\operatorname{Argmin}_z\theta(z)$ may be understood as a set of local minimizers or even the set of stationary points. Infeasibility detection is well known difficult in the nonconvex optimization case. Indeed, for a nonconvex problem, infeasibility detection has many of the difficulties inhered in global optimization since even if an algorithm identifies an infeasible point where constraint violations are locally minimized, there may exist feasible points in other regions of the space of decision variables.

Now we give a simple example to explain the above concepts. Consider the simple quadratic programming problem

$$\begin{aligned} \min \quad & x_1^2+x_2^2 \\ \text{s.t.} \quad & x_1+x_2-1\le 0, \\ & -x_1-x_2+2\le 0. \end{aligned} \tag{1.3}$$

It is easy to find that the feasible region is empty. We consider the minimization problem over the set of points with least constraint violation. We can regard the least violation for the constraints as the optimal value of the following problem

$$\begin{aligned} \min \quad & \frac{1}{2}(y_1^2+y_2^2) \\ \text{s.t.} \quad & x_1+x_2-1+y_1\le 0, \\ & -x_1-x_2+2+y_2\le 0. \end{aligned} \tag{1.4}$$

Then the set of points with the least violation is given by

$$S=\{x:(x,y) \text{ solves Problem (1.4)}\}.$$

It is not difficult to obtain

$$S=\{(x_1,x_2):x_1+x_2-3/2=0\}.$$

Therefore the minimum point of the objective over the violation set is $(3/4,3/4)$. It is not difficult to verify that $\varrho(t)=\frac{1}{2}t^2$ for $t\ge 0$ and

$$\theta(x)=\frac{1}{2}\left\{[x_1+x_2-1]_+^2+[-x_1-x_2+2]_+^2\right\}$$

is an infeasibility measure in Definition 1.1 and

$$\text{Argrmin}_x\theta(x)=S.$$

A popular method for the minimization with least violation constraint is the penalty method. Define the penalty function as

$$P_c(x)=x_1^2+x_2^2+c\left\{[x_1+x_2-1]_+^2+[-x_1-x_2+2]_+^2\right\},$$

where $[t]_+=\max\{0,t\}$ for $t\in\Re$. Obviously P_c is a smooth convex function with

$$\nabla P_c(x_1,x_2)=\begin{bmatrix} 2x_1+2c\{[x_1+x_2-1]_+-[-x_1-x_2+2]_+\} \\ 2x_2+2c\{[x_1+x_2-1]_+-[-x_1-x_2+2]_+\} \end{bmatrix}.$$

By solving $\nabla P_c(x)=0$, we obtain the minimizer of $P_c(x)$ is

$$x^*(c)=\left(\frac{3c}{1+4c},\frac{3c}{1+4c}\right)^T. \tag{1.5}$$

Thus we have

$$\lim_{c\to\infty} x^*(c)=\left(\frac{3}{4},\frac{3}{4}\right)^T;$$

namely, the limit of the minimizer of P_c approaches the optimal solution. This example gives us an intuition of the penalty method, this paper will discuss the penalty method for solving an approximate solution to Problem (1.2). On the other hand, from the above example, for finite $c>0$, the minimizer of P_c never coincides with the optimal solution. Therefore, it is significant to find a method, different from the penalty method, for solving the minimization optimization problem over the set of points with least violation constraint. In this paper, we will propose a smoothing method for finding a stationary point of Problem (1.2).

The rest of this paper is organized as follows. In Section 2, for taking $\varrho(t)=t^2/2$ and $\|\cdot\|$ as the standard Euclidean norm, we reformulate the mathematical model of the minimization problem with the smallest constraint violation for a general nonlinear programming problem. For nonlinear programming with possible inconsistent constraints, we prove M-stationary property of a local minimizer of the MPCC problem associated with the minimization problem over the set of the least constraint violation. Especially, an elegant necessary optimality condition, named as L-stationary condition, is established from the classical optimality theory of Lipschitz continuous optimization. In Section 3, we will explain the penalty method can find an approximate solution to Problem (1.2) and present the proximal method for solving the penalized problem. In Section 4, we propose the smoothing Fischer-Burmeister function method for solving the MPCC problem associated with the optimization problem over the set of the least constraint violation. It is demonstrated that, when the positive smoothing parameter approaches to 0, any point in the outer limit of the KKT-point mapping is an L-stationary point of the MPCC problem. Some discussions are made in the last section.

2 Necessary optimality conditions

2.1 Mathematical model

For $\varrho(t)=\frac{1}{2}t^2$ for $t\geq 0$, the least violation for the constraint is defined as the optimal value of the following problem

$$\begin{array}{ll} \min & \frac{1}{2}\|(h(x),y)\|^2 \\ \text{s.t.} & g(x)+y\geq 0_p. \end{array} \tag{2.1}$$

The set of points with the least violation is given by

$$S=\{x:(x,y) \text{ solves Problem (2.1)}\}.$$

Our problem is to minimize f over S; namely,

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & (x,y) \text{ solves} \\ & \begin{cases} \min\limits_{w,z} & \frac{1}{2}[\|h(w)\|^2+\|z\|^2] \\ \text{s.t.} & g(w)+z\geq 0. \end{cases} \end{aligned} \tag{2.2}$$

Denote the lower level problem of Problem (2.2) by $\mathrm{P_L}$; namely,

$$(\mathrm{P_L}) \qquad \begin{aligned} \min_{w,z} \quad & \frac{1}{2}[\|h(w)\|^2+\|z\|^2] \\ \text{s.t.} \quad & g(w)+z\geq 0. \end{aligned} \tag{2.3}$$

Associated with the above $\varrho(t)$, we have that

$$\begin{aligned} \theta(w) &= \frac{1}{2}\|h(w)\|^2+\min_z\left\{\frac{1}{2}\|z\|^2 : g(w)+z\geq 0\right\} \\ &= \frac{1}{2}\left[\|h(w)\|^2+\|[g(w)]_-\|^2\right], \end{aligned} \tag{2.4}$$

where $[z]_-=([z_1]_-,\cdots,[z_p]_-)^T$ for $z\in\Re^p$ with $[t]_-=\min\{0,t\}$ for $t\in\Re$. It is easy to verify that $\theta(w)$ is an infeasibility measure for Problem (1.1). Then the set of optimal solutions of Problem $(\mathrm{P_L})$ can be expressed as

$$\mathrm{Argmin}\,(\mathrm{P_L})=\mathrm{Argmin}_x\theta(x)=\mathrm{Argmin}_x\frac{1}{2}\left[\|h(x)\|^2+\|[g(x)]_-\|^2\right].$$

Therefore Problem (2.2) can equivalently be expressed as

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x\in\mathrm{argmin}_w\frac{1}{2}\left[\|h(w)\|^2+\|[g(w)]_-\|^2\right]. \end{aligned} \tag{2.5}$$

It is easy to obtain that $\theta(x)$ is differentiable if h and g are differentiable with

$$\nabla\theta(x)=\mathcal{J}h(x)^Th(x)+\mathcal{J}g(x)^T[g(x)]_-. \tag{2.6}$$

Remark 2.1. Although Problem (2.2) is a bilevel programming, the lower level problem $\mathrm{P_L}$ has an explicit solution. This simplifies this bi-level optimization problem as a optimization problem with a Lipschitz continuous equality constraint. Here we should point out that there are some new developments of numerical algorithms for bilevel optimization, see for instance the paper [10].

In the next two subsections, we will discuss convex nonlinear programming with least constraint violation and non-convex nonlinear programming with least constraint violation, respectively. Both problems initiate from mathematical model (2.5) and use formula (2.6).

2.2 Convex NLP with least constraint violation

In this subsection, we consider the following convex nonlinear programming problem

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & Ax-b=0, \\ & g(x)\geq 0, \end{array} \tag{2.7}$$

where f is a differentiable convex function, $A\in\Re^{q\times n}$ and $b\in\Re^q$, and each g_i $(i=1,\cdots,p)$ is a concave differentiable function. In this case, we have that $\theta(x)$ is a convex function, and Problem (2.5) is a convex optimization problem which is reduced to

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & A^T(Ax-b)+\mathcal{J}g(x)^T[g(x)]_-=0. \end{array} \tag{2.8}$$

Although Problem (2.8) is a convex optimization problem, we can not handle the constraints easily because they are nonsmooth equalities. We have to transform the constraints to smoothing constraints and then construct numerical algorithms. Noting that the constraint in (2.8), by introducing an artificial vector $y\in\Re^p$, can be expressed as

$$\begin{array}{l} A^T(Ax-b)-\mathcal{J}g(x)^Ty=0, \\ 0\leq y\perp g(x)+y\geq 0. \end{array}$$

Defining $z=g(x)+y$, the above system can be rewritten as

$$F(x,y,z)=0, \quad (y,z)\in\Omega, \tag{2.9}$$

where

$$F(x,y,z)=\left[\begin{array}{l} A^T(Ax-b)-\mathcal{J}g(x)^Ty \\ g(x)+y-z \end{array}\right] \quad \text{and} \quad \Omega=\{(y,z): 0\leq y\perp z\geq 0\}. \tag{2.10}$$

Therefore, Problem (2.5) is equivalently expressed as

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & F(x,y,z)=0, \\ & (y,z)\in\Omega. \end{array} \tag{2.11}$$

The Jacobian of F at (x,y,z) is of the form

$$\mathcal{J}F(x,y,z)=\begin{bmatrix} A^TA-\sum_{j=1}^{p}y_j\nabla^2 g_j(x) & -\mathcal{J}g(x)^T & 0 \\ \mathcal{J}g(x) & I_p & -I_p \end{bmatrix}. \tag{2.12}$$

Define

$$\Omega=\{(y,z)\in\Re^p\times\Re^p: 0\le y\perp z\ge 0\}.$$

We use Φ to denote the feasible set of Problem (2.11); namely,

$$\Phi=\{(x,y,z)\in\Re^n\times\Omega: F(x,y,z)=0\}.$$

Then Problem (2.11) is simplified as an MPCC problem

$$\min f(x) \quad \text{s.t. } (x,y,z)\in\Phi. \tag{2.13}$$

In the following, we derive the tangent cone, the regular normal cone and the normal cone of Φ at $(x,y,z)\in\Phi$, which are useful in developing S-stationary conditions and M-stationary conditions for Problem (2.13).

The tangent cone of Φ at (x,y,z) denoted by $T_\Phi(x,y,z)$, the regular normal cone of Φ at (x,y,z) denoted by $\widehat{N}_\Phi(x,y,z)$ and the normal cone of Φ at (x,y,z) denoted by $N_\Phi(x,y,z)$, are defined respectively by

$$T_\Phi(x,y,z)=\left\{(d_x,d_y,d_z):\begin{array}{l}\exists t_k\searrow 0,\exists(d_x^k,d_y^k,d_z^k)\to(d_x,d_y,d_z)\\ \text{satisfying } (x,y,z)+t_k(d_x^k,d_y^k,d_z^k)\in\Phi\end{array}\right\};$$

$$\widehat{N}_\Phi(x,y,z)=\left\{(v_x,v_y,v_z):\begin{array}{l}\langle(v_x,v_y,v_z),(x',y',z')-(x,y,z)\rangle\\ \le \mathrm{o}(\|(x',y',z')-(x,y,z)\|),(x',y',z')\in\Phi\end{array}\right\};$$

$$N_\Phi(x,y,z)=\left\{(v_x,v_y,v_z):\begin{array}{l}\exists(x^k,y^k,z^k)\xrightarrow{\Phi}(x,y,z),\exists(v_x^k,v_y^k,v_z^k)\to(v_x,v_y,v_z)\\ \text{satisfying } (v_x^k,v_y^k,v_z^k)\in\widehat{N}_\Phi(x^k,y^k,z^k)\end{array}\right\}.$$

Let $\omega=\{(\zeta_1,\zeta_2)\in\Re^2_+:\zeta_1\zeta_2=0\}$. For Θ with complementarity constraints, we have the following lemma about the variational geometry of Θ at a point $(\bar{a},\bar{b})\in\Theta$.

