MPCC Strategies for Nonsmooth Nonlinear Programs

K. X. Wang and L. T. Biegler

College of Control Science and Engineering
Zhejiang University, Hangzhou, Zhejiang China 310027
Chemical Engineering Department
Carnegie Mellon University, Pittsburgh, PA USA 15213
email: biegler@cmu.edu

March 25, 2022

Abstract. This paper develops solution strategies for large-scale nonsmooth optimization problems. We transform nonsmooth programs into equivalent mathematical programs with complementarity constraints (MPCCs), and devise NLP-based strategies for their solution. For this purpose, two NLP formulations based on complementarity relaxations are put forward, one of which applies a parameterized formulation and operates with a bounding algorithm, with the aim of taking advantage of the NLP sensitivities in search for the solution; and the other relates closely to the well-studied Lin-Fukushima formulation. Relations between the solutions of these NLPs and of the MPCC is revealed by sensitivity analysis. With appropriate assumptions, the resulting solution of the NLP formulations are proved to be C- and M-stationary for the MPCCs in the limit. Numerical performance of the proposed formulations, and the formulations by Lin & Fukushima and by Scholtes are studied and compared, with selected examples from the MacMPEC collection and two large-scale distillation cases.

1 Introduction

We consider the nonsmooth system written as

$$c(x, y, q) = 0,\ y_j = \max(0, x_j),\ j = 1, \ldots, n_x, \tag{1}$$

where $x, y \in \mathbb{R}^{n_x}$ and $q \in \mathbb{R}^{n_q}$. The related optimization problem is given by

$$\begin{aligned} \min \quad & f(x, y, q) && \text{(2a)} \\ \text{s.t.} \quad & c(x, y, q) = 0 && \text{(2b)} \\ & y_j - \max(0, x_j) = 0,\ j = 1, \ldots, n_x, && \text{(2c)} \end{aligned}$$

where $f : \mathbb{R}^{2n_x+n_q} \to \mathbb{R}$ and $c : \mathbb{R}^{2n_x+n_q} \to \mathbb{R}^{n_c}$ are twice continuously differentiable functions, and their second-order derivatives are Lipschitz continuous. For simplicity, we assume that any inequalities in the original optimization problem can be transformed to barrier terms in f.

Based on concepts of generalized derivatives and nonsmooth equation solving of Clarke [9], Barton and coworkers [2, 30, 43, 46] have developed a powerful conceptual framework for the solution of engineering models with nonsmooth elements. In addition to providing algorithmic differentiation tools [30] and modeling strategies, they have demonstrated these approaches on nontrivial process systems, including thermodynamic models with phase transitions. On the other hand, nonsmooth optimization is difficult for large-scale nonlinear systems. Modern large-scale NLP algorithms [37] exploit exact first and second derivatives from the optimization and apply Newton-based approaches to solve for the KKT conditions. Since the solution of nonsmooth optimization problems do not satisfy KKT conditions, NLP methods do not apply directly to these systems.

An alternate approach to solve these optimization problems is through reformulation of the nonsmooth model to Mathematical Programs with Complementarity Constraints (MPCCs). Based on active research over the past four decades, NLP-based solution strategies have been developed to find stationary points of MPCCs. Moreover, more recent NLP methods incorporate exact Hessian information, which allows them to verify convergence to locally optimal solutions that satisfy second order sufficient conditions (SOSCs). In addition, Griewank, Walther, Hegerhorst-Schultchen et al. [17, 18, 20–22] have studied related *abs-normal* NLPs, which are also equivalent to MPCCs.

Nonlinear complementarity systems that are closely related to nonsmooth systems (1) are given by

$$c(z) = 0,\ 0 \le G(z) \perp H(z) \ge 0. \tag{3}$$

Defining

$$z = \begin{bmatrix} x \\ y \\ q \end{bmatrix}, \quad \begin{bmatrix} c(z) \\ y - G(z) \\ y - x - H(z) \end{bmatrix} = 0,$$

and noting that

$$0 \le y \perp y - x \ge 0 \quad \Leftrightarrow \quad y_j = \max(0, x_j),\ j = 1, \ldots, n_x,$$

shows the equivalence of (3) to (1), and also (2) to the following MPCC:

$$\begin{aligned} \text{MPCC}: \quad \min \quad & f(x, y, q) && \text{(4a)} \\ \text{s.t.} \quad & c(x, y, q) = 0 && \text{(4b)} \\ & 0 \le y \perp y - x \ge 0. && \text{(4c)} \end{aligned}$$

The purpose of this paper is to develop NLP-based frameworks for the solution of large-scale, nonsmooth optimization problems. The proposed approach is designed to take advantage of efficient, off-the-shelf NLP solvers frequently used for smooth systems optimization, where exact first and second order derivatives can be exploited. These solvers apply Newton-based algorithms to the KKT conditions and also provide dual information for parametric sensitivity of the optimum. The incentive to use these solvers is particularly great in the case of large-scale applications.

The remainder of this section presents NLP and MPCC preliminaries as well as a review of NLP-based MPCC and smoothing approaches. Section 2 develops solution strategies for the nonsmooth optimization equivalent MPCCs, including NLP formulations and the applicable algorithm. Convergence properties of the proposed approaches are analyzed in Section 3. Numerical studies in Section 4 verify the theoretical results and also demonstrate some practical issues.

1.1 Preliminaries

We first present essential NLP concepts with a parametric formulation, which is closely related to the strategies devised in Section 2.

Consider the general NLP problem

$$\min_{z} f(z) \quad \text{s.t. } c(z,p)=0,\ g(z,p)\le 0, \tag{5}$$

where $z \in \mathbb{R}^{n_z}$, $f : \mathbb{R}^{n_z} \to \mathbb{R}$, $c : \mathbb{R}^{n_z} \to \mathbb{R}^{n_c}$, $g : \mathbb{R}^{n_z} \to \mathbb{R}^{n_g}$, and $p \in \mathbb{R}^{n_p}$ is a fixed parameter. To characterize the solution of (5) we define its KKT point.

Definition 1.1. *(KKT, [37]) Karush–Kuhn–Tucker (KKT) conditions for Problem (5) are given by*

$$\begin{aligned} \nabla f(\bar{z}) + \nabla c(\bar{z},p)\bar{\lambda} + \nabla g(\bar{z},p)\bar{\mu} = 0, \\ c(\bar{z},p)=0, \quad 0 \le \bar{\mu} \perp g(\bar{z},p) \le 0, \end{aligned} \tag{6}$$

for some multipliers $\bar{\lambda}$ and $\bar{\mu}$, where $\bar{z}$ is a KKT point. We also define $\mathcal{L}(z,\lambda,\mu,p) = f(z) + c(z,p)^T\lambda + g(z,p)^T\mu$ as the Lagrange function of (5).

A constraint qualification (CQ) is required so that a KKT point is necessary for a local minimizer. For Problem (5) the following CQ is widely invoked.

Definition 1.2. *(LICQ, [37]) The linear independence constraint qualification (LICQ) holds at a feasible point z of (5) when the gradient vectors*

$$\nabla c_i(z,p), i=1,\ldots,n_c \quad \textit{and} \quad \nabla g_j(z,p),\ \forall j \in I_g(z) \tag{7}$$

are linearly independent, where $I_g(z) = \{j | g_j(z,p)=0\}$. LICQ at a KKT point also implies that the associated multipliers satisfying (6) are unique.

Theorem 1.3. *(SOSC, [13]) A KKT point $\bar{z}$ with multipliers $\bar{\lambda}$ and $\bar{\mu}$ is a strict local optimum of (5), if the following second-order sufficient conditions (SOSC) hold:*

$$d^T \nabla_{zz}\mathcal{L}\left(\bar{z},\bar{\lambda},\bar{\mu},p\right) d > 0 \tag{8}$$

for all $d \ne 0$, such that

$$\begin{aligned} &\nabla c_i(\bar{z},p)^T d = 0, \quad i=1,\ldots,n_c \\ &\nabla g_j(\bar{z},p)^T d = 0, \quad \textit{for all } \bar{\mu}_j > 0 \textit{ and } j \in I_g(\bar{z}) \\ &\nabla g_j(\bar{z},p)^T d \le 0, \quad \textit{for all } \bar{\mu}_j = 0 \textit{ and } j \in I_g(\bar{z}). \end{aligned} \tag{9}$$

Definition 1.4. *(Strict Complementarity, [13]) At a KKT point $\bar{z}$ of* (5) *and the associated multipliers $\bar{\lambda}$ and $\bar{\mu}$ satisfying (6), the strict complementarity condition (SC) is defined by $\bar{\mu}_j - g_j(\bar{z}, p) > 0$ for each $j \in I_g(\bar{z})$.*

Theorem 1.5. *(Sensitivity Results, [13]) Let $\bar{z}(p_0)$ be a KKT point of Problem* (5) *with nominal parameter value p_0, and assume that SC, LICQ and SOSC hold at $\bar{z}(p_0)$. Further let the functions f, c, g in (5) be at least $\ell + 1$ times continuously differentiable in z and ℓ times continuously differentiable in p. Then*

- *$\bar{z}(p_0)$ is an isolated minimizer, and the associated multipliers $\bar{\lambda}(p_0)$ and $\bar{\mu}(p_0)$ are unique;*
- *for p in a neighborhood of p_0, the set of active constraints remains unchanged;*
- *for p in a neighborhood of p_0, there exists an ℓ times continuously differentiable function $s(p) = \big(\bar{z}(p), \bar{\lambda}(p), \bar{\mu}(p)\big)$, that corresponds to a locally unique minimum for* (5)*;*
- *there exist finite Lipschitz constants $L_s, L_f > 0$, such that*

$$\|s(p) - s(p_0)\| \leq L_s \|p - p_0\| \text{ and } |f(\bar{z}(p)) - f(\bar{z}(p_0))| \leq L_f \|p - p_0\|. \tag{10}$$

We now generalize these properties to stationarity conditions of MPCC (4). Necessary conditions for a local minimizer of an MPCC are described by the concept of *B-stationarity.* A point $z^* = (x^*, y^*, q^*)$ is B-stationary, if it is feasible to MPCC (4) and $d = 0$ is a solution to the following linear program with complementarity constraints (LPCC):

$$\min_{d} \quad \nabla f(z^*)^T d \tag{11a}$$

$$\text{s.t.} \quad c(z^*) + \nabla c(z^*)^T d = 0 \tag{11b}$$

$$0 \leq y^* + d_y \perp (y^* + d_y) - (x^* + d_x) \geq 0. \tag{11c}$$

Verification of B-stationarity may require the solution of 2^m linear programs, where m is the cardinality of the biactive set $I_1(z^*) \cap I_2(z^*)$, with

$$\begin{aligned} I_1(z^*) &= \left\{ j | y_j^* = 0 \right\}, \\ I_2(z^*) &= \left\{ j | y_j^* - x_j^* = 0 \right\}. \end{aligned} \tag{12}$$

More stationarity concepts are developed by using *weak stationarity.* Given a feasible point z^* of MPCC (4), if there exist multipliers $\lambda^*, \sigma_1^*, \sigma_2^*$ satisfying

$$\nabla f(z^*) + \sum_{i=1}^{n_c} \lambda_i^* \nabla c_i(z^*) - \sum_{j \in I_1(z^*)} \sigma_{1j}^* \begin{bmatrix} 0 \\ e_j \\ 0 \end{bmatrix} - \sum_{j \in I_2(z^*)} \sigma_{2j}^* \begin{bmatrix} -e_j \\ e_j \\ 0 \end{bmatrix} = 0, \tag{13}$$

where e_j is a vector with the jth element being 1 and other elements being 0, then z^* is weakly stationary. Furthermore, z^* satisfies

- C-stationarity, if $\sigma_{1j}^* \sigma_{2j}^* \geq 0$ for all $j \in I_1(z^*) \cap I_2(z^*)$;

- M-stationarity, if $\sigma_{1j}^*, \sigma_{2j}^* > 0$ or $\sigma_{1j}^* \sigma_{2j}^* = 0$ for all $j \in I_1(z^*) \cap I_2(z^*)$;
- S-stationarity (i.e., strong stationarity), if $\sigma_{1j}^*, \sigma_{2j}^* \geq 0$ for all $j \in I_1(z^*) \cap I_2(z^*)$.

In particular, a strongly stationary point z^* of MPCC (4) solves the relaxed NLP given by

$$\begin{array}{rll} \text{RNLP}: & \min & f(x,y,q) \\ & \text{s.t.} & c(x,y,q) = 0 \\ & & y_j = 0,\ y_j - x_j > 0, \quad j \in I_1(z^*) \setminus I_2(z^*) \\ & & y_j > 0,\ y_j - x_j = 0, \quad j \in I_2(z^*) \setminus I_1(z^*) \\ & & y_j \geq 0,\ y_j \geq x_j, \qquad\ \ j \in I_1(z^*) \cap I_2(z^*). \end{array} \tag{14}$$

In general, S-stationarity is stronger than B-stationarity. A notable exception is their equivalence in the presence of MPCC-LICQ [40], namely, the following set of gradients is linearly independent at z^*:

$$\{\nabla c_i(z^*) \,|\, i = 1, \ldots, n_c\} \cup \left\{ \begin{bmatrix} 0 \\ e_j \\ 0 \end{bmatrix} \,|\, j \in I_1(z^*) \right\} \cup \left\{ \begin{bmatrix} -e_j \\ e_j \\ 0 \end{bmatrix} \,|\, j \in I_2(z^*) \right\}. \tag{15}$$

1.2 Previous Work on NLP-based MPCC Solution

Transforming a large-scale nonsmooth system into the equivalent MPCC, we usually arrive at a model that contains a small part of complementarity conditions, while the remaining are regular smooth (nonlinear) equations. Dealing with the complementarity structure explicitly can be expensive, which has motivated the adaptation of well-developed NLP methods for MPCC solution. However, complementarity conditions, for example (4c), introduce a combinatorial structure to the equivalent NLP and generally result in failure of standard regularity assumptions on the NLP constraints, posing challenges for regular NLP algorithms even to find a feasible point. Starting from seminal monographs on MPCCs [11, 36], a rich, comprehensive framework has been developed to characterize MPCC solutions and algorithmic strategies.

