Mathematical Programs with Complementarity Constraints and a Non-Lipschitz Objective: Optimality and Approximation

Lei Guo* and Xiaojun Chen†

September 16, 2019

Abstract

We consider a class of mathematical programs with complementarity constraints (MPCC) where the objective function involves a non-Lipschitz sparsity-inducing term. Due to the existence of the non-Lipschitz term, existing constraint qualifications for locally Lipschitz MPCC cannot ensure that necessary optimality conditions hold at a local minimizer. In this paper, we present necessary optimality conditions and MPCC-tailed qualifications for the non-Lipschitz MPCC. The qualifications are related to the constraints and non-Lipschitz term, which ensure that local minimizers satisfy these necessary optimality conditions. Moreover, we present an approximation method for solving the non-Lipschitz MPCC and establish its convergence. Finally, we use numerical examples of sparse solutions of linear complementarity problems and the second-best road pricing problem in transportation science to illustrate the effectiveness of our approximation method for solving the non-Lipschitz MPCC.

Keywords: Mathematical program with complementarity constraints, non-Lipschitz continuity, sparse solution, optimality condition, approximation

MSC2010 Classification: 49M20, 90C26, 90C33, 90C59

1 Introduction

Sparse solutions of systems of equalities and inequalities have been extensively studied for many important problems in signal processing, image science, statistical and machine leaning [2, 4–7, 9, 10, 12, 14, 15, 19, 33, 41, 42, 49]. In this paper, we are interested in sparse solutions of complementarity problems that is a system of inequalities with a complementarity equality. Sparse solutions of linear complementarity problems (LCP) are solutions of the following problem [13]:

$$\begin{aligned} \min \quad & \|x\|_0 \\ \text{s.t.} \quad & x \geq 0,\ Mx + q \geq 0,\ x^T(Mx+q) = 0, \end{aligned} \tag{1.1}$$

*School of Business, East China University of Science and Technology, Shanghai 200237, P.R. China. E-mail: lguo@ecust.edu.cn. This author's work was supported by the NSFC Grant (Nos. 11771287, 71632007 and 11401379).

†Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong. Email: maxjchen@polyu.edu.hk. The author's work was supported in part by Hong Kong Research Grants Council PolyU153016/16p.

where $M \in \mathbb{R}^{n\times n}, q \in \mathbb{R}^n$, and $\|x\|_0$ denotes the number of nonzero entries in x.

The non-Lipschitz function $\|x\|_p^p := \sum_{i=1}^n |x_i|^p$ $(0 < p < 1)$ has been used as a continuous approximation of $\|x\|_0$. In particular, it has been shown that there is $\bar{p} \in (0, 1)$ such that solutions of the following problem

$$\begin{aligned} \min \quad & \|x\|_p^p \\ \text{s.t.} \quad & x \geq 0,\ Mx + q \geq 0,\ x^T(Mx + q) = 0 \end{aligned} \tag{1.2}$$

with any $p \in [0, \bar{p}]$ are solutions of problem (1.1) in [13]. Since LCP can characterize the optimality conditions of linear and convex quadratic programming, problem (1.2) can be used to find sparse solutions of linear and convex quadratic programming problems.

In this paper, we consider the following problem:

$$\begin{aligned} \min \quad & F(x) := f(x) + \|Dx\|_p^p \\ \text{s.t.} \quad & G(x) \geq 0,\ H(x) \geq 0,\ G(x)^T H(x) = 0, \end{aligned} \tag{1.3}$$

where $p \in (0, 1)$, $D \in \mathbb{R}^{r\times n}$, and $f : \mathbb{R}^n \to \mathbb{R}$, $G, H : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable. The feasible region of problem (1.3) can characterize optimal solutions of convex programming problems or equilibria of equilibrium systems [34].

Problem (1.3) is a class of mathematical programs with complementarity constraints (MPCC), which has numerous applications in economics and engineering [34]. However, all existing results in the MPCC literature assume local Lipschitz continuity of functions, which are not applicable to problem (1.3) due to the existence of non-Lipschitz term $\|Dx\|_p^p$. The research in theoretical and numerical treatments for locally Lipschitz MPCC focuses mainly on the difficulty that Mangasarian-Fromovitz constraint qualification (MFCQ) is violated at every feasible point; see [39, 47]. Several stationarity concepts such as Clarke stationarity, Mordukhovich stationarity, and strong stationarity, and associated MPCC-tailed constraint qualifications, and perturbation analysis for parametric MPCC have been studied; see, e.g., [21–23, 28, 37, 39, 46, 47] and the references therein for more discussions. Moreover, many numerical methods have been proposed to solve locally Lipschitz MPCC such as regularization (or relaxation) methods [24, 40], sequential quadratic programming methods [18], augmented Lagrangian methods [27], interior point methods [32], active set methods [20, 26], and penalty function methods [25].

Optimization problems involving the non-Lipschitz term $\|\cdot\|_p^p$ $(0 < p < 1)$ and linear constraints or box constrains have been widely used to induce a sparse solution in regression, feature selection in machine learning, edge preserving image restoration, compressed sensing in signal processing, and Markowitz portfolios [4–6]. Optimality conditions and lower bound theory have been established for these problems [14]. Moreover, smoothing regularization methods [4], smoothing trust region methods [12], smoothing SQP methods [30], iteratively re-weighted minimization methods [8, 10, 15, 33, 49], and penalty function methods [11] are proposed for solving these problems, in which $\|\cdot\|_p^p$ is approximated by locally Lipschitz functions. However, existing theory and algorithms for convexly-constrained optimization with a non-Lipschitz objective function are not applicable to problem (1.3) due to the existence of complementarity constraints.

The smoothly clipped absolute deviation (SCAD) [17] is often used as a sparsity-inducing function in the literature. It has been shown that the SCAD function may be seen as a surrogate for ℓ_0 quasi-norm under two conditions ([31, Theorem 3.2 and Corollary 3.2]). However, the conditions are difficult to hold for problem (1.1). On the other hand, we have the existence of $\bar{p} \in (0,1)$ such that the solution set of problem (1.2) with $p \in (0,\bar{p})$ is contained in the solution set of problem (1.1) without conditions.

This paper aims to derive necessary optimality conditions of problem (1.3) and propose associated qualifications, and develop an approximation method for finding points satisfying these necessary optimality conditions. Our main contributions are summarized as follows.

- We investigate the applicability of the basic qualification (BQ) at $x^* \in \mathcal{F}$:

$$-\partial^\infty F(x^*) \cap \mathcal{N}_{\mathcal{F}}(x^*) = \{0\}, \tag{1.4}$$

  where $\mathcal{F} := \{x \in \mathbb{R}^n : G(x) \geq 0,\ H(x) \geq 0,\ G(x)^T H(x) = 0\}$ is the feasible region of problem (1.3), $\partial^\infty F(x^*)$ is the horizon subdifferential of F at x^*, and $\mathcal{N}_{\mathcal{F}}(x^*)$ is the limiting normal cone to $\mathcal{F}$ at x^*. The BQ ensures the validity of the sum rule of the subdifferentials of the objective function F and the indicator function $\delta_{\mathcal{F}}$ (e.g., [38, Corollary 10.9]), which plays an important role in developing necessary optimality conditions for problem (1.3) as shown in Section 2. We provide conditions that imply the failure of the BQ due to the existence of the non-Lipschitz term. In particular, the conditions imply that the BQ fails at any feasible point x with zero components for problem (1.2).

- Motivated by the stationarity concepts of locally Lipschitz MPCC, we present the Clarke (C-), Mordukhovich (M-), and strong (S-) stationarity conditions for problem (1.3). Moreover, we propose two MPCC-tailed qualifications which guarantee that M-stationarity and S-stationarity are necessary for local optimality of problem (1.3), respectively. To the best of our knowledge, there are no necessary optimality conditions for non-Lipschitz MPCC in the literature.

- We propose an approximation method for solving problem (1.3). At each step of the method, the non-Lipschitz term is approximated by a locally Lipschitz function and the complementarity constraints are relaxed such that certain constraint qualifications hold. We show in Theorem 4.1 that any accumulation point of the sequence generated by our method is C-stationary under MPCC linear independence (LI) qualification. Moreover, we present weak second-order necessary conditions (WSONC) for the approximation problems, and show in Theorem 4.2 that the accumulation point is also M-stationary (stronger than C-stationary) if MPCC-LI qualification holds and the approximation problems satisfy WSONC. We also provide a sufficient condition such that the accumulation point is S-stationary (stronger than M-stationary) in Theorem 4.3.

The rest of the paper is organized as follows. In Section 2, we study the applicability of the BQ and derive necessary optimality conditions for problem (1.3). In Section 3, we introduce the approximation problems of problem (1.3) and present second-order necessary conditions for these approximation problems. In Section 4, we propose an approximation method for solving problem (1.3) and establish its convergence. In Section 5, we present numerical results of our approximation method for solving problem (1.2) and a second-best road pricing problem in transportation science.

1.1 Notation and terminology

The following notation will be used throughout this paper. For any given $x \in \mathbb{R}^n$, $\|x\|$ denotes the Euclidean norm of x and $\mathrm{Diag}(x)$ the diagonal matrix whose ith diagonal entry is x_i. We let $\mathcal{B}_\delta(x)$ denote an open ball centered at x with positive radius δ and $|\mathcal{J}|$ the number of elements of a finite set $\mathcal{J}$. We let e denote a vector of all ones and e_i a unit vector whose i-th component is 1 with approximate dimension, and let $I_{r\times r}$ be a $r \times r$ identity matrix. We let $\nabla F(x)$ stand for the transposed Jacobian of a smooth function F at x. Given a matrix $A \in \mathbb{R}^{r\times l}$ and an index $i \in \{1,\dots,r\}$, A_i denotes the transpose of the ith row of A. For any $\theta \in \mathbb{R}$, we let

$$\mathrm{sign}\,\theta = \begin{cases} 1 & \text{if } \theta > 0, \\ [-1,1] & \text{if } \theta = 0, \\ -1 & \text{if } \theta < 0. \end{cases}$$

The feasible region of problem (1.3) is assumed to be nonempty and can be rewritten as

$$\mathcal{F} = \{x \in \mathbb{R}^n : (G(x), H(x)) \in \mathcal{C}^m \}, \tag{1.5}$$

where $\mathcal{C}^m := \{(a,b) \in \mathbb{R}^m \times \mathbb{R}^m : a \geq 0, b \geq 0, a^T b = 0\}$ and we write $\mathcal{C}$ instead of $\mathcal{C}^1$. For a given $x \in \mathcal{F}$, we define the following index sets:

$$\begin{aligned} &\mathcal{I}_0(x) := \{i = 1,\dots,r : D_i^T x = 0\}, \qquad \mathcal{I}_{\neq}(x) := \{i = 1,\dots,r : D_i^T x \neq 0\}, \\ &\mathcal{I}_G(x) := \{i = 1,\dots,m : G_i(x) = 0\}, \quad \mathcal{I}_H(x) := \{i = 1,\dots,m : H_i(x) = 0\}, \\ &\mathcal{I}_{+0}(x) := \mathcal{I}_H(x)\backslash\mathcal{I}_G(x), \quad \mathcal{I}_{00}(x) := \mathcal{I}_G(x)\cap\mathcal{I}_H(x), \quad \mathcal{I}_{0+}(x) := \mathcal{I}_G(x)\backslash\mathcal{I}_H(x). \end{aligned}$$

It is obvious that $\{\mathcal{I}_{+0}(x), \mathcal{I}_{00}(x), \mathcal{I}_{0+}(x)\}$ is a partition of $\{1,\dots,m\}$.

Given a closed set Ω and a point $x^* \in \Omega$, the regular normal cone of Ω at x^* is a closed and convex cone defined as

$$\hat{\mathcal{N}}_\Omega(x^*) := \{d : d^T(x - x^*) \leq o(\|x - x^*\|) \quad \forall x \in \Omega\},$$

where $o(\cdot)$ means that $o(\alpha)/\alpha \to 0$ as $\alpha \in \mathbb{R}_+ \to 0$, which is actually the polar of the tangent cone of Ω at x^*, and the limiting normal cone of Ω at x^* is a closed cone defined as

$$\mathcal{N}_\Omega(x^*) := \{d : \exists x^k \in \Omega, x^k \to x^*, \exists d^k \in \hat{\mathcal{N}}_\Omega(x^k) \text{ s.t. } d^k \to d\}.$$

For a continuous function $\varphi : \mathbb{R}^d \to \mathbb{R}$ and a point $x^* \in \mathbb{R}^d$, the regular subdifferential of φ at x^* is defined as

$$\hat{\partial}\varphi(x^*) := \{v : \varphi(x) \geq \varphi(x^*) + v^T(x - x^*) + o(\|x - x^*\|) \quad \forall x \in \mathbb{R}^d\},$$

the limiting subdifferential of φ at x^* is defined as

$$\partial\varphi(x^*) := \{v : \exists x^k \to x^*, v^k \in \hat{\partial}\varphi(x^k) \text{ s.t. } v^k \to v\},$$

and the horizon subdifferential of φ at x^* is defined as

$$\partial^\infty\varphi(x^*) := \{v : \exists x^k \to x^*, v^k \in \hat{\partial}\varphi(x^k) \text{ and } t_k \to 0 \text{ with } t_k \geq 0 \text{ s.t. } t_k v^k \to v\}.$$

It is well-known that when φ is continuously differentiable at x^*, then $\partial\varphi(x^*) = \{\nabla\varphi(x^*)\}$; see, e.g., [38, Exercise 8.8], and moreover, φ is Lipschitz around x^* if and only if $\partial^\infty\varphi(x^*) = \{0\}$ by [38, Theorem 9.13].

The following lemma will be employed in Section 2.