Lemma 2.1. *For $(\bar{a},\bar{b})\in\Omega$, the tangent cone, the regular normal cone and normal cone of Ω at $(\bar{a},\bar{b})$ are calculated by*

$$T_\Omega(\bar{a},\bar{b})=\bigotimes_{i=1}^{p}T_\omega(\bar{a}_i,\bar{b}_i),\quad \widehat{N}_\Omega(\bar{a},\bar{b})=\bigotimes_{i=1}^{p}\widehat{N}_\omega(\bar{a}_i,\bar{b}_i)\quad \textit{and}\quad N_\Omega(\bar{a},\bar{b})=\bigotimes_{i=1}^{p}N_\omega(\bar{a}_i,\bar{b}_i),$$

where

$$\begin{aligned}
&\bigotimes_{i=1}^{p} T_\omega(\bar{a}_i,\bar{b}_i) = \left\{ (u,v) \,|\, (u_i,v_i) \in T_\omega(\bar{a}_i,\bar{b}_i),\ i=1,\cdots,p \right\},\\
&\bigotimes_{i=1}^{p} \widehat{N}_\omega(\bar{a}_i,\bar{b}_i) = \left\{ (u,v) \,|\, (u_i,v_i) \in \widehat{N}_\omega(\bar{a}_i,\bar{b}_i),\ i=1,\cdots,p \right\},\\
&\bigotimes_{i=1}^{p} N_\omega(\bar{a}_i,\bar{b}_i) = \left\{ (u,v) \,|\, (u_i,v_i) \in N_\omega(\bar{a}_i,\bar{b}_i),\ i=1,\cdots,p \right\},
\end{aligned}$$

and

$$T_\omega(\bar{a}_i,\bar{b}_i) = \begin{cases} \Re\times\{0\}, & \text{if } a_i>0,\ b_i=0;\\ \{0\}\times\Re, & \text{if } a_i=0,\ b_i>0;\\ \omega, & \text{if } a_i=0,\ b_i=0, \end{cases}$$

$$\widehat{N}_\omega(\bar{a}_i,\bar{b}_i) = \begin{cases} \{0\}\times\Re, & \text{if } a_i>0,\ b_i=0;\\ \Re\times\{0\}, & \text{if } a_i=0,\ b_i>0;\\ \Re_-\times\Re_-, & \text{if } a_i=0,\ b_i=0, \end{cases}$$

$$N_\omega(\bar{a}_i,\bar{b}_i) = \begin{cases} \{0\}\times\Re, & \text{if } a_i>0,\ b_i=0;\\ \Re\times\{0\}, & \text{if } a_i=0,\ b_i>0;\\ (\Re\times\{0\})\bigcup(\{0\}\times\Re)\bigcup(\Re_-\times\Re_-), & \text{if } a_i=0,\ b_i=0. \end{cases}$$

For deriving the tangent cone, the regular normal cone and the normal cone of Φ at $(x,y,z)\in\Phi$, we need the following assumption.

Define

$$H(x,y) := A^TA - \sum_{j=1}^{p} y_j \nabla^2 g_j(x). \tag{2.14}$$

The nonsingularity of $H(x,y)$ is equivalent to that the Jacobian $\mathcal{J}F(x,y,z)$ given by (2.12) is of full row rank.

For $(y,z)\in\Omega$, define

$$\alpha := \{i : y_i>0,\ z_i=0\},\quad \beta := \{i : y_i=z_i=0\},\quad \gamma := \{i : y_i=0,\ z_i>0\}. \tag{2.15}$$

Proposition 2.1. *Let $(x,y,z)\in\Phi$ be a given point.*

(i) *One has that*

$$\widehat{N}_\Phi(x,y,z) \supseteq \left\{ \begin{pmatrix} H(x,y)\eta_1 + \mathcal{J}g(x)^T\eta_2 \\ -\mathcal{J}g(x)\eta_1 + \eta_2 + \xi_a \\ -\eta_2 + \xi_b \end{pmatrix} : \begin{array}{l} (\eta_1,\eta_2)\in\Re^{n+p} \\ (\xi_a,\xi_b)\in\widehat{N}_\Omega(y,z) \end{array} \right\}. \tag{2.16}$$

(ii) *If the matrix $H(x,y)$ is nonsingular, then*

$$T_\Phi(x,y,z)=\left\{ d\in\Re^n\times\Re^p\times\Re^p:\ \begin{array}{l} H(x,y)d_x-\mathcal{J}g(x)^Td_y=0\\ \mathcal{J}g(x)d_x+d_y-d_z=0\\ (d_y,d_z)\in T_\Omega(y,z)\end{array}\right\}. \tag{2.17}$$

(iii) *If $H(x,y)+\mathcal{J}g_\alpha^T\mathcal{J}g_\alpha$ is positively definite, then*

$$N_\Phi(x,y,z)\subseteq\left\{ \begin{pmatrix} H(x,y)\eta_1+\mathcal{J}g(x)^T\eta_2\\ -\mathcal{J}g(x)\eta_1+\eta_2+\xi_a\\ -\eta_2+\xi_b\end{pmatrix}:\ \begin{array}{l}(\eta_1,\eta_2)\in\Re^{n+p}\\ (\xi_a,\xi_b)\in N_\Omega(y,z)\end{array}\right\}. \tag{2.18}$$

Proof. It follows from Theorem 6.14 of [17] that

$$\widehat{N}_\Phi(x,y,z)\supseteq\mathcal{J}F(x,y,z)^*(\Re^n\times\Re^p\times\Re^p)+\widehat{N}_{\Re^n\times\Omega}(x,y,z),$$

where

$$\widehat{N}_{\Re^n\times\Omega}(x,y,z))=\{0_n\}\times\widehat{N}_\Omega(y,z).$$

Thus we obtain the inclusion (2.16). Now we prove (2.17) when $H(x,y)$ is nonsingular. Since the Jacobian $\mathcal{J}F(x,y,z)$ given by (2.12) is of full row rank, we now prove the following equality (see Exercise 6.7 of [17] for similar result):

$$T_\Phi(x,y,z)=\left\{d:\,(d_y,d_z)\in T_\Omega(y,z):\,\mathcal{J}F(x,y,z)d=0\right\}. \tag{2.19}$$

It is obvious that the set in the left hand-side is contained in the right hand-side and hence we only need to prove the opposite inclusion. For any $d=(d_x,d_y,d_z)$ satisfying $(d_y,d_z)\in T_\Omega(y,z)$, $\mathcal{J}F(x,y,z)d=0$, one has that there exist $d^k=(d_x^k,d_y^k,d_z^k)\to d$ and $t_k\searrow 0$ such that

$$(x,(y,z))+t_k(d_x^k,(d_y^k,d_z^k))\in\Re^n\times\Omega.$$

It follows from Lemma 2.1 that $[d_y]_i[d_z]_i=0$ for $i=1,\cdots,p$. For α, β and γ defined by (2.15). Let

$$\beta_a:=\left\{i\in\beta:\,[d_y]_i>0,[d_z]_i=0\right\},\ \beta_b:=\left\{i\in\beta:\,[d_y]_i=[d_z]_i=0\right\},\ \beta_c:=\left\{i\in\beta:\,[d_y]_i=0,[d_z]_i>0\right\}$$

and

$$\Gamma_d:=\left\{(y,z)\in\Re^p\times\Re^p:\ \begin{array}{l}(y_{\alpha\cup\beta_a},z_{\alpha\cup\beta_a})\in\Re_+^{|\alpha|+|\beta_a|}\times\{0_{|\alpha|+|\beta_a|}\}\\ (y_{\beta_c\cup\gamma},z_{\beta_c\cup\gamma})\in\{0_{|\beta_c|+|\gamma|}\}\times\Re_+^{|\beta_c|+|\gamma|}\\ (y_{\beta_b},z_{\beta_b})=(0_{|\beta_b|},0_{|\beta_b|})\end{array}\right\}.$$

Then Γ_d is a convex set and $\Gamma_d \subset \Omega$. Since the Jacobian $\mathcal{J}F(x,y,z)$ given by (2.12) is of full row rank, it follows from Theorem 2.87 of [3] that there exist a neighborhood $\mathcal{V}$ of (x,y,z) and a positive constant κ such that

$$\begin{aligned}&\text{dist}\Big((x',y',z')),\, [\Re^n \times \Gamma_d] \cap F^{-1}(0)\Big)\\ \le &\kappa \big\| F(x',y',z'), \Pi_{\Re^n\times\Gamma_d}(x',y',z')\big\|, \quad (x',y',z')\in\mathcal{V}.\end{aligned}$$

Noticing that for $(x^k,y^k,z^k)=(x,y,z)+t_kd^k$, and

$$\begin{aligned}&\Big([y]^k_{\alpha\cup\beta_a},[z]^k_{\alpha\cup\beta_a}\Big)\in\Re_+^{|\alpha|+|\beta_a|}\times\{0_{|\alpha|+|\beta_a|}\},\\ &\Big([y]^k_{\beta_c\cup\gamma},[z]^k_{\beta_c\cup\gamma}\Big)\in\{0_{|\beta_c|+|\gamma|}\}\times\Re_+^{|\beta_c|+|\gamma|},\\ &\Big([y]^k_{\beta_b},[z]^k_{\beta_b}\Big)=\Big(t_k[d^k_y]_{\beta_b},t_k[d^k_z]_{\beta_b}\Big)=\text{o}(t_k),\end{aligned}$$

we have that

$$\begin{aligned}&\text{dist}\Big((x,y,z)+t_kd^k,\Phi\Big)\\ =&\text{dist}\Big((x,y,z)+t_kd^k,[\Re^n\times\Omega]\cap F^{-1}(0)\Big)\\ =&\text{dist}\Big((x,y,z)+t_kd^k,[\Re^n\times\Gamma_d]\cap F^{-1}(0)\Big)\\ \le&\kappa\Big[\Big\|(t_k[d^k_y]_{\beta_b},t_k[d^k_z]_{\beta_b})\Big\|+\Big\|F(x^k,y^k,z^k)\Big\|\Big]\\ =&\kappa\Big\|F(x,y,z)+t_k\mathcal{J}F(x,y,z)d^k+\int_0^1[\mathcal{J}F((x,y,z)+st_kd^k)-\mathcal{J}F(x,y,z)]\text{d}st_kd^k]\Big\|\\ &+\kappa\Big\|(t_k[d^k_y]_{\beta_b},t_k[d^k_z]_{\beta_b})\Big\|\\ =&\text{o}(t_k),\end{aligned}$$

which implies that $d\in T_\Phi(x,y,z)$. Therefore we obtain equality (2.19).