Fukushima and Pang [16] study the behavior of a sequence generated by a smoothing continuation method for MPCCs using the Fischer-Burmeister smoothing function. They show that under the linear independence constraint qualifications and an additional condition called the asymptotic weak nondegeneracy, the limit of KKT points satisfying the second-order necessary conditions for the perturbed problems is a B-stationary point of the original MPCC.

Scheel and Scholtes [40] characterize B-, C-, M- and S-stationarity for MPCCs, develop second-order optimality conditions and present some stability results for MPCCs. These properties also relate to penalty formulations and relaxed NLP problems. Scholtes [41] considers a sequence of stationary points of parametric NLPs which relax MPCC solutions with a vanishing positive t:

$$\text{REG}(t): \quad \min \quad f(x,y,q) \tag{16a}$$
$$\text{s.t.} \quad c(x,y,q) = 0 \tag{16b}$$
$$y \geq 0, y - x \geq 0 \tag{16c}$$
$$y_j(y_j - x_j) \leq t,\ j = 1, \ldots, n_x. \tag{16d}$$

He shows that stationary points are C-stationary if MPCC-LICQ qualification holds; they are M-stationary if, in addition, an approaching subsequence satisfies second-order necessary conditions, and they are B-stationary if, in addition, an upper level strict complementarity condition holds. These complement the results of [16]. It is further shown that every local minimizer of the MPCC which satisfies the linear independence, upper level strict complementarity, and a second-order optimality condition, can be embedded into a locally unique piecewise smooth curve of local minimizers of the parametric NLP (16).

Lin and Fukushima [35] present a bounding scheme for MPCCs using two-sided relaxations of complementarity constraints. This approach also provides lower bounds on the MPCC solution. Under mild assumptions they show that their approach converges to C-stationary points, with additional sufficient conditions for M-stationarity. Several additional two-sided relaxation approaches have been developed, including a nonsmooth relaxation [26], a local-support approach [44], and a relaxation of both complementarity and nonnegativity [10]. These approaches also converge to C-stationary points, under reasonable assumptions.

Moreover, Guo et al. [19] present an approach where the MPCC conditions (for C/M/S stationarity) are reformulated as smooth equations with box constraints. A modified Levenberg–Marquardt method is developed to solve these constrained equations. The method is shown to be locally and superlinearly convergent, and sufficient conditions are given for local error bounds.

Ralph and Wright [39] describe properties of regularized and penalized NLP formulations for MPCCs, and focus on properties of these formulations near MPCC local solutions, where strong stationarity and second-order sufficient conditions are satisfied. Existence and uniqueness of solutions for these formulations are investigated, and estimates are obtained for the distance of these solutions to MPCC solutions.

Hoheisel, Kanzow and Schwartz [23] provide a theoretical and numerical comparison of several relaxed MPCC schemes. In particular, they improve the convergence properties of several existing relaxation methods for MPCCs, show which CQs are satisfied by relaxed problems and present a numerical comparison of all relaxation schemes based on the MacMPEC test problem collection [33]. Kanzow and Schwartz [29] also consider relaxation methods for MPCCs, based on solving a sequence of nonlinear programs depending on a vanishing parameter. Most of these relaxation methods can obtain C-stationary points, although M-stationary points can be obtained with stronger second-order conditions. Moreover, Hoheisel et al. [23] and Schwartz [42], generalize the definition of stationarity conditions for MPCCs, through relaxed constraint qualifications (i.e., MPCC-LICQ $\implies$ MPCC-MFCQ $\implies$ MPCC-GCQ) to define appropriate stationarity conditions. In particular, it has been shown that a local minimum that satisfies MPCC-LICQ also satisfies strong stationarity (see also [15, 40]), while a local minimum that satisfies MPCC-GCQ also satisfies M-stationarity (see also [14]). Weaker necessary conditions with MPCC-GCQ include C-, A- and weak stationarity. Also, they show that the C-stationarity results of [41] and [35] hold under the weaker MPCC-MFCQ condition.

More recently, Hegerhorst-Schultchen and coworkers [20–22] considered abs-normal NLPs, developed in [17, 18], which are given by

$$\min f(x, |z|) \;\text{ s.t. }\; c_E(x, |z|) = 0,\; c_I(x, |z|) \geq 0,\; c_Z(x, |z|) - z = 0.$$

The objective and constraint functions are level-one nonsmooth functions, as in (2). The abs-normal NLP can be reformulated to the MPCC:

$$\min f(x, u+v) \;\text{ s.t. }\; c_E(x, u+v) = 0,\; c_I(x, u+v) \geq 0,\; c_Z(x, u+v) - u - v = 0,\; 0 \leq u \perp v \geq 0,$$

and can also be written equivalently using max operators, i.e., $|z_i| = \max(0, z_i) + \max(0, -z_i)$. An extensive analysis of this problem is provided in [22], where optimality conditions are characterized with extensions of KKT conditions and kink qualifications that replace constraint qualifications. Moreover, the relationships of abs-normal NLPs to MPCCs in terms of specialized constraint qualifications are provided in detail as well as stationary point properties relating to first and second order optimality conditions.

1.3 Smoothing the Max Operator

NCP-functions φ represent complementarity (4c) as

$$\varphi(y_j, y_j - x_j) = 0 \quad \text{if and only if} \quad y_j \geq 0, y_j - x_j \geq 0, y_j(y_j - x_j) = 0. \tag{17}$$

The functions are usually Lipschitz-continuous but not differentiable at $(y_j, y_j - x_j) = (0, 0)$; therefore their perturbed smooth approximations are often used. A typical example is the perturbed Fischer-Burmeister function:

$$\varphi^t_{\text{FB}}(y_j, y_j - x_j) = y_j + (y_j - x_j) - \sqrt{y_j^2 + (y_j - x_j)^2 + t^2},$$

where the smoothing factor $t \geq 0$, and $t = 0$ recovers the property (17). A wide variety of NCP-functions have been developed and applied to reformulate MPCCs and approximate the solution by solving a sequence of NLPs with their smoothing factor tending to zero [12, 16, 25, 29, 34]. In this research, we reformulate the max operator in (1) with a class of parametric smooth NCP-functions generated from distribution density functions $\delta(\xi), \xi \in \mathbb{R}$ [8], which approximate a Dirac delta function, the twice derivative of the max function.

Assume the density function $\delta(\xi)$ satisfies the following properties.

Assumption 1.6. *The function $\delta(\xi)$ is smooth and has infinite support, i.e.,*

$$\delta(\xi) > 0,\ \forall \xi \in \mathbb{R};$$

and

$$\int_{-\infty}^{\infty} \delta(\xi)\mathrm{d}\xi = 1,\ \int_{-\infty}^{\infty} |\xi|\delta(\xi)\mathrm{d}\xi < \infty. \tag{18}$$

The shape of the density function can be parametrized by ϵ and we define the parametrized density as

$$d(\xi, \epsilon) = \frac{2}{\epsilon}\delta(2\xi/\epsilon), \tag{19}$$

the smoothed step function as

$$s(\xi, \epsilon) = \int_{-\infty}^{\xi} d(\bar{\xi}, \epsilon)\mathrm{d}\bar{\xi}, \tag{20}$$

and the smoothed max function as

$$m(\xi, \epsilon) = \int_{-\infty}^{\xi} s(\bar{\xi}, \epsilon) \mathrm{d}\bar{\xi}. \tag{21}$$

Function (21) satisfies the following properties.

Proposition 1.7. *[8, Proposition 2.2] Let Assumptions 1.6 hold, then:*

1. *$m(\xi, \epsilon)$ is continuously smooth.*
2. *The following inequalities hold.*

$$0 \leq m(\xi, \epsilon) - \max(0, \xi) \leq \kappa\epsilon/2, \text{ where } \kappa = \int_{-\infty}^{0} |\xi| \delta(\xi) \mathrm{d}\xi; \tag{22a}$$

$$m(\xi, \epsilon) > \xi; \tag{22b}$$

$$0 < \nabla_\xi m(\xi, \epsilon) < 1. \tag{22c}$$

3. *$m(\xi, \epsilon)$ is strictly increasing with ξ and strictly convex.*
4. *$m(0, \epsilon) = \kappa\epsilon/2$.*
5. *$\max_\xi [m(\xi, \epsilon) - \max(0, \xi)] = m(0, \epsilon) = \kappa\epsilon/2$.*

Proof. Items 1 – 4 follow directly from Proposition 2.2 in [8]. Item 5 is proved as follows. Defining $\phi_+(\xi) = m(\xi; \epsilon) - \xi$ for $\xi > 0$ and $\phi_-(\xi) = m(\xi; \epsilon)$ for $\xi < 0$, we have from Item 2 that $\nabla\phi_+(\xi) < 0$ and $\nabla\phi_-(\xi) > 0$. Hence $\phi_+(0) > \phi_+(\xi_1) > \phi_+(\xi_2)$ for $0 < \xi_1 < \xi_2$ and $\phi_-(0) > \phi_-(\xi_1) > \phi_-(\xi_2)$ for $0 > \xi_1 > \xi_2$. Since $\phi_-(0) = \phi_+(0)$ and $m(\xi, \epsilon) - \max(0, \xi)$ is absolutely continuous, it has its maximum value at $\xi = 0$. □

We consider two popular examples of $m(\xi, \epsilon)$ that satisfy Proposition 1.7:

- The smoothed square root function [1,6] with

$$m(\xi, \epsilon) = \frac{1}{2}(\xi + \sqrt{\xi^2 + \epsilon^2}), \tag{23}$$

  where $\kappa = 1$ and $\nabla_\xi m(0, \epsilon) = s(0, \epsilon) = 1/2$.

- The neural network function [7] with

$$m(\xi, \epsilon) = \xi + \frac{\epsilon}{2} \log(1 + e^{-2\xi/\epsilon}), \tag{24}$$

  where $\kappa = \log 2$ and $\nabla_\xi m(0, \epsilon) = s(0, \epsilon) = 1/2$.

By defining the vector function $h^\epsilon(x)$ with elements $h_j^\epsilon(x) = m(x_j, \epsilon),\ j = 1, \ldots, n_x$, we can rewrite (2) as the modified NLP:

$$\min \quad f(x, y, q) \tag{25a}$$

$$\text{s.t.} \quad c(x, y, q) = 0 \tag{25b}$$

$$y - h^\epsilon(x) = 0. \tag{25c}$$

2 Solution Strategies

To solve (2) we develop a strategy that considers the errors of the smoothed function $h^\epsilon(x)$, which is embedded within a parametric form of NLP (25). This modification, i.e., NLP (27), can be adapted to provide a local upper bound to (2). A relaxed smooth problem, i.e., NLP (28), is also formulated, which provides a local lower bound to (2). These two problems are the basis for the proposed algorithm, which finds a solution of (2) by solving a sequence of NLPs with decreasing ϵ and, additionally, adaptively determined parameters. This section provides background concepts and properties for the NLP formulations and the resulting algorithm.

2.1 Smooth NLP Formulations for Problem (2)

The KKT conditions for (25) can be written as

$$\nabla_x f(x,y,q) + \nabla_x c(x,y,q)\lambda - \nabla_x h^\epsilon(x) u = 0 \quad (26a)$$
$$\nabla_y f(x,y,q) + \nabla_y c(x,y,q)\lambda + u = 0 \quad (26b)$$
$$\nabla_q f(x,y,q) + \nabla_q c(x,y,q)\lambda = 0 \quad (26c)$$
$$c(x,y,q) = 0,\ y - h^\epsilon(x) = 0. \quad (26d)$$

Assume that NLP (25) satisfies LICQ and SOSC at its local solution $\bar{z} = (\bar{x}, \bar{y}, \bar{q})$. Defining the parametric program:

$$\min \quad f(x,y,q) \quad (27a)$$
$$s.t. \quad c(x,y,q) = 0 \quad (27b)$$
$$y - h^\epsilon(x) + p = 0, \quad (27c)$$

the same KKT conditions (26) apply, with the addition of parameter $p \in \mathbb{R}^{n_x}$ in the last equation. We know from Theorem 1.5, that for $\epsilon > 0$ sufficiently small and $p_j \in [0, \kappa\epsilon/2]$ for all $j = 1, \ldots, n_x$, there exists a local solution $z(\epsilon)$ within an ϵ-ball of $\bar{z}$; and we can easily obtain sensitivity information from the KKT conditions for (27). Note that the interval for p_j is sufficient to find a feasible point of the original problem (2), since $0 \leq h^\epsilon_j(x) - \max(0, x_j) \leq \kappa\epsilon/2$ (Item 2, Proposition 1.7).

We also consider the relaxed NLP given by

$$\min \quad f(x,y,q) \quad (28a)$$
$$\text{s.t.} \quad c(x,y,q) = 0 \quad (28b)$$
$$-(\kappa\epsilon/2)e \leq y - h^\epsilon(x) \leq 0, \quad (28c)$$

where $e^T = [1, 1, \ldots, 1]$. Here we assume that (28) satisfies SC, LICQ and SOSC at its KKT points. Note that the feasible region of (28) contains the feasible region of (2) and the solution of (28) therefore locally provides a lower bound to (2). The KKT conditions for

(28) are given by

$$\begin{aligned}
\nabla_x f(x,y,q) + \nabla_x c(x,y,q)\lambda - \nabla_x h^\epsilon(x)(u_U - u_L) &= 0 && \text{(29a)}\\
\nabla_y f(x,y,q) + \nabla_y c(x,y,q)\lambda + (u_U - u_L) &= 0 && \text{(29b)}\\
\nabla_q f(x,y,q) + \nabla_q c(x,y,q)\lambda &= 0 && \text{(29c)}\\
0 \le u_L \perp y - h^\epsilon(x) + (\kappa\epsilon/2)e \ge 0 & && \text{(29d)}\\
0 \le u_U \perp y - h^\epsilon(x) \le 0. & && \text{(29e)}
\end{aligned}$$

Note that u in (26) is replaced by $u_U - u_L$ in (29). Moreover, since the constraints (28c) cannot be active simultaneously for $\epsilon > 0$, we have $0 \le u_L \perp u_U \ge 0$.

To derive ϵ-bounds for a solution of (2), we note that for $\epsilon > 0$ there exist some values of $p_j \in [0, \kappa\epsilon/2],\ j = 1, \dots, n_x$, with which a solution of (27) is feasible to (2) and hence forms an upper bound to (2). On the other hand, a solution of (2) is feasible to Problem (28); hence (28) can provide a lower bound to (2). Moreover, problems (27) and (28) are closely related, so that for $\epsilon > 0$ sufficiently small, we can make sensitivity corrections with respect to p at the solution of (27), which approximates the solution of (28). These observations can be summarized by the following proposition.