Lemma 1.1 *Let $x \in \mathcal{F}$. Then for any $(u, v) \in \mathbb{R}^m \times \mathbb{R}^m$ with $u_i = 0$ $i \in \mathcal{I}_{+0}(x)$, $v_i = 0$ $i \in \mathcal{I}_{0+}(x)$, $u_i \geq 0, v_i \geq 0$ $i \in \mathcal{I}_{00}(x)$, we have that*

$$-\nabla G(x)u - \nabla H(x)v \in \mathcal{N}_{\mathcal{F}}(x).$$

Proof. It follows from [38, Theorem 6.14] and the definition of normal cones that

$$(\nabla G(x), \nabla H(x))\hat{\mathcal{N}}_{\mathcal{C}^m}(G(x), H(x)) \subseteq \hat{\mathcal{N}}_{\mathcal{F}}(x) \subseteq \mathcal{N}_{\mathcal{F}}(x).$$

By [38, Proposition 6.41], it is easy to see that

$$\hat{\mathcal{N}}_{\mathcal{C}^m}(G(x), H(x)) = \hat{\mathcal{N}}_{\mathcal{C}}(G_1(x), H_1(x)) \times \ldots \times \hat{\mathcal{N}}_{\mathcal{C}}(G_m(x), H_m(x)),$$

where the regular normal cone $\hat{\mathcal{N}}_C(G_i(x), H_i(x))$ can be directly calculated as (e.g., [45, Proposition 2.7])

$$\hat{\mathcal{N}}_C(G_i(x), H_i(x)) = \left\{ (u_i, v_i) : \begin{array}{ll} u_i \in \mathbb{R}, v_i = 0 & \text{if } G_i(x) = 0 < H_i(x) \\ u_i = 0, v_i \in \mathbb{R} & \text{if } G_i(x) > 0 = H_i(x) \\ u_i \leq 0, v_i \leq 0 & \text{if } G_i(x) = H_i(x) = 0 \end{array} \right\}.$$

Then the desired result follows immediately from the above three formulas. ■

2 Necessary optimality conditions

To the best of our knowledge, the developed necessary optimality conditions for MPCC in the literature assume local Lipschitz continuity and even continuous differentiability of all involved functions. Since the objective function of problem (1.3) is not locally Lipschitz, all the existing results are not applicable. In this section, we will develop necessary optimality conditions for problem (1.3).

Clearly we can equivalently rewrite problem (1.3) as "min $F(x) + \delta_{\mathcal{F}}(x)$", where $\delta_{\mathcal{F}}$ is the indicator function of $\mathcal{F}$. By Fermat's rule (e.g., [38, Theorem 10.1]), the most general stationarity, $0 \in \partial(F + \delta_{\mathcal{F}})(x^*)$, holds at a local minimizer x^* of problem (1.3) without requiring any condition. A sufficient condition for a more explicit and useful stationarity, $0 \in \partial F(x^*) + \mathcal{N}_{\mathcal{F}}(x^*)$, to hold is the BQ (1.4) at x^*.

We now investigate the applicability of the BQ for problem (1.3). Consider a special case of problem (1.3)

$$\begin{aligned} \min \quad & \|Dx\|_p^p \\ \text{s.t.} \quad & x \geq 0,\ Mx + q \geq 0,\ x^T(Mx+q) = 0, \end{aligned} \tag{2.1}$$

where $D = [I_{r\times r}, 0_{r\times r_2}] \in \mathbb{R}^{r\times n}$ with $n = r + r_2$. Problem (2.1) is more general than problem (1.2) since these two problems are the same when $r_2 = 0$. Let $y := Dx$. It is clear that when $y > 0$, the objective function $\phi(\cdot) := \|\cdot\|_p^p$ is Lipschitz around x since $\partial^\infty \phi(y) = \{0\}$, and thus the BQ holds since $0 \in \mathcal{N}_{\mathcal{F}}(x)$. On the other hand, ϕ is not Lipschitz around x since $\partial^\infty \phi(y) \neq \{0\}$ when $y \not> 0$. In this case, the BQ fails as shown in the following.

Proposition 2.1 *For problem* (2.1)*, the BQ fails at all feasible point x with zero components in $\{1, \ldots, r\}$.*

Proof. Let $y := Dx$ and recall the definition of D in problem (2.1). The horizon subdifferential of ϕ at x can be easily derived as follows

$$\partial^\infty \phi(y) = \{(a, b) \in \mathbb{R}^r \times \mathbb{R}^{r_2} : a_i = 0 \text{ if } y_i > 0, b = 0\}. \tag{2.2}$$

Applying Lemma 1.1 to the reformulation $\mathcal{F} = \{x : (x, Mx + q) \in \mathcal{C}^n\}$ implies that

$$S := \left\{ -u - M^T v : \begin{array}{l} u_i = 0\ i \in \mathcal{I}_{+0}(x), v_i = 0\ i \in \mathcal{I}_{0+}(x) \\ u_i \geq 0, v_i \geq 0\ i \in \mathcal{I}_{00}(x) \end{array} \right\} \subseteq \mathcal{N}_{\mathcal{F}}(x).$$

Then to get the desired result that $-\partial^\infty \phi(y) \cap \mathcal{N}_{\mathcal{F}}(x) \neq \{0\}$, it suffices to show that $-\partial^\infty \phi(y) \cap S \neq \{0\}$. Let $i_0 \in \{1, \ldots, r\}$ be an index such that $y_{i_0} = x_{i_0} = 0$ and $e_{i_0} \in \mathbb{R}^n$. It is clear that $e_{i_0} \in \partial^\infty \phi(y)$ by (2.2). On the other hand, it is easy to see that $-e_{i_0} = -u - v$, with $u = e_{i_0}$ and $v = 0$, belongs to S. Thus we have $0 \neq -e_{i_0} \in -\partial^\infty \phi(y) \cap S$. The proof is complete. ∎

From Proposition 2.1, it is reasonable to infer that the BQ is difficult to hold for problem (1.3). We next investigate how the BQ works for problem (1.3). When D is of full row rank, we have the following result.

Proposition 2.2 *Let $x^* \in \mathcal{F}$ and $\mathcal{I}_0^* := \mathcal{I}_0(x^*) \neq \emptyset$. Assume that D is of full row rank. If there exists a vector $(\lambda, u, v) \in \mathbb{R}^{|\mathcal{I}_0^*|} \times \mathbb{R}^m \times \mathbb{R}^m$ such that*

$$\begin{aligned} & -\nabla G(x^*)u - \nabla H(x^*)v + \sum_{i \in \mathcal{I}_0^*} \lambda_i D_i = 0,\ \lambda \neq 0, \\ & u_i = 0\ i \in \mathcal{I}_{+0}(x^*),\ v_i = 0\ i \in \mathcal{I}_{0+}(x^*),\ u_i \geq 0, v_i \geq 0\ i \in \mathcal{I}_{00}(x^*), \end{aligned} \tag{2.3}$$

then the BQ fails at x^* for problem (1.3).

Proof. Using the reformulation (1.5) and Lemma 1.1, it follows that for all (u,v) with $\mu_i = 0\ i \in \mathcal{I}_{+0}(x^*),\ \nu_i = 0\ i \in \mathcal{I}_{0+}(x^*),\ \mu_i \geq 0, \nu_i \geq 0\ i \in \mathcal{I}_{00}(x^*)$, we have that

$$-\nabla G(x^*)u - \nabla H(x^*)v \in \mathcal{N}_{\mathcal{F}}(x^*).$$

This together with the assumption implies that there is $\lambda \neq 0$ such that

$$-\sum_{i\in\mathcal{I}_0^*} \lambda_i D_i \in \mathcal{N}_{\mathcal{F}}(x^*). \tag{2.4}$$

Since D is of full row rank, it is clear that $\sum_{i\in\mathcal{I}_0^*} \lambda_i D_i \neq 0$ and moreover, by [38, Exercise 10.7 and Corollary 10.9], it follows that

$$\sum_{i\in\mathcal{I}_0^*} \lambda_i D_i \in \partial^\infty F(x^*) = \{ \sum_{i\in\mathcal{I}_0^*} \mu_i D_i : \mu_i \in \mathbb{R}\ i \in \mathcal{I}_0^* \}.$$

These and (2.4) imply that

$$0 \neq -\sum_{i\in\mathcal{I}_0^*} \lambda_i D_i \in -\partial^\infty F(x^*) \cap \mathcal{N}_{\mathcal{F}}(x^*).$$

Thus the BQ fails at x^* and the proof is complete. ■

For an arbitrary matrix $D \in \mathbb{R}^{r\times n}$, one can introduce a variable $z = Dx$ and make the non-Lipschitz term separable, that is, reformulating problem (1.3) as

$$\begin{array}{ll} \min & \Psi(x,z) := f(x) + \|z\|_p^p \\ \text{s.t.} & Dx - z = 0, \\ & G(x) \geq 0,\ H(x) \geq 0,\ G(x)^T H(x) = 0. \end{array} \tag{2.5}$$

It is not difficult to verify that problem (1.3) and problem (2.5) are equivalent in the sense that x^* is a local minimizer of problem (1.3) if and only if (x^*, Dx^*) is a local minimizer of problem (2.5). Using the same proof technique as in Proposition 2.2, the following result follows immediately.

Proposition 2.3 *Let $x^* \in \mathcal{F}$ and $\mathcal{I}_0^* := \mathcal{I}_0(x^*) \neq \emptyset$. If there exists a vector $(\lambda, u, v) \in \mathbb{R}^{|\mathcal{I}_0^*|} \times \mathbb{R}^m \times \mathbb{R}^m$ such that (2.3) holds, then the BQ fails at (x^*, Dx^*) for problem (2.5).*

As shown in Propositions 2.1, 2.2, and 2.3, it is very likely that the BQ fails at a local minimizer of problem (1.3). However, it should be noted that if MPCC-LI qualification as defined in the following holds, then condition (2.3) never hold. This fact motivates us to define the following qualifications. These qualifications are actually the standard MPCC-LICQ and MPCC-RCPLD in the literature ([21, 40]) for the system

$$\{x \in \mathbb{R}^d : D_i^T x = 0\ i \in \mathcal{I}_0(x^*), (G(x), H(x)) \in \mathcal{C}^m\}.$$

Since these conditions are not only related to the constraints but also the objective function, we call them qualifications.

Definition 2.1 *Let $x^* \in \mathcal{F}$. (i) We say that MPCC-LI qualification holds at x^* for problem* (1.3) *if the following family of gradients is linearly independent:*

$$\big\{\nabla G_i(x^*) : i \in \mathcal{I}_G(x^*)\big\} \cup \big\{\nabla H_i(x^*) : i \in \mathcal{I}_H(x^*)\big\} \cup \big\{D_i : i \in \mathcal{I}_0(x^*)\big\}.$$

(ii) Let $I_1 \subseteq \mathcal{I}_0(x^*)$, $I_2 \subseteq \mathcal{I}_{0+}(x^*)$, and $I_3 \subseteq \mathcal{I}_{+0}(x^*)$ be such that $\mathcal{G}(x^*, I_1, I_2, I_3)$ is a basis for $\operatorname{span}\mathcal{G}(x^*, \mathcal{I}_0(x^*), \mathcal{I}_{0+}(x^*), \mathcal{I}_{+0}(x^*))$. *We say that MPCC relaxed constant positive linear dependence (RCPLD) qualification holds at x^* if there exists $\delta > 0$ such that*

- *$\mathcal{G}(x^*, \mathcal{I}_0(x^*), \mathcal{I}_{0+}(x^*), \mathcal{I}_{+0}(x^*))$ has the same rank for all $x \in \mathcal{B}_\delta(x^*)$;*
- *for any $I_4, I_5 \subseteq \mathcal{I}_{00}(x^*)$, if there exist $\lambda \in \mathbb{R}^{|I_1|}$, $u \in \mathbb{R}^{|I_2 \cup I_4|}$, and $v \in \mathbb{R}^{|I_3 \cup I_5|}$ such that $(\lambda, u, v) \neq 0$, $u_i v_i = 0$ or $u_i > 0, v_i > 0$ for any $i \in \mathcal{I}_{00}(x^*)$, and*

$$\sum_{i \in I_1} \lambda_i D_i - \sum_{i \in I_2 \cup I_4} u_i \nabla G_i(x^*) - \sum_{i \in I_3 \cup I_5} v_i \nabla H_i(x^*) = 0$$

then for any $x \in \mathcal{B}_\delta(x^*)$, the following family of gradients is linearly dependent:

$$\{D_i : i \in I_1\} \cup \{\nabla G_i(x) : i \in I_2 \cup I_4\} \cup \{\nabla H_i(x) : i \in I_3 \cup I_5\},$$

where $\mathcal{G}(x, I_1, I_2, I_3) := \{D_i : i \in I_1\} \cup \{\nabla G_i(x) : i \in I_2\} \cup \{\nabla H_i(x) : i \in I_3\}$.

Throughout this paper, we call these two conditions in Definition 2.1 MPCC-LI qualification and MPCC-RCPLD qualification respectively in contrast with MPCC-LICQ and MPCC-RCPLD for locally Lipschitz MPCC in the literature.

Motivated by the stationarity concepts for locally Lipschitz MPCC (see, e.g., [39,47]), we define the stationarity conditions for problem (1.3) as follows.