Now we prove formula (2.18) when $H(x,y)$ is positively definite. We consider the following conditions

$$\left(\begin{array}{l} H(x,y)\eta_1+\mathcal{J}g(x)^T\eta_2\\ -\mathcal{J}g(x)\eta_1+\eta_2+\xi_a\\ -\eta_2+\xi_b\end{array}\right)=0,\quad (\xi_a,\xi_b)\in N_\Omega(y,z). \tag{2.20}$$

The above conditions imply

$$\left(\begin{array}{l} H(x,y)\eta_1+\mathcal{J}g(x)^T\xi_b\\ -\mathcal{J}g(x)\eta_1+\xi_a+\xi_b\end{array}\right)=0,\quad (\xi_a,\xi_b)\in N_\Omega(y,z).$$

Multiplying η_i^T to both sides of the first line of the above relations, we obtain from the second line of the above relations that

$$\eta_1^TH(x,y)\eta_1+[\xi_a+\xi_b]^T\xi_b=0,$$

which implies, from the positive definiteness of $H(x,y)$, and $[\xi_a]_i[\xi_b]_i \geq$ for $i=1,\cdots,p$, that $\eta_1=0$ and $\xi_b=0$. Therefore, we obtain from (2.20) that $\eta_1=0$, $\eta_2=0$ and $(\xi_a,\xi_b)=(0,0)$. Thus we have the basic constraint qualification in Theorem 6.14 of [17] is satisfied and we have that

$$N_\Phi(x,y,z) \subseteq \mathcal{J}F(x,y,z)^T(\Re^n \times \Re^p \times \Re^p) + N_{\Re^n \times \Omega}(x,y,z),$$

where

$$N_{\Re^n \times \Omega}(x,y,z)) = \{0_n\} \times N_\Omega(y,z).$$

Thus we obtain formula (2.18). The proof is completed. □

From the above lemma, we can easily develop the necessary optimality conditions for a local minimizer of Problem (2.11). For this purpose, we define

$$\alpha^* := \{i : y_i^* > 0 = z_i^*\}, \quad \beta^* := \{i : y_i^* = 0 = z_i^*\}, \quad \gamma^* := \{i : y_i^* = 0 < z_i^*\}.$$

Theorem 2.1 (M-stationary point)**.** *Let (x^*,y^*,z^*) be a local minimizer of Problem* (2.11)*. Let $H(x^*,y^*)$ is positively definite. Then there exist $\lambda^* \in \Re^n$, $[\xi_a]^*_{\beta^*} \in \Re^{|\beta^*|}$ and $[\xi_b]^*_{\beta^*} \in \Re^{|\beta^*|}$ satisfying*

$$\Big\{[\xi_a]_i^*[\xi_b]_i^* = 0\Big\} \quad or \quad \Big\{[\xi_a]_i^* \leq 0 \ and \ [\xi_b]_i^* \leq 0\Big\} \quad \forall i \in \beta^*, \tag{2.21}$$

such that

$$\begin{aligned} &\nabla f(x^*) + \left[H(x^*,y^*) + \mathcal{J}g_{\alpha^*}(x^*)^T \mathcal{J}g_{\alpha^*}(x^*)\right]\lambda^* + \mathcal{J}g_{\beta^*}(x^*)^T[\xi_b]^*_{\beta^*} = 0, \\ &\mathcal{J}g_{\beta^*}(x^*)\lambda^* + [\xi_a]^*_{\beta^*} + [\xi_b]^*_{\beta^*} = 0. \end{aligned} \tag{2.22}$$

Proof. We have from Theorem 6.12 of [17] that

$$0 \in \nabla_{x,y,z} f(x^*) + N_\Phi(x^*,y^*,z^*).$$

From Proposition 2.1 (iii) that there exist $(\eta_1^*,\eta_2^*) \in \Re^{n+p}$ and $(\xi_a^*,\xi_b^*) \in N_\Omega(y^*,z^*)$ such that

$$\begin{pmatrix} \nabla f(x^*) \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} H(x^*,y^*)\eta_1^* + \mathcal{J}g(x^*)^T\eta_2^* \\ -\mathcal{J}g(x^*)\eta_1^* + \eta_2^* + \xi_a^* \\ -\eta_2^* + \xi_b^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

which implies

$$\begin{pmatrix} \nabla f(x^*) \\ 0 \end{pmatrix} + \begin{pmatrix} H(x^*,y^*)\eta_1^* + \mathcal{J}g(x^*)^T\xi_b^* \\ -\mathcal{J}g(x^*)\eta_1^* + \xi_a^* + \xi_b^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Since $[\xi_b^*]_{\gamma^*} = 0$ and $[\xi_a^*]_{\alpha^*} = 0$, we have from the second line of the above equalities that $[\xi_b^*]_{\alpha^*} = \mathcal{J}g_{\alpha^*}(x^*)\eta_1^*$. Thus we obtain from the first line of the above equalities that

$$\nabla f(x^*) + H(x^*,y^*)\eta_1^* + \mathcal{J}g_{\alpha^*}(x^*)^T\mathcal{J}g_{\alpha^*}(x^*)\eta_1^* + \mathcal{J}g_{\beta^*}(x^*)^T[\xi_b^*]_{\beta^*} = 0.$$

Denote $\eta_1^* = \lambda^*$, we obtain (2.22), where $[\xi_a]^*_{\beta^*} \in \Re^{|\beta^*|}$ and $[\xi_b]^*_{\beta^*} \in \Re^{|\beta^*|}$ satisfy (2.21). The proof is completed. □

Define

$$G(x,y,z):=\begin{bmatrix} A^T(Ax-b)-\mathcal{J}g(x)^Ty \\ g(x)+y-z \\ \min\{y,z\} \end{bmatrix}. \tag{2.23}$$

Then Problem (2.11) is expressed as

$$\begin{aligned} \min_{x,y,z} \quad & f(x) \\ \text{s.t.} \quad & G(x,y,z)=0. \end{aligned} \tag{2.24}$$

Noticing that G is a Lipschitz continuous mapping, Problem (2.24) is a Lipschitz continuous optimization problem. So we may use the optimality conditions for Lipschitz continuous optimization developed in Clarke (1983). This leads to the so-called C-stationary point. We say that the point (x^*,y^*,z^*) is a C-stationary point if there exist $\lambda^*\in\Re^n$, $[\xi_a]^*_{\beta^*}\in\Re^{|\beta^*|}$ and $[\xi_b]^*_{\beta^*}\in\Re^{|\beta^*|}$ satisfying

$$[\xi_a]^*_i[\xi_b]^*_i\ge 0 \quad \forall i\in\beta^*, \tag{2.25}$$

such that

$$\begin{aligned} &\nabla f(x^*)+\left[H(x^*,y^*)+\mathcal{J}g_{\alpha^*}(x^*)^T\mathcal{J}g_{\alpha^*}(x^*)\right]\lambda^*+\mathcal{J}g_{\beta^*}(x^*)^T[\xi_b]^*_{\beta^*}=0,\\ &\mathcal{J}g_{\beta^*}(x^*)\lambda^*+[\xi_a]^*_{\beta^*}+[\xi_b]^*_{\beta^*}=0. \end{aligned} \tag{2.26}$$

In the following, we will see that for Problem (2.24), we may obtain a better result than conditions in (2.26) by using the optimality conditions for Lipschitz continuous optimization given by [19].

Proposition 2.2 (Fritz-John stationary point)**.** *Let (x^*,y^*,z^*) be a local minimizer of Problem (2.24). Then there exist $\lambda_0^*\in\Re_+,\lambda^*\in\Re^n$ and $[v_b]_{\beta^*}\in\Re^{|\beta^*|}$ satisfying*

$$[v_b]_i\in[0,1], \quad i\in\beta^* \tag{2.27}$$

such that

$$\lambda_0^*\nabla f(x^*)+\left[H(x^*,y^*)+\mathcal{J}g_{\alpha^*}(x^*)^T\mathcal{J}g_{\alpha^*}(x^*)+\mathcal{J}g_{\beta^*}(x^*)^T\mathrm{Diag}([v_b]_{\beta^*})\mathcal{J}g_{\beta^*}(x^*)\right]\lambda^*=0, \tag{2.28}$$

where $\mathrm{Diag}(v)=\mathrm{Diag}(v_1,\cdots,v_m)$ *for* $v\in\Re^m$.

Proof. The generalized Lagrangian of (2.24) is

$$\begin{aligned} &L^g(x,y,z,\eta_0,\eta_1,\eta_2,\xi)\\ =&\eta_0f(x)+\left\langle\eta_1,A^T(Ax-b)-\mathcal{J}g(x)^Ty\right\rangle+\langle\eta_2,g(x)+y-z\rangle+\langle\xi,\min\{y,z\}\rangle. \end{aligned}$$

It follows from the necessary optimality conditions for Lipschitz continuous optimization in [19] that there exist nonzero vectors $(\eta_0^*,\eta_1^*,\eta_2^*,\eta_3^*,\xi^*)$ with $\eta_0^*\geq 0$ such that

$$0\in\partial_c L(x^*,y^*,z^*,\eta_0^*,\eta_1^*,\eta_2^*,\xi^*),$$

where ∂_c denotes the Clarke generalized Jacobian. Noting that

$$\partial_c \min\{[y]^*,[z]^*\}=[\mathrm{Diag}(v_a)\quad \mathrm{Diag}(v_b)], \tag{2.29}$$

where $v_a\in\Re^p$ and $v_b\in\Re^p$ satisfy

$$\begin{aligned} &[v_a]_i=0, \quad [v_b]_i=1, &&\text{if } i\in\alpha^*;\\ &[v_a]_i=1, \quad [v_b]_i=0, &&\text{if } i\in\gamma^*;\\ &[v_a]_i=t, \quad [v_b]_i=1-t, &&\text{for some } t\in[0,1] \text{ if } i\in\beta^*. \end{aligned} \tag{2.30}$$

Then we get from $0\in\partial_c L(x^*,y^*,z^*,\eta_0^*,\eta_1^*,\eta_2^*,\xi^*)$ that there exist $v_a\in\Re^p$ and $v_b\in\Re^p$ satisfying (2.30) such that

$$\begin{pmatrix} \eta_0^*\nabla f(x^*)+H(x^*,y^*)\eta_1^*+\mathcal{J}g(x^*)^T\eta_2^* \\ -\mathcal{J}g(x^*)\eta_1^*+\eta_2^*+\mathrm{Diag}(v_a)\xi^* \\ -\eta_2^*+\mathrm{Diag}(v_b)\xi^* \end{pmatrix}=0.$$

This set of equalities can be simplified as

$$\begin{pmatrix} \eta_0^*\nabla f(x^*)+H(x^*,y^*)\eta_1^*+\mathcal{J}g(x^*)^T\mathrm{Diag}(v_b)\xi^* \\ -\mathcal{J}g(x^*)\eta_1^*+\mathrm{Diag}(v_b)\xi^*+\mathrm{Diag}(v_a)\xi^* \end{pmatrix}=0. \tag{2.31}$$

In view of (2.30), we have

$$\mathrm{Diag}(v_b)+\mathrm{Diag}(v_a)=I_p.$$

So we get from (2.31) that $\xi^*=\mathcal{J}g(x^*)\eta_1^*$. Substituting this expression back to the first equation in (2.31), we obtain

$$\eta_0^*\nabla f(x^*)+H(x^*,y^*)\eta_1^*+\mathcal{J}g(x^*)^T\mathrm{Diag}(v_b)\mathcal{J}g(x^*)\eta_1^*=0.$$

Setting $\eta_0^*=\lambda_0^*$ and $\eta_1^*=\lambda^*$, using (2.30) again, we obtain (2.28) where $[v_b]_{\beta^*}$ satisfies (2.27). □

From Proposition 2.2, we obtain an elegant set of necessary optimality conditions as follows.

Theorem 2.2. *Let* (x^*,y^*,z^*) *be a local minimizer of Problem* (2.24). *Suppose that the matrix*

$$\Big[H(x^*,y^*)+\mathcal{J}g_{\alpha^*}(x^*)^T\mathcal{J}g_{\alpha^*}(x^*)\Big] \tag{2.32}$$

is positively definite. Then there exist $\lambda^*\in\Re^n$ *and* $[v_b]_{\beta^*}\in\Re^{|\beta^*|}$ *satisfying*

$$[v_b]_i\in[0,1],\quad i\in\beta^* \tag{2.33}$$

such that

$$\nabla f(x^*)+\left[H(x^*,y^*)+\mathcal{J}g_{\alpha^*}(x^*)^T\mathcal{J}g_{\alpha^*}(x^*)+\mathcal{J}g_{\beta^*}(x^*)^T\mathrm{Diag}([v_b]_{\beta^*})\mathcal{J}g_{\beta^*}(x^*)\right]\lambda^*=0. \tag{2.34}$$

Definition 2.1. *We say that* (x^*,y^*) *is a L-stationary point for Problem* (2.24) *if there exists* λ^* *such that* (2.34) *is satisfied and* (2.34) *is called L-stationary condition.*

Let us introduce the following notation

$$\mathcal{S}^*=\left\{(x^*,y^*,\lambda^*)\in\Re^{2n}:\begin{array}{c}\exists[v_b]_{\beta^*}\in\Re^{|\beta^*|}:\ [v_b]_i\in[0,1],\ i\in\beta^*\text{ such that}\\ \nabla f(x^*)+\Big[H(x^*,y^*)+\mathcal{J}g_{\alpha^*}(x^*)^T\mathcal{J}g_{\alpha^*}(x^*)\\ +\mathcal{J}g_{\beta^*}(x^*)^T\mathrm{Diag}([v_b]_{\beta^*})\mathcal{J}g_{\beta^*}(x^*)\Big]\lambda^*=0\end{array}\right\}. \tag{2.35}$$

In the next section, we will propose a smoothing function method to generate a sequence of $\{(x^k,y^k,\lambda^k)\}$, whose any cluster point is an element of $\mathcal{S}^*$.