Proposition 2.1. *Let LICQ and SOSC hold at local solutions of (25) and (27), and SC, LICQ and SOSC hold at local solutions of (28), for $\epsilon > 0$ suitably small and $p_j \in [0, \kappa\epsilon/2],\ j = 1, \dots, n_x$. Then the following statements hold.*

1. *A solution of (25) along with a correction based on linearization of the KKT conditions provides an ϵ^2-approximate solution to (27) with $p \neq 0$.*

2. *A solution of (25) (or (27)) with sensitivity corrections provides an ϵ^2-approximate solution to (28).*

3. *Let $z(p)$ be a KKT point for (27), then the sensitivity of $f(z(p))$ with respect to p, i.e., $\frac{\mathrm{d}f(z(p))}{\mathrm{d}p}$, is given directly by $u(p)$.*

Proof. Each claim is proved as follows.

1. Theorem 1.5 allows the KKT conditions of (25) to be expanded in a Taylor series. Define $s(p) = (z(p), \lambda(p), u(p))$ as the primal-dual solution of (27) and $s(0) = (z(0), \lambda(0), u(0))$ as the primal-dual solution of (25). For the Lagrangian $\mathcal{L} = f(z) + c(z)^T\lambda + (y - h^\epsilon(x) + p)^T u$, represent the KKT conditions (26) as $\nabla_s \mathcal{L}(s(0), 0) = 0$ and the KKT conditions for (27) as $\nabla_s \mathcal{L}(s(p), p) = 0$. Using the solution of (25), the sensitivity correction to approximate the solution of (27) is derived from the following Taylor expansion:

   $$\nabla_s \mathcal{L}(s(0),0) + \left(\nabla_{sp}\mathcal{L}(s(0),0)^T + \nabla_{ss}\mathcal{L}(s(0),0)\frac{\mathrm{d}s}{\mathrm{d}p}^T\right) p + O(\|p\|^2) = \nabla_s \mathcal{L}(s(p),p).$$

   Since $\nabla_s \mathcal{L}(s(0),0) = \nabla_s \mathcal{L}(s(p),p) = 0$, we have

   $$\lim_{\|p\|\to 0} \left(\nabla_{sp}\mathcal{L}(s(0),0)^T + \nabla_{ss}\mathcal{L}(s(0),0)\frac{\mathrm{d}s}{\mathrm{d}p}^T\right)\frac{p}{\|p\|} = 0,$$

which implies that

$$\frac{\mathrm{d}s}{\mathrm{d}p}^T = -\nabla_{ss}\mathcal{L}(s(0),0)^{-1}\nabla_{sp}\mathcal{L}(s(0),0)^T. \tag{30}$$

Also since LICQ and SOSC hold, $\nabla_{ss}\mathcal{L}$ is nonsingular and bounded in ϵ neighborhood of solutions of (25) and (27). Applying Taylor's theorem, and noting that $\nabla_{sp}\mathcal{L}(s(0),0)^T p = \nabla_s\mathcal{L}(s(0),p)$ and $\nabla_s\mathcal{L}(s(p),p) = 0$, we can derive

$$\begin{aligned}
s(0)+\frac{\mathrm{d}s}{\mathrm{d}p}^T p - s(p) &= s(0) - s(p) - \nabla_{ss}\mathcal{L}(s(0),0)^{-1}\nabla_s\mathcal{L}(s(0),p)\\
&= \nabla_{ss}\mathcal{L}(s(0),0)^{-1}[\nabla_{ss}\mathcal{L}(s(0),0)(s(0)-s(p)) - (\nabla_s\mathcal{L}(s(0),p) - \nabla_s\mathcal{L}(s(p),p))]\\
&= \nabla_{ss}\mathcal{L}(s(0),0)^{-1}\left(\nabla_{ss}\mathcal{L}(s(0),0) - \nabla_{ss}\mathcal{L}(s',p)\right)(s(0)-s(p)),
\end{aligned}$$

where $s' = s(p) + \tau(s(0)-s(p))$ for some $\tau \in (0,1)$. Hence,

$$\left\| s(0) + \frac{\mathrm{d}s}{\mathrm{d}p}^T p - s(p) \right\| \le C_{\mathrm{inv}} L_{\mathrm{kkt}} \|s(0)-s(p)\|^2 = O(\|p\|^2) = O(\epsilon^2), \tag{31}$$

where L_{kkt} is the Lipschitz constant for $\nabla_{ss}\mathcal{L}$, and C_{inv} is the upper bound on the norm of its inverse; and we have used the fourth item of Theorem 1.5 and $p_j \in [0, \kappa\epsilon/2]$ for the equalities.

2. From the solution of Problem (28) given by (z^-, λ, u_L, u_U), define $p^- = h^\epsilon(x^-) - y^-$ and $u = u_U - u_L$. Substituting p^- for p in Problem (27) leads to the same solution z^-, with the multipliers (λ, u).

   We also note that $p_j^- \in [0, \kappa\epsilon/2]$ because of (28c). To show that the solutions of (25) and (27) with sensitivity corrections are ϵ^2-approximate to (28), or equivalently, to (27) with parameter p^-, we apply the result from Item 1 and note that

$$\begin{aligned}
s(0) + \frac{\mathrm{d}s}{\mathrm{d}p}^T p^- - s(p^-) &= O(\|p^-\|^2) = O(\epsilon^2),\\
s(p) + \frac{\mathrm{d}s}{\mathrm{d}p}^T (p^- - p) - s(p^-) &= O(\|p - p^-\|^2) = O(\epsilon^2).
\end{aligned}$$

3. Consider the Lagrange function for (27) at two KKT points and two parameter values $p \neq p'$. Applying Taylor's theorem with $p(\tau) = p + \tau(p'-p)$ ($\tau \in [0,1]$) leads to

$$\begin{aligned}
f(z(p')) - f(z(p)) &= \mathcal{L}(s(p'),p') - \mathcal{L}(s(p),p)\\
&= \int_0^1 \left(\nabla_p\mathcal{L}(s(p(\tau)),p(\tau))^T + \nabla_s\mathcal{L}(s(p(\tau)),p(\tau))^T\frac{\mathrm{d}s}{\mathrm{d}p}^T\right)(p'-p)\mathrm{d}\tau\\
&= \int_0^1 u(p(\tau))^T(p'-p)\mathrm{d}\tau,
\end{aligned} \tag{32}$$

where we used $\nabla_s\mathcal{L}(s(p(\tau)),p(\tau)) = 0$ to deduce the last equality. As $p' \to p$, we obtain that $\mathrm{d}f(z(p))/\mathrm{d}p = u(p)$.

□

2.2 Bounding Algorithm for Problem (2)

From the above properties, for any parameters $p_j, p'_j \in [0, \kappa\epsilon/2]$ $(j = 1, \ldots, n_x)$ where $\epsilon > 0$, and the corresponding solutions $z(p)$ and $z(p')$ of (27), it is straightforward to show that

$$f(z(p')) = f(z(p)) + \frac{\mathrm{d}f(z(p))}{\mathrm{d}p}(p' - p) + O(\|p' - p\|^2),$$

and hence that

$$f(z(p)) + \sum_{j=1}^{n_x} |u_j(p)(p'_j - p_j)| + |O(\epsilon^2)| \geq f(z(p')) \geq f(z(p)) - \sum_{j=1}^{n_x} |u_j(p)(p'_j - p_j)| - |O(\epsilon^2)|. \tag{33}$$

This can be specified for a (local) solution z^- of (28) and z^* of (2) as

$$f(z(p)) + \frac{\kappa\epsilon}{2} \sum_{j=1}^{n_x} |u_j(p)| + |O(\epsilon^2)| \geq f(z^*) \geq f(z^-) \geq f(z(p)) - \frac{\kappa\epsilon}{2} \sum_{j=1}^{n_x} |u_j(p)| - |O(\epsilon^2)|. \tag{34}$$

To isolate the solution of (2), we devise the Bounding Algorithm. For a sequence of smoothing parameters $\{\epsilon^k\}$ tending to zero, NLP (27) is solved to local solutions $\{z^k\}$. At each solution $z^k = z(p^k)$, the sensitivities $u(p^k)$ are applied to estimate both the upper bound of $f(z^*)$ and $f(z^-)$ in (34) based on the current objective value $f(z(p^k))$, as well as to update the parameters p^{k+1} as shown in (35) for the subsequent formulation (27) to be solved. At its convergence, the Bounding Algorithm arrives at an ϵ_{tol}-approximate solution of (2), for the prescribed tolerance ϵ_{tol}.

A sensitivity analysis plays an important role in the course of exploring a MPCC solution. As shown in Step 3 of the algorithm, possible improvement of the objective value at a solution z^k is examined based on sensitivities $\frac{\mathrm{d}f(z^k)}{\mathrm{d}p_j^k}$ $(j = 1, \ldots, n_x)$, given by $u_j^k = u_j(p^k)$. In the case of $u_j^k < 0$ (or $u_j^k > 0$), increasing (or decreasing) the value of p_j^k can lead to decrease in the objective function. If these p_j^k stay within $[0, \kappa\epsilon/2]$ after the adaptation, then they are recorded by the sets J_0 and J_ϵ, which are used in (35) to prescribe the parameters for the next solution.

In fact, the adaptation of parameters p_j changes the active bounds of the underlying problem (28). Specifically, for all $j \in J_0$, the constraints $y_j - h_j^\epsilon(x)$ leave their upper bounds at z^k and instead come to the lower bounds at z^{k+1}; and the inverse happens for all $j \in J_\epsilon$. NLP (27) is solved repeatedly with such possible 'jumping' of p_j between 0 and $\kappa\epsilon/2$ as ϵ tending to zero, in the hope that this will identify the correct active set and approximate the solution of (2) better in subsequent steps.

Bounding Algorithm: A procedure to isolate the solution of (2)

Specify initial smoothing factor $\epsilon^0 > 0$, reducing factor $\gamma \in (0,1)$, initial point $z^0 = (x^0, y^0, q^0)$, solution tolerance $\epsilon_{\text{tol}} > 0$. Set initial paramater $p^0 \leftarrow 0$, counter $k \leftarrow 0$.

(Optional) For some $\tilde{\epsilon} \in (\epsilon_{\text{tol}}, \epsilon^0)$, starting from ϵ^0 and z^0, solve a sequence of problems (25) with $\epsilon \to \tilde{\epsilon}$, to obtain a local solution $\tilde{z} = (\tilde{x}, \tilde{y}, \tilde{q})$. Set $\epsilon^0 \leftarrow \tilde{\epsilon}$, $z^0 \leftarrow \tilde{z}$, and $p_j^0 \leftarrow \tilde{y}_j - \max(0, \tilde{x}_j)$.

while $\epsilon^k \geq \epsilon_{\text{tol}}$ **do**

Step 1: Solve NLP (27) with p^k to obtain primal variables $z^k = (x^k, y^k, q^k)$ and dual variables (λ^k, u^k).

Step 2: Approximate the upper bound of (2) with

$$f^{up} = f(z^k) + (\kappa\epsilon^k) \sum_{j\in\{1,\ldots,n_x\}} |u_j^k|.$$

Step 3: Approximate the objective of Problem (28) as follows.

- Let $J_0 = \{j | p_j^k = 0 \text{ and } u_j^k < 0\}$; setting $p_j^k = \frac{\kappa\epsilon^k}{2}$ for $j \in J_0$ would reduce $f(z^k)$.
- Let $J_\epsilon = \{j | p_j^k = \frac{\kappa\epsilon^k}{2} \text{ and } u_j^k > 0\}$; setting $p_j^k = 0$ for $j \in J_\epsilon$ would reduce $f(z^k)$.
- The objective value with the above p^k adaptation would be (approximately)

$$f^{low} = f(z^k) - (\kappa\epsilon^k) \sum_{j\in J_0 \cup J_\epsilon} |u_j^k|.$$

Step 4: Set $\epsilon^{k+1} \leftarrow \gamma\epsilon^k$, correct the parameters as

$$p_j^{k+1} = \begin{cases} \kappa\epsilon^{k+1}/2, & j \in J_0; \\ 0, & j \in J_\epsilon; \\ \gamma p_j^k, & \text{otherwise.} \end{cases} \tag{35}$$

Step 5: Set $k \leftarrow k+1$ and go to Step 1.

end

2.3 Relation to Lin-Fukushima Regularization

We reveal the equivalence between formulation (28) and the well-studied regularization proposed by Lin and Fukushima in [35]. As a consequence, convergence results derived for the latter should be also applicable to (28). However, at the same time, we can see technical differentiation between these two relaxations, which motivates further discussion on convergence in the next section.

According to the method of [35], MPCC (4) is approximated by

$$\text{LF}(t): \quad \min \quad f(x,y,q) \tag{36a}$$

$$s.t. \quad c(x,y,q) = 0 \tag{36b}$$

$$\Psi_{L,j}(z) = (y_j + t)(y_j - x_j + t) \geq t^2, \; j = 1,\ldots,n_x \tag{36c}$$

$$\Psi_{U,j}(z) = y_j(y_j - x_j) \leq t^2, \; j = 1,\ldots,n_x, \tag{36d}$$

where t is a positive parameter. Suppose MPCC-LICQ (15) holds at a feasible point $\hat{z}$ of

MPCC (4). It has been proved that in a neighborhood $\mathcal{N}(\hat{z})$ of $\hat{z}$, standard LICQ holds at every feasible point $z \in \mathcal{N}(\hat{z})$ of NLP (36) for $t > 0$ sufficiently small.