Definition 2.2 *Let $x^* \in \mathcal{F}$. (i) We say that x^* is a weakly (W-) stationary point of problem* (1.3) *if there exist $\lambda \in \mathbb{R}^r$, $\mu \in \mathbb{R}^m$, and $\nu \in \mathbb{R}^m$ such that*

$$\begin{aligned} &\nabla f(x^*) + D^T\lambda - \nabla G(x^*)\mu - \nabla H(x^*)\nu = 0, \\ &\lambda_i = p|D_i^T x^*|^{p-1}\operatorname{sign}(D_i^T x^*) \quad i \in \mathcal{I}_{\neq}(x^*), \\ &\mu_i = 0 \quad i \in \mathcal{I}_{+0}(x^*), \quad \nu_i = 0 \quad i \in \mathcal{I}_{0+}(x^*). \end{aligned} \tag{2.6}$$

(ii) We say that x^* is a C-stationary point of problem (1.3) *if there exist $\lambda \in \mathbb{R}^r$, $\mu \in \mathbb{R}^m$, and $\nu \in \mathbb{R}^m$ satisfying* (2.6) *and*

$$\mu_i \nu_i \geq 0 \quad i \in \mathcal{I}_{00}(x^*).$$

(iii) We say that x^* is an M-stationary point of problem (1.3) *if there exist $\lambda \in \mathbb{R}^r$, $\mu \in \mathbb{R}^m$, and $\nu \in \mathbb{R}^m$ satisfying* (2.6) *and*

$$\mu_i > 0, \quad \nu_i > 0 \quad \text{or} \quad \mu_i \nu_i = 0 \quad i \in \mathcal{I}_{00}(x^*).$$

(iv) We say that x^* is an S-stationary point of problem (1.3) *if there exist $\lambda \in \mathbb{R}^r$, $\mu \in \mathbb{R}^m$, and $\nu \in \mathbb{R}^m$ satisfying* (2.6) *and*

$$\mu_i \geq 0, \quad \nu_i \geq 0 \quad i \in \mathcal{I}_{00}(x^*).$$

A W-stationary point x^* is said to satisfy upper lever strict complementarity (ULSC) if there exist $\lambda \in \mathbb{R}^r$, $\mu \in \mathbb{R}^m$, and $\nu \in \mathbb{R}^m$ satisfying (2.6) and $\mu_i \nu_i \neq 0$ for all $i \in \mathcal{I}_{00}(x^*)$.

We now give some comments on these stationarity conditions in Definition 2.2. One can easily see that the following relations hold:

$$\text{S-stationarity} \Rightarrow \text{M-stationarity} \Rightarrow \text{C-stationarity} \Rightarrow \text{W-stationarity.}$$

It should be also pointed out that these stationarity conditions in Definition 2.2 are actually the same as those of applying directly the stationarity conditions for locally Lipschitz MPCC to problem (1.3). Take the W-stationarity for an example. The standard W-stationarity for locally Lipschitz MPCC (see, e.g., [39]) at x^* says that there exist multipliers $(\mu, \nu) \in \mathbb{R}^m \times \mathbb{R}^m$ such that

$$\begin{aligned} & 0 \in \nabla f(x^*) + \sum_{i=1}^{r} \partial(|\cdot|^p)(D_i^T x^*) D_i - \nabla G(x^*)\mu - \nabla H(x^*)\nu, \\ & \mu_i = 0 \ i \in \mathcal{I}_{+0}(x^*), \ \nu_i = 0 \ i \in \mathcal{I}_{0+}(x^*), \end{aligned}$$

which are the same as (2.6) as long as we note that the limiting subdifferential of the function $|t|^p$ is $\mathbb{R}$ at $t = 0$ and $\{p|t|^{p-1}\text{sign}\,(t)\}$ otherwise.

Now we are ready to establish necessary optimality for problem (1.3).

Theorem 2.1 *Let $x^* \in \mathcal{F}$ be a local minimizer of problem* (1.3). *(i) If MPCC-RCPLD qualification holds at x^*, then x^* is an M-stationary point of problem* (1.3). *(ii) If MPCC-LI qualification holds at x^*, then x^* is an S-stationary point of problem* (1.3).

Proof. As discussed before Proposition 2.3, (x^*, z^*) with $z^* := Dx^*$ is a local minimizer of problem (2.5). Let $\mathcal{I}_0^* := \mathcal{I}_0(x^*)$ and $\mathcal{I}_{\neq}^* := \mathcal{I}_{\neq}(x^*)$. Then it is easy to verify that $(x^*, z^*_{\mathcal{I}_{\neq}^*})$ is a local minimizer of the restricted problem

$$\begin{aligned} \min \quad & f(x) + \sum_{i \in \mathcal{I}_{\neq}^*} |z_i|^p \\ \text{s.t.} \quad & D_i^T x - z_i = 0 \ i \in \mathcal{I}_{\neq}^*, \ D_i^T x = 0 \ i \in \mathcal{I}_0^*, \\ & (G(x), H(x)) \in \mathcal{C}^m. \end{aligned} \tag{2.7}$$

Denote by Ξ the feasible region of problem (2.7). Since $z_i^* \neq 0$ for all $i \in \mathcal{I}_{\neq}^*$, it is easy to see that the objective function of problem (2.7) is continuously differentiable at $(x^*, z^*_{\mathcal{I}_{\neq}^*})$. Then problem (2.7) is a smooth MPCC.

(i) When MPCC-RCPLD qualification holds at x^*, it is not difficult to verify that MPCC-RCPLD holds at $(x^*, z^*_{\mathcal{I}_{\neq}^*}) \in \Xi$. Then $(x^*, z^*_{\mathcal{I}_{\neq}^*})$ is an M-stationary point of problem (2.7) by [21, Corollary 4.1]. The desired M-stationarity of x^* for problem (1.3) follows immediately.

(ii) When MPCC-LI qualification holds at x^*, it is not difficult to verify that MPCC-LICQ holds at $(x^*, z^*_{\mathcal{I}_{\neq}^*}) \in \Xi$. Then $(x^*, z^*_{\mathcal{I}_{\neq}^*})$ is an S-stationary point of problem (2.7) by [39, Theorem 2]. The desired S-stationarity of x^* for problem (1.3) follows immediately.

The proof is complete. ■

3 Approximation problems

The non-Lipschitz term and complementarity constraints make problem (1.3) difficult to solve. In this section, we propose an approximation to problem (1.3) as follows

$$(P_{\epsilon,\sigma}) \qquad \begin{array}{rl} \min & F_\epsilon(x) := f(x) + \varphi_\epsilon(x) \\ \text{s.t.} & G(x) \geq 0,\ H(x) \geq 0,\ G(x)^T H(x) \leq \sigma. \end{array}$$

Here the non-Lipschitz function $\|Dx\|_p^p$ is approximated by a locally Lipschitz function

$$\varphi_\epsilon(x) := \sum_{i=1}^{r} (|D_i^T x| + \epsilon_i)^p,$$

where $\epsilon_i > 0$ $i = 1, \ldots, r$; see, e.g., [33, 49], and the feasible region $\mathcal{F}$ is approximated by

$$\mathcal{F}_\sigma := \{x \in \mathbb{R}^n : G(x) \geq 0,\ H(x) \geq 0,\ G(x)^T H(x) \leq \sigma\},$$

where $\sigma > 0$. The set $\mathcal{F}_\sigma$ is the so-called Scholtes' relaxation or regularization of $\mathcal{F}$ [40], and has good numerical approximation properties [24] in which numerical comparison of several relaxation methods was investigated. The kind of Scholtes' relaxation was also used for solving mathematical programs with vanishing constraints [1]. As shown in Proposition 3.1, problem $(P_{\epsilon,\sigma})$ is easier to satisfy some constraint qualifications compared with problem (1.3).

Recall that MPCC-LI qualification is actually the standard MPCC-LICQ for the following mixed complementarity system:

$$\left\{x : D_i^T x = 0\ i \in \mathcal{I}_0(x^*),\ G(x) \geq 0,\ H(x) \geq 0,\ G(x)^T H(x) = 0\right\}.$$

Then the following result follows directly from [40, Lemma 2.1].

Proposition 3.1 *Let MPCC-LI qualification hold at $x^* \in \mathcal{F}$. Then there exist $\sigma_0 > 0$ and $\delta_0 > 0$ such that for any $\sigma \in (0, \sigma_0]$ and $x \in \mathcal{B}_{\delta_0}(x^*)$, the standard LICQ holds at x for the following system*

$$\left\{x : D_i^T x = 0\ i \in \mathcal{I}_0(x^*),\ G(x) \geq 0,\ H(x) \geq 0,\ G(x)^T H(x) \leq \sigma\right\}.$$

In particular, the standard LICQ holds at $x \in \mathcal{F}_\sigma \cap \mathcal{B}_{\delta_0}(x^*)$ with $\sigma \in (0, \sigma_0]$.

In the following, we will study the distance between optimal solution sets of problem (1.3) and problem $(P_{\epsilon,\sigma})$.

The following simple example shows that for any given $\sigma > 0$ and $\epsilon > 0$, optimal solutions of problem $(P_{\epsilon,\sigma})$ are different from those of problem (1.3). But we observe that optimal solutions of problem $(P_{\epsilon,\sigma})$ tend to those of problem (1.3) as $\sigma \downarrow 0$ and $\epsilon \downarrow 0$.

Example 3.1 *Consider a non-Lipschitz MPCC in* $\mathbb{R}^2$*:* $p = 0.5$*,* $f(x) \equiv 0$*,* $D = I_{2\times 2}$*,* $G(x) = 1 - x_1$ *and* $H(x) = 1 - x_2$*. Direct calculation implies that the feasible region of problem* (1.3) *is*

$$\{x \in \mathbb{R}^2 : x_1 \le 1,\ x_2 = 1 \quad \text{or} \quad x_2 \le 1, x_1 = 1\}.$$

Then it is easy to see that the solution set of problem (1.3) *is* $\mathcal{S} = \{(0,1)^T, (1,0)^T\}$*. Moreover, the feasible region of problem* $(P_{\epsilon,\sigma})$ *with* $0 < \sigma < 1$ *and* $\epsilon > 0$ *is*

$$\{x \in \mathbb{R}^2 : x_1 \le 1, x_2 \le 1, (1 - x_1)(1 - x_2) \le \sigma\}.$$

It is not difficult to verify that the solution set of problem $(P_{\epsilon,\sigma})$ *is*

$$\mathcal{S}_{\epsilon,\sigma} = \begin{cases} \{(0, 1-\sigma)^T\} & \text{if } \epsilon_1 < \epsilon_2, \\ \{(0, 1-\sigma)^T, (1-\sigma, 0)^T\} & \text{if } \epsilon_1 = \epsilon_2, \\ \{(1-\sigma, 0)^T\} & \text{if } \epsilon_1 > \epsilon_2. \end{cases}$$

It is clear that the Pompeiu-Hausdorff distance between $\mathcal{S}_{\epsilon,\sigma}$ *and* $\mathcal{S}$ *is positive but it converges to 0 as* $\sigma \downarrow 0$ *and* $\epsilon \downarrow 0$*. Here the Pompeiu-Hausdorff distance between* X *and* Y *is defined as* $\max\{\sup_{x\in X} \inf_{y\in Y} \|x - y\|, \sup_{y\in X} \inf_{x\in Y} \|x - y\|\}$*.*

As shown in the following result, optimal solutions of problem $(P_{\epsilon,\sigma})$ approach to those of problem (1.3) as $\sigma \downarrow 0$ and $\epsilon \downarrow 0$.

Theorem 3.1 *Let* $x_{\epsilon,\sigma}$ *be an optimal solution of problem* $(P_{\epsilon,\sigma})$ *for any* $\sigma > 0$ *and* $\epsilon > 0$*, and let* x^* *be an arbitrary accumulation point of* $\{x_{\epsilon,\sigma}\}$ *as* $\sigma \downarrow 0$ *and* $\epsilon \downarrow 0$*. Then* x^* *is an optimal solution of problem* (1.3)*.*

Proof. First it is easy to show that $x^* \in \mathcal{F}$. Since $\mathcal{F} \subseteq \mathcal{F}_\sigma$ for any $\sigma > 0$, then by the optimality of $x_{\epsilon,\sigma}$, we have that

$$f(x_{\epsilon,\sigma}) + \sum_{i=1}^{r} (|D_i^T x_{\epsilon,\sigma}| + \epsilon_i)^p \le f(x) + \sum_{i=1}^{r} (|D_i^T x| + \epsilon_i)^p \quad \forall x \in \mathcal{F}.$$

Upon taking limits on both sides of the above inequality as $\sigma \downarrow 0$ and $\epsilon \downarrow 0$, we have that $f(x^*) + \|Dx^*\|_p^p \le f(x) + \|Dx\|_p^p$ for any $x \in \mathcal{F}$. This proof is complete. ∎

3.1 Second-order necessary optimality conditions for problem $(P_{\epsilon,\sigma})$

The convergence analysis in Section 4 requires second-order necessary conditions for approximation problems. Thus in this subsection, we develop the second-order necessary optimality for problem $(P_{\epsilon,\sigma})$, in which f, G, H are all assumed to be twice continuously differentiable. By Fermat's rule (e.g., [38, Theorem 10.1]), it is easy to see that any local minimizer x^* of problem $(P_{\epsilon,\sigma})$ satisfies the first-order necessary optimality conditions:

$$0 \in \nabla f(x^*) + p \sum_{i=1}^{r} D_i (|D_i^T x^*| + \epsilon_i)^{p-1} \text{sign}\,(D_i^T x^*) + \mathcal{N}_{\mathcal{F}_\sigma}(x^*).$$