Remark 2.2. The condition that matrix (2.32) is positively definite is not strict because it holds if either A^TA is positively definite or $\nabla^2c_i(x^*)$ is positively definite and $y_i^*>0$ for some index i since $y^*\geq 0$.

2.3 Non-convex NLP with least constraint violation

In this subsection, we consider the non-convex nonlinear programming problem (1.1), namely the problem of the following form

$$\begin{array}{ll}\min & f(x)\\ \text{s.t.} & h(x)=0,\\ & g(x)\geq 0,\end{array}$$

where $f:\Re^n\to\Re$, $h:\Re^n\to\Re^q$ and $g:\Re^n\to\Re^p$. The optimization problem with least constraint violation for Problem (1.1) is Problem (2.5). Since solving Problem (2.5) is very difficult, it is a natural idea that we use $[\nabla\theta]^{-1}(0)$ to replace $\mathrm{argmin}_w\theta(w)$ in Problem (2.5), namely the problem of the form

$$\begin{array}{ll}\min & f(x)\\ \text{s.t.} & \mathcal{J}h(x)^Th(x)+\mathcal{J}g(x)^T[g(x)]_-=0.\end{array} \tag{2.36}$$

Problem (2.36) is a relaxed problem of Problem (2.5) because

$$\operatorname{argmin}_x \theta(x) \subseteq [\nabla\theta]^{-1}(0) = \{x\colon \nabla\theta(x)=0\} = \Big\{x\colon \mathcal{J}h(x)^T h(x) + \mathcal{J}g(x)^T [g(x)]_- = 0\Big\}.$$

Noting that the constraint in (2.36), by introducing an artificial vector $y \in \Re^p$, can be expressed as

$$\begin{aligned}&\mathcal{J}h(x)^T h(x) - \mathcal{J}g(x)^T y = 0,\\ &0 \le y \perp g(x) + y \ge 0.\end{aligned}$$

Defining $z := g(x)+y$, the above system can be rewritten as

$$\mathcal{F}(x,y,z) = 0, \quad (y,z) \in \Omega, \tag{2.37}$$

where

$$\mathcal{F}(x,y,z) = \begin{bmatrix} \mathcal{J}h(x)^T h(x) - \mathcal{J}g(x)^T y \\ g(x) + y - z \end{bmatrix} \quad \text{and} \quad \Omega = \{(y,z)\colon 0 \le y \perp z \ge 0\}. \tag{2.38}$$

Therefore, Problem (2.36) is equivalently expressed as

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & \mathcal{F}(x,y,z) = 0, \\ & (y,z) \in \Omega. \end{aligned} \tag{2.39}$$

If h and g are twice continuously differentiable, then the Jacobian of F at (x,y,z) is of the form

$$\mathcal{J}\mathcal{F}(x,y,z) = \begin{bmatrix} \sum_{j=1}^{q} h_j(x)\nabla^2 h_j(x) + \mathcal{J}h(x)^T \mathcal{J}h(x) - \sum_{j=1}^{p} y_j \nabla^2 g_j(x) & -\mathcal{J}g(x)^T & 0 \\ \mathcal{J}g(x) & I_p & -I_p \end{bmatrix}. \tag{2.40}$$

Problem (2.39) is simplified as an MPCC problem

$$\min f(x) \quad \text{s.t. } (x,y,z) \in \Phi, \tag{2.41}$$

where Φ can be expressed as

$$\Phi = \{(x,y,z) \in \Re^n \times \Omega\colon\ \mathcal{F}(x,y,z) = 0\}.$$

Define

$$\mathcal{H}(x,y) = \sum_{j=1}^{q} h_j(x)\nabla^2 h_j(x) + \mathcal{J}h(x)^T \mathcal{J}h(x) - \sum_{j=1}^{p} y_j \nabla^2 g_j(x). \tag{2.42}$$

The nonsingularity of $\mathcal{H}(x,y)$ is equivalent to that the Jacobian $\mathcal{J}\mathcal{F}(x,y,z)$ given by (2.40) is of full row rank.

Different from the conclusions of Proposition 2.1 for the convex nonlinear programming problem, we have the following results about the variational geometry of Φ for the non-convex nonlinear programming problem.

Proposition 2.3. *Let* $(x,y,z)\in\Phi$ *be a given point.*

(i) *One has that*

$$\widehat{N}_\Phi(x,y,z)\supseteq\left\{\begin{pmatrix}\mathcal{H}(x,y)\eta_1+\mathcal{J}g(x)^T\eta_2\\ -\mathcal{J}g(x)\eta_1+\eta_2+\xi_a\\ -\eta_2+\xi_b\end{pmatrix}:\begin{array}{l}(\eta_1,\eta_2)\in\Re^{n+p}\\ (\xi_a,\xi_b)\in\widehat{N}_\Omega(y,z)\end{array}\right\}. \tag{2.43}$$

(ii) *Let*

$$\left.\begin{array}{r}[\mathcal{H}(x,y)+\mathcal{J}g_\alpha(x)^T\mathcal{J}g_\alpha(x)]\eta+\mathcal{J}g_\beta(x)^T[\xi_b]_\beta=0\\ -\mathcal{J}g_\beta(x)\eta+[\xi_a]_\beta+[\xi_b]_\beta=0\\ i\in\beta:\Big\{[\xi_a]_i[\xi_b]_i=0\Big\}\ \text{or}\ \Big\{[\xi_a]_i\le 0\ \text{and}\ [\xi_b]_i\le 0\Big\}\end{array}\right\}\Longrightarrow\eta=0,\ [\xi_a]_\beta=[\xi_b]_\beta=0, \tag{2.44}$$

be satisfied. Then

$$N_\Phi(x,y,z)\subseteq\left\{\begin{pmatrix}\mathcal{H}(x,y)\eta_1+\mathcal{J}g(x)^T\eta_2\\ -\mathcal{J}g(x)\eta_1+\eta_2+\xi_a\\ -\eta_2+\xi_b\end{pmatrix}:\begin{array}{l}(\eta_1,\eta_2)\in\Re^{n+p}\\ (\xi_a,\xi_b)\in N_\Omega(y,z)\end{array}\right\}. \tag{2.45}$$

(iii) *If the matrix* $\mathcal{H}(x,y)+\mathcal{J}g_\alpha(x)^T\mathcal{J}g_\alpha(x)$ *is positively definite, then the inclusion* (2.45) *holds.*

Proof. The inclusion (2.43) comes from Theorem 6.14 of [17]. Now we prove formula (2.45) when condition (2.44) holds. We consider the following conditions

$$\begin{pmatrix}\mathcal{H}(x,y)\eta_1+\mathcal{J}g(x)^T\eta_2\\ -\mathcal{J}g(x)\eta_1+\eta_2+\xi_a\\ -\eta_2+\xi_b\end{pmatrix}=0,\quad(\xi_a,\xi_b)\in N_\Omega(y,z). \tag{2.46}$$

The above conditions imply

$$\begin{pmatrix}\mathcal{H}(x,y)\eta_1+\mathcal{J}g(x)^T\xi_b\\ -\mathcal{J}g(x)\eta_1+\xi_a+\xi_b\end{pmatrix}=0,\quad(\xi_a,\xi_b)\in N_\Omega(y,z).$$

Since $[\xi_a]_\alpha=0,[\xi_b]_\gamma=0$ and for $i\in\beta$:

$$\Big\{[\xi_a]_i[\xi_b]_i=0\Big\}\quad\text{or}\quad\Big\{[\xi_a]_i\le 0\ \text{and}\ [\xi_b]_i\le 0\Big\},$$

we have

$$\begin{aligned}&[\mathcal{H}(x,y)+\mathcal{J}g_\alpha(x)^T\mathcal{J}g_\alpha(x)]\eta+\mathcal{J}g_\beta(x)^T[\xi_b]_\beta=0,\\ &-\mathcal{J}g_\beta(x)\eta+[\xi_a]_\beta+[\xi_b]_\beta=0.\end{aligned}$$

Therefore, we obtain from (2.44) that $\eta=0$ and $(\xi_a,\xi_b)=(0,0)$. Thus we have that the basic constraint qualification in Theorem 6.14 of [17] is satisfied and we have that

$$N_\Phi(x,y,z)\subseteq \mathcal{J}\mathcal{F}(x,y,z)^*(\Re^n\times\Re^p\times\Re^p)+N_{\Re^n\times\Omega}(x,y,z),$$

where

$$N_{\Re^n\times\Omega}(x,y,z))=\{0_n\}\times N_\Omega(y,z).$$

Thus we obtain formula (2.45).

If the matrix $\mathcal{H}(x,y)+\mathcal{J}g_\alpha(x)^T\mathcal{J}g_\alpha(x)$ is positively definite, then we can easily check that the basic constraint qualification in Theorem 6.14 of [17] is satisfied and formula (2.45) holds, this proves (iii). The proof is completed. □

From the above lemma, we can easily develop the necessary optimality conditions for a local minimizer of Problem (2.39). For this purpose, we define

$$\alpha^*:=\{i: y_i^*>0=z_i^*\},\quad \beta^*:=\{i: y_i^*=0=z_i^*\},\quad \gamma^*:=\{i: y_i^*=0<z_i^*\}.$$

Similar to the proof of Theorem 2.1, we can prove the following result.

Theorem 2.3 (M-stationary point). *Let (x^*,y^*,z^*) be a local minimizer of Problem (2.11). Let (2.44) be satisfied at (x^*,y^*) or $\mathcal{H}(x^*,y^*)+\mathcal{J}g_{\alpha^*}(x^*)^T\mathcal{J}g_{\alpha^*}(x^*)$ is positively definite. Then there exist $\lambda^*\in\Re^n$, $[\xi_a]^*_{\beta^*}\in\Re^{|\beta^*|}$ and $[\xi_b]^*_{\beta^*}\in\Re^{|\beta^*|}$ satisfying*

$$\Big\{[\xi_a]^*_i[\xi_b]^*_i=0\Big\}\quad or\quad \Big\{[\xi_a]^*_i\le 0\ \ and\ \ [\xi_b]^*_i\le 0\Big\}\quad \forall i\in\beta^*, \tag{2.47}$$

such that

$$\begin{aligned}&\nabla f(x^*)+\big[\mathcal{H}(x^*,y^*)+\mathcal{J}g_{\alpha^*}(x^*)^T\mathcal{J}g_{\alpha^*}(x^*)\big]\lambda^*+\mathcal{J}g_{\beta^*}(x^*)^T[\xi_b]^*_{\beta^*}=0,\\ &\mathcal{J}g_{\beta^*}(x^*)\lambda^*+[\xi_a]^*_{\beta^*}+[\xi_b]^*_{\beta^*}=0.\end{aligned} \tag{2.48}$$

Define

$$\mathcal{G}(x,y,z):=\begin{bmatrix}\mathcal{J}h(x)^Th(x)-\mathcal{J}g(x)^Ty\\ g(x)+y-z\\ \min\{y,z\}\end{bmatrix}. \tag{2.49}$$

Then Problem (2.39) is expressed as

$$\begin{aligned}\min_{x,y,z}\ \ & f(x)\\ \text{s.t.}\ \ & \mathcal{G}(x,y,z)=0.\end{aligned} \tag{2.50}$$

Like Proposition 2.2, we can obtain the following Fritz-John optimality condition.