With this constraint qualification, convergence results have been established for MPCC (4). Specifically, for a sequence $\{t^k\}$ with $\lim_{k\to\infty} t^k = 0$, suppose the stationary points $\{z^k\}$ of (36) have an accumulation point z^* where MPCC-LICQ holds, then z^* is C-stationary for MPCC (4). Furthermore, if every z^k meets additional second-order conditions

$$d^T \nabla_{zz} \mathcal{L}_{\mathrm{LF}}(z^k, \lambda^k, \mu_U^k, \mu_L^k) d \geq -\alpha_{\mathrm{LF}}^k \|d\|^2, \tag{37}$$

for the Lagrangian $\mathcal{L}_{\mathrm{LF}} = f(z^k) + c(z^k)^T\lambda^k - (\Psi_L(z^k) - (t^k)^2)^T \mu_L^k + (\Psi_U(z^k) - (t^k)^2)^T \mu_U^k$, for the bounded sequence $\{\alpha_{LF}^k\}$ of positive constants, and all the directions d (chosen to be bounded) in the set

$$\mathcal{D}_{\mathrm{LF}}(z^k) = \left\{ d \middle| \begin{array}{ll} \nabla c_i(z^k)^T d = 0, & i = 1, \dots, n_c \\ \nabla \Psi_{L,j}(z^k)^T d = 0, & \forall j \in I_{\Psi_L}(z^k, t^k) = \{j \mid \Psi_{L,j}(z^k) = (t^k)^2\} \\ \nabla \Psi_{U,j}(z^k)^T d = 0, & \forall j \in I_{\Psi_U}(z^k, t^k) = \{j \mid \Psi_{U,j}(z^k) = (t^k)^2\} \end{array} \right\},$$

then z^* is M-stationary for MPCC (4). These convergence properties have been extended in later studies. C-stationarity is proved under a weaker MPCC-MFCQ assumption on z^* [24]. In addition, when the sequence of NLPs (36) is only solved approximately, C-stationarity of z^* still holds as analyzed in [27, 28].

We establish the relation between NLPs (28) and (36), based on the smoothed square root function (23). Then from the upper bound of (28c) we have

$$\begin{aligned} & y_j - \frac{x_j + \sqrt{x_j^2 + \epsilon^2}}{2} \leq 0 \\ \Leftrightarrow \quad & y_j + (y_j - x_j) \leq \sqrt{x_j^2 + \epsilon^2} = \sqrt{(y_j - (y_j - x_j))^2 + \epsilon^2} \\ \Leftrightarrow \quad & y_j(y_j - x_j) \leq \epsilon^2/4; \end{aligned} \tag{38}$$

and from the lower bound we have

$$\begin{aligned} & y_j - \frac{x_j + \sqrt{x_j^2 + \epsilon^2}}{2} \geq -\epsilon/2 \\ \Leftrightarrow \quad & y_j + (y_j - x_j) + \epsilon \geq \sqrt{x_j^2 + \epsilon^2} = \sqrt{(y_j - (y_j - x_j))^2 + \epsilon^2} \\ \Leftrightarrow \quad & y_j(y_j - x_j) + \frac{\epsilon}{2}(y_j + (y_j - x_j)) \geq 0 \\ \Leftrightarrow \quad & (y_j + \epsilon/2)(y_j - x_j + \epsilon/2) \geq \epsilon^2/4. \end{aligned} \tag{39}$$

With $\epsilon = 2t$, inequalities (38) and (39) are identical to $\Psi_{U,j}(z)$ and $\Psi_{L,j}(z)$, respectively, and thus leads to the same relaxation.

On the other hand, the complementary elements y_j and $y_j - x_j$ do not present explicitly in (28c), unlike in the functions (36c)-(36d). Hence we cannot compare Lagrange gradients and Hessians of these two reformulations of complementarities directly, while such comparisons are needed for the test on constraint qualifications and stationarity properties.

3 Convergence Analysis

Suppose that an infinite sequence of stationary points $\{z^k\}$ of NLP (27) or NLP (28) is generated with a sequence $\{\epsilon^k\}$ of positive scalars tending to zero and, additionally for (27), a sequence $\{p^k\}$ of parameters determined by the Bounding Algorithm. This section analyzes stationarity of limit points of $\{z^k\}$, for the nonsmooth problem equivalent MPCC reformulation (4). The discussion follows the convergence analysis in [16], which is related to Problem (25).

Sections 3.1 and 3.2 develop convergence results for the square root function (23) smoothed problems. Section 3.3 extends these results to problems based on the neural network function (24).

3.1 Constraint Qualification of Subproblems

To facilitate and generalize the analysis, we apply the MPCC notation in (3) and denote $G_j(z) = y_j, H_j(z) = y_j - x_j$. For $\epsilon > 0$, define the function

$$\Phi_j^\epsilon(z) = y_j - h_j^\epsilon(x) = G_j(z) - h_j^\epsilon(G_j(z) - H_j(z)).$$

At a feasible point z of NLP (27), we have $\Phi_j^\epsilon(z) + p_j = 0$ $(j = 1, \dots, n_x)$, where the value of p_j is determined by the Bounding Algorithm. On the other hand, the equation $\Phi_j^\epsilon(z) + p_j = 0$ also holds for any feasible point z of NLP (28), where $p_j = 0$ if z locates on the upper bound of the constraint (28c), $p_j = \kappa\epsilon/2$ if z on the lower bound, and $p_j \in (0, \kappa\epsilon/2)$ if z in the interior. Based on this observation, the following presentation is applicable to both of the NLPs.

With the square root function (23), we have that (recall $\kappa = 1$)

$$\begin{aligned}
\Phi_j^\epsilon(z) + p_j &= \frac{1}{2}\left(2y_j - x_j - \sqrt{x_j^2 + \epsilon^2} + 2p_j\right) \\
&= \frac{1}{2}\left(G_j(z) + H_j(z) - \sqrt{(G_j(z) - H_j(z))^2 + \epsilon^2} + 2p_j\right), \\
\nabla_G \Phi_j^\epsilon(z) &= \frac{1}{2} - \frac{G_j(z) - H_j(z)}{2\sqrt{(G_j(z) - H_j(z))^2 + \epsilon^2}}, \\
\nabla_H \Phi_j^\epsilon(z) &= \frac{1}{2} + \frac{G_j(z) - H_j(z)}{2\sqrt{(G_j(z) - H_j(z))^2 + \epsilon^2}}, \\
\nabla_{GG} \Phi_j^\epsilon(z) = \nabla_{HH} \Phi_j^\epsilon(z) &= \frac{-2(G_j(z) + p_j)(H_j(z) + p_j)}{((G_j(z) - H_j(z))^2 + \epsilon^2)^{3/2}}, \\
\nabla_{GH} \Phi_j^\epsilon(z) = \nabla_{HG} \Phi_j^\epsilon(z) &= \frac{2(G_j(z) + p_j)(H_j(z) + p_j)}{((G_j(z) - H_j(z))^2 + \epsilon^2)^{3/2}}.
\end{aligned} \tag{40}$$

At a point z such that $\Phi_j^\epsilon(z) + p_j = 0$, it follows that

$$2(y_j + p_j) - x_j = (x_j^2 + \epsilon^2)^{1/2} > 0 \tag{41a}$$

$$\implies \quad ((y_j + p_j) + (y_j - x_j + p_j))^2 = ((y_j + p_j) - (y_j - x_j + p_j))^2 + \epsilon^2$$

$$\implies \quad (y_j + p_j)(y_j - x_j + p_j) = \epsilon^2/4. \tag{41b}$$

This leads to the equivalence:

$$\Phi_j^0(z) = 0 \quad \Leftrightarrow \quad \begin{cases} G_j(z) = y_j = 0 \text{ or } H_j(z) = y_j - x_j = 0, \\ G_j(z)H_j(z) = y_j(y_j - x_j) = 0. \end{cases} \tag{42a}$$

$$\Phi_j^\epsilon(z) + p_j = 0 \quad \Leftrightarrow \quad \begin{cases} G_j(z) + p_j = y_j + p_j > 0, \\ H_j(z) + p_j = y_j - x_j + p_j > 0, \\ (G_j(z) + p_j)(H_j(z) + p_j) = (y_j + p_j)(y_j - x_j + p_j) = \epsilon^2/4. \end{cases} \tag{42b}$$

Here, (42a) is the limit of (41b) at $\epsilon = 0$ (and thus $p = 0$), which recovers the complementarities in MPCC (4). On the other hand, (42b) characterizes the status when $\epsilon > 0$, which follows by noting from (41b) that $y_j + p_j$ and $y_j - x_j + p_j$ must have the same sign, and that if they are both negative then (41a) is violated. It follows from (42b) that

$$\begin{aligned} \sqrt{(G_j(z) - H_j(z))^2 + \epsilon^2} &= \sqrt{((G_j(z) + p_j) - (H_j(z) + p_j))^2 + \epsilon^2} \\ &= \sqrt{(G_j(z) + p_j)^2 + (H_j(z) + p_j)^2 + 2(G_j(z) + p_j)(H_j(z) + p_j)} \\ &= |G_j(z) + H_j(z) + 2p_j| = G_j(z) + H_j(z) + 2p_j. \end{aligned}$$

This simplifies the derivatives in (40) at a point $\Phi_j^\epsilon(z) + p_j = 0$ as follows:

$$\begin{aligned} \nabla_G \Phi_j^\epsilon(z) &= \frac{H_j(z) + p_j}{G_j(z) + H_j(z) + 2p_j}, \\ \nabla_H \Phi_j^\epsilon(z) &= \frac{G_j(z) + p_j}{G_j(z) + H_j(z) + 2p_j}, \\ \nabla_{GG} \Phi_j^\epsilon(z) = \nabla_{HH} \Phi_j^\epsilon(z) &= \frac{-2(G_j(z) + p_j)(H_j(z) + p_j)}{(G_j(z) + H_j(z) + 2p_j)^3}, \\ \nabla_{GH} \Phi_j^\epsilon(z) = \nabla_{HG} \Phi_j^\epsilon(z) &= \frac{2(G_j(z) + p_j)(H_j(z) + p_j)}{(G_j(z) + H_j(z) + 2p_j)^3}. \end{aligned} \tag{43}$$

Thus the gradient of $\Phi_j^\epsilon(z)$ is given by

$$\begin{aligned} \nabla \Phi_j^\epsilon(z) &= \frac{H_j(z) + p_j}{G_j(z) + H_j(z) + 2p_j} \nabla G_j(z) + \frac{G_j(z) + p_j}{G_j(z) + H_j(z) + 2p_j} \nabla H_j(z) \\ &= \frac{y_j - x_j + p_j}{2y_j - x_j + 2p_j} \begin{bmatrix} 0 \\ e_j \\ 0 \end{bmatrix} + \frac{y_j + p_j}{2y_j - x_j + 2p_j} \begin{bmatrix} -e_j \\ e_j \\ 0 \end{bmatrix}. \end{aligned} \tag{44}$$

Now consider $\epsilon = 0$ (and thus $p = 0$). At any feasible point $\hat{z}$ of the MPCC, with $\hat{x}_j = \hat{y}_j = 0$ ($j \in I_1(\hat{z}) \cap I_2(\hat{z})$), the function Φ_j^0 is not differentiable. In this case, the Clarke generalized gradient is defined as

$$\partial \Phi_j^0(\hat{z}) = \left\{ r \middle| r = \lim_{k \to \infty} \nabla \Phi_j^0(z^k), \text{ with } z^k \to \hat{z} \text{ and } \nabla \Phi_j^0(z^k) \text{ exist} \right\}. \tag{45}$$

It is worth noting, following from (44), that any accumulation point $\hat{r}$ of $\{\nabla \Phi_j^\epsilon(z)\}$ for $j \in I_1(\hat{z}) \cap I_2(\hat{z})$ is contained in the set

$$\mathcal{G}_j(\hat{z}) = \left\{ r \middle| r = \hat{\xi}_j \begin{bmatrix} 0 \\ e_j \\ 0 \end{bmatrix} + \hat{\eta}_j \begin{bmatrix} -e_j \\ e_j \\ 0 \end{bmatrix}, \ (\hat{\xi}_j, \hat{\eta}_j) \in \mathcal{B} \right\}, \tag{46}$$

where $\mathcal{B} = \left\{(\hat{\xi}_j, \hat{\eta}_j) | (1-\hat{\xi}_j)^2 + (1-\hat{\eta}_j)^2 \le 1\right\}$, and hence is represented by

$$\hat{r} = \hat{\xi}_j \nabla G_j(\hat{z}) + \hat{\eta}_j \nabla H_j(\hat{z}) = \hat{\xi}_j \begin{bmatrix} 0 \\ e_j \\ 0 \end{bmatrix} + \hat{\eta}_j \begin{bmatrix} -e_j \\ e_j \\ 0 \end{bmatrix}, \tag{47}$$

for some $(\hat{\xi}_j, \hat{\eta}_j)$ satisfying $(1-\hat{\xi}_j)^2 + (1-\hat{\eta}_j)^2 \le 1$. This leads to the following results.

Theorem 3.1. *Suppose MPCC-LICQ holds at a feasible point $\hat{z}$ of MPCC (4). Then in a neighborhood $\mathcal{U}(\hat{z})$ of $\hat{z}$, LICQ holds at every feasible point $z \in \mathcal{U}(\hat{z})$ of NLP (27), for any $\epsilon > 0$ sufficiently small and $p_j \in [0, \epsilon/2], j = 1, \dots, n_x$.*

Proof. We note that the proof is not restricted to the feasible points of NLP (27) whose parameters p_j are set by the Bounding Algorithm. Instead, $p_j \in [0, \epsilon/2]$ is the only requirement on the parameters.