Note that in problem $(P_{\epsilon,\sigma})$, $(|D_i^T x| + \epsilon_i)^p$ $i = 1,\ldots,r$, are not continuously differentiable. This makes the chain rule of (second-order) subdifferentials fail. Thus, second-order necessary optimality conditions of problem (3.1) cannot be obtained by directly employing the existing results in the literature. In what follows, we investigate how to develop second-order necessary optimality of problem (3.1) by utilizing the characterization of function $F_\epsilon(x)$. For simplicity of notation, in this subsection we consider problem $(P_{\epsilon,\sigma})$ in the following form:

$$\begin{array}{ll} \min & F_\epsilon(x) \\ \text{s.t.} & g^\sigma(x) \le 0, \end{array} \tag{3.1}$$

where $g^\sigma : \mathbb{R}^n \to \mathbb{R}^{2m+1}$ is a twice continuously differentiable function with the form

$$g^\sigma(x) := (-G(x)^T, -H(x)^T, G(x)^T H(x) - \sigma)^T. \tag{3.2}$$

Proposition 3.2 *Assume that x^* is a local minimizer of problem* (3.1)*. If the gradients*

$$\{\nabla g_i^\sigma(x^*), D_j : i \in \mathcal{I}_g^*, j \in \mathcal{I}_0^*\}$$

are linearly independent where $\mathcal{I}_{g^\sigma}^* := \{i : g_i^\sigma(x^*) = 0\}$, $\mathcal{I}_0^* := \mathcal{I}_0(x^*)$, then x^* is a KKT point, i.e., there exists a unique nonnegative multiplier $\mu \in \mathbb{R}^{|\mathcal{I}_{g^\sigma}^*|}$ such that

$$0 \in \nabla f(x^*) + p \sum_{i=1}^r D_i(|D_i^T x^*| + \epsilon_i)^{p-1} \text{sign}\,(D_i^T x^*) + \sum_{i \in \mathcal{I}_{g^\sigma}^*} \mu_i \nabla g_i^\sigma(x^*), \tag{3.3}$$

and weak second-order necessary condition (WSONC) holds at x^*, i.e.,

$$d^T \big[\nabla^2 f(x^*) + \sum_{i \in \mathcal{I}_{g^\sigma}^*} \mu_i \nabla^2 g_i^\sigma(x^*) + p(p-1) \sum_{i=1}^r (|D_i^T x^*| + \epsilon_i)^{p-2} D_i D_i^T\big] d \ge 0 \quad \forall d \in C_w(x^*), \tag{3.4}$$

where $C_w(x^*)$ is the critical subspace of problem (3.1) *defined as*

$$C_w(x^*) := \left\{ d \in \mathbb{R}^n : \ D_i^T d = 0 \ i \in \mathcal{I}_0^*, \quad \nabla g_j^\sigma(x^*)^T d = 0 \ j \in \mathcal{I}_{g^\sigma}^* \right\}.$$

Proof. First we observe that (x^*, t^*) with $t_i^* := |D_i^T x^*|$, $i = 1,\ldots,r$ is a local minimizer of the following problem

$$\begin{array}{ll} \min\limits_{x,t} & f(x) + \sum\limits_{i=1}^r (t_i + \epsilon_i)^p \\ \text{s.t.} & g^\sigma(x) \le 0, \\ & Dx - t \le 0, \ -Dx - t \le 0, \end{array} \tag{3.5}$$

where all the involved functions are twice continuously differentiable. We next show that LICQ holds at (x^*, t^*) for problem (3.5). Let $\mathcal{I}_{\neq}^* := \mathcal{I}_{\neq}(x^*)$ and

$$\sum_{i \in \mathcal{I}_{g^\sigma}^*} \mu_i \begin{pmatrix} \nabla g_i^\sigma(x^*) \\ 0 \end{pmatrix} + \sum_{i \in \mathcal{I}_0^*} u_i \begin{pmatrix} D_i \\ -e_i \end{pmatrix} - \sum_{i \in \mathcal{I}_0^*} v_i \begin{pmatrix} D_i \\ e_i \end{pmatrix} + \sum_{i \in \mathcal{I}_{\neq}^*} \lambda_i \begin{pmatrix} D_i \text{sign}(D_i^T x^*) \\ -e_i \end{pmatrix} = 0. \tag{3.6}$$

Since $\mathcal{I}_0^* \cap \mathcal{I}_{\neq}^* = \emptyset$, it is easy to see that $\lambda_i = 0$ for all $i \in \mathcal{I}_{\neq}^*$. Thus, (3.6) simplifies to

$$\sum_{i\in\mathcal{I}_{g^\sigma}^*} \mu_i \nabla g_i^\sigma(x^*) + \sum_{i\in\mathcal{I}_0^*} (u_i - v_i) D_i = 0, \tag{3.7}$$

$$\sum_{i\in\mathcal{I}_0^*} (u_i + v_i) e_i = 0. \tag{3.8}$$

Since $\{\nabla g_i^\sigma(x^*), D_j : i \in \mathcal{I}_{g^\sigma}^*, j \in \mathcal{I}_0^*\}$ are linearly independent, by (3.7) it follows that $\mu_i = 0$ for all $i \in \mathcal{I}_{g^\sigma}^*$ and $u_i = v_i$ for all $i \in \mathcal{I}_0^*$. This together with (3.8) implies that $u_i = v_i = 0$ for all $i \in \mathcal{I}_0^*$. Thus, all the coefficients $(\mu_i : i \in \mathcal{I}_{g^\sigma}^*)$, $(u_i, v_i : i \in \mathcal{I}_0^*)$, and $(\lambda_i : i \in \mathcal{I}_{\neq}^*)$ are equal to zero. Thus LICQ holds at (x^*, t^*) for problem (3.5). Then, WSONC of problem (3.5) are satisfied at (x^*, t^*) (see, e.g., [35, Theorem 4]), i.e., there exists a unique nonnegative multiplier vector (μ, u, v) such that

$$\nabla f(x^*) + \sum_{i=1}^{r} (u_i - v_i) D_i + \sum_{i\in\mathcal{I}_{g^\sigma}^*} \mu_i \nabla g(x^*) = 0, \tag{3.9}$$

$$p(t_i^* + \epsilon_i)^{p-1} - (u_i + v_i) = 0 \quad i = 1, \ldots, r, \tag{3.10}$$

$$(t_i^* - D_i^T x^*) u_i = 0, \ (t_i^* + D_i^T x^*) v_i = 0 \quad i = 1, \ldots, r, \tag{3.11}$$

and

$$(d^T, h^T) \begin{bmatrix} \nabla^2 f(x^*) + \sum\limits_{i\in\mathcal{I}_{g^\sigma}^*} \mu_i \nabla^2 g_i^\sigma(x^*) & 0 \\ 0 & \mathrm{Diag}\,(T^*) \end{bmatrix} \begin{pmatrix} d \\ h \end{pmatrix} \geq 0 \quad \forall \begin{pmatrix} d \\ h \end{pmatrix} \in C_w(x^*, t^*), \tag{3.12}$$

where $T^* := p(p-1)\big((|D_i^T x^*| + \epsilon_i)^{p-2} : i = 1, \ldots, r\big)$ and

$$C_w(x^*, t^*) := \left\{ \begin{pmatrix} d \\ h \end{pmatrix} \in \mathbb{R}^n \times \mathbb{R}^r : \begin{array}{ll} \nabla g_i^\sigma(x^*)^T d = 0 & i \in \mathcal{I}_{g^\sigma}^* \\ D_i^T d = h_i = 0 & i \in \mathcal{I}_0^* \\ \mathrm{sign}\,(D_i^T x^*) D_i^T d - h_i = 0 & i \in \mathcal{I}_{\neq}^* \end{array} \right\}. \tag{3.13}$$

It follows from (3.10) and the nonnegtiveness of multipliers u, v that

$$u_i - v_i \in p(|D_i^T x^*| + \epsilon_i)^{p-1}[-1, 1] \quad i \in \mathcal{I}_0^*.$$

Moreover, by (3.10) and (3.11), it is not difficult to verify that

$$u_i - v_i = p(|D_i^T x^*| + \epsilon_i)^{p-1} \mathrm{sign}\,(D_i^T x^*) \quad i \in \mathcal{I}_{\neq}^*.$$

These last two relations and the definition of sign$(\cdot)$ imply that

$$u_i - v_i \in p(|D_i^T x^*| + \epsilon_i)^{p-1} \mathrm{sign}\,(D_i^T x^*) \quad i = 1, \ldots, r.$$

Then (3.3) follows immediately from (3.9) and the above relation. Moreover, it follows from (3.12) that

$$d^T[\nabla^2 f(x^*) + \sum_{i\in\mathcal{I}_{g^\sigma}^*} \mu_i \nabla^2 g_i^\sigma(x^*)] d + p(p-1) \sum_{i=1}^{r} (|D_i^T x^*| + \epsilon_i)^{p-2} (h_i)^2 \geq 0 \ \forall \begin{pmatrix} d \\ h \end{pmatrix} \in C_w(x^*, t^*). \tag{3.14}$$

By (3.13), any direction $\begin{pmatrix} d \\ h \end{pmatrix} \in C_w(x^*, t^*)$ satisfies

$$D_i^T d = h_i = 0 \ i \in \mathcal{I}_0^*, \quad \text{sign}\,(D_i^T x^*) D_i^T d - h_i = 0 \ i \in \mathcal{I}_{\neq}^*,$$

which together with (3.14) implies (3.4) immediately. The proof is complete. ■

The following theorem can be obtained by applying Propositions 3.1 and 3.2 to problem $(P_{\epsilon,\sigma})$. For simplicity of notation, we let $\psi_\sigma(x) := G(x)^T H(x) - \sigma$.

Theorem 3.2 *Let $\epsilon > 0$. Let x^σ be a local minimizer of problem $(P_{\epsilon,\sigma})$ for any $\sigma > 0$. Assume that MPCC-LI qualification holds at $x^* \in \mathcal{F}$. Then there exist $\bar{\delta} > 0$ and $\bar{\sigma} > 0$ such that if $x^\sigma \in \mathcal{B}_{\bar{\delta}}(x^*)$ and $\sigma \in (0, \bar{\sigma})$, then x^σ is a KKT point of problem $(P_{\epsilon,\sigma})$, i.e., there exist $\alpha \in \mathbb{R}^m$, $\beta \in \mathbb{R}^m$, and $\gamma \in \mathbb{R}$ such that*

$$\begin{array}{l} 0 \in \nabla f(x^\sigma) + p \sum\limits_{i=1}^{r} (|D_i^T x^\sigma| + \epsilon_i)^{p-1} \text{sign}\,(D_i^T x^\sigma) D_i \\ \qquad\qquad -\nabla G(x^\sigma)\alpha - \nabla H(x^\sigma)\beta + [\nabla G(x^\sigma) H(x^\sigma) + \nabla H(x^\sigma) G(x^\sigma)]\gamma, \\ \alpha \geq 0, \ G(x^\sigma)^T \alpha = 0, \quad \beta \geq 0, \ H(x^\sigma)^T \beta = 0, \quad \gamma \geq 0, \ \psi_\sigma(x^\sigma)\gamma = 0, \end{array}$$

and moreover, the WSONC holds at x^σ, i.e,

$$d^T(\Phi_\sigma + M_\sigma) d \geq 0 \quad \forall d \in C_w(x^\sigma),$$

where

$$\Phi_\sigma := \quad \nabla^2 f(x^\sigma) + p(p-1) \sum_{i=1}^{r} (|D_i^T x^\sigma| + \epsilon_i)^{p-2} D_i D_i^T, \tag{3.15}$$

$$\begin{array}{rl} M_\sigma := & -\sum\limits_{i=1}^{m} \alpha_i \nabla^2 G_i(x^\sigma) - \sum\limits_{i=1}^{m} \beta_i \nabla^2 H_i(x^\sigma) + \gamma \sum\limits_{i=1}^{m} [G_i(x^\sigma) \nabla^2 H_i(x^\sigma) + H_i(x^\sigma) \nabla^2 G_i(x^\sigma)] \\ & +\gamma \sum\limits_{i=1}^{m} [\nabla G_i(x^\sigma) \nabla H_i(x^\sigma)^T + \nabla H_i(x^\sigma) \nabla G_i(x^\sigma)^T], \end{array} \tag{3.16}$$

and

$$C_w(x^\sigma) := \left\{ d \in \mathbb{R}^n : \begin{array}{l} \nabla G_i(x^\sigma)^T d = 0 \ i \in \mathcal{I}_G(x^\sigma) \\ \nabla H_i(x^\sigma)^T d = 0 \ i \in \mathcal{I}_H(x^\sigma) \\ \nabla \psi_\sigma(x^\sigma)^T d = 0 \text{ if } \psi_\sigma(x^\sigma) = 0 \end{array} \right\}. \tag{3.17}$$

Proof. First we may choose $\delta_0 > 0$ such that $\mathcal{I}_0(x^\sigma) \subseteq \mathcal{I}_0(x^*)$ for all $x^\sigma \in \mathcal{B}_{\delta_0}(x^*)$. Then by Proposition 3.1, there exist $\bar{\sigma} \in (0, \infty)$ and $\bar{\delta} \in (0, \delta_0)$ such that if $x^\sigma \in \mathcal{B}_{\bar{\delta}}(x^*)$ and $\sigma \in (0, \bar{\sigma}]$, LICQ holds at x^σ for the following system

$$\left\{ x \in \mathcal{F}_\sigma : \ D_i^T x = 0 \ i \in \mathcal{I}_0(x^\sigma) \right\}.$$

Then the desired results follow directly from Proposition 3.2 by setting $g(x)$ as in (3.2). ■

4 Convergence of an approximation method

In Section 3, we have shown that optimal solutions of approximation problems approach to those of problem (1.3). However, it may be difficult to find an exact optimal solution of problem $(P_{\epsilon,\sigma})$. Thus it is also necessary to investigate the convergence behavior of approximate stationary points of problem $(P_{\epsilon,\sigma})$ as $\epsilon \downarrow 0, \sigma \downarrow 0$.

Algorithm 4.1 *Let $\{\sigma_k\} \downarrow 0$ and $\{\epsilon^k\} \downarrow 0$, $\{\varepsilon_1^k\}$ and $\{\varepsilon_2^k\}$ be sequences of nonnegative parameters approaching to 0, and $\{\zeta^k\}$ be a sequence of error parameters satisfying $\|\zeta^k\| \leq \varepsilon_1^k$. Choose an arbitrary point $x^{0,0} \in \mathbb{R}^n$ and set $k = 0$.*

(i) Solve problem $(P_{\epsilon^k,\sigma_k})$ with initial point $x^{k,0}$ to get $\tilde{x}^k$ such that there exists x^k satisfying $\|x^k - \tilde{x}^k\| \leq \varepsilon_2^k$ and

$$\zeta^k \in \nabla f(x^k) + p\sum_{i=1}^{r}(|D_i^T x^k| + \epsilon_i^k)^{p-1}\text{sign}\,(D_i^T x^k)D_i + \mathcal{N}_{\mathcal{F}_{\sigma_k}}(x^k). \tag{4.1}$$

(ii) Set $x^{k+1,0} = \tilde{x}^k$, $k = k+1$, and go to Step (i).