Proposition 2.4 (Fritz-John stationary point)**.** *Let* (x^*,y^*,z^*) *be a local minimizer of Problem* (2.50)*. Then there exist* $\lambda_0^*\in\Re_+$*,* $\lambda^*\in\Re^n$ *and* $[v_b]_{\beta^*}\in\Re^{|\beta^*|}$ *satisfying*

$$[v_b]_i\in[0,1],\quad i\in\beta^* \tag{2.51}$$

such that

$$\lambda_0^*\nabla f(x^*)+\left[\mathcal{H}(x^*,y^*)+\mathcal{J}g_{\alpha^*}(x^*)^T\mathcal{J}g_{\alpha^*}(x^*)+\mathcal{J}g_{\beta^*}(x^*)^T\mathrm{Diag}([v_b]_{\beta^*})\mathcal{J}g_{\beta^*}(x^*)\right]\lambda^*=0, \tag{2.52}$$

where $\mathrm{Diag}(v)=\mathrm{Diag}(v_1,\cdots,v_m)$ *for* $v\in\Re^m$*.*

From Proposition 2.4, we obtain an elegant set of necessary optimality conditions as follows.

Theorem 2.4. *Let* (x^*,y^*,z^*) *be a local minimizer of Problem* (2.50)*. Suppose that the matrix*

$$\left[\mathcal{H}(x^*,y^*)+\mathcal{J}g_{\alpha^*}(x^*)^T\mathcal{J}g_{\alpha^*}(x^*)\right] \tag{2.53}$$

is positively definite. Then there exist $\lambda^*\in\Re^n$ *and* $[v_b]_{\beta^*}\in\Re^{|\beta^*|}$ *satisfying*

$$[v_b]_i\in[0,1],\quad i\in\beta^* \tag{2.54}$$

such that

$$\nabla f(x^*)+\left[\mathcal{H}(x^*,y^*)+\mathcal{J}g_{\alpha^*}(x^*)^T\mathcal{J}g_{\alpha^*}(x^*)+\mathcal{J}g_{\beta^*}(x^*)^T\mathrm{Diag}([v_b]_{\beta^*})\mathcal{J}g_{\beta^*}(x^*)\right]\lambda^*=0. \tag{2.55}$$

Definition 2.2. *We say that* (x^*,y^*) *is a L-stationary point for Problem* (2.50) *if there exists* λ^* *such that* (2.55) *is satisfied and* (2.55) *is called L-stationary condition.*

In the next section, we will show that any accumulation point of the sequence of points generated by the smoothing Fischer-Burmeister function method, under certain conditions, satisfies L-stationary condition in Definition 2.2.

3 The penalty method

A natural penalized problem for Problem (1.2) is

$$\min f(x)+\lambda\theta(x), \tag{3.1}$$

where $\lambda>0$ is a penalty parameter.

Let S^* be a solution set of Problem (1.2):

$$S^*=\mathrm{Argmin}\Big\{f(x):x\in\mathrm{Argmin}\theta\Big\}.$$

Assumption 3.1. Assume that for any $\alpha\in\Re$, the level set

$$\mathrm{lev}_\alpha f=\{x\in\Re^n: f(x)\le\alpha\}$$

is a non-empty and bounded set.

It follows from Assumption 3.1 that

$$f_*:=\inf_x f(x)>-\infty \tag{3.2}$$

and the solution set of Problem (1.2) is non-empty, namely $S^*\neq\emptyset$.

Proposition 3.1. *Consider Problem* (2.5). *Let f,h and g be continuous functions defined on $\Re^n$ and Assumption* 3.1 *be satisfied. Then*

- (i) *The solution set of Problem* (3.1) *is non-empty and compact.*
- (ii) *For any solution x_λ of Problem* (3.1), *$\lambda\to f(x_\lambda)$ is increasing in $\lambda>0$ and $\lambda\to\theta(x_\lambda)$ is decreasing in $\lambda>0$.*
- (iii) *For any $x^*\in S^*$, $\lim_{\lambda\to\infty} f(x_\lambda)=f(x^*)$ and $\lim_{\lambda\to\infty}\theta(x_\lambda)=\theta(x^*)$.*
- (iv) *Let the solution set of Problem* (3.1) *be $S(\lambda)$. The outer limit of $S(\lambda)$ satisfies*

$$\limsup_{\lambda\to\infty} S(\lambda)\subset S^*. \tag{3.3}$$

Proof. The assertion (i) is obvious. Let $\lambda_i>0$, $i=1,2$ and $\lambda_2>\lambda_1$. One has that

$$f(x_{\lambda_1})+\lambda_2\theta(\lambda_1)\ge f(x_{\lambda_2})+\lambda_2\theta(\lambda_2)$$

and

$$f(x_{\lambda_2})+\lambda_1\theta(\lambda_2)\ge f(x_{\lambda_1})+\lambda_1\theta(\lambda_1).$$

Summing the above two inequalities, we obtain $\theta(\lambda_2)\le\theta(\lambda_1)$. From the second inequality,

$$f(x_{\lambda_2})-f(x_{\lambda_1})\ge\lambda_1[\theta(\lambda_1)-\theta(\lambda_2)]\ge 0.$$

This proves (ii).

Now we prove (iii). For any $x^*\in S^*$, we have for $\lambda>0$,

$$f(x^*)+\lambda\theta(x^*)\ge f(x_\lambda)+\lambda\theta(\lambda). \tag{3.4}$$

From (3.4) that

$$\theta(\lambda)-\theta(x^*)\le\frac{1}{\lambda}[f(x^*)-f(x_\lambda)]\le\frac{1}{\lambda}[f(x^*)-f_*],$$

which implies $\theta(\lambda)\to\theta(x^*)$ as $\lambda\to\infty$. As $\theta(\lambda)$ is decreasing when $\lambda>0$, we have $\theta(x_\lambda)\geq\theta(x^*)$. Thus from (3.4) that

$$f(x_\lambda)\geq f(x^*). \tag{3.5}$$

Since $\{x_\lambda:\lambda>0\}$ is bounded, we have that it has a cluster point. Suppose $\overline{x}$ be a cluster point, there exists $\lambda_k\to\infty$ such that $x_{\lambda_k}\to\overline{x}$. Form the increasing property of $f(x_\lambda)$, we have that $f(x_{\lambda_k})\to f(\overline{x})$. It follows from $\theta(x_{\lambda_k})\to\theta(\overline{x})$ and $\theta(x_{\lambda_k})\to\theta(x^*)$ that $\overline{x}\in\operatorname{argmin}\theta$. Therefore we have that $f(\overline{x})\geq f(x^*)$. From (3.5), $f(x_{\lambda_k})\geq f(x^*)$ and this implies $f(\overline{x})\geq f(x^*)$. Therefore we obtain $\overline{x}\in S^*$ and $f(x_\lambda)\to\theta(x^*)$ as $\lambda\to\infty$.

For any point $\overline{x}\in\limsup_{\lambda\to\infty}S(\lambda)$. There exists a sequence $\lambda_k\to\infty$ and $x^k\in S(\lambda_k)$ such that $x^k\to\overline{x}$. Just like the above demonstration for (iii), we can easily obtain $\overline{x}\in S^*$. The proof is completed. □

From Proposition 3.1, we know that for $x^*\in S^*$, both $f(x_\lambda)$ and $\theta(x_\lambda)$ are close to $f(x^*)$ and $\theta(x^*)$, respectively when $\lambda>0$ is quite large. Now we are in a position to construct a numerical algorithm for solving Problem (3.1) when $\lambda>0$ is very large.

In the following discussions, we may assume that g are twice continuously differentiable over a bounded convex open set $\mathcal{O}$ containing $\mathrm{lev}_{\alpha_0}f$ for $\alpha_0=f(x^0)$ and f is just lower semicontinuous over this open set. Under this assumption, we can check that g and $\mathcal{J}g$ are Lipschitz continuous over $\mathcal{O}$, this implies that $\nabla\theta(x)=\mathcal{J}g(x)^T[g(x)]_-$ is Lipschitz continuous over $\mathcal{O}$, we denote the Lipschitz constant of $\nabla\theta$ by L_θ. We present the proximal gradient method for Problem (3.1) as follows:

The proximal gradient method

Step 0. Given $x^0\in\Re^n$, $\delta>0$, $c_0>0$. Set $k=0$.

Step 1. Compute $\nabla\theta(x^k)=\mathcal{J}g(x^k)^T[g(x^k)]_-$ and

$$x^{k+1}\in\operatorname{argmin}\Big\{f(x)+\lambda\nabla\theta(x^k)^T(x-x^k)+\frac{c_k}{2}\|x-x^k\|^2\Big\}. \tag{3.6}$$

Step 2. Update c_k by $c_{k+1}=c_k$ or $c_{k+1}=c_k+\delta$.

Step 3. Set $k:=k+1$ and go to Step 1.

We discuss the convergence properties of the proximal gradient method in the following two cases: the convex case and the nonconvex case.

Case A. When f is a l.s.c. convex function, $h(x)=Ax-b$ and every g_i is a twice continuously differentiable convex function.

In view of proximal mapping $P_{c_k^{-1}f}$, we may express x^{k+1} by (3.6) as

$$x^{k+1}=P_{c_k^{-1}f}\Big(x^k-c_k^{-1}\lambda\nabla\theta(x^k)\Big). \tag{3.7}$$

It follows from Theorem 20.21 of [1] that if c_k is chosen as

$$c_k \equiv L_\theta,$$

then for any $x^* \in S^*$, the sequence $\{x^k\}$ generated by the proximal gradient method satisfies

$$f(x^k)+\lambda\theta(x^k)-[f(x^*)+\lambda\theta(x^*)] \le \frac{L_\theta\|x^0-x^*\|^2}{2k}. \tag{3.8}$$

Case B. When f is a l.s.c. function, $h(x)$ and every g are twice continuously differentiable functions.

In view of proximal mapping $P_{c_k^{-1}f}$, we may express x^{k+1} by (3.6) as

$$x^{k+1} \in P_{c_k^{-1}f}\Big(x^k - c_k^{-1}\lambda\nabla\theta(x^k)\Big). \tag{3.9}$$

It follows from Lemma 2 of [2] that for $x \in \mathcal{O}$ and

$$\begin{aligned} &x^+ \in P_{c^{-1}f}(x-c^{-1}\nabla\theta(x)),\\ &f(x^+)+\lambda\theta(x^+) \le f(x)+\lambda\theta(x)-\frac{1}{2}(c-L_\theta)\|x^+-x\|^2. \end{aligned} \tag{3.10}$$

Proposition 3.2. *Let $c_k \equiv 2L_\theta$ and $\{x^k\}$ be generated by the proximal gradient method. Suppose that Assumption* 3.1 *is satisfied. Then*

(i) *The sequence $\{x^k\}$ is bounded and the sequence $\{f(x^k)+\lambda\theta(x^k)\}$ is decreasing.*

(ii) *The following series is convergent*

$$\sum_{k=0}^{\infty}\|x^{k+1}-x^k\|^2 < +\infty.$$

(iii) *Any accumulation $\overline{x}$ of $\{x^k\}$ is a stationary point of Problem* (3.1), *namely*

$$0 \in \partial f(\overline{x})+\lambda\nabla\theta(\overline{x}). \tag{3.11}$$

Proof. Denote $\psi_\lambda := f+\lambda\theta$ and $\alpha_0 := \psi_\lambda(x^0)$. Then we have

$$\{x^k\} \subset \mathrm{lev}_{\alpha_0}\psi_\lambda.$$

For any $x \in \mathrm{lev}_{\alpha_0}\psi_\lambda$, we have

$$\psi_\lambda(x) \le \alpha_0$$

and from $\theta(x) \ge 0$, we obtain $f(x) \le \alpha_0$ and thus

$$\mathrm{lev}_{\alpha_0}\psi_\lambda \le \mathrm{lev}_{\alpha_0} f,$$

which is bounded from Assumption 3.1. Therefore the sequence $\{x^k\}$ is bounded. From (3.10), we have

$$\psi_\lambda(x^k)-\psi_\lambda(x^{k+1})\geq\frac{L_\theta}{2}\|x^{k+1}-x^k\|^2, \tag{3.12}$$

which implies that the sequence $\{f(x^k)+\lambda\theta(x^k)\}$ is decreasing.