It follows from the Lipschitz continuity of ∇c in (4), and the gradient of Φ_j^ϵ characterized by (44) and (46), that

$$\begin{aligned}
\lim_{\epsilon \to 0} \nabla c_i(z) &= \nabla c_i(\hat{z}), && i = 1, \dots, n_c, \\
\lim_{\epsilon \to 0} \nabla \Phi_j^\epsilon(z) &= \nabla G_j(\hat{z}) = \begin{bmatrix} 0 \\ e_j \\ 0 \end{bmatrix}, && j \notin I_2(\hat{z}), \\
\lim_{\epsilon \to 0} \nabla \Phi_j^\epsilon(z) &= \nabla H_j(\hat{z}) = \begin{bmatrix} -e_j \\ e_j \\ 0 \end{bmatrix}, && j \notin I_1(\hat{z}), \\
\lim_{\epsilon \to 0} \operatorname{dist}(\nabla \Phi_j^\epsilon(z), \mathcal{G}_j(\hat{z})) &= 0, && j \in I_1(\hat{z}) \cap I_2(\hat{z}),
\end{aligned} \tag{48}$$

where $\operatorname{dist}(\nabla \Phi_j^\epsilon(z), \mathcal{G}_j(\hat{z}))$ is the minimal distance between $\nabla \Phi_j^\epsilon(z)$ and the set $\mathcal{G}_j(\hat{z})$. For NLP (27), consider the equation

$$\begin{aligned}
0 &= \sum_{i=1}^{n_c} \lambda_i \nabla c_i(z) + \sum_{j=1}^{n_x} u_j \nabla \Phi_j^\epsilon(z) \\
&= \sum_{i=1}^{n_c} \lambda_i \nabla c_i(z) + \sum_{j \notin I_2(\hat{z})} u_j \nabla \Phi_j^\epsilon(z) + \sum_{j \notin I_1(\hat{z})} u_j \nabla \Phi_j^\epsilon(z) + \sum_{j \in I_1(\hat{z}) \cap I_2(\hat{z})} u_j \nabla \Phi_j^\epsilon(z).
\end{aligned} \tag{49}$$

In view of the limits in (48), and the MPCC-LICQ assumption at $\hat{z}$, we can conclude that, for $\epsilon > 0$ sufficiently small, $\lambda_i = u_j = 0$ ($i = 1 \dots n_c$, $j = 1 \dots n_x$). □

Theorem 3.2. *Suppose MPCC-LICQ holds at a feasible point $\hat{z}$ of MPCC (4). Then in a neighborhood $\mathcal{U}(\hat{z})$ of $\hat{z}$, LICQ holds at every feasible point $z \in \mathcal{U}(\hat{z})$ of NLP (28), for any $\epsilon > 0$ sufficiently small.*

Proof. For NLP (28), define the following sets for the active inequality constraints:

$$\begin{aligned}
I_L(z, \epsilon) &= \{j \mid \Phi_j^\epsilon(z) = -\epsilon/2\}, \\
I_U(z, \epsilon) &= \{j \mid \Phi_j^\epsilon(z) = 0\}.
\end{aligned}$$

Consider the equation

$$
\begin{aligned}
0 &= \textstyle\sum_{i=1}^{n_c} \lambda_i \nabla c_i(z) &&- \sum_{j\in I_L(z,\epsilon)} u_{L,j} \nabla \Phi_j^\epsilon(z) + \sum_{j\in I_U(z,\epsilon)} u_{U,j} \nabla \Phi_j^\epsilon(z) \\
&= \textstyle\sum_{i=1}^{n_c} \lambda_i \nabla c_i(z) &&- \sum_{j\in I_L(z,\epsilon)} u_{L,j} \left[\tfrac{H_j(z)+\epsilon/2}{G_j(z)+H_j(z)+\epsilon} \nabla G_j(z) + \tfrac{G_j(z)+\epsilon/2}{G_j(z)+H_j(z)+\epsilon} \nabla H_j(z) \right] \\
& &&+ \sum_{j\in I_U(z,\epsilon)} u_{U,j} \left[\tfrac{H_j(z)}{G_j(z)+H_j(z)} \nabla G_j(z) + \tfrac{G_j(z)}{G_j(z)+H_j(z)} \nabla H_j(z) \right],
\end{aligned} \tag{50}
$$

where we have used the derivatives in (43), and recall that $p_j = -\epsilon/2, \forall j \in I_L(z,\epsilon)$, and $p_j = 0, \forall j \in I_U(z,\epsilon)$. In view of the MPCC-LICQ assumption at $\hat{z}$, and the relation $I_L(z,\epsilon) \cup I_U(z,\epsilon) \subseteq I_1(\hat{z}) \cup I_2(\hat{z})$ holding for $\epsilon > 0$ sufficiently small, we obtain from (50) that

$$
\begin{aligned}
\lambda_i &= 0, && i = 1,\ldots,n_c \\
u_{L,j} \left[\tfrac{H_j(z)+\epsilon/2}{G_j(z)+H_j(z)+\epsilon} \right] = 0,\ u_{L,j} \left[\tfrac{G_j(z)+\epsilon/2}{G_j(z)+H_j(z)+\epsilon} \right] &= 0, && j \in I_L(z,\epsilon) \\
u_{U,j} \left[\tfrac{H_j(z)}{G_j(z)+H_j(z)} \right] = 0,\ u_{U,j} \left[\tfrac{G_j(z)}{G_j(z)+H_j(z)} \right] &= 0, && j \in I_U(z,\epsilon).
\end{aligned}
$$

Note from (42b) that

$$
\begin{aligned}
\tfrac{H_j(z)+\epsilon/2}{G_j(z)+H_j(z)+\epsilon} > 0,\ \tfrac{G_j(z)+\epsilon/2}{G_j(z)+H_j(z)+\epsilon} &> 0, \quad \forall j \in I_L(z,\epsilon), \\
\tfrac{H_j(z)}{G_j(z)+H_j(z)} > 0,\ \tfrac{G_j(z)}{G_j(z)+H_j(z)} &> 0, \quad \forall j \in I_U(z,\epsilon).
\end{aligned}
$$

Therefore, the solution of (50) is $\lambda_i = u_{L,j} = u_{U,j} = 0$ ($i = 1 \ldots n_c$, $j = 1 \ldots n_x$). □

3.2 Stationarity Properties

Consider a sequence of the stationary points of NLP (27) or NLP (28), generated with $\epsilon \to 0$ and, additionally for (27), a vector p adapted by the Bounding Algorithm. As a consequence of the properties (42), limit points of the sequence are feasible to MPCC (4). Now we investigate stationarity properties of the limit points for the MPCC.

3.2.1 C-Stationarity

Theorem 3.3. *For a sequence of positive scalars $\epsilon^k \to 0$, apply the Bounding Algorithm to NLP (27), such that the parameter p^k is updated whenever ϵ^k is updated. Let the generated sequence of stationary points $z^k \to z^*$. Suppose that MPCC-LICQ holds at z^*. Then z^* is a C-stationary point of MPCC (4).*

Proof. Rewrite the conditions (26a)-(26c) as follows:

$$
0 = \nabla f(z^k) + \nabla c(z^k)\lambda^k + \sum_{j\notin I_2(z^*)} u_j^k \nabla \Phi_j^\epsilon(z^k) + \sum_{j\notin I_1(z^*)} u_j^k \nabla \Phi_j^\epsilon(z^k) + \sum_{j\in I_1(z^*)\cap I_2(z^*)} u_j^k \nabla \Phi_j^\epsilon(z^k),
$$

where the multipliers λ^k, u^k are unique for every $\epsilon^k > 0$ sufficiently small. According to (48), we have at z^* that

$$\begin{aligned} 0 = \nabla f(z^*) + \nabla c(z^*)\lambda^* + \sum_{j \notin I_2(z^*)} u_j^* \nabla G_j(z^*) + \sum_{j \notin I_1(z^*)} u_j^* \nabla H_j(z^*) \\ + \sum_{j \in I_1(z^*) \cap I_2(z^*)} \left(u_j^* \xi_j^* \nabla G_j(z^*) + u_j^* \eta_j^* \nabla H_j(z^*) \right), \end{aligned} \tag{51}$$

where

$$\lambda^* = \lim_{\epsilon^k \to 0} \lambda^k, \; u_j^* = \lim_{\epsilon^k \to 0} u_j^k,$$

and

$$r^* = \xi_j^* \nabla G_j(z^*) + \eta_j^* \nabla H_j(z^*)$$

for some (ξ_j^*, η_j^*) satisfying $(1 - \xi_j^*)^2 + (1 - \eta_j^*)^2 \le 1$, such that r^* is an accumulation point of $\{\nabla \Phi_j^\epsilon(z^k)\}$ for every $j \in I_1(z^*) \cap I_2(z^*)$. Given the MPCC-LICQ assumption at z^*, the multipliers associated with z^* are unique, which are the limit points of the unique NLP multipliers λ^k, u^k. With the following settings:

$$\begin{aligned} \sigma_{1j}^* &= -u_j^*, & j &\notin I_2(z^*), \\ \sigma_{2j}^* &= -u_j^*, & j &\notin I_1(z^*), \\ \sigma_{1j}^* &= -u_j^* \xi_j^*, & j &\in I_1(z^*) \cap I_2(z^*), \\ \sigma_{2j}^* &= -u_j^* \eta_j^*, & j &\in I_1(z^*) \cap I_2(z^*), \end{aligned} \tag{52}$$

the point z^* satisfies the weak stationarity conditions (13). For the biactive set, since $\xi_j^*, \eta_j^* \ge 0$, then $\sigma_{1j}^* \sigma_{2j}^* = (u_j^*)^2 \xi_j^* \eta_j^* \ge 0$, which proves C-stationarity of z^*. □

Examining C-stationarity based on NLP (28) is not so straightforward. For the inequality relaxation (28c) of the complementarity conditions, not only the active lower and upper bounds, but also the interior needs to considered, so as to derive a complete set of the MPCC multipliers.

Theorem 3.4. *For a sequence of positive scalars $\epsilon^k \to 0$, let the sequence of NLP (28)'s stationary points $z^k \to z^*$. Suppose that MPCC-LICQ holds at z^*. Then z^* is a C-stationary point of MPCC (4).*

Proof. It follows from the conditions (29) that

$$\begin{aligned} 0 &= \nabla f(z^k) + \nabla c(z^k)\lambda^k - \sum_{j \in I_L(z^k, \epsilon^k)} u_{L,j}^k \nabla \Phi_j^\epsilon(z^k) + \sum_{j \in I_U(z^k, \epsilon^k)} u_{U,j}^k \nabla \Phi_j^\epsilon(z^k) \\ &= \nabla f(z^k) + \nabla c(z^k)\lambda^k \\ &\quad - \sum_{j \in I_L(z^k, \epsilon^k)} u_{L,j}^k \left[\tfrac{H_j(z^k) + \epsilon^k/2}{G_j(z^k) + H_j(z^k) + \epsilon^k} \nabla G_j(z^k) + \tfrac{G_j(z^k) + \epsilon^k/2}{G_j(z^k) + H_j(z^k) + \epsilon^k} \nabla H_j(z^k) \right] \\ &\quad + \sum_{j \in I_U(z^k, \epsilon^k)} u_{U,j}^k \left[\tfrac{H_j(z^k)}{G_j(z^k) + H_j(z^k)} \nabla G_j(z^k) + \tfrac{G_j(z^k)}{G_j(z^k) + H_j(z^k)} \nabla H_j(z^k) \right], \end{aligned}$$

where the multipliers λ^k, u_L^k, u_U^k are unique for every $\epsilon^k > 0$ sufficiently small. By setting

$$
\begin{aligned}
\sigma_{1j}^k &= u_{L,j}^k \left[\tfrac{H_j(z^k)+\epsilon^k/2}{G_j(z^k)+H_j(z^k)+\epsilon^k}\right],\ j \in I_L(z^k,\epsilon^k) \text{ and } j \in I_1(z^*),\\
\sigma_{2j}^k &= u_{L,j}^k \left[\tfrac{G_j(z^k)+\epsilon^k/2}{G_j(z^k)+H_j(z^k)+\epsilon^k}\right],\ j \in I_L(z^k,\epsilon^k) \text{ and } j \in I_2(z^*),\\
\sigma_{1j}^k &= -u_{U,j}^k \left[\tfrac{H_j(z^k)}{G_j(z^k)+H_j(z^k)}\right],\ j \in I_U(z^k,\epsilon^k) \text{ and } j \in I_1(z^*),\\
\sigma_{2j}^k &= -u_{U,j}^k \left[\tfrac{G_j(z^k)}{G_j(z^k)+H_j(z^k)}\right],\ j \in I_U(z^k,\epsilon^k) \text{ and } j \in I_2(z^*),
\end{aligned}
\tag{53}
$$

we rewrite the above equations as

$$
\begin{aligned}
0 &= \nabla f(z^k) + \sum\nolimits_{i=1}^{n_c} \lambda_i^k \nabla c_i(z^k)\\
&- \sum_{\substack{j\in I_L(z^k,\epsilon^k)\\ j\notin I_2(z^*)}} \sigma_{1j}^k \left[\nabla G_j(z^k) + \tfrac{G_j(z^k)+\epsilon^k/2}{H_j(z^k)+\epsilon^k/2}\nabla H_j(z^k)\right] - \sum_{\substack{j\in I_L(z^k,\epsilon^k)\\ j\notin I_1(z^*)}} \sigma_{2j}^k \left[\nabla H_j(z^k) + \tfrac{H_j(z^k)+\epsilon^k/2}{G_j(z^k)+\epsilon^k/2}\nabla G_j(z^k)\right]\\
&- \sum_{\substack{j\in I_L(z^k,\epsilon^k)\\ j\in I_1(z^*)\cap I_2(z^*)}} \left[\sigma_{1j}^k \nabla G_j(z^k) + \sigma_{2j}^k \nabla H_j(z^k)\right]\\
&- \sum_{\substack{j\in I_U(z^k,\epsilon^k)\\ j\notin I_2(z^*)}} \sigma_{1j}^k \left[\nabla G_j(z^k) + \tfrac{G_j(z^k)}{H_j(z^k)}\nabla H_j(z^k)\right] - \sum_{\substack{j\in I_U(z^k,\epsilon^k)\\ j\notin I_1(z^*)}} \sigma_{2j}^k \left[\nabla H_j(z^k) + \tfrac{H_j(z^k)}{G_j(z^k)}\nabla G_j(z^k)\right]\\
&- \sum_{\substack{j\in I_U(z^k,\epsilon^k)\\ j\in I_1(z^*)\cap I_2(z^*)}} \left[\sigma_{1j}^k \nabla G_j(z^k) + \sigma_{2j}^k \nabla H_j(z^k)\right].
\end{aligned}
\tag{54}
$$

Denote (54) as $\nabla f(z^k) + A(z^k)\omega^k = 0$, where ω^k contains all the multipliers $\lambda_i^k, \sigma_{1j}^k$, and σ_{2j}^k, while their corresponding vectors are columns of matrix $A(z^k)$. Now consider the following augmented system:

$$
\nabla f(z^k) + \bar{A}(z^k)\bar{\omega}^k = \nabla f(z^k) + \left[A(z^k) \vdots \overbrace{\underbrace{-\nabla G_j(z^k)}_{j\in I_1(z^*)} \vdots \underbrace{-\nabla H_j(z^k)}_{j\in I_2(z^*)}}^{\forall j\notin I_L(z^k,\epsilon^k)\cup I_U(z^k,\epsilon^k)} \right] \begin{bmatrix} \omega^k \\ 0 \\ 0 \end{bmatrix} = 0.
$$