We give some comments on Step (i) of Algorithm 4.1. The point $\tilde{x}^k$ required in Step (i) of Algorithm 4.1 can be seen as a generalization of an approximate stationary point since when $\varepsilon_2^k = 0$, it follows that $\tilde{x}^k = x^k$ is an approximate stationary point of problem $(P_{\epsilon^k,\sigma_k})$. On the other hand, by Ekeland's variational principal (e.g., [38, Proposition 1.43]), the point $\tilde{x}^k$ can be used to characterize approximate optimal solutions of problem $(P_{\epsilon^k,\sigma_k})$ as shown in the proof of Theorem A.2. Recall that $\psi_\sigma(x) := G(x)^T H(x) - \sigma$.

Theorem 4.1 *Let $\{\tilde{x}^k\}$ be a sequence generated by Algorithm 4.1 and x^* be an arbitrary accumulation point of $\{\tilde{x}^k\}$. Suppose further that MPCC-LI qualification holds at x^*. Then x^* is a C-stationary point of problem* (1.3).

Proof. Without loss of generality, we assume that $\tilde{x}^k \to x^*$ as $k \to \infty$. By the implementation process of Algorithm 4.1, there exists $x^k \in \mathcal{F}_{\sigma_k}$ satisfying (4.1) such that $\|x^k - \tilde{x}^k\| \to 0$ as $k \to \infty$, which together with the fact that $\tilde{x}^k \to x^*$ implies that $x^k \to x^*$ as $k \to \infty$. Then by the continuity of functions G, H, it is not difficult to see that $x^* \in \mathcal{F}$.

By Proposition 3.1 and the relations that $\sigma_k \to 0$ and $x^k \to x^*$, it follows that LICQ holds at $x^k \in \mathcal{F}_{\sigma_k}$ when k is sufficiently large. This together with e.g. [38, Theorem 6.14] implies that

$$\mathcal{N}_{\mathcal{F}_{\sigma_k}}(x^k) \subseteq \left\{ \begin{array}{l} -\nabla G(x^k)\alpha^k - \nabla H(x^k)\beta^k + \\ {[\nabla G(x^k)H(x^k) + \nabla H(x^k)G(x^k)]\gamma^k} \end{array} : \begin{array}{l} \alpha^k \geq 0,\ G(x^k)^T\alpha^k = 0 \\ \beta^k \geq 0,\ H(x^k)^T\beta^k = 0 \\ \gamma^k \geq 0,\ \psi_{\sigma_k}(x^k)\gamma^k = 0 \end{array} \right\}. \tag{4.2}$$

It then follows from (4.1) that for all k sufficiently large, there exist α^k, β^k, and γ^k such that

$$\begin{aligned}\zeta^k \in \nabla f(x^k) + p\sum_{i=1}^{r}(|D_i^T x^k| + \epsilon_i^k)^{p-1}\text{sign}\,(D_i^T x^k)D_i \\ -\nabla G(x^k)(\alpha^k - \gamma^k H(x^k)) - \nabla H(x^k)(\beta^k - \gamma^k G(x^k)),\end{aligned} \tag{4.3}$$

$$\alpha^k \geq 0, G(x^k)^T\alpha^k = 0,\ \beta^k \geq 0, H(x^k)^T\beta^k = 0,\ \gamma^k \geq 0, \psi_{\sigma_k}(x^k)\gamma^k = 0. \tag{4.4}$$

For simplicity, we denote $\mathcal{I}_0^* := \mathcal{I}_0(x^*)$, $\mathcal{I}_{\neq}^* := \mathcal{I}_{\neq}(x^*)$, $\mathcal{I}_{0+}^* := \mathcal{I}_{0+}(x^*)$, $\mathcal{I}_{00}^* := \mathcal{I}_{00}(x^*)$, and $\mathcal{I}_{+0}^* := \mathcal{I}_{+0}(x^*)$. By the continuity of functions G, H, it follows from (4.4) that for all k sufficiently large, $\alpha_i^k = 0$ for all $i \in \mathcal{I}_{+0}^*$ and $\beta_i^k = 0$ for all $i \in \mathcal{I}_{0+}^*$. Then after a suitable rearrangement of terms, (4.3) can be rewritten as

$$\begin{aligned}&\zeta^k - \nabla f(x^k) - \sum_{i\in\mathcal{I}_{\neq}^*} \lambda_i^k D_i - \sum_{i\in\mathcal{I}_{+0}^*} \gamma^k H_i(x^k)\nabla G_i(x^k) - \sum_{i\in\mathcal{I}_{0+}^*} \gamma^k G_i(x^k)\nabla H_i(x^k)\\ &= \sum_{i\in\mathcal{I}_0^*} \lambda_i^k D_i - \sum_{i\in\mathcal{I}_{0+}^*\cup\mathcal{I}_{00}^*} \mu_i^k \nabla G_i(x^k) - \sum_{i\in\mathcal{I}_{+0}^*\cup\mathcal{I}_{00}^*} \nu_i^k \nabla H_i(x^k),\end{aligned} \tag{4.5}$$

where

$$\begin{aligned}&\mu^k := \alpha^k - \gamma^k H(x^k),\ \nu^k := \beta^k - \gamma^k G(x^k),\\ &\lambda_i^k := p(|D_i^T x^k| + \epsilon_i^k)^{p-1}\text{sign}\,(D_i^T x^k) \quad i \in \mathcal{I}_{\neq}^*, \quad \lambda_i^k \in \mathbb{R}\ i \in \mathcal{I}_0^*.\end{aligned} \tag{4.6}$$

We next show the convergence of the sequences $\{\lambda_i^k : i \in \mathcal{I}_0^*\}$, $\{\mu^k\}$, and $\{\nu^k\}$ by considering two cases. (a) Assume that $\{\gamma^k\}$ is bounded. Then $\mu_i^k = -\gamma^k H_i(x^k) \to 0$ for all $i \in \mathcal{I}_{+0}^*$ and $\nu_i^k = -\gamma^k G_i(x^k) \to 0$ for all $i \in \mathcal{I}_{0+}^*$. Then the left hand side of (4.5) tends to

$$-\nabla f(x^*) - \sum_{i\in\mathcal{I}_{\neq}^*} p|D_i^T x^*|^{p-1}\text{sign}\,(D_i^T x^*)D_i$$

as $k \to \infty$. Then by MPCC-LI qualification at x^*, it is not hard to verify that $\{\lambda_i^k : i \in \mathcal{I}_0^*\}$, $\{\mu_i^k : i \in \mathcal{I}_{0+}^* \cup \mathcal{I}_{00}^*\}$, and $\{\nu_i^k : i \in \mathcal{I}_{+0}^* \cup \mathcal{I}_{00}^*\}$ converge to some $\{\lambda_i^* : i \in \mathcal{I}_0^*\}$, $\{\mu_i^* : i \in \mathcal{I}_{0+}^* \cup \mathcal{I}_{00}^*\}$, and $\{\nu_i^* : i \in \mathcal{I}_{+0}^* \cup \mathcal{I}_{00}^*\}$, respectively. Then taking limits on both sides of (4.5) implies that

$$\begin{aligned}\nabla f(x^*) + \sum_{i\in\mathcal{I}_{\neq}^*} p|D_i^T x^*|^{p-1}\text{sign}\,(D_i^T x^*)D_i + \sum_{i\in\mathcal{I}_0^*} \lambda_i^* D_i&\\ - \sum_{i\in\mathcal{I}_{0+}^*\cup\mathcal{I}_{00}^*} \mu_i^*\nabla G_i(x^*) - \sum_{i\in\mathcal{I}_{+0}^*\cup\mathcal{I}_{00}^*} \nu_i^*\nabla H_i(x^*) &= 0.\end{aligned} \tag{4.7}$$

(b) Assume that $\{\gamma^k\}$ is unbounded. Then dividing (4.5) by γ^k implies that

$$\begin{aligned}&\frac{\zeta^k - \nabla f(x^k) - \sum_{i\in\mathcal{I}_{\neq}^*}\lambda_i^k D_i}{\gamma^k} - \sum_{i\in\mathcal{I}_{+0}^*} \nabla G_i(x^k)H_i(x^k) - \sum_{i\in\mathcal{I}_{0+}^*} \nabla H_i(x^k)G_i(x^k)\\ &= \sum_{i\in\mathcal{I}_0^*} D_i\frac{\lambda_i^k}{\gamma^k} - \sum_{i\in\mathcal{I}_{0+}^*\cup\mathcal{I}_{00}^*} \nabla G_i(x^k)\frac{\mu_i^k}{\gamma^k} - \sum_{i\in\mathcal{I}_{+0}^*\cup\mathcal{I}_{00}^*} \nabla H_i(x^k)\frac{\nu_i^k}{\gamma^k}.\end{aligned} \tag{4.8}$$

By the definition of $\mathcal{I}_{\neq}^*, \mathcal{I}_{+0}^*$, and $\mathcal{I}_{0+}^*$, it is easy to see that $\lambda_i^k \to p|D_i^T x^*|^{p-1}\text{sign}\,(D_i^T x^*)D_i$ for all $i \in \mathcal{I}_{\neq}^*$, $G_i(x^k) \to 0$ for all $i \in \mathcal{I}_{0+}^*$, and $H_i(x^k) \to 0$ for all $i \in \mathcal{I}_{+0}^*$. These together with the boundedness of $\{\nabla f(x^k)\}$ and $\{\zeta^k\}$ imply that the left hand side of (4.8) tends to 0 as $k \to \infty$. Then by MPCC-LI qualification at x^*, one can easily have that the coefficients of gradients on the right hand side of (4.8) tend to 0. In particular, $\frac{\mu_i^k}{\gamma^k} \to 0$ for any $i \in \mathcal{I}_{0+}^*$ and $\frac{\nu_i^k}{\gamma^k} \to 0$ for any $i \in \mathcal{I}_{+0}^*$ as $k \to \infty$. If $i \in \mathcal{I}_{0+}^*$, then $H_i(x^*) > 0$. Then by the fact that $\frac{\mu_i^k}{\gamma^k} = \frac{\alpha_i^k}{\gamma^k} - H_i(x^k) \to 0$, it follows that the equality $\alpha_i^k = 0$ never holds on an infinite subsequence for all $i \in \mathcal{I}_{0+}^*$. This together with (4.4) means that for all k sufficiently large, $G_i(x^k) = 0$ for any $i \in \mathcal{I}_{0+}^*$. By symmetry, one can also have that for all k sufficiently large, $H_i(x^k) = 0$ for any $i \in \mathcal{I}_{+0}^*$. Then $\mu_i^k = -\gamma^k H_i(x^k) = 0$ for all $i \in \mathcal{I}_{+0}^*$ and $\nu_i^k = -\gamma^k G_i(x^k) = 0$ for all $i \in \mathcal{I}_{0+}^*$. These two relations and (4.5) imply that

$$\zeta^k - \nabla f(x^k) - \sum_{i \in \mathcal{I}_{\neq}^*} \lambda_i^k D_i = \sum_{i \in \mathcal{I}_0^*} \lambda_i^k D_i - \sum_{i \in \mathcal{I}_{0+}^* \cup \mathcal{I}_{00}^*} \mu_i^k \nabla G_i(x^k) - \sum_{i \in \mathcal{I}_{+0}^* \cup \mathcal{I}_{00}^*} \nu_i^k \nabla H_i(x^k). \quad (4.9)$$

Due to the boundedness of the left hand side of the above equation, by MPCC-LI qualification at x^*, it is not hard to verify that $\{\lambda_i^k : i \in \mathcal{I}_0^*\}$, $\{\mu_i^k : i \in \mathcal{I}_{0+}^* \cup \mathcal{I}_{00}^*\}$, and $\{\nu_i^k : i \in \mathcal{I}_{+0}^* \cup \mathcal{I}_{00}^*\}$ converge to some $\{\lambda_i^* : i \in \mathcal{I}_0^*\}$, $\{\mu_i^* : i \in \mathcal{I}_{0+}^* \cup \mathcal{I}_{00}^*\}$, and $\{\nu_i^* : i \in \mathcal{I}_{+0}^* \cup \mathcal{I}_{00}^*\}$, respectively. Taking limits on both sides of (4.9) as $k \to \infty$ implies

$$\nabla f(x^*) + \sum_{i \in \mathcal{I}_{\neq}^*} p|D_i^T x^*|^{p-1} \text{sign}\,(D_i^T x^*) D_i + \sum_{i \in \mathcal{I}_0^*} \lambda_i^* D_i - \sum_{i \in \mathcal{I}_{0+}^* \cup \mathcal{I}_{00}^*} \mu_i^* \nabla G_i(x^*) - \sum_{i \in \mathcal{I}_{+0}^* \cup \mathcal{I}_{00}^*} \nu_i^* \nabla H_i(x^*) = 0. \quad (4.10)$$

The inequality $\mu_i^* \nu_i^* \geq 0$ follows immediately since

$$\begin{aligned} \mu_i^k \nu_i^k &= (\alpha_i^k - \gamma^k H_i(x^k))(\beta_i^k - \gamma^k G_i(x^k)) \\ &= \alpha_i^k \beta_i^k - \gamma^k \alpha_i^k G_i(x^k) - \gamma^k \beta_i^k H_i(x^k) + (\gamma^k)^2 G_i(x^k) H_i(x^k) \\ &= \alpha_i^k \beta_i^k + (\gamma^k)^2 G_i(x^k) H_i(x^k) \geq 0, \end{aligned}$$

where the last equality follows from (4.4). This together with (4.7) and (4.10) implies that x^* is a C-stationary point. ■

If we can use Algorithm 4.1 to find an approximate WSONC point $\tilde{x}^k$ for all large $k \geq 0$ i.e., there exists x^k satisfying $\|x^k - \tilde{x}^k\| \leq \varepsilon_2^k$ and x^k is an approximate WSONC point of problem $(P_{\epsilon^k, \sigma_k})$, Theorem 4.1 can be improved as shown in Theorem 4.2. Recall that x^k is an approximate WSONC point of problem $(P_{\epsilon^k, \sigma_k})$ if there exist α^k, β^k, and γ^k such that (4.3) and (4.4) hold, and

$$d^T(\Phi_k + M_k)d \geq -\tau_k \quad \forall d \in C_w(x^k),$$

where $\tau_k \geq 0$ is a tolerance, and Φ_k, M_k, and $C_w(x^k)$ are respectively defined as (3.15), (3.16), and (3.17) with ϵ^k, σ^k, x^k, α^k, β^k, and γ^k in place of ϵ, σ, x^σ, α, β, and γ, respectively.