As $\{x^k\}$ is bounded, we obtain

$$\gamma_0:=\inf_k\{\psi_\lambda(x^k)\}>-\infty.$$

Making a summation for the inequality (3.12), we obtain

$$\psi_\lambda(x^0)-\gamma_*\geq\frac{L_\theta}{2}\sum_{k=1}^{\infty}\|x^{k+1}-x^k\|^2,$$

namely the assertion (ii) holds.

From the definition of x^{k+1}, we have from the optimality condition for Problem (3.6) that

$$0\in\partial f(x^{k+1})+\lambda\nabla\theta(x^k)+L_\theta(x^{k+1}-x^k). \tag{3.13}$$

Let $\overline{x}$ be an accumulation point of $\{x^k\}$. There exists a sequence of indices $\{k_i\}$ such that $x^{k_i}\to\overline{x}$. From (ii), we know $\|x^{k+1}-x^k\|\to 0$ and thus $x^{k_i+1}\to\overline{x}$. For $k=k_i$, we obtain from (3.13) that

$$0\in\partial f(x^{k_i+1})+\lambda\nabla\theta(x^{k_i})+L_\theta(x^{k_i+1}-x^{k_i}).$$

From the outer semicontinuity of ∂f, taking the limit for $i\to\infty$, we obtain the inclusion (3.11), namely (iii) holds. The proof is completed. □

4 The smoothing Fischer-Burmeister function method

4.1 The convex NLP with least constraint violation

In this subsection, we consider Problem (2.7). Assume that the constraints in Problem (2.7) are inconsistent. In this case, the optimization problem with least constraint violation is equivalent to Problem (2.13) if h is an affine mapping and g_i is a smooth concave function for $i=1,\cdots,p$. We will present a smoothing function method to solve the MPCC problem (2.11), namely the following problem

$$\begin{aligned}\min_{x,y,z}\quad & f(x)\\ \text{s.t.}\quad & F(x,y,z)=0,\\ & (y,z)\in\Omega,\end{aligned}$$

where F is defined by formula (2.38). It is well known that, for such a problem, it is not suitable to treat it as a traditional NLP problem because, as explained in [12, Example

3.1.1 and Example 3.1.2], even the basic constraint qualification (namely, the tangent cone is equal to the linearized cone at an optimal solution) does not hold. To overcome this difficulty, various relaxation approaches have been proposed dealing with the complementarity constraints. Facchinei et al. (1999) [8] and Fukushima and Pang (1999) [9] used $\psi_\varepsilon(a,b)=0$ to approximate the complementarity relation: $0\le a$, $0\le b$, $ab=0$, where $\psi_\varepsilon(a,b)$ is the smoothing Fischer-Burmeister function

$$\psi_\varepsilon(a,b):=a+b-\sqrt{a^2+b^2+2\varepsilon^2}. \tag{4.1}$$

Other relaxations of the complementarity relation can be found in for example Scholtes [18], which uses

$$a\ge 0,\quad b\ge 0,\quad ab\le\varepsilon,$$

and Lin and Fukushima [11], which uses

$$(a+\varepsilon)(b+\varepsilon)\ge\varepsilon^2 \quad\text{and}\quad ab\le\varepsilon^2.$$

In this section, we shall use $\psi_\varepsilon(a,b)=0$ to approximate the complementarity relation, where $\phi_\varepsilon(a,b)$ is the smoothing Fischer-Burmeister function defined by (4.1).

Define

$$\Psi_\varepsilon(y,z):=\begin{bmatrix}\psi_\varepsilon(y_1,z_1)\\ \vdots\\ \psi_\varepsilon(y_p,z_p)\end{bmatrix} \tag{4.2}$$

and

$$\Omega(\varepsilon):=\Big\{(y,z)\in\Re^p\times\Re^p:\ \Psi_\varepsilon(y,z)=0\Big\}. \tag{4.3}$$

Then if $(y,z)\in\Omega(\varepsilon)$, we have

$$y>0,\quad z>0\quad\text{and}\quad y_iz_i=\varepsilon^2,\quad i=1,\cdots,p.$$

Obviously, $\psi_0(a,b)=0$ if and only if $0\le a$, $0\le b$, $ab=0$. Therefore $\Omega(0)=\Omega$.

For any $(y,z)\in\Re^{2p}$, we have

$$\mathcal{J}_{y,z}\Psi_\varepsilon(y,z)=\big[\mathcal{J}_y\Psi_\varepsilon(y,z)\quad \mathcal{J}_z\Psi_\varepsilon(y,z)\big],$$

where

$$\mathcal{J}_y\Psi_\varepsilon(y,z)=\begin{bmatrix}1-\dfrac{[y]_1}{\sqrt{[y]_1^2+[z]_1^2+2\varepsilon^2}} & & \\ & \ddots & \\ & & 1-\dfrac{[y]_p}{\sqrt{[y]_p^2+[z]_p^2+2\varepsilon^2}}\end{bmatrix}$$

and

$$\mathcal{J}_z\Psi_\varepsilon(y,z)=\begin{bmatrix} 1-\dfrac{[z]_1}{\sqrt{[y]_1^2+[z]_1^2+2\varepsilon^2}} & & \\ & \ddots & \\ & & 1-\dfrac{[z]_p}{\sqrt{[y]_p^2+[z]_p^2+2\varepsilon^2}} \end{bmatrix}.$$

Let $(y,z)\in\Omega_\varepsilon$. Then for $i=1,\cdots,p$,

$$[y]_i+[z]_i-\sqrt{[y]_i^2+[z]_i^2+2\varepsilon^2}=0,$$

we have $[y]_i>0$, $[z]_i>0$ and $[y]_i[z]_i=\varepsilon^2$. Thus

$$\begin{aligned}1-\frac{[y]_i}{\sqrt{[y]_i^2+[z]_i^2+2\varepsilon^2}}&=1-\frac{[y]_i}{\sqrt{[y]_i^2+[z]_i^2+2[y]_i[z]_i}}\\&=1-\frac{[y]_i}{[y]_i+[z]_i}\\&=\frac{[z]_i}{[y]_i+[z]_i},\end{aligned}$$

and in turn we obtain

$$1-\frac{[y]_i}{\sqrt{[y]_i^2+[z]_i^2+2\varepsilon^2}}=\frac{[z]_i}{[y]_i+[z]_i},\quad 1-\frac{[z]_i}{\sqrt{[y]_i^2+[z]_i^2+2\varepsilon^2}}=\frac{[y]_i}{[y]_i+[z]_i}. \tag{4.4}$$

Obviously, for any $\varepsilon>0$, both $\mathcal{J}_y\Psi_\varepsilon(y,z)$ and $\mathcal{J}_z\Psi_\varepsilon(y,z)$ are nonsingular matrices. We can easily obtain the following conclusion.

Lemma 4.1. *Let $\varepsilon>0$. Then for any $(y,z)\in\Omega(\varepsilon)$, the linear independence constraint qualification (LICQ) holds and the tangent cone of $\Omega(\varepsilon)$ at (y,z) is*

$$T_{\Omega(\varepsilon)}(y,z)=\left\{(\triangle y,\triangle z)\in\Re^{2m}:\ \mathcal{J}_{y,z}\Psi_\varepsilon(y,z)(\triangle y,\triangle z)=0\right\}, \tag{4.5}$$

and the normal cone of $\Omega(\varepsilon)$ at (y,z) is

$$N_{\Omega(\varepsilon)}(y,z)=\widehat{N}_{\Omega(\varepsilon)}(y,z)=\mathcal{J}_{y,z}\Psi_\varepsilon(y,z)^T\Re^p. \tag{4.6}$$

We use the following problem, denoted by P_ε, to approximate Problem (2.11):

$$\begin{aligned}\min_{x,y,z}\quad & f(x)\\ \text{s.t.}\quad & F(x,y,z)=0,\\ & (y,z)\in\Omega(\varepsilon),\end{aligned} \tag{4.7}$$

where $\Omega(\varepsilon)$ is defined by (4.3). Furthermore, we use $\Phi(\varepsilon)$ to denote the feasible set for Problem (4.7); namely,

$$\Phi(\varepsilon)=\{(x,y,z)\in\Re^n\times\Omega(\varepsilon): F(x,y,z)=0\}. \tag{4.8}$$

Define

$$G_\varepsilon(x,y,z):=\begin{bmatrix} F(x,y,z)\\ \Psi_\varepsilon(y,z)\end{bmatrix}. \tag{4.9}$$

Then $\Phi(\varepsilon)$ is expressed as

$$\Phi(\varepsilon)=\{(x,y,z)\in\Re^n\times\Re^p\times\Re^p: G_\varepsilon(x,y,z)=0\}.$$

By some calculations, we obtain

$$\mathcal{J}G_\varepsilon(x,y,z)=\begin{bmatrix} H(x,y) & -\mathcal{J}g(x)^T & 0\\ \mathcal{J}g(x) & I & -I\\ 0 & \mathcal{J}_y\Psi_\varepsilon(y,z) & \mathcal{J}_z\Psi_\varepsilon(y,z). \end{bmatrix}. \tag{4.10}$$

Similarly to the proof of Proposition 2.1, we can establish the following result.

Proposition 4.1. *For $(x,y,z)\in\Phi(\varepsilon)$, if*

$$H(x,y)+\mathcal{J}g(x)^T\mathcal{J}_z\Psi_\varepsilon(y,z)\mathcal{J}g(x) \tag{4.11}$$

is nonsingular, then $\mathcal{J}G_\varepsilon(x,y,z)$ is full of row rank. In this case,

$$T_{\Phi(\varepsilon)}(x,y,z)=\{d\in\Re^n\times\Re^p\times\Re^p: \mathcal{J}G_\varepsilon(x,y,z)d=0\} \tag{4.12}$$

and

$$N_{\Phi(\varepsilon)}(x,y,z)=\widehat{N}_{\Phi(\varepsilon)}(x,y,z)=\mathcal{J}G_\varepsilon(x,y,z)^T\Re^{n+p+p}. \tag{4.13}$$

Proof. Let us check that $\mathcal{J}F_\varepsilon(x,y,z)^T$ is of full rank in column. For $\xi_1\in\Re^n$, $\xi_2\in\Re^p$ and $\xi_3\in\Re^p$, consider

$$\mathcal{J}G_\varepsilon(x,y,z)^T\begin{pmatrix}\xi_1\\ \xi_2\\ \xi_3\end{pmatrix}=0.$$

It is equivalent to

$$\begin{bmatrix} H(x,y)\xi_1+\mathcal{J}g(x)^T\xi_2\\ -\mathcal{J}g(x)\xi_1+\xi_2+\mathcal{J}_y\Psi_\varepsilon(y,z)\xi_3\\ -\xi_2+\mathcal{J}_z\Psi_\varepsilon(y,z)\xi_3\end{bmatrix}=0. \tag{4.14}$$

Noting that

$$\mathcal{J}_y\Psi_\varepsilon(y,z)+\mathcal{J}_z\Psi_\varepsilon(y,z)=I_p,$$

one has from (4.14) that

$$\begin{aligned}\xi_2 &= \mathcal{J}_z\Psi_\varepsilon(y,z)\xi_3,\\ \xi_3 &= \mathcal{J}g(x)\xi_1,\end{aligned} \tag{4.15}$$

and

$$\Big[H(x,y)+\mathcal{J}c(x)^T\mathcal{J}_z\Psi_\varepsilon(y,z)\mathcal{J}c(x)\Big]\xi_1=0. \tag{4.16}$$