In the limit, $\bar{A}(z^k)$ converges to a matrix $\bar{A}(z^*)$, which has columns

$$
\{\nabla c_i(z^*)\,|\, i = 1,\ldots,n_c\} \cup \{-\nabla G_j(z^*)\,|\, j \in I_1(z^*)\} \cup \{-\nabla H_j(z^*)\,|\, j \in I_2(z^*)\}; \tag{55}
$$

and $\bar{\omega}^k$ converges to a vector $\bar{\omega}^*$, whose elements are as follows (refer to (53)):

$$\lambda_i^* \;=\; \lim_{\epsilon^k \to 0} \lambda_i^k, \quad i = 1, \ldots, n_x, \tag{56a}$$

$$\sigma_{1j}^* \;=\; \begin{cases} u_{L,j}^* = \lim_{\epsilon^k \to 0} u_{L,j}^k \geq 0, & j \in I_L^0 \text{ and } j \in I_1(z^*) \setminus I_2(z^*) \\ u_{L,j}^* \xi_j^* = \xi_j^* \lim_{\epsilon^k \to 0} u_{L,j}^k \geq 0, & j \in I_L^0 \text{ and } j \in I_1(z^*) \cap I_2(z^*) \\ -u_{U,j}^* = -\lim_{\epsilon^k \to 0} u_{U,j}^k \leq 0, & j \in I_U^0 \text{ and } j \in I_1(z^*) \setminus I_2(z^*) \\ -u_{U,j}^* \xi_j^* = -\xi_j^* \lim_{\epsilon^k \to 0} u_{U,j}^k \leq 0, & j \in I_U^0 \text{ and } j \in I_1(z^*) \cap I_2(z^*) \\ 0, & j \notin (I_L^0 \cup I_U^0) \text{ and } j \in I_1(z^*), \end{cases} \tag{56b}$$

$$\sigma_{2j}^* \;=\; \begin{cases} u_{L,j}^* = \lim_{\epsilon^k \to 0} u_{L,j}^k \geq 0, & j \in I_L^0 \text{ and } j \in I_2(z^*) \setminus I_1(z^*) \\ u_{L,j}^* \eta_j^* = \eta_j^* \lim_{\epsilon^k \to 0} u_{L,j}^k \geq 0, & j \in I_L^0 \text{ and } j \in I_1(z^*) \cap I_2(z^*) \\ -u_{U,j}^* = -\lim_{\epsilon^k \to 0} u_{U,j}^k \leq 0, & j \in I_U^0 \text{ and } j \in I_2(z^*) \setminus I_1(z^*) \\ -u_{U,j}^* \eta_j^* = -\eta_j^* \lim_{\epsilon^k \to 0} u_{U,j}^k \leq 0, & j \in I_U^0 \text{ and } j \in I_1(z^*) \cap I_2(z^*) \\ 0, & j \notin (I_L^0 \cup I_U^0) \text{ and } j \in I_2(z^*), \end{cases} \tag{56c}$$

for the sets

$$I_L^0 = \{j \,|\, j \in I_L(z^k, \epsilon^k) \text{ for every } \epsilon^k > 0 \text{ sufficiently small}\},$$
$$I_U^0 = \{j \,|\, j \in I_U(z^k, \epsilon^k) \text{ for every } \epsilon^k > 0 \text{ sufficiently small}\},$$

and some (ξ_j^*, η_j^*) satisfying $(1-\xi_j^*)^2 + (1-\eta_j^*)^2 \leq 1$, such that $r^* = \xi_j^* \nabla G_j(z^*) + \eta_j^* \nabla H_j(z^*)$ is an accumulation point of $\{\nabla \Phi_j^\epsilon(z^k)\}$ for every $j \in I_1(z^*) \cap I_2(z^*)$. Given the MPCC-LICQ assumption at z^*, the matrix $\bar{A}(z^*)$ has full column rank, and the multipliers $\lambda_i^*, \sigma_{1j}^*, \sigma_{2j}^*$ in (56) are unique, which are converged from the unique NLP multipliers λ^k, u_L^k, u_U^k. Then z^* is C-stationary because

$$\sigma_{1j}^* \sigma_{2j}^* = \begin{cases} (u_{L,j}^*)^2 \xi_j^* \eta_j^* \geq 0, & j \in I_L^0 \text{ and } j \in I_1(z^*) \cap I_2(z^*) \\ (u_{U,j}^*)^2 \xi_j^* \eta_j^* \geq 0, & j \in I_U^0 \text{ and } j \in I_1(z^*) \cap I_2(z^*) \\ 0, & j \notin (I_L^0 \cup I_U^0) \text{ and } j \in I_1(z^*) \cap I_2(z^*). \end{cases}$$

□

In the light of the expressions (52) and (56), the limit point z^* is B-stationary, or equivalently, strongly stationary since MPCC-LICQ is assumed, if (i) $u_j^* \leq 0$ for all $j \in I_1(z^*) \cap I_2(z^*)$, (ii) $I_U^0 = \emptyset$, or (iii) $u_{U,j}^* = 0$ for all $j \in I_U^0 \cap I_1(z^*) \cap I_2(z^*)$, which indicates the violation of SC at every stationary point z^k of (28) for small positive ϵ^k. Condition (ii) represents the case that for $\epsilon^k > 0$ small enough, the relaxed constraints $-\epsilon^k/2 \leq \Phi_j^\epsilon(z^k) \leq 0$ do not have active upper bounds, namely, $u_{U,j}^k = 0$, $\forall j \in I_1(z^*) \cap I_2(z^*)$. From the point of view of the parameterized constraints $\Phi_j^\epsilon(z^k) + p_j^k = 0$, in this case $p_j^k = \epsilon^k/2$, and $u_j^k \leq 0$ (since $u_j^k = u_{U,j}^k - u_{L,j}^k$), for all $j \in I_1(z^*) \cap I_2(z^*)$.

3.2.2 M-Stationarity

Now we consider the second-order conditions at stationary points z^k of the NLPs, with the aim of exploring stronger results beyond C-stationarity in the limit.

Define the Lagrangian for NLP (27) as $\mathcal{L}_{\rm BA}(z, \lambda, u) = f(z) + c(z)^T\lambda + (\Phi^\epsilon(z) + p)^T u$. The Hessian of the Lagrangian is given by

$$\nabla_{zz}\mathcal{L}_{\rm BA}(z, \lambda, u) = \nabla_{zz} f(z) + \sum_{i=1}^{n_c} \lambda_i \nabla_{zz} c_i(z) + \sum_{j=1}^{n_x} u_j \nabla_{zz}\Phi_j^\epsilon(z). \tag{57}$$

Theorem 3.5. *For a sequence of positive scalars $\epsilon^k \to 0$, apply the Bounding Algorithm to NLP (27), such that the parameter p^k is updated whenever ϵ^k is updated. Let the generated sequence of stationary points $z^k \to z^*$. In addition to the assumptions of Theorem 3.3, suppose that the reduced Hessian of the Lagrangian at each z^k is bounded below when $\epsilon^k > 0$ suitably small, in the sense that*

$$d^T \nabla_{zz}\mathcal{L}_{\rm BA}(z^k, \lambda^k, u^k) d \geq -\alpha_{\rm BA}^k \|d\|^2, \ \forall d \in \mathcal{D}_{\rm BA}(z^k), \tag{58}$$

for the bounded sequence $\{\alpha_{\rm BA}^k\}$ of positive constants, and

$$\mathcal{D}_{\rm BA}(z^k) = \left\{ d \middle| \begin{array}{ll} \nabla c_i(z^k)^T d = 0, & i = 1, \dots, n_c \\ \nabla \Phi_j^\epsilon(z^k)^T d = 0, & j = 1, \dots, n_x \end{array} \right\}. \tag{59}$$

Then z^* is an M-stationary point of MPCC (4).

Proof. For the purpose of deriving a contradiction, suppose z^* is not M-stationary. According to Theorem 3.3, there exists some index $j_0 \in I_1(z^*) \cap I_2(z^*)$ such that

$$\begin{aligned} \sigma_{1j_0}^* &= -u_{j_0}^* \xi_{j_0}^* < 0, \\ \sigma_{2j_0}^* &= -u_{j_0}^* \eta_{j_0}^* < 0. \end{aligned} \tag{60}$$

This implies $u_{j0}^*, \xi_{j_0}^*, \eta_{j0}^* > 0$, because $(1 - \xi_{j0}^*)^2 + (1 - \eta_{j0}^*)^2 \leq 1$.

We can choose a direction d^k such that

$$\begin{aligned} \nabla c_i(z^k)^T d^k &= 0, \ i = 1, \dots, n_c, \\ \nabla G_j(z^k)^T d^k &= 0, \ j \in I_1(z^*) \text{ and } j \neq j_0, \\ \nabla H_j(z^k)^T d^k &= 0, \ j \in I_2(z^*) \text{ and } j \neq j_0, \\ \nabla G_{j_0}(z^k)^T d^k &= d_G = \nabla_H \Phi_{j_0}^\epsilon(z^k), \\ \nabla H_{j_0}(z^k)^T d^k &= d_H = -\nabla_G \Phi_{j_0}^\epsilon(z^k). \end{aligned} \tag{61}$$

The direction d^k is well-defined for $\epsilon^k > 0$ sufficiently small, because the MPCC-LICQ assumption at z^* guarantees linear independence of the coefficient vectors of d^k, and the right-hand sides of the last two equations are confined by the set $\mathcal{G}_{j0}(z^*)$ in the limit. Note from (43) and (44) that $d^k \in \mathcal{D}_{\rm BA}(z^k)$.

Contribution of the constraint $\Phi_j^\epsilon(z^k) + p_j^k = 0$ to $(d^k)^T \nabla_{zz}\mathcal{L}_{\text{BA}}(z^k, \lambda^k, u^k) d^k$ is that

$$
\begin{aligned}
& u_j^k (d^k)^T \nabla_{zz}\Phi_j^\epsilon(z^k) d^k \\
= \; & u_j^k (d^k)^T \big[\nabla_G \Phi_j^\epsilon(z^k) \nabla_{zz} G_j(z^k) + \nabla_H \Phi_j^\epsilon(z^k) \nabla_{zz} H_j(z^k) \\
+ \; & \nabla_{GG}\Phi_j^\epsilon(z^k) \nabla G_j(z^k) \nabla G_j(z^k)^T + \nabla_{GH}\Phi_j^\epsilon(z^k) \nabla G_j(z^k) \nabla H_j(z^k)^T \\
+ \; & \nabla_{HG}\Phi_j^\epsilon(z^k) \nabla H_j(z^k) \nabla G_j(z^k)^T + \nabla_{HH}\Phi_j^\epsilon(z^k) \nabla H_j(z^k) \nabla H_j(z^k)^T \big] d^k.
\end{aligned} \tag{62}
$$

This term is zero for all $j \neq j_0$, which follows from the definition of d^k and $\nabla_{zz} G_j(z^k) = \nabla_{zz} H_j(z^k) = 0$ for MPCC (4). For $j = j_0$, we derive the following from (43):

$$
\begin{aligned}
& u_{j_0}^k (d^k)^T \nabla_{zz}\Phi_{j_0}^\epsilon(z^k) d^k \\
= \; & u_{j_0}^k \left[\nabla_{GG}\Phi_{j_0}^\epsilon(z^k) d_G^2 + 2\nabla_{GH}\Phi_{j_0}^\epsilon(z^k) d_G d_H + \nabla_{HH}\Phi_{j_0}^\epsilon(z^k) d_H^2\right] \\
= \; & \frac{-2u_{j_0}^k (G_{j_0}(z^k) + p_{j_0}^k)(H_{j_0}(z^k) + p_{j_0}^k)}{(G_{j_0}(z^k) + H_{j_0}(z^k) + 2p_{j_0}^k)^3} (d_G - d_H)^2 \\
= \; & \frac{-2u_{j_0}^k (G_{j_0}(z^k) + p_{j_0}^k)(H_{j_0}(z^k) + p_{j_0}^k)}{(G_{j_0}(z^k) + H_{j_0}(z^k) + 2p_{j_0}^k)^3} \\
= \; & \frac{-2u_{j_0}^k}{G_{j_0}(z^k) + H_{j_0}(z^k) + 2p_{j_0}^k} \nabla_G \Phi_{j_0}^\epsilon(z^k) \nabla_H \Phi_{j_0}^\epsilon(z^k).
\end{aligned} \tag{63}
$$

As $\epsilon^k \to 0$, $u_{j_0}^k$, $\nabla_G \Phi_{j_0}^\epsilon(z^k)$, $\nabla_H \Phi_{j_0}^\epsilon(z^k)$ converge to $u_{j_0}^*, \xi_{j_0}^*, \eta_{j_0}^*$, which are positive and bounded, while $G_{j_0}(z^k)$, $H_{j_0}(z^k)$, $p_{j_0}^k$ tend to zero. As a consequence,

$$
u_{j_0}^k (d^k)^T \nabla_{zz}\Phi_{j_0}^\epsilon(z^k) d^k \to -\infty. \tag{64}
$$

Since all other terms in $(d^k)^T \nabla_{zz}\mathcal{L}_{\text{BA}}(z^k, \lambda^k, u^k) d^k$ are bounded, then (64) yields the contradiction to (58). Hence the assumption must be false and z^* is M-stationary. □

To investigate M-stationarity based on NLP (28), define the Lagrangian $\mathcal{L}_{\text{MLF}}(z, \lambda, u_L, u_U) = f(z) + c(z)^T\lambda - (\Phi^\epsilon(z) + \epsilon/2)^T u_L + \Phi_j^\epsilon(z)^T u_U$. The Hessian of the Lagrangian is given by

$$
\nabla_{zz}\mathcal{L}_{\text{MLF}}(z, \lambda, u_L, u_U) = \nabla_{zz} f(z) + \sum_{i=1}^{n_c} \lambda_i \nabla_{zz} c_i(z) - \sum_{j=1}^{n_x} u_{L,j} \nabla_{zz}\Phi_j^\epsilon(z) + \sum_{j=1}^{n_x} u_{U,j} \nabla_{zz}\Phi_j^\epsilon(z).
$$

Theorem 3.6. *For a sequence of positive scalars $\epsilon^k \to 0$, let the sequence of NLP (28)'s stationary points $z^k \to z^*$. In addition to the assumptions of Theorem 3.4, suppose that the reduced Hessian of the Lagrangian at each z^k is bounded below when ϵ^k suitably small, in the sense that*

$$
w^T \nabla_{zz}\mathcal{L}_{\text{MLF}}(z^k, \lambda^k, u_L^k, u_U^k) w \geq -\alpha_{\text{MLF}}^k \|w\|^2, \; \forall w \in \mathcal{D}_{\text{MLF}}(z^k), \tag{65}
$$

for the bounded sequence $\{\alpha_{\text{MLF}}^k\}$ of positive constants, and

$$
\mathcal{D}_{\text{MLF}}(z^k) = \left\{ w \middle| \begin{array}{ll} \nabla c_i(z^k)^T w = 0, & i = 1, \dots, n_c \\ \nabla \Phi_j^\epsilon(z^k)^T w = 0, & \forall j \in I_L(z^k, \epsilon^k) \cup I_U(z^k, \epsilon^k) \end{array} \right\}. \tag{66}
$$

Then z^* is an M-stationary point of MPCC (4).