Theorem 4.2 *Assume that f, G, H are twice continuously differentiable. In addition to the assumptions of Theorem 4.1, we also assume that for all large k, approximate WSONC of problem $(P_{\epsilon^k, \sigma_k})$ holds at x^k. Let $\tau_k \leq \tau_0$ for some positive number τ_0. Then x^* is an M-stationary point of problem* (1.3).

Proof. In the same way as the proof of Theorem 4.1, we can show that x^* is a C-stationary point. Assume to the contrary that x^* is not an M-stationary point. There must exist $i_0 \in \mathcal{I}_{00}^*$ such that $\mu_{i_0}^* < 0$ and $\nu_{i_0}^* < 0$. Then by (4.6), we have

$$\alpha_{i_0}^k - \gamma^k H_{i_0}(x^k) \to \mu_{i_0}^* < 0,\ \beta_{i_0}^k - \gamma^k G_{i_0}(x^k) \to \nu_{i_0}^* < 0. \tag{4.11}$$

This together with (4.4) implies that for all k sufficient large,

$$\alpha_{i_0}^k = \beta_{i_0}^k = 0,\ H_{i_0}(x^k) > 0,\ G_{i_0}(x^k) > 0. \tag{4.12}$$

Then (4.11) simplifies to

$$-\gamma^k H_{i_0}(x^k) \to \mu_{i_0}^* < 0,\ -\gamma^k G_{i_0}(x^k) \to \nu_{i_0}^* < 0. \tag{4.13}$$

Let

$$A_k := \begin{bmatrix} D_i^T & i \in \mathcal{I}_0^* \\ \nabla G_i(x^k)^T & i = i_0 \\ \nabla G_i(x^k)^T & i \in \mathcal{I}_{00}^* \backslash \{i_0\} \cup \mathcal{I}_G(x^k) \\ \frac{G_i(x^k)}{H_i(x^k)} \nabla H_i(x^k)^T + \nabla G_i(x^k)^T & i \in \mathcal{I}_{0+}^* \backslash \mathcal{I}_G(x^k) \\ \nabla H_i(x^k)^T & i = i_0 \\ \nabla H_i(x^k)^T & i \in \mathcal{I}_{00}^* \backslash \{i_0\} \cup \mathcal{I}_H(x^k) \\ \frac{H_i(x^k)}{G_i(x^k)} \nabla G_i(x^k)^T + \nabla H_i(x^k)^T & i \in \mathcal{I}_{+0}^* \backslash \mathcal{I}_H(x^k) \end{bmatrix}$$

and

$$z^k = \begin{bmatrix} 0 & i \in \mathcal{I}_0^* \\ 1 & i = i_0 \\ 0 & i \in \mathcal{I}_{00}^* \backslash \{i_0\} \\ 0 & i \in \mathcal{I}_{0+}^* \\ -\frac{H_{i_0}(x^k)}{G_{i_0}(x^k)} & i = i_0 \\ 0 & i \in \mathcal{I}_{00}^* \backslash \{i_0\} \\ 0 & i \in \mathcal{I}_{+0}^* \end{bmatrix}.$$

Since $\mathcal{I}_G(x^k) \subseteq \mathcal{I}_{0+}^* \cup \mathcal{I}_{00}^* \backslash \{i_0\}$ and $\mathcal{I}_H(x^k) \subseteq \mathcal{I}_{+0}^* \cup \mathcal{I}_{00}^* \backslash \{i_0\}$ for all k sufficiently large, we can have that $(\mathcal{I}_{00}^* \backslash \{i_0\} \cup \mathcal{I}_G(x^k)) \cup (\mathcal{I}_{0+}^* \backslash \mathcal{I}_G(x^k)) = \mathcal{I}_{0+}^* \cup \mathcal{I}_{00}^* \backslash \{i_0\}$ and $(\mathcal{I}_{00}^* \backslash \{i_0\} \cup \mathcal{I}_H(x^k)) \cup (\mathcal{I}_{0+}^* \backslash \mathcal{I}_H(x^k)) = \mathcal{I}_{0+}^* \cup \mathcal{I}_{00}^* \backslash \{i_0\}$. Then by MPCC-LI qualification at x^*, it follows that the limit point of the matrix sequence $\{A_k\}$ has full row rank. Thus, A_k has full row rank for all k sufficiently large. Let $d^k := A_k^T (A_k A_k^T)^{-1} z^k$, which clearly satisfies $A_k d^k = z^k$. Moreover, by (4.13), the sequence $\{z^k\}$ is convergent since $\frac{H_{i_0}(x^k)}{G_{i_0}(x^k)} \to \frac{\mu_{i_0}^*}{\nu_{i_0}^*}$ as $k \to \infty$. It then follows that $\{d^k\}$ is bounded. By the choice of d^k, it easily follows that

$$\begin{aligned} \nabla \psi_{\sigma_k}(x^k)^T d^k &= \sum_{i=1}^{m} [G_i(x^k) \nabla H_i(x^k)^T d^k + H_i(x^k) \nabla G_i(x^k)^T d^k] \\ &= G_{i_0}(x^k) \nabla H_{i_0}(x^k)^T d^k + H_{i_0}(x^k) \nabla G_{i_0}(x^k)^T d^k \\ &= -H_{i_0}(x^k) + H_{i_0}(x^k) \\ &= 0. \end{aligned} \tag{4.14}$$

Moreover, one can have that $\mathcal{I}_0(x^k) \subseteq \mathcal{I}_0^*$ for all k sufficiently large. Thus, it follows from the fact that $A_k d^k = z^k$ and (4.14) that $d^k \in C_w(x^k)$ for all k sufficiently large.

We next show that $\varphi_k(d^k) := (d^k)^T(\Phi_k + M_k)d^k \to -\infty$ as $k \to \infty$. This contradicts the approximate WSONC at x^k when k is sufficiently large. This contradiction implies that x^* is an M-stationary point. By the definition of Φ_k and the fact that $D_i^T d^k = 0$ for all $i \in \mathcal{I}_0^*$, we have that

$$(d^k)^T\Phi_k d^k = \quad (d^k)^T\nabla^2 f(x^k)d^k + p(p-1)\sum_{i\in\mathcal{I}_{\neq}^*}(|D_i^T x^k| + \epsilon_i^k)^{p-2}(D_i^T d^k)^2.$$

It is not hard to verify that the sequence $\{(d^k)^T\Phi_k d^k\}$ is bounded due to the boundedness of $\{d^k\}$ and $\{x^k\}$, and the relation that $|D_i^T x^k| \to |D_i^T x^*| > 0$ for all $i \in \mathcal{I}_{\neq}^*$. By the definition of M_k, it follows that

$$\begin{aligned}(d^k)^T M_k d^k &:= 2\gamma^k \sum_{i=1}^m \nabla G_i(x^k)^T d^k \nabla H_i(x^k)^T d^k \\ &-(d^k)^T\left[\sum_{i=1}^m \alpha_i^k \nabla^2 G_i(x) + \sum_{i=1}^m \beta_i^k \nabla^2 H_i(x) - \gamma^k \sum_{i=1}^m [G_i(x)\nabla^2 H_i(x) + H_i(x)\nabla^2 G_i(x)]\right] d^k.\end{aligned}$$

After a rearrangement of terms, by (4.6), we have that

$$\begin{aligned}(d^k)^T M_k d^k &= 2\gamma^k \sum_{i=1}^m \nabla G_i(x^k)^T d^k \nabla H_i(x^k)^T d^k - \sum_{i=1}^m (\alpha_i^k - \gamma^k H_i(x^k))(d^k)^T\nabla^2 G_i(x)d^k \\ &\quad -\sum_{i=1}^m (\beta_i^k - \gamma^k G_i(x^k))(d^k)^T\nabla^2 H_i(x)d^k \\ &= -\sum_{i=1}^m \mu_i^k (d^k)^T\nabla^2 G_i(x^k)d^k - \sum_{i=1}^m \nu_i^k (d^k)^T\nabla^2 H_i(x^k)d^k \qquad (4.15) \\ &\quad +2\gamma^k \sum_{i=1}^m \nabla G_i(x^k)^T d^k \nabla H_i(x^k)^T d^k. \qquad (4.16)\end{aligned}$$

In the proof of Theorem 4.1, we have shown that $\{u^k\}$ and $\{v^k\}$ have limits which imply that $\{u^k\}$ and $\{v^k\}$ are bounded. Recall that $\{d^k\}$ is bounded. Thus, it is easy to see that the terms in (4.15) are all bounded. At the same time, by the choice of d^k, we have

$$\begin{aligned}&\gamma^k \sum_{i=1}^m \nabla G_i(x^k)^T d^k \nabla H_i(x^k)^T d^k \\ &\quad = -\sum_{i\in\mathcal{I}_{0+}^*} \frac{G_i(x^k)}{H_i(x^k)}(\nabla H_i(x^k)^T d^k)^2 - \sum_{i\in\mathcal{I}_{+0}^*} \frac{H_i(x^k)}{G_i(x^k)}(\nabla G_i(x^k)^T d^k)^2 \\ &\qquad +\gamma^k \nabla G_{i_0}(x^k)^T d^k \nabla H_{i_0}(x^k)^T d^k \\ &\quad = -\sum_{i\in\mathcal{I}_{0+}^*} \frac{G_i(x^k)}{H_i(x^k)}(\nabla H_i(x^k)^T d^k)^2 - \sum_{i\in\mathcal{I}_{+0}^*} \frac{H_i(x^k)}{G_i(x^k)}(\nabla G_i(x^k)^T d^k)^2 \qquad (4.17) \\ &\qquad + \frac{-\gamma^k H_{i_0}(x^k)}{G_{i_0}(x^k)}. \qquad (4.18)\end{aligned}$$

We observe that the terms in (4.17) tend to 0 as $k \to \infty$. Moreover, since $G_{i_0}(x^k) \downarrow 0$ and $-\gamma^k H_{i_0}(x^k) \to u^*_{i_0} < 0$, the term in (4.18) tends to $-\infty$ as $k \to \infty$. Thus the term in (4.16) tends to $-\infty$ and then $(d^k)^T M_k d^k \to -\infty$ as $k \to \infty$. This together with the boundedness of $\{(d^k)^T \Phi_k d^k\}$ implies that $\varphi_k(d^k) \to -\infty$ as $k \to \infty$. The proof is complete. ∎

The proof for the following result is standard as that in [40].

Theorem 4.3 *If, in addition to the assumptions of Theorem 4.2, ULSC holds at x^*, then x^* is an S-stationary point of problem* (1.3).

Proof. In Theorem 4.2, we have shown that x^* is an M-stationary point. Moreover, the multipliers (μ, ν) associated with x^* are unique by MPCC-LI qualification at x^*. These facts together with the ULSC assumption imply that x^* is an S-stationary point immediately. ∎

5 Numerical simulations

In this section we conduct numerical experiments to test the performance of the proposed Algorithm 4.1. All computations are performed on a Lenovo laptop (1.80 GHz-2.40 GHz, 7.92GB RAM) with MATLAB R2018a.

5.1 Sparse LCP solution

Consider problem (1.2) with $p \in (0, 1]$. The case with $p = 1$ will be used as a benchmark against cases with $0 < p < 1$. We first generate M and q such that the LCP in the constraints of problem (1.2) has a solution z with sparsity being $s = n/10$. In particular, we generate M by setting $M = H * \mathrm{Diag}(|w|) * H^T$ where the entries of matrix H are randomly chosen from the standard normal distribution, and the $n/5$ components with random positions of vector w are also randomly chosen from the standard normal distribution and other components of vector w are equal to 0. Let *ind* be a row vector containing s unique integers selected randomly from $\{1, 2 \ldots, n\}$. Then let z be a vector whose components in *ind* are generated by the standard normal distribution truncated by $0.1 * e$ and other components are equal to 0. Let q be a vector that is equal to $-M * z$ for indices in *ind* and equal to the maximum of $-M * z$ and 0 for indices not in *ind*. From the construction process, it is easy to see that z (with sparsity being s) is a solution to the LCP in the constraints of problem (1.2). Our goal is to find an approximate solution x^k such that $\|x^k\|_0 \le \|z\|_0$ by using Algorithm 4.1 to solve problem (1.2). The initial point of using Algorithm 4.1 to solve problem (1.2) is obtained by solving the quadratic programming problem:

$$\begin{aligned} \min \quad & x^T M x + q^T x \\ \text{s.t.} \quad & x \ge 0, \ Mx + q \ge 0. \end{aligned}$$

We set $\sigma_k = 10^{-3-2k}$ and $\epsilon^k = 10^{-3-k}$. In Step (i) of Algorithm 4.1, we use the active-set method implemented in KNITRO [29] with the default setting to solve problem $(P_{\epsilon^k, \sigma_k})$. Algorithm 4.1 is terminated once the solution x^k of the k-th approximation problem satisfies

$$\|\min(x^k, Mx^k + q)\| \le 10^{-6} \quad \text{and} \quad \|x^k - x^{k-1}\| / \max\{\|x^{k-1}\|, 1\} \le 10^{-8}.$$

In the tests below, we set $n = 500$ and $n = 1000$, and generate 100 random instances for each such n. The computational results reported in Tables 1 and 2 are averaged over the 100 instances, where we report the number of nonzero entries ($\|x^k\|_0$), the deviation of sparsity (**Dev**=$\max\{\|x^k\|_0 - s, 0\}$) compared with s, and the residual (**Res**=$\|\min\{x^k, Mx^k + q\}\|$) at the approximate solution x^k, and the CPU time (**CPU**) in seconds. From Tables 1 and 2, we observe that our method can find a sparse solution successfully for almost all instances when $0 < p < 1$ and moreover, it can produce sparser solutions when using $0 < p < 1$ than using $p = 1$.