From the assumption that the matrix of (4.11) is nonsingular, we obtain from (4.16) that $\xi_1=0$ and in turn from (4.15) that $\xi_3=0$ and $\xi_2=0$. Thus $\mathcal{J}G_\varepsilon(x,y,z)$ is full of row rank and hence (4.12) and (4.13) follow from Chapter 6 of [17]. □

Lemma 4.2. *For* $\Omega(\varepsilon)$ *defined by* (4.3)*, we have*

$$\lim_{\varepsilon\searrow 0}\Omega(\varepsilon)=\Omega(0). \tag{4.17}$$

Proof. For any $(y,z)\in\limsup\limits_{\varepsilon\searrow 0}\Omega(\varepsilon)$, there exist $\varepsilon_k\searrow 0$ and $([y]^k,[z]^k)\in\Omega(\varepsilon_k)$ such that $([y]^k,[z]^k)\to(y,z)$. The inclusion $([y]^k,[z]^k)\in\Omega(\varepsilon_k)$ implies that

$$[y]^k+[z]^k-\sqrt{([y]^k)^2+([z]^k)^2+2\varepsilon_k^2}=0.$$

Then, letting $k\to\infty$, we have

$$y+z-\sqrt{[y]^2+[z]^2}=0;$$

namely, $\psi_0(y,z)=0$ and $(y,z)\in\Omega(0)$. Therefore we have

$$\limsup_{\varepsilon\searrow 0}\Omega(\varepsilon)\subset\Omega(0).$$

For any $(y,z)\in\Omega(0)$, let

$$I_+=\{i\colon [y]_i>0\},\quad J_+=\{i\colon [z]_i>0\},\quad I_0=\{1,\cdots,m\}\setminus\Big(I_+\cup J_+\Big).$$

For any $\varepsilon>0$ defined $(y(\varepsilon),z(\varepsilon))$ by

$$([y]_i(\varepsilon),[z]_i(\varepsilon)):=\begin{cases}([y]_i,\varepsilon^2/[y]_i), & \text{if } i\in I_+;\\ (\varepsilon^2/[z]_i,[z]_i), & \text{if } i\in J_+;\\ (\varepsilon,\varepsilon), & \text{if } i\in I_0.\end{cases} \tag{4.18}$$

Then $\psi_\varepsilon([y]_i(\varepsilon),[z]_i(\varepsilon))=0$ for $i=1,\cdots,m$. Thus $\Psi_\varepsilon(y(\varepsilon),z(\varepsilon))=0$ or, equivalently, $(y(\varepsilon),z(\varepsilon)\in\Omega(\varepsilon)$. Obviously, $(y(\varepsilon),z(\varepsilon)\to(y,z)$. This implies that

$$\liminf_{\varepsilon\searrow 0}\Omega(\varepsilon)\supset\Omega(0).$$

Therefore $\Omega(\varepsilon)\to\Omega(0)$ as $\varepsilon\searrow 0$. □

Corollary 4.1. *Let* $\Phi(\varepsilon)$ *be defined by* (4.8). *Then*

$$\Phi(\varepsilon)\to\Phi \quad as \quad \varepsilon\searrow 0.$$

Proof. The result can be obtained by noting that $\Phi(\varepsilon)$ and Φ can be expressed as

$$\Phi(\varepsilon)=\{(x,y,z)\in\Re^n\times\Re^p\times\Re^p : F(x,y,z)=0\}\cap\Re^n\times\Omega(\varepsilon)=F^{-1}(0)\cap\Re^n\times\Omega(\varepsilon)$$

and

$$\Phi=\{(x,y,z)\in\Re^n\times\Re^p\times\Re^p : F(x,y,z)=0\}\cap\Re^n\times\Omega=F^{-1}(0)\cap\Re^n\times\Omega,$$

respectively. □

Now denote the optimal value and the (global) solution set of Problem P_ε by $\kappa(\varepsilon)$ and $S(\varepsilon)$, respectively; namely,

$$\begin{aligned}&\kappa(\varepsilon):=\inf\{f(x):(x,y,z)\in\Phi(\varepsilon)\},\\&S(\varepsilon):=\text{Argmin}\{f(x):(x,y,z^I)\in\Phi(\varepsilon)\}.\end{aligned}$$

Theorem 4.1. *Let* f *be level-bounded; namely, the level set of* f *is bounded. Let* P_ε *be defined by* (4.7), *and* $\kappa(\varepsilon)$ *and* $S(\varepsilon)$ *be its optimal value and solution set, respectively. Then the function* $\kappa(\varepsilon)$ *is continuous at* 0 *with respect to* $\Re_+$ *and the set-valued mapping* $S(\varepsilon)$ *is outer semi-continuous at* 0 *with respect to* $\Re_+$.

Proof. As f is level-bounded, we have $\kappa(\varepsilon)$ is finite and $S(\varepsilon)\neq\emptyset$ for any $\varepsilon\geq 0$. Let

$$\widehat{f}_\varepsilon(x,y,z):=f(x)+\delta_{\Phi(\varepsilon)}(x,y,z),$$

where $\delta_{\Phi(\varepsilon)}$ is the indicator function of $\Phi(\varepsilon)$. From Lemma 4.2, $\Phi(\varepsilon)\to\Omega(0)$ as $\varepsilon\searrow 0$, $\widehat{f}_\varepsilon$ epi-converges to $\widehat{f}_0$. The level-boundedness of $\widehat{f}_\varepsilon$ is easily verified for $\varepsilon\geq 0$. Therefore, we have from Theorem 7.41 of Rockafellar and Wets (1998) that the function $\kappa(\varepsilon)$ is continuous at 0 with respect to $\Re_+$ and the set-valued mapping $S(\varepsilon)$ is outer semi-continuous at 0 with respect to $\Re_+$. The proof is completed. □

If $(x,y,z)\in\Phi(\varepsilon)$ is a local minimizer of P_ε and $\mathcal{J}G_\varepsilon(x,y,z)$ is of full row rank, then there exists a vector $\xi\in\Re^{n+2p}$ such that

$$\nabla_{x,y,z}f(x)+\mathcal{J}G_\varepsilon(x,y,z)^T\xi=0,$$

which is reduced to

$$\nabla f(x)+\Big[H(x,y)+\mathcal{J}g(x)^T\mathcal{J}_z\Psi_\varepsilon(y,z)\mathcal{J}g(x)\Big]\xi_1=0.$$

This leads to the following definition.

Definition 4.1. *We say* $(x,y,z)\in\Phi(\varepsilon)$ *is a stationary point of* P_ε *if there exists a vector* $\lambda\in\Re^n$

$$\nabla f(x)+\Big[H(x,y)+\mathcal{J}g(x)^T\mathcal{J}_z\Psi_\varepsilon(y,z)\mathcal{J}g(x)\Big]\lambda=0. \tag{4.19}$$

The following theorem is about the convergence of the stationary points for P_ε, which shows that a cluster point of stationary points for P_ε is related to the condition (2.34) when $\varepsilon\searrow 0$.

Theorem 4.2. *Let* $(x(\varepsilon),y(\varepsilon),z(\varepsilon))\in\Re^{n+2p}$ *be a stationary point for* P_ε *for* $\varepsilon>0$*, with multiplier* $\lambda(\varepsilon)\in\Re^n$*. Then for any*

$$(x^*,y^*,z^*,\lambda^*)\in\limsup_{\varepsilon\searrow 0}\{(x(\varepsilon),y(\varepsilon),z(\varepsilon),\lambda(\varepsilon))\},$$

one has that $(x^*,y^*,\lambda^*)\in\mathcal{S}^*$*, where* $\mathcal{S}^*$ *is defined by* (2.35).

Proof. Let

$$(x^*,y^*,z^*,\lambda^*)\in\limsup_{\varepsilon\searrow 0}\{(x(\varepsilon),y(\varepsilon),z(\varepsilon),\lambda(\varepsilon))\}.$$

Then there exists a sequence $\varepsilon_k\searrow 0$ and $(x^k,y^k,z^k,\lambda^k)=(x(\varepsilon_k),y(\varepsilon_k),z(\varepsilon_k),\lambda(\varepsilon_k))$ such that $(x^k,y^k,z^k,\lambda^k)\to(x^*,y^*,z^*,\lambda^*)$ with

$$\nabla_x f(x^k)+\Big[H(x^k,y^k)+\mathcal{J}g(x^k)^T\mathcal{J}_z\Psi_\varepsilon(y^k,z^k)\mathcal{J}g(x^k)\Big]\lambda^k=0. \tag{4.20}$$

It follows from Lemma 4.2 that $(y^*,z^*)\in\Omega$. Define

$$\alpha:=\{i:y_i^*>0=z_i^*\},\quad \beta:=\{i:y_i^*=0=z_i^*\},\quad \gamma:=\{i:y_i^*=0<z_i^*\}.$$

Noting that

$$\mathcal{J}_z\Psi_\varepsilon(y^k,z^k)=\begin{bmatrix}\dfrac{y_1^k}{z_1^k+y_1^k} & & \\ & \ddots & \\ & & \dfrac{y_p^k}{z_p^k+y_p^k}\end{bmatrix},$$

we have

$$\frac{y_i^k}{z_i^k+y_i^k}\to\begin{cases}1, & i\in\alpha;\\ 0, & i\in\gamma.\end{cases}$$

For $i\in\beta$, since $\frac{y_i^k}{z_i^k+y_i^k}\in(0,1)$, it has an cluster point $\eta_i\in[0,1]$. Thus there exists $\{k_m:m\in\mathbf{N}\}$ such that

$$\frac{y_i^{k_m}}{z_i^{k_m}+y_i^{k_m}}\to\begin{cases}1, & i\in\alpha;\\ \eta_i, & i\in\beta;\\ 0, & i\in\gamma.\end{cases} \tag{4.21}$$

Taking the limit for $k=k_m$, $m\to\infty$ in (4.20), we obtain

$$\nabla f(x^*)+\left[H(x^*,y^*)+\mathcal{J}g_\alpha(x^*)^T\mathcal{J}g_\alpha(x^*)+\mathcal{J}g_\beta(x^*)^T\mathrm{Diag}(\eta_\beta)\mathcal{J}g_\beta(x^*)\right]\lambda^*=0 \tag{4.22}$$

with $\eta_i\in[0,1]$ for $i\in\beta$. Thus (x^*,y^*,z^*,λ^*) satisfied (2.34) and $(x^*,y^*,\lambda^*)\in\mathcal{S}^*$. The proof is completed. ☐

Theorem 4.3. *Let* $(x(\varepsilon),y(\varepsilon),z(\varepsilon))\in\Re^{n+2p}$ *be a local minimizer of* P_ε *for* $\varepsilon>0$*. Let*

$$(x^*,y^*,z^*)\in\limsup_{\varepsilon\searrow 0}\{(x(\varepsilon),y(\varepsilon),z(\varepsilon))\}.$$

If the matrix

$$\left[H(x^*,y^*)+\mathcal{J}g_\alpha(x^*)^T\mathcal{J}g_\alpha(x^*)\right] \tag{4.23}$$

is positively definite, then there exists a vector $\lambda^*\in\Re^n$ *such that* $(x^*,y^*,\lambda^*)\in\mathcal{S}^*$*.*