Proof. The proof is similar to that of Theorem 3.5. Suppose for contradiction that z^* is not M-stationary. It follows from Theorem 3.4 that there exists some index $j_0 \in I_1(z^*) \cap I_2(z^*)$ such that $\sigma^*_{1j_0} < 0$ and $\sigma^*_{2j_0} < 0$. In view of (56), we have $j_0 \in I^0_U \cap I_1(z^*) \cap I_2(z^*)$ and $u^*_{U,j0}, \xi^*_{j_0}, \eta^*_{j0} > 0$.

We can choose a direction w^k such that

$$\begin{aligned} \nabla c_i(z^k)^T w^k &= 0,\ i = 1, \ldots, n_c, \\ \nabla G_j(z^k)^T w^k &= 0,\ j \in I_1(z^*) \text{ and } j \neq j_0, \\ \nabla H_j(z^k)^T w^k &= 0,\ j \in I_2(z^*) \text{ and } j \neq j_0, \\ \nabla G_{j_0}(z^k)^T w^k &= w_G = \nabla_H \Phi^\epsilon_{j_0}(z^k), \\ \nabla H_{j_0}(z^k)^T w^k &= w_H = -\nabla_G \Phi^\epsilon_{j_0}(z^k). \end{aligned} \tag{67}$$

The direction w^k is well-defined because of the MPCC-LICQ assumption and the bounded right-hand sides. Also note that $w^k \in \mathcal{D}_{\mathrm{MLF}}(z^k)$.

As before, $(w^k)^T \nabla_{zz} \Phi^\epsilon_j(z^k) w^k = 0$ for all $j \neq j_0$. While for $j = j_0$, we have that

$$\begin{aligned} & u^k_{U,j_0} (w^k)^T \nabla_{zz} \Phi^\epsilon_{j_0}(z^k) w^k \\ = \;& u^k_{U,j_0} \left[\nabla_{GG} \Phi^\epsilon_{j_0}(z^k) w_G^2 + 2 \nabla_{GH} \Phi^\epsilon_{j_0}(z^k) w_G w_H + \nabla_{HH} \Phi^\epsilon_{j_0}(z^k) w_H^2 \right] \\ = \;& \frac{-2 u^k_{U,j_0} G_{j_0}(z^k) H_{j_0}(z^k)}{(G_{j_0}(z^k) + H_{j_0}(z^k))^3} (w_G - w_H)^2 \\ = \;& \frac{-2 u^k_{U,j_0} G_{j_0}(z^k) H_{j_0}(z^k)}{(G_{j_0}(z^k) + H_{j_0}(z^k))^3} \\ = \;& \frac{-2 u^k_{U,j_0}}{G_{j_0}(z^k) + H_{j_0}(z^k)} \nabla_G \Phi^\epsilon_{j_0}(z^k) \nabla_H \Phi^\epsilon_{j_0}(z^k) \quad \to -\infty, \end{aligned} \tag{68}$$

because $u^k_{U,j_0}, \nabla_G \Phi^\epsilon_{j_0}(z^k), \nabla_H \Phi^\epsilon_{j_0}(z^k)$ converge to $u^*_{U,j_0}, \xi^*_{j_0}, \eta^*_{j_0}$, which are all positive and bounded, while $G_{j_0}(z^k)$ and $H_{j_0}(z^k)$ tend to zero. This brings the desired contradiction, since all other terms in $(w^k)^T \nabla_{zz} \mathcal{L}_{\mathrm{MLF}}(z^k, \lambda^k, u^k_L, u^k_U) w^k$ are bounded. Therefore we conclude that z^* is M-stationary. □

3.3 Extension to Neural Network Functions

This section extends the convergence results for the reformulation using the square root function to that using the neural network function. For simplicity, we only present outlines of the extension pertinent to NLP (27); the results for NLP (28) can be extended similarly.

With the neural network function (24), the function Φ^ϵ_j enforces the following properties (Item 2, Proposition 1.7; see also [8, Proposition 3.3]):

$$\max(0, -G_j(z)) \leq \kappa\epsilon/2,\ \max(0, -H_j(z)) \leq \kappa\epsilon/2,\ \max(0, G_j(z) H_j(z)) \leq \epsilon^2/2. \tag{69}$$

Obviously, the complementarity conditions of MPCC (4) are satisfied more accurately as $\epsilon \to 0$. The smoothed constraint function and its gradients are given below (recall $p_j = \kappa\epsilon/2$

and $\kappa = \log 2$):

$$\begin{aligned}\Phi_j^\epsilon(z) + p_j &= y_j - x_j - \frac{\epsilon}{2}\log(1 + e^{-2x_j/\epsilon}) + p_j \\ &= H_j(z) - \frac{\epsilon}{2}\log(1 + e^{-2(G_j(z)-H_j(z))/\epsilon}) + p_j, \\ \nabla_G \Phi_j^\epsilon(z) &= \frac{e^{-2(G_j(z)-H_j(z))/\epsilon}}{1 + e^{-2(G_j(z)-H_j(z))/\epsilon}}, \\ \nabla_H \Phi_j^\epsilon(z) &= \frac{1}{1 + e^{-2(G_j(z)-H_j(z))/\epsilon}}, \\ \nabla_{GG}\Phi_j^\epsilon(z) = \nabla_{HH}\Phi_j^\epsilon(z) &= \frac{-2e^{-2(G_j(z)-H_j(z))/\epsilon}}{\epsilon(1 + e^{-2(G_j(z)-H_j(z))/\epsilon})^2}, \\ \nabla_{GH}\Phi_j^\epsilon(z) = \nabla_{HG}\Phi_j^\epsilon(z) &= \frac{2e^{-2(G_j(z)-H_j(z))/\epsilon}}{\epsilon(1 + e^{-2(G_j(z)-H_j(z))/\epsilon})^2}.\end{aligned} \tag{70}$$

At a point z such that $\Phi_j^\epsilon(z) + p_j = 0$, we have

$$e^{2(H_j(z)+p_j)/\epsilon} = 1 + e^{-2(G_j(z)-H_j(z))/\epsilon}. \tag{71}$$

Substituting (71) into (70), we can simplify the derivatives to

$$\begin{aligned}\nabla_G \Phi_j^\epsilon(z) &= e^{-2(G_j(z)+p_j)/\epsilon}, \\ \nabla_H \Phi_j^\epsilon(z) &= e^{-2(H_j(z)+p_j)/\epsilon}, \\ \nabla_{GG}\Phi_j^\epsilon(z) = \nabla_{HH}\Phi_j^\epsilon(z) &= \frac{-2}{\epsilon} e^{-2(G_j(z)+H_j(z)+2p_j)/\epsilon}, \\ \nabla_{GH}\Phi_j^\epsilon(z) = \nabla_{HG}\Phi_j^\epsilon(z) &= \frac{2}{\epsilon} e^{-2(G_j(z)+H_j(z)+2p_j)/\epsilon}.\end{aligned} \tag{72}$$

Thus the gradient of Φ_j^ϵ is given by

$$\begin{aligned}\nabla \Phi_j^\epsilon(z) &= \nabla_G \Phi_j^\epsilon \nabla G_j(z) + \nabla_H \Phi_j^\epsilon \nabla H_j(z) \\ &= e^{-2(G_j(z)+p_j)/\epsilon} \begin{bmatrix} 0 \\ e_j \\ 0 \end{bmatrix} + e^{-2(H_j(z)+p_j)/\epsilon} \begin{bmatrix} -e_j \\ e_j \\ 0 \end{bmatrix}.\end{aligned} \tag{73}$$

Denote $\hat{z}$ as a feasible point of MPCC (4), and $\mathcal{U}(\hat{z})$ as a neighborhood of $\hat{z}$. With the MPCC-LICQ assumption at $\hat{z}$, we firstly verify that LICQ holds at every feasible point $z \in \mathcal{U}(\hat{z})$ of NLP (27), for any $\epsilon > 0$ sufficiently small and $p_j \in [0, \frac{\epsilon \log 2}{2}], j = 1, \ldots, n_x$. Multiplying both sides of (71) with $e^{-2(H_j(z)+p_j)/\epsilon}$ leads to

$$e^{-2(G_j(z)+p_j)/\epsilon} + e^{-2(H_j(z)+p_j)/\epsilon} = 1. \tag{74}$$

Combining this with (72), we deduce that for any $j \notin I_1(\hat{z})$, $\lim_{\epsilon\to 0} \nabla_G \Phi_j^\epsilon(z) = 0$ and $\lim_{\epsilon\to 0} \nabla_H \Phi_j^\epsilon(z) = 1$; for any $j \notin I_2(\hat{z})$, $\lim_{\epsilon\to 0} \nabla_G \Phi_j^\epsilon(z) = 1$ and $\lim_{\epsilon\to 0} \nabla_H \Phi_j^\epsilon(z) = 0$. Hence the limits in (48) still hold and the rest of the proof for Theorem 3.1 directly applies.

For a sequence of positive scalars $\epsilon^k \to 0$, apply the Bounding Algorithm to NLP (27) formulated in the neural network function. Let the generated sequence of stationary points

$z^k \to z^*$, and suppose that MPCC-LICQ holds at z^*. Then C-stationarity of z^* for MPCC (4) immediately follows from the proof of Theorem 3.3.

With the additional second-order conditions (58), we can prove M-stationarity of z^* for MPCC (4) by contradiction. Here we again assume (60) holds for some index $j_0 \in I_1(z^*) \cap I_2(z^*)$, and choose the direction $d^k \in \mathcal{D}_{\text{BA}}(z^k)$ given by (61). Then we have

$$\begin{aligned} & u_{j_0}^k (d^k)^T \nabla_{zz} \Phi_{j_0}^\epsilon (z^k) d^k \\ = \; & u_{j_0}^k \left[\nabla_{GG} \Phi_{j_0}^\epsilon (z^k) d_G^2 + 2 \nabla_{GH} \Phi_{j_0}^\epsilon (z^k) d_G d_H + \nabla_{HH} \Phi_{j_0}^\epsilon (z^k) d_H^2 \right] \\ = \; & \frac{-2 u_{j_0}^k}{\epsilon^k} e^{-2(G_{j_0}(z^k) + H_{j_0}(z^k) + 2p_{j_0})/\epsilon^k} \left(e^{-2(G_{j_0}(z^k)+p_{j_0})/\epsilon^k} + e^{-2(H_{j_0}(z^k)+p_{j_0})/\epsilon^k} \right)^2 \\ = \; & \frac{-2 u_{j_0}^k}{\epsilon^k} \nabla_G \Phi_{j_0}^\epsilon (z^k) \nabla_H \Phi_{j_0}^\epsilon (z^k) \quad \to -\infty \end{aligned} \tag{75}$$

as ϵ tending to zero, where we have used (72) and (74) to derive the last equality, and the limit is obtained by noting that $u_{j_0}^k, \nabla_G \Phi_{j_0}^\epsilon(z^k), \nabla_H \Phi_{j_0}^\epsilon(z^k)$ converge to positive and finite values $u_{j_0}^*, \xi_{j_0}^*, \eta_{j_0}^*$. Since all other terms in $(d^k)^T \nabla_{zz} \mathcal{L}_{\text{BA}}(z^k, \lambda^k, u^k) d^k$ are bounded, then (75) contradicts the boundedness in (58). Thus z^* must be M-stationary.

4 Numerical Results

To demonstrate the performance of the above methods, this section provides numerical results for MPCCs drawn from two sources. The first set is selected from the MacMPEC test set, and the second is a set of distillation case studies. Five MPCC formulations are considered for the numerical study.

- Regularized (REG) formulation (16) of Scholtes.
- The Lin-Fukushima (LF) formulation (36) using the smooth square root function.
- NCP formulation (25) using the smooth square root function.
- The Bounding Algorithm (BA), based on (27) using the smooth square root function.
- The modified Lin-Fukushima (MLF) formulation (28) using the smooth square root function.

CONOPT is chosen to solve the reformulated NLPs, since it is a Newton-based active set method, which converges the sequence of problems with $\epsilon^k \to 0$ quickly by taking advantage of the results of preceding solutions. The following characteristics of CONOPT are beneficial to this numerical study. (i) It checks for directions of negative curvature, in order to confirm whether second order necessary conditions are satisfied. (ii) By partitioning the problem variables into basic, nonbasic, and superbasic variables, it frequently applies Newton's method to the basic variables, and checks for degeneracy of the basis Jacobian. If no degeneracies are flagged at the solution, then the variable partition corresponds to satisfaction of the LICQ for the NLPs.

In our initial numerical experiments we also applied the neural network smoothing function, but it did not perform as well, or as stably, as the square root function. As we can see from the expression (70) of the function $\Phi_j^\epsilon(z) + p_j$, when $G_j(z)$ is (nearly) zero, $H_j(z) > 0$, and ϵ is small, the exponential term becomes very large and can lead to numerical errors. In GAMS we frequently encountered overflows as $\epsilon \le 10^{-2}$, where the argument in the exponential term exceeds 320 and the corresponding functions are undefined. A typical case vulnerable to such errors is the complementarity problems converted from bilevel programs with inequality constraints in the lower-level optimization, where the complementary elements are the lower-level constraints and their multipliers. An easy way to deal with such errors is to switch the role of $G_j(z)$ and $H_j(z)$ in (70). However, this can only be done by a trial and error solution strategy, based on whether $G_j(z)$ or $H_j(z)$ is active at the solution; this is unknown a priori. Because this leads to a solution strategy that may not allow meaningful comparisons, our numerical study focuses only on the square root function.