Table 1: Results for random instances with $n = 500$

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$\|x^k\|_0$	51	51	51	50.5	50	50	50	50.5	50.5	87.4
Dev	1	1	1	0.5	0	0	0	0.5	0.5	37.4
Res	1e-12	1e-12	1e-12	1e-12	1e-12	2e-12	2e-12	2e-11	1e-10	4e-9
CPU	3.0	2.8	2.9	2.4	2.2	2.2	2.1	2.1	3.2	8.5

Table 2: Results for random instances with $n = 1000$

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$\|x^k\|_0$	100	100	100	100	100	100	100	100	101	150.8
Dev	0	0	0	0	0	0	0	0	1	50.8
Res	8e-12	7e-12	8e-12	8e-12	8e-12	8e-12	9e-12	1e-11	2e-08	2e-08
CPU	17.5	17.5	15.5	14.9	14.9	14.4	14.3	14.5	22.0	59.9

5.2 Second-best road pricing problem

We are given a directed network of nodes and links $(\mathcal{V}, \mathcal{A})$ representing the road network of some city. Let $\mathcal{W}$ denote the set of all origin-destination (OD) pairs. For certain pairs of nodes $w = (w_s, w_t) \in \mathcal{W}$, we are given an amount of flow d_w that flows from w_s to w_t. Let $d = (d_w : w \in \mathcal{W})$ denote the vector of all OD demands. Let $\mathcal{R} = \cup_{w\in\mathcal{W}}\mathcal{R}_w$ denote the set of all paths in the network, where $\mathcal{R}_w$ denotes the set of all paths connecting OD pair $w \in \mathcal{W}$. Let $\Delta = [\delta_{ar}] \in \{0,1\}^{|\mathcal{A}|\times|\mathcal{R}|}$ denote the link/path incidence matrix where δ_{ar} is equal to 1 if link $a \in \mathcal{A}$ is on path $r \in \mathcal{R}$ and 0 otherwise. Let $\Lambda = [\eta_{wr}] \in \{0,1\}^{|\mathcal{W}|\times|\mathcal{R}|}$ denote the OD/path incidence matrix, where η_{wr} is equal to 1 if path $r \in \mathcal{R}_w$ and 0 otherwise.

Let f_{rw} denote the flow variable on path $r \in \mathcal{R}_w$ and $f = (f_{rw} : r \in \mathcal{R}_w, w \in \mathcal{W})$ the vector of path flows. Let v_a denote the flow variable on link $a \in \mathcal{A}$ and $v = (v_a : a \in \mathcal{A})$ the vector of link flows. Let μ_w denote the generalized travel cost between OD pair $w \in \mathcal{W}$ at equilibrium and $\mu = (\mu_w : w \in \mathcal{W})$ the vector of generalized OD travel costs. Let $t_a(v)$ denote the travel cost function for a given link $a \in \mathcal{A}$ and $t(v) = (t_a(v) : a \in \mathcal{A})$ the vector of link travel costs, depending on link flows v. Then Wardrop's user equilibrium (UE) flows satisfy the flow conservation conditions $\Lambda f - d = 0,\ \triangle f - v = 0,\ f \geq 0$, and

$$\triangle^T t(v) - \Lambda^T \mu \geq 0,\ \ f \geq 0,\ \ (\triangle^T t(v) - \Lambda^T \mu)^T f = 0.$$

We refer the reader to the monograph [43] for detailed discussions.

In general, the UE flow is not the same as the system optimum (SO) flow that is the optimal solution of minimizing the total network travel cost $t(v)^T v$ subject to the flow conservation conditions. Then various road pricing schemes are explored to decentralize the SO flow pattern into an UE flow pattern in a general network. The assumptions of the first-best road pricing scheme are generally not met in reality due to the fact that we cannot charge on *all links* in view of high operating cost and poor public acceptance. As an improvement to first-best road pricing scheme, the general second-best road pricing model in the literature is to choose *a subset of links* for toll charges to minimize the total travel cost while taking account of the route choice behavior of network users; see the monograph [43]. However, it is difficult and even almost impossible to predetermine an appropriate set of toll links especially when the set of links $\mathcal{A}$ is huge.

Let $u = (u_a : a \in \mathcal{A})$ denote the vector of link tolls in which u_a denotes the toll charging on link $a \in \mathcal{A}$. By introducing the sparsity-induced function $\|u\|_p^p$, we give an alternative approach that minimizes the total travel cost and the number of toll links simultaneously. In particular, we propose a new second-best road pricing model as follows

$$\begin{array}{ll} \min\limits_{u,v,f,\mu} & t(v)^T v + \tau \|u\|_p^p \\ \text{s.t.} & \Lambda f - d = 0,\ \triangle f - v = 0, \\ & \triangle^T (t(v) + u) - \Lambda^T \mu \geq 0,\ f \geq 0,\ (\triangle^T (t(v) + u) - \Lambda^T \mu)^T f = 0, \end{array} \tag{5.1}$$

where $\tau > 0$ is used to tune the tradeoff between the total network travel cost and the number of toll links. Letting $G(u, v, f, \mu) := \triangle^T (t(v) + u) - \Lambda^T \mu$ and $H(u, v, f, \mu) := f$, problem (5.1) is a special case of problem (1.3) with nonlinear complementarity constraints [1].

We apply Algorithm 4.1 to solve the second-best road pricing problem (5.1) with $p \in (0, 1)$. The test examples are based on a road network as shown in Figure 1, which is first adopted in [36] and then is extensively used in transportation community. This network contains 13 nodes, 19 directed links, and 4 OD movements $1 \to 2$, $1 \to 3$, $4 \to 2$, and $4 \to 3$. The link cost function $t_a(v_a)$ is assumed to follow a Bureau of Public Roads (BPR) function:

$$t_a(v_a) = t_a^0 \Big(1 + b_a \Big(\frac{v_a}{c_a} \Big)^{n_a} \Big),$$

where $b_a \equiv 0.15$ and $n_a \equiv 4$ are given as in [44, 48]. The free-flow travel time t_a^0 and the link capacity c_a are the same as those used in [44]. The entries of the OD demand vector d are chosen from a normal distribution with mean 1000 and standard deviation 400.

[1] We check that all the results presented in the paper have evident valid counterparts in the presence of additional usual equality and inequality constraints. We omit the usual constraints of problem (1.3) for simplifying the analysis since all the essential difficulties are associated with the complementarity constraints and the non-Lipschitz term in the objective.

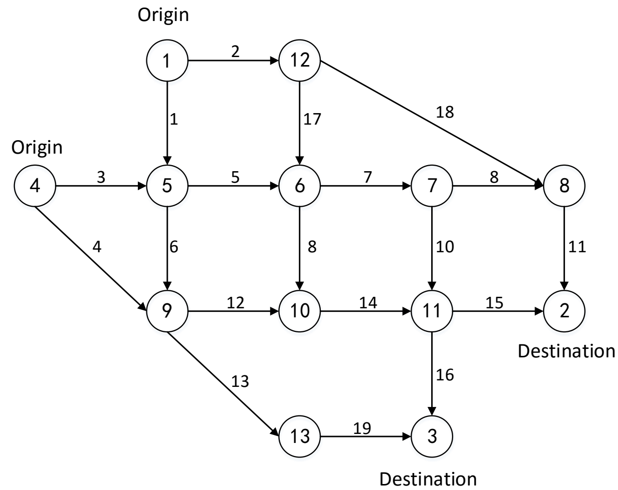

Figure 1: Nguyen and Dupuis's road network

The initial point in our method for solving problem (5.1) is chosen as (u^0, v^0, f^0, μ^0) where $u^0 = 0$ and (v^0, f^0, μ^0) is an arbitrary UE solution. The parameters are the same as those mentioned above. In Table 3, we report numerical results including the vector of link tolls (u^k) and the number of toll links $(\|u^k\|_0)$ at the approximate solution, the standardized total network travel cost (**TotalCost** $= (C_\tau - C_0)/C_0$) where C_τ is the total network cost at the approximate solution when the tradeoff is τ, and the CPU time (**CPU**) in seconds. Since we are concerned about the impact of choice of τ on the number of toll links and the total network travel cost, we fix $p = 0.5$ and set $\tau = 0, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 2$. It should be noted that when setting $\tau = 0$, the obtained solution is actually the first-best road pricing strategy that makes the SO flow to be the same as the UE flow, which can be seen as a benchmark against cases with $\tau > 0$. As shown in Table 3, the scheme with $\tau = 0$ leads to charge on almost all links, which is not a good decision in view of high operating cost and managing difficulty. From Table 3, we observe that when we choose $\tau = 2$, then $u = (0, 0, 0, 5.40, 3.24, 0, 3.54, 0, 0, 0, 3.18, 0, 0, 0, 0, 0, 0, 5.64, 0)^T$ as the vector of link tolls, and the total network cost only increases 0.014% compared with the first-best pricing case, while the number of toll links decreases 68.74 % (only 5 links are for toll charges). The numerical results demonstrate the effectiveness of the non-Lipschitz MPCC model (5.1) and our approximation method.

Acknowledgments

The authors are grateful to the associate editor and two referees for their helpful comments and Dr. Yafeng Liu for helpful discussions.

Table 3: Results for second-best road pricing problem (5.1) with $p = 0.5$

τ	0	1e-4	1e-3	1e-2	1e-1	1	2
	7.24	2.40	2.40	2.40	2.40	2.38	0
	1.17	0	0	0	0	0	0
	5.65	0	0	0	0	0	0
	10.45	4.80	4.80	4.80	4.80	4.79	5.40
	2.34e-4	2.40	2.40	2.40	2.40	2.39	3.24
	1.94e-7	0	0	0	0	0	0
	0.66	4.00	4.00	4.00	4.00	3.99	3.54
	3.20	0	0	0	0	0	0
	11.05	1.33e-4	6.93e-4	0	0	0	0
u^k	6.55	0	0	0	0	0	0
	0	3.20	3.20	3.20	3.20	3.19	3.18
	0.80	2.58e-3	0	0	0	0	0
	5.16	3.20	3.20	3.20	3.20	3.16	0
	0	0	0	0	0	0	0
	1.30	0	0	0	0	0	0
	1.15	0	0	0	0	0	0
	1.27	0	0	0	0	0	0
	15.38	6.40	6.40	6.40	6.40	6.38	5.64
	0	0	0	0	0	0	0
$\|u^k\|_0$	16	9	8	7	7	7	5
TotalCost	0	3.60e-12	7.34e-12	7.35e-12	2.69e-10	2.77e-08	1.44e-04
CPU	0.4	0.4	0.3	0.3	0.3	0.2	0.2

References

[1] W. Achtziger, C. Kanzow, T. Hoheisel, *On a relaxation method for mathematical programs with vanishing constraints, GAMM-Mitt., 35 (2012), 110–130.*

[2] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), 183–202.

[3] S.J. Bi, X.L. Liu and S.H. Pan, *Exact penalty decomposition method for zeronorm minimization based on MPEC formulation*, SIAM J. Sci. Comput., (36) 2014: A1451–A1477.

[4] W. Bian and X. Chen, *Linearly constrained non-Lipschitz optimization for image restoration*, SIAM J. Imaging Sci., 8 (2015), 2294–2322.

[5] J. Brodie, I. Daubechies, C. De Mol, D. Giannone and I. Loris, *Sparse and stable Markowitz portfolios*, Proc. Nat. Acad. Sci., 106 (2009), 12267–12272.

[6] A.M. Bruckstein, D.L. Donoho and M. Elad, *From sparse solutions of systems of equations to sparse modeling of signals and images*, SIAM Rev., 51 (2009), 34–81.

[7] O.P. Burdakov, C. Kanzow and A. Schwartz, *Mathematical programs with cardinality constraints: reformulation by complementarity-type conditions and a regularization method*, SIAM J. Optim., 26 (2016), 397–425.

[8] E. Candès, M. Wakin and S. Boyd, *Enhancing sparsity by reweighted ℓ_1 minimization*, J. Fourier Anal. Appl., 14 (2008), 877–905.

[9] M. Červinka, C. Kanzow and A. Schwartz, *Constraint qualifications and optimality conditions for optimization problems with cardinality constraints*, Math. Program., 160 (2016), 353–377.

[10] R. Chartrand, *Exact reconstruction of sparse signals via nonconvex minimization*, IEEE Signal Process. Lett., 14 (2007), 707–710.

[11] X. Chen, Z. Lu and T.K. Pong, *Penalty methods for a class of non-Lipschitz optimization problems*, SIAM J. Optim., 26 (2016), 1465–1492.

[12] X. Chen, L. Niu and Y. Yuan, *Optimality conditions and a smoothing trust region newton method for non-Lipschitz optimization*, SIAM J. Optim., 23 (2013), 1528–1552.

[13] X. Chen and S. Xiang, *Sparse solutions of linear complementarity problems*, Math. Program., 159 (2016), 539–556.

[14] X. Chen, F. Xu and Y. Ye, *Lower bound theory of nonzero entries in solutions of l_2-l_p minimization*, SIAM J. Sci. Comput., 32 (2010), 2832–2852.

[15] I. Daubechies, R. DeVore, M. Fornasier and C.S. Güntürk, *Iteratively reweighted least squares minimization for sparse recovery*, Commun. Pure. Appl. Math., 63 (2010), 1–38.

[16] A. d'Aspremont, *A semidefinite representation for some minimum cardinality problems*, IEEE C. Dec. Contr., 5 (2003), 4985–4990.

[17] J. Fan and R. Li, *Variable selection via nonconcave penalized likelihood and its oracle properties*, J. Amer. Stat. Assoc., 96 (2001), pp. 1348–1360.