Proof. For

$$(x^*,y^*,z^*)\in\limsup_{\varepsilon\searrow 0}\{(x(\varepsilon),y(\varepsilon),z(\varepsilon)))\},$$

there exists a sequence $\varepsilon_k\searrow 0$ and $\left(x^k,y^k,z^k\right)=(x(\varepsilon_k),y(\varepsilon_k),z(\varepsilon_k))$ such that $\left(x^k,y^k,z^k\right)\to(x^*,y^*,z^*)$. Since the matrix in (4.23) is positively definite, the matrix

$$\left[H(x^k,y^k)+\mathcal{J}g(x^k)^T\mathcal{J}_z\Psi_\varepsilon(y^k,z^k)\mathcal{J}g(x^k)\right] \tag{4.24}$$

is positively definite for k large enough. Since (x^k,y^k,z^k) is a local minimizer of $\mathrm{P}_{\varepsilon_k}$, there exists a unique vector $\lambda^k\in\Re^n$ such that

$$\nabla f(x^k)+\left[H(x^k,y^k)+\mathcal{J}g(x^k)^T\mathcal{J}_z\Psi_\varepsilon(y^k,z^k)\mathcal{J}g(x^k)\right]\lambda^k=0.$$

Then

$$\lambda^k=-\left[H(x^k,y^k)+\mathcal{J}g(x^k)^T\mathcal{J}_z\Psi_\varepsilon(y^k,z^k)\mathcal{J}g(x^k)\right]^{-1}\nabla f(x^k)$$

and λ^k has a cluster point λ^* such that there exists $[v_b]_\beta\in\Re^{|\beta|}$ satisfying

$$[v_b]_i\in[0,1],\quad i\in\beta$$

and

$$\nabla f(x^*)+\left[H(x^*,y^*)+\mathcal{J}g_\alpha(x^*)^T\mathcal{J}g_\alpha(x^*)+\mathcal{J}g_\beta(x^*)^T\mathrm{Diag}([v_b]_\beta)\mathcal{J}g_\beta(x^*)\right]\lambda^*=0.$$

The proof is completed. ☐

From the above theorem, we see that the smoothing Fischer-Burmeister function method works for the convex nonlinear programming with inconsistent constraints. Specifically, when the positive smoothing parameter of the method approaches to zero, any point in the outer limit of the KKT-point mapping is an L-stationary point of the equivalent MPCC problem.

4.2 The non-convex NLP with least constraint violation

For the non-convex nonlinear programming problem with least constraint violation, we can also develop the smoothing Fischer-Burmeister function method as in the Subsection 4.1. As in the previous section, we may use the following problem, denoted by $\mathcal{P}_\varepsilon$, to approximate Problem (2.39):

$$\begin{aligned} \min_{x,y,z} \quad & f(x) \\ \text{s.t.} \quad & \mathcal{F}(x,y,z)=0, \\ & (y,z)\in\Omega(\varepsilon), \end{aligned} \tag{4.25}$$

where $\Omega(\varepsilon)$ is defined by (4.3). Furthermore, we use $\Upsilon(\varepsilon)$ to denote the feasible set for Problem (4.25); namely,

$$\Upsilon(\varepsilon):=\{(x,y,z)\in\Re^n\times\Omega(\varepsilon):\mathcal{F}(x,y,z)=0\}. \tag{4.26}$$

Define

$$\mathcal{G}_\varepsilon(x,y,z):=\begin{bmatrix}\mathcal{F}(x,y,z)\\ \Psi_\varepsilon(y,z)\end{bmatrix}. \tag{4.27}$$

Then $\Upsilon(\varepsilon)$ is expressed as

$$\Upsilon(\varepsilon)=\{(x,y,z)\in\Re^n\times\Re^p\times\Re^p:\mathcal{G}_\varepsilon(x,y,z)=0\}.$$

By some calculations, we obtain

$$\mathcal{J}\mathcal{G}_\varepsilon(x,y,z)=\begin{bmatrix}\mathcal{H}(x,y) & -\mathcal{J}g(x)^T & 0\\ \mathcal{J}g(x) & I & -I\\ 0 & \mathcal{J}_y\Psi_\varepsilon(y,z) & \mathcal{J}_z\Psi_\varepsilon(y,z).\end{bmatrix}. \tag{4.28}$$

Definition 4.2. *We say $(x,y,z)\in\Upsilon(\varepsilon)$ is a stationary point of $\mathcal{P}_\varepsilon$ if there exists a vector $\lambda\in\Re^n$*

$$\nabla f(x)+\Big[\mathcal{H}(x,y)+\mathcal{J}g(x)^T\mathcal{J}_z\Psi_\varepsilon(y,z)\mathcal{J}g(x)\Big]\lambda=0. \tag{4.29}$$

The following theorem is about the convergence of the stationary points for $\mathcal{P}_\varepsilon$, which shows that a cluster point of stationary points for $\mathcal{P}_\varepsilon$ is related to the condition (2.55) when $\varepsilon\searrow 0$. The proof is the same as that of Theorem 4.2.

Theorem 4.4. *Let $(x(\varepsilon),y(\varepsilon),z(\varepsilon))\in\Re^{n+2p}$ be a stationary point for $\mathcal{P}_\varepsilon$ for $\varepsilon>0$, with multiplier $\lambda(\varepsilon)\in\Re^n$. Then for any*

$$(x^*,y^*,z^*,\lambda^*)\in\limsup_{\varepsilon\searrow 0}\{(x(\varepsilon),y(\varepsilon),z(\varepsilon),\lambda(\varepsilon))\},$$

one has that (x^*,y^*,z^*,λ^*) satisfies condition (2.55), namely (x^*,y^*) is a L-stationary point.

Like Theorem 4.3, we may demonstrate the following convergence theorem for the smoothing method.

Theorem 4.5. *Let* $(x(\varepsilon),y(\varepsilon),z(\varepsilon))\in\Re^{n+2p}$ *be a local minimizer of* P_ε *for* $\varepsilon>0$. *Let*

$$(x^*,y^*,z^*)\in\limsup_{\varepsilon\searrow 0}\{(x(\varepsilon),y(\varepsilon),z(\varepsilon))\}.$$

If the matrix

$$\Big[\mathcal{H}(x^*,y^*)+\mathcal{J}g_\alpha(x^*)^T\mathcal{J}g_\alpha(x^*)\Big] \tag{4.30}$$

is positively definite, then there exists a vector $\lambda^*\in\Re^n$ *such that* (x^*,y^*,z^*,λ^*) *satisfies condition* (2.55), *namely* (x^*,y^*) *is a L-stationary point.*

5 Discussions

This paper established the optimization model with least constraint violation to model nonlinear programming problems with possible inconsistent constraints. If the constraints are consistent, the model is reduced to the original problem. When the constraints in a convex nonlinear programming problem are possible inconsistent, the model is reformulated as an MPEC problem. When the non-convex constraints in an nonlinear programming problem are possible inconsistent, a relaxed model is reformulated as an MPEC problem, too. For the nonlinear programming problem with possible inconsistent convex constraints, M-stationary property for the equivalent MPCC problem is proved for a local minimizer. Importantly, the so-called L-stationary point is proposed, from the optimality theory for Lipschitz continuous optimization. For the nonlinear programming problem with possible inconsistent non-convex constraints, the M-stationary condition as well as L-stationary property are also developed for the equivalent MPCC problem.

The smoothing Fischer-Burmeister function method is constructed to solve the equivalent MPCC problem for the nonlinear programming problem with possible convex constraints, or to solve the relaxed MPCC problem for the nonlinear programming problem with possible non-convex constraints, and in either case any accumulation point of the sequence generated by the smoothing function method is an L-stationary point.

There are many topics left to investigate for optimization with least constraint violation. When we do not know whether the feasible region is nonempty or not, the optimization problem with least violation is always feasible, this is its advantage. However, if the constrained problem is feasible, then the model involves the infeasibility measure $\theta(x)$, which is usually smooth but not twice differentiable even functions of the original problem are all twice differentiable, this brings computational difficulties. The penalty method for solving the optimization problem with least constraint violation is proposed, properties of the penalty method are developed and the convergence of the proximal gradient method for the penalized optimization problem is analyzed. The smoothing Fischer-Burmeister function method only deals with the case when the constraints are inconsistent for convex nonlinear programming, it has nothing to do with the original problem when it is feasible. Is it possible for us to propose a unified algorithm, which can solve the optimization problem with least violation no matter when the original problem is either infeasible or feasible? Another question is as follows. The smoothing function algorithm can only cope with the nonlinear programming problem with least constraint violation. Can we construct algorithms to deal with other conic optimization problems, for example nonlinear semidefinite optimization problem?

Acknowledgments

Y.-H. Dai was supported by the Natural Science Foundation of China (Nos. 11991020, 12021001, 11631013, 11971372 and 11991021), the Strategic Priority Research Program of Chinese Academy of Sciences (No. XDA27000000) and Beijing Academy of Artificial Intelligence (BAAI). L. Zhang was supported by the Natural Science Foundation of China (Nos. 11971089 and 11731013).

The authors are very grateful to Professors Ya-Xiang Yuan, Xinwei Liu and Zhongwen Chen for their useful discussions and comments. They also thank the referees for their valuable comments, which have helped to improve the quality of this paper.

References

[1] A. Beck, First-Order Methods in Optimization, SIAM, Philadelphia, 2017.

[2] J. Bolte, S. Sabach and M. Teboulle, Proximal alternating linearized minimization for nonconvex and nonsmooth problems, Math. Program., Ser. A, 146 (2014), 459-494.

[3] J. F. Bonnans and A. Shapiro, Perturbation Analysis of Optimization Problems, New York, Springer, 2000.

[4] R. H. Byrd, F. E. Curtis and J. Nocedal, Infeasibility detection and SQP methods for nonlinear optimization, SIAM J. Optim., 20:5 (2010), 2281-2299.

[5] J. V. Burke, F. E. Curtis and H. Wang, A sequential quadratic optimization algorithm with rapid infeasibility detection, SIAM J. Optim., 24 (2014), 839-872.

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

[7] Y. H. Dai, X. W. Liu and J. Sun, A primal-dual interior-point method capable of rapidly detecting infeasibility for nonlinear programs, Journal of Industrial and Management Optimization, 16:2 (2020), 1009-1035.

[8] F. Facchinei, H. Jiang and L. Qi, A smoothing method for mathematical programs with equilibrium constraints, Math. Prog., 85 (1999), 107-134.

[9] M. Fukushima and J. S. Pang, Convergence of a smoothing continuation method for mathematical problems with complementarity constraints, Ill-Posed Variational Problems and Regularization Techniques, Lecture Notes in Economics and Mathematical Systems, Vol. 477, Théra, M. and Tichatschke, R. (eds.), Springer-Verlag, Berlin/Heidelberg, 1999, 105-116.

[10] R. S. Liu, P. Mu, X. M. Yuan, S. Z. Zeng and J. Zhang, A generic first-order algorithmic framework for bi-level programming beyond lower level singleton, 2020, arxiv2006.04045.

[11] G. H. Lin and M. Fukushima, A modified relaxation scheme for mathematical programs with complementarity constraints, Ann. Oper. Res., 133 (2005), 63-84.

[12] Z. Q. Luo, J. S. Pang and D. Ralph, Mathematical Programs with Equilibrium Constraints, Cambridge University Press, 1996.
[13] J. Nocedal and S. J. Wright, Numerical Optimization, Springer Press, 1999.
[14] S. M. Robinson, Strongly regular generalized equations, Mathematics of Operations Research, 5 (1980), 43-62.
[15] S. M. Robinson, Some continuity properties of polyhedral multifunctions, Mathematical Programming Study, 14 (1981), 206-214.
[16] S. M. Robinson, Generalized equations and their solutions, Part II: Applications to nonlinear programming, Mathematical Programming Study, 19 (1982), 200-221.
[17] R. T. Rockafellar and R. J.-B. Wets, Variational Analysis, Berlin-Heidelberg: Springer-Verlag, 1998.
[18] S. Scholtes, Convergence properties of a regularization scheme for mathematical programs with complementarity constraints, SIAM J. Optim., 11 (2001), 918-936.
[19] W. Schirotzek, Nonsmooth Analysis, Berlin Heidelberg: Springer-Verlag, 2007.