[18] R. Fletcher, S. Leyffer, D. Ralph and S. Scholtes, *Local convergence of SQP methods for mathematical programs with equilibrium constraints*, SIAM J. Optim., 17 (2006), 259–286.

[19] S. Foucart and M. Lai, *Sparsest solutions of underdetermined linear systems via ℓ_q minimization for* $0 < q \leq 1$, Appl. Comput. Harmon. Anal., 26 (2009), 395–407.

[20] M. Fukushima and P. Tseng, *An implementable active-set algorithm for computing a B-stationary point of the mathematical program with linear complementarity constraints*, SIAM J. Optim., 12 (2002), 724–739.

[21] L. Guo and G.H. Lin, *Notes on some constraint qualifications for mathematical programs with equilibrium constraints*, J. Optim. Theory Appl., 156 (2013), 600–616.

[22] L. Guo, G.H. Lin and J.J. Ye, *Stability analysis for parametric mathematical programs with geometric constraints and its applications*, SIAM J. Optim., 22 (2012), 1151–1176.

[23] L. Guo, G.H. Lin, J.J. Ye and J. Zhang, *Sensitivity analysis for parametric mathematical programs with equilibrium constraints*, SIAM J. Optim., 24 (2014), 1206–1237.

[24] T. Hoheisel, C. Kanzow and A. Schwartz, *Theoretical and numerical comparison of relaxation methods for mathematical programs with complementarity constraints*, Math. Program., 137 (2013), 257–288.

[25] X. Hu and D. Ralph, *Convergence of a penalty method for mathematical programming with equilibrium constraints*, J. Optim. Theory Appl., 123 (2004), 365–390

[26] A.F. Izmailov and M.V. Solodov, *An active-set Newton method for mathematical programs with complementarity constraints*, SIAM J. Optim., 19 (2008), 1003–1027.

[27] A.F. Izmailov, M.V. Solodov and E.I. Uskov, *Global convergence of augmented Lagrangian methods applied to optimization problems with degenerate constraints, including problems with complementarity constraints*, SIAM J. Optim., 22 (2012), 1579–1606.

[28] C. Kanzow and A. Schwartz, *Mathematical programs with equilibrium constraints: enhanced Fritz John conditions, new constraint qualifications and improved exact penalty results*, SIAM J. Optim., 20 (2010), 2730–2753.

[29] KNITRO, https://www.artelys.com/en/optimization-tools/knitro (2016).

[30] Y. Liu, S. Ma, Y-H. Dai and S. Zhang, *A smoothing SQP framework for a class of composite ℓ_q minimization over polyhedron*, Math. Program., 158 (2016), 467–500.

[31] Y.L. Liu, S.H. Pan and S.J. Bi, *Equivalent Lipschitz surrogates for zero-norm and rank optimization problems*, J. Global Optim., 72 (2018), 679–704.

[32] X. Liu and J. Sun, *Generalized stationary points and an interior-point method for mathematical programs with equilibrium constraints*, Math. Program., 101 (2004), 231–261.

[33] Z. Lu, *Iterative reweighted minimization methods for ℓ_p regularized unconstrained nonlinear programming*, Math. Program., 147 (2014), 277–307.

[34] Z.Q. Luo, J.S. Pang and D. Ralph, Mathematical Programs with Equilibrium Constraints, Cambridge University Press, Cambridge, United Kingdom, 1996.

[35] G.P. McCormick, *Second order conditions for constrained minima*, SIAM J. Appl. Math., 15 (1967), 641–652.

[36] S. Nguyen and C. Dupuis, *An efficient method for computing traffic equilibria in networks with asymmetric transportation costs*, Transport. Sci., 18 (1984), 185–202.

[37] J.V. Outrata, *Optimality conditions for a class of mathematical programs with equilibrium constraints*, Math. Oper. Res., 24 (1999), 627–644.

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

[39] H.S. Scheel and S. Scholtes, *Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity*, Math. Oper. Res., 25 (2000), 1–22.

[40] S. Scholtes, *Convergence properties of a regularization scheme for mathematical programs with complementarity constraints*, SIAM J. Optim., 11 (2001), 918–936.

[41] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. Royal. Statist. Soc. Ser. B, 58 (1996), 267–288.

[42] S.J. Wright, D.N. Robert and M.T. Figueiredo, *Sparse reconstruction by separable approximation*, IEEE Trans. Signal Proces., 57 (2009), 2479–2493.

[43] H. Yang and H. Huang, Mathematical and Economic Theory of Road Pricing, Elsevier, Amsterdam, 2005.

[44] Y.F. Yin, S.M. Madanat and X.Y. Lu, *Robust improvement schemes for road networks under demand uncertainty*, Eur. J. Oper. Res., 198 (2009), 470–479.

[45] J.J. Ye, *Optimality conditions for optimization problems with complementarity constraints*, SIAM J. Optim., 9 (1999), 374–387.

[46] J.J. Ye, *Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints*, J. Math. Anal. Appl., 307 (2005), 350–369.

[47] J.J. Ye, D.L. Zhu and Q.J. Zhu, *Exact penalization and necessary optimality conditions for generalized bilevel programming problems*, SIAM J. Optim., 7 (1997), 481–507.

[48] C. Zhang, X. Chen and A. Sumalee, *Robust Wardrop's user equilibrium assignment under stochastic demand and supply: expected residual minimization approach*, Transport Res. B-Meth., 45 (2011), 534–552.

[49] Y. Zhao and D. Li, *Reweighted ℓ_1-minimization for sparse solutions to underdetermined linear systems*, SIAM J. Optim., 22 (2012), 1065–1088.

A An iteratively reweighted ℓ_1 minimization method for solving problem $(P_{\epsilon,\sigma})$

Algorithm A.1 *Choose an arbitrary initial point $y^0 \in \mathbb{R}^n$ and set $\imath = 0$.*

Step 1) *Solve the weighted ℓ_1 minimization problem*

$$\begin{array}{ll} \min & f(y) + p\sum_{i=1}^{r} w_i^{\imath} |D_i^T y| \\ \text{s.t.} & G(y) \geq 0,\ H(y) \geq 0,\ G(y)^T H(y) \leq \sigma \end{array} \tag{A.1}$$

to get $y^{\imath+1}$, where $w_i^{\imath} = (|D_i^T y^{\imath}| + \epsilon_i)^{p-1}$ for all $i = 1, \ldots, r$.

Step 2) *Set $\imath = \imath + 1$ and go to Step 1).*

The proof of the following theorem uses the techniques in [33] for iteratively reweighted ℓ_1 methods for unconstrained minimization problems. For completeness, we give a brief proof.

Theorem A.1 *Any accumulation point y^* of the sequence $\{y^{\imath}\}$ generated by Algorithm A.1 is a stationary point of problem $(P_{\epsilon,\sigma})$.*

Proof. Let q be such that $1/p + 1/q = 1$, and let

$$\Upsilon_\epsilon(y, w) := f(y) + p\sum_{i=1}^{r} \left[w_i(|D_i^T y| + \epsilon_i) - \frac{w_i^q}{q} \right].$$

It is easy to verify that for any $y \in \mathbb{R}^n$ and $\epsilon > 0$,

$$F_\epsilon(y) = \min_{w \geq 0} \Upsilon_\epsilon(y, w), \tag{A.2}$$

and for any $\imath \geq 0$,

$$w^{\imath} = \text{Arg}\min_{w \geq 0} \Upsilon_\epsilon(y^{\imath}, w), \qquad y^{\imath+1} \in \text{Arg}\min_{y \in \mathcal{F}_\sigma} \Upsilon_\epsilon(y, w^{\imath}). \tag{A.3}$$

It then follows that

$$F_\epsilon(y^{\imath+1}) = \Upsilon_\epsilon(y^{\imath+1}, w^{\imath+1}) \leq \Upsilon_\epsilon(y^{\imath+1}, w^{\imath}) \leq \Upsilon_\epsilon(y^{\imath}, w^{\imath}) = F_\epsilon(y^{\imath}). \tag{A.4}$$

Hence, the sequence of $\{F_\epsilon(y^{\imath})\}_{\imath \geq 0}$ is nonincreasing. Let $y^{\imath} \to y^*$ as $\imath \in K \to \infty$. This together with the continuity of F_ϵ and the monotonicity of $\{F_\epsilon(y^{\imath})\}_{\imath \geq 0}$ implies that $F_\epsilon(y^{\imath}) \to$

$F_\epsilon(y^*)$ as $\imath \to \infty$. Moreover, it is easy to see that $w^\imath \to (|D_i^T y^*| + \epsilon_i)^{p-1}$ as $\imath \in K \to \infty$ for all $i = 1, \ldots, r$. Then by (A.4), we have that $\Upsilon_\epsilon(y^{\imath+1}, w^\imath) \to F_\epsilon(y^*) = \Upsilon_\epsilon(y^*, w^*)$. It follows from the second relation in (A.3) that $\Upsilon_\epsilon(y^{\imath+1}, w^\imath) \le \Upsilon_\epsilon(y, w^\imath)$ for all $y \in \mathcal{F}_\sigma$. Upon taking limits on both sides of this inequality as $\imath \in K \to \infty$, we have that $\Upsilon_\epsilon(y^*, w^*) \le \Upsilon_\epsilon(y, w^*)$ for all $y \in \mathcal{F}_\sigma$. This means that $y^* \in \operatorname{Arg}\min_{y \in \mathcal{F}_\sigma} \Upsilon_\epsilon(y, w^*)$. By Fermat's rule (see, e.g., [38, Theorem 10.1]), it follows that

$$0 \in \nabla f(y^*) + p \sum_{i=1}^r D_i(|D_i^T y^*| + \epsilon_i)^{p-1} \operatorname{sign}(D_i^T y^*) + \mathcal{N}_{\mathcal{F}_\sigma}(y^*),$$

which is the stationary condition of problem $(P_{\epsilon,\sigma})$ at y^*. The proof is complete. ∎

Although any accumulation point of the sequence generated by Algorithm A.1 is a stationary point of problem $(P_{\epsilon,\sigma})$, all iteration points may not be approximate stationary points since $\|y^\imath - y^{\imath+1}\|$ may not converge to 0 as $\imath \to \infty$. However, a weak approximate stationary point can be obtained as follows.

Theorem A.2 *Assume that the sequence $\{y^\imath\}$ generated by Algorithm A.1 has a bounded subsequence. Then for any $\varepsilon > 0$, there exist $\tilde{y}^\imath$ and ζ satisfying $\|y^\imath - \tilde{y}^\imath\| \le \varepsilon$ and $\|\zeta\| \le \varepsilon$ such that*

$$\zeta \in \nabla f(\tilde{y}^\imath) + p \sum_{i=1}^r D_i(|D_i^T \tilde{y}^\imath| + \epsilon_i)^{p-1} \operatorname{sign}(D_i^T \tilde{y}^\imath) + \mathcal{N}_{\mathcal{F}_\sigma}(\tilde{y}^\imath).$$

Proof. Without loss of generality, we assume that the whole sequence $\{y^\imath\}$ is bounded. Let $\varepsilon_0 > 0$. Since the sequence $\{F_\epsilon(y^\imath)\}_{\imath \ge 0}$ is convergent, it follows that $\|F_\epsilon(y^\imath) - F_\epsilon(y^{\imath+1})\| \le \varepsilon_0$ when $\imath$ is sufficiently large. Then by (A.4), we have

$$0 \le \Upsilon_\epsilon(y^\imath, w^\imath) - \Upsilon_\epsilon(y^{\imath+1}, w^\imath) \le \varepsilon_0.$$

This means that $y^\imath$ is an ε_0-optimal solution to the problem of minimizing $\Upsilon_\epsilon(y, w^\imath)$ on $\mathcal{F}_\sigma$. Then by Ekeland's variational principal, there exists $\tilde{y}^\imath$ such that $\|\tilde{y}^\imath - y^\imath\| \le \sqrt{\varepsilon_0}$, $\Upsilon_\epsilon(\tilde{y}^\imath, w^\imath) \le \Upsilon_\epsilon(y^\imath, w^\imath)$, and $\tilde{y}^\imath$ is the unique minimizer of the problem

$$\min \ \Upsilon_\epsilon(y, w^\imath) + \sqrt{\varepsilon_0}\|y - \tilde{y}^l\| \quad \text{s.t. } y \in \mathcal{F}_\sigma.$$

Then by Fermat's rule (see, e.g., [38, Theorem 10.1]), we have that

$$0 \in \nabla f(\tilde{y}^\imath) + p \sum_{i=1}^r w_i^\imath D_i \operatorname{sign}(D_i^T \tilde{y}^\imath) + \sqrt{\varepsilon_0}\mathbb{B} + \mathcal{N}_{\mathcal{F}_\sigma}(\tilde{y}^\imath),$$

where $\mathbb{B}$ stands for the closed unit ball of $\mathbb{R}^n$. This and the definition of $w^\imath$ imply that

$$\begin{aligned} 0 \in \nabla f(\tilde{y}^\imath) &+ p \sum_{i=1}^r D_i(|D_i^T \tilde{y}^\imath| + \epsilon_i)^{p-1} \operatorname{sign}(D_i^T \tilde{y}^\imath) + \mathcal{N}_{\mathcal{F}_\sigma}(\tilde{y}^\imath) + \sqrt{\varepsilon_0}\mathbb{B} \\ &+ p \sum_{i=1}^r D_i[(|D_i^T y^\imath| + \epsilon_i)^{p-1} - (|D_i^T \tilde{y}^\imath| + \epsilon_i)^{p-1}] \operatorname{sign}(D_i^T \tilde{y}^\imath). \end{aligned} \tag{A.5}$$

Since $\{y^\imath\}$ is bounded, we can find an ε_0 such that when $\|\tilde{y}^\imath - y^\imath\| \le \sqrt{\varepsilon_0} \le \varepsilon/2$, the norm of the term in (A.5) is bounded above by $\varepsilon/2$. Then the desired result follows immediately. ∎