4.1 MacMPEC Results

Out of 133 problems from the MacMPEC collection [33], 15 problems are selected that have nonempty biactive sets at their solutions. Such problems are of interest because nonempty biactive set complicates the analysis of MPCC stationarity; and biactive elements often pose numerical difficulties for solving the reformulated NLPs, hence they can be employed to test performance of different NLP formulations. The above five MPCC formulations are implemented in GAMS and applied to these problems. The outer loop is controlled by $\epsilon^0 = 0.25, \epsilon_{\text{tol}} = 10^{-6}$, and the reducing factor $\gamma = 0.1$; for BA, $\tilde{\epsilon} = 0.01$ is chosen to start bounding. The resulting sequence of NLPs are solved by GAMS solver CONOPT4, where the optimality tolerance is 10^{-7}. Biactive elements are recognized at the last NLP solution $\bar{z}$, if both $G_j(\bar{z})$ and $H_j(\bar{z})$ are no more than 10^{-5}.

The following demonstrates performance of the MPCC formulations in three parts, namely, for 11 problems from the selected set that converge to strongly stationary points, for 3 problems that do not converge to strongly stationary solutions, and a discussion of problem ralph2 for special discrimination of strong stationarity.

4.1.1 Examples with Strongly Stationary Solutions

Starting from the default initial points, we solve 11 problems of the selected set to strongly stationary points by using BA, MLF, and NCP formulations. General information at the solutions is shown in Table 1. On the other hand, REG solves 10 of them (except bilevel1), while LF solves 9 of them (except bilevel1 and outrata31), to the same solutions. For bilevel1, REG and LF converge to a strongly stationary point which has an empty biactive set. For outrata31, LF converges to a point infeasible to the original MPCC.

For those problems all the formulations converge to the same points, Table 2 shows the iteration counts and complementarity residuals at the final solutions $\bar{z}$, from which we can see REG takes the fewest iterations to converge. Multipliers of the reformulated constraints corresponding to the biactive elements are given in Table 3. As discussed at the end of Section 3.2.1, strong stationarity of the solutions can be identified by BA with nonzero parameters $\bar{p}_j > 0$, or by MLF with inactive upper bounds, i.e., $\bar{u}_{U,j} = 0$ for all $j \in I_1(\bar{z}) \cap I_2(\bar{z})$. In

Problem	$f(\bar{z})$	$\bar{z}$	$I_1(\bar{z}) \cap I_2(\bar{z})$
bard2m	-6598	(6.3, 2.7, 12.4, 18.6, 0, 9, 31, 0, -8, -8, 0, -13.33)	3
bilevel1	5	(25, 30, 5, 10, 0, 0, 0, 0, 0, 0)	6
df1	0	(1, 0)	1
ex9.2.3	5	(25, 30, 5, 10, 0, 0, 0, 0, 0, 5, 0, 15, 15, 20, 10)	2
ex9.2.8	1.5	(0.25, 0, 0, 0, 0, 1)	1
ex9.2.9	2	(2, 6, 0, 2, 0, 0, 0, 6, 0)	3
kth1	0	(0, 0)	1
outrata31	3.21	(2.68, 1.49, 0, 0.66, 4.06)	3
qpec1	80	(-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)	11, 12, 13, 14, 15, 16, 17, 18, 19, 20
scholtes2	15	(0, 2, 0)	1
sl1	0	(2.01, 0, 10, 0.01, 0, 0, 0.04, 0)	3

Table 1: MacMPEC examples converged to strongly stationary points

addition, REG multipliers $\bar{\nu}_{GH,j} = 0$ for all $j \in I_1(\bar{z}) \cap I_2(\bar{z})$ also indicate strong stationarity of the solution [41].

Note that LF often leads to very large multipliers for the biactive elements. To check linear independence of the constraints, consider the following equation (in contrast to (50) for MLF):

$$\begin{aligned} 0 = \sum_{i=1}^{n_c} \lambda_i \nabla c_i(z) - \sum_{j \in I_{\Psi_L}(z,t)} \mu_{L,j} \left[(H_j(z) + t) \nabla G_j(z) + (G_j(z) + t) \nabla H_j(z) \right] \\ + \sum_{j \in I_{\Psi_U}(z,t)} \mu_{U,j} \left[H_j(z) \nabla G_j(z) + G_j(z) \nabla H_j(z) \right]. \end{aligned} \tag{76}$$

For $j \in I_1(z) \cap I_2(z)$, the coefficients of $\nabla G_j(z)$ and $\nabla H_j(z)$ in the brackets all converge to zero as $t \to 0$, although they are not exactly zero when $t > 0$. As a consequence, even if z is close to a feasible point of the MPCC where MPCC-LICQ holds, the multipliers $\mu_{L,j}$ and $\mu_{U,j}$ are not necessarily zero to satisfy (76) with $t > 0$ sufficiently small. From a practical point of view, numerically dependent systems arise from the LF formulation for very small positive t.

Further, we solve the problems in Table 2 to a smaller tolerance $\epsilon_{\text{tol}} \leq 10^{-12}$. All the formulations except LF converge to the same solutions as before more accurately, with the complementarity residuals vanishing to zero. With ϵ (or t) getting smaller, REG usually only requires a few iterations for convergence of each NLP, while for the NCP-based formulations (i.e., BA/MLF/NCP), the number of iterations in solving each NLP does not always decrease. On the other hand, LF converges more accurately only for ex9.2.9. For df1, qpec1, and sl1, complementarity residuals at $\bar{z}$ cannot be decreased below 1.08e-5, 1.97e-5, and 2.7e-7, respectively. For bard2m, ex9.2.3, ex9.2.8, and scholtes2, the multipliers $\bar{\mu}_L$ and $\bar{\mu}_U$

Problem	BA	LF	MLF	NCP	REG
bard2m	61/1.25e-6	47/1.25e-6	59/1.25e-6	61/1.5e-7	9/8.88e-16
df1	9/9.16e-9	26/1.08e-5	25/2.6e-7	9/1.24e-6	7/0
ex9.2.3	58/1.25e-6	121/3.93e-16	69/1.25e-6	50/1.02e-6	17/1.7e-7
ex9.2.8	38/9.34e-16	21/1.56e-12	13/1.56e-10	85/1.73e-12	14/2.5e-6
ex9.2.9	28/1.25e-6	17/1.25e-6	12/1.25e-6	40/1.25e-6	7/0
kth1	40/3.37e-11	17/1.92e-10	12/6.38e-12	57/1.25e-6	7/0
qpec1	28/1.25e-6	124/1.25e-6	42/1.25e-6	28/1.25e-6	56/5.55e-17
scholtes2	43/2.3e-7	57/3.01e-6	59/2.3e-7	49/1.02e-6	10/0
sl1	33/1.25e-6	38/1.25e-6	31/1.25e-6	29/8.8e-7	6/2.22e-16

Table 2: Total iterations/complementarity residuals (ψ) in strongly stationary cases. The iteration count is the sum of iterations in consecutive CONOPT4 solutions till the outer loop converges. The complementarity residual is calculated by $\psi = \max_j(\min(G_j(\bar{z}), H_j(\bar{z}))), j \in I_1(\bar{z}) \cap I_2(\bar{z})$.

become significantly more ill-conditioned as t decreases, deflecting the previous solutions to points without biactive elements, with the complementarity residuals also become very small. However, these points are infeasible to the original MPCC, with one of the $G_j(\bar{z})$ and $H_j(\bar{z})$ negative and the other close to zero. Finally, kth1 fails in NLP solution when $t \leq 10^{-12}$. Profiles for the last five problems are presented in Figure 1.

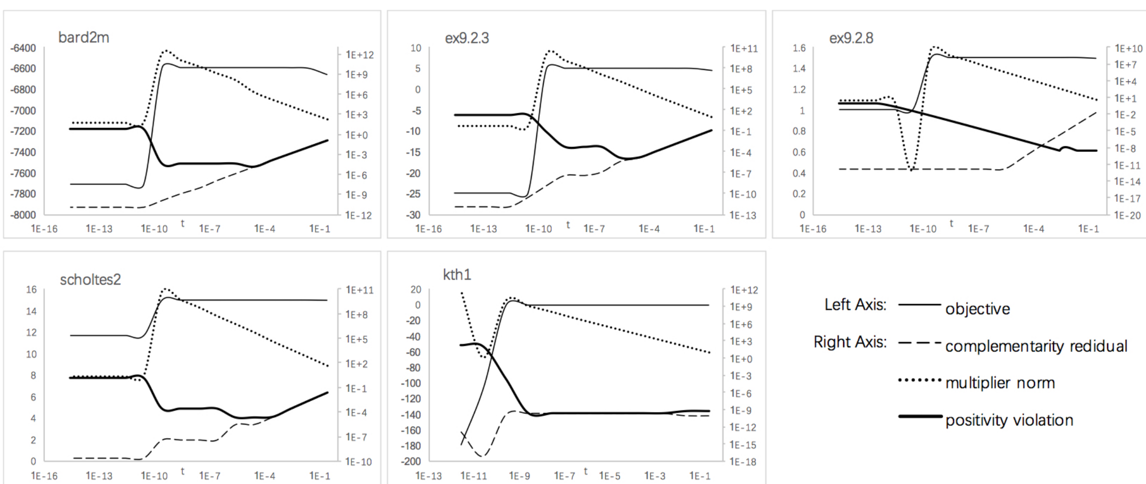

Figure 1: LF profiles as t decreases from 0.25 to 10^{-15}, where complementarity residual $= \max_j(\min(G_j(z^k), H_j(z^k)))$, positivity violation $= \max_j(0, -G_j(z^k), -H_j(z^k))$, for $j = 1, \ldots, n_x$; multiplier norm $= \max(\|\mu_L^k\|_\infty, \|\mu_U^k\|_\infty)$.

4.1.2 Examples without Strongly Stationary Solutions

Starting from the default initial points, all the five formulations solve problems ex9.2.2, qpec2, and scholtes4 to the solutions as shown in Table 4. Iteration counts and complemen-

Problem	BA		LF		MLF		NCP	REG		
	$\bar{u}$	$\bar{p}$	$\bar{\mu}_L$	$\bar{\mu}_U$	$\bar{u}_L$	$\bar{u}_U$	$\bar{u}_{\text{NCP}}$	$\bar{\nu}_G$	$\bar{\nu}_H$	$\bar{\nu}_{GH}$
bard2m	-144.00	1.25e-6	1.17e8	1.14e8	144.00	0	-144.00	2.00	142.00	0
df1	0	1.25e-6	0	0	0	0	-2.5e-6	0	0	0
ex9.2.3	-2.50	1.25e-6	1.35e6	1.30e6	2.50	0	-2.50	1.00	1.50	0
ex9.2.8	-0.50	1.25e-6	4.00e5	0	1.00	0	-2.10	0.50	2.00	0
ex9.2.9	-2.00	1.25e-6	0	0	0	0	-2.00	0	1.00	0
kth1	-2.00	1.25e-6	8.00e5	0	2.00	0	-2.00	1.00	1.00	0
qpec1	-4.00	1.25e-6	1.60e6	0	4.00	0	-4.00	4.00	-	0
	-4.00	1.25e-6	1.60e6	0	4.00	0	-4.00	4.00	-	0
	-4.00	1.25e-6	1.60e6	0	4.00	0	-4.00	4.00	-	0
	-4.00	1.25e-6	1.60e6	0	4.00	0	-4.00	4.00	-	0
	-4.00	1.25e-6	1.60e6	0	4.00	0	-4.00	4.00	-	0
	-4.00	1.25e-6	1.60e6	0	4.00	0	-4.00	4.00	-	0
	-4.00	1.25e-6	1.60e6	0	4.00	0	-4.00	4.00	-	0
	-4.00	1.25e-6	1.60e6	0	4.00	0	-4.00	4.00	-	0
	-4.00	1.25e-6	1.60e6	0	4.00	0	-4.00	4.00	-	0
	-4.00	1.25e-6	1.60e6	0	4.00	0	-4.00	4.00	-	0
scholtes2	-10.00	1.25e-6	5.09e6	4.97e6	10.00	0	-10.00	6.00	4.00	0
sl1	0	1.25e-6	0	0	0	0	-1.33e-6	0	0	0

Table 3: Multipliers of biactive elements in strongly stationary cases. REG does not have $\bar{\nu}_H$ for qpec1, in that G_j and H_j have identical expressions in this example.

tarity residuals at the final solutions $\bar{z}$ are given by Table 5. Table 6 gives multipliers of the biactive elements at $\bar{z}$.

For these problems, NCP-based formulations outperform the others, in terms of iteration counts and complementarity residuals; and MLF is the only formulation that identifies the two biactive pairs of ex9.2.2 precisely. The behavior of LF is about the same as in the strongly stationary examples. However, REG does not exhibit superior iteration counts as before, and it cannot converge accurately in terms of the complementarity residuals. When $\epsilon_{\text{tol}} = 10^{-12}$ is applied, all the formulations converge to the same solutions as in Table 4; and LF and REG still cannot converge accurately. For LF, the complementarity residuals stay at their previous levels; for REG, the nonzero multipliers $\bar{\nu}_{GH}$ in Table 6 increase to $O(10^4)$ or larger, and the complementarity residuals cannot decrease below 10^{-5}.

Problem	$f(\bar{z})$	$\bar{z}$	$I_1(\bar{z}) \cap I_2(\bar{z})$
ex9.2.2	100	(10, 10, 0, 0, 0, 0, 0, 10, 10, 0)	1, 4
qpec2	45	(1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)	11, 12, 13, 14, 15, 16, 17, 18, 19, 20
scholtes4	0	(0, 0, 0)	1

Table 4: MacMPEC examples converged to C-stationary points

Problem	BA	LF	MLF	NCP	REG
ex9.2.2	40/1.25e-6	30/7.80e-6	24/1.25e-6	48/7.2e-7	28/9.13e-4
qpec2	26/1.25e-6	115/1.92e-5	32/1.25e-6	26/1.25e-6	211/1.58e-3
scholtes4	14/1.25e-6	12/1.11e-5	12/1.25e-6	14/1.25e-6	12/1.58e-3

Table 5: Total iterations/complementarity residuals (ψ) in C-stationary cases. The iteration count is the sum of iterations in consecutive CONOPT4 solutions till the outer loop converges. The complementarity residual is calculated by $\psi = \max_j(\min(G_j(\bar{z}), H_j(\bar{z}))), j \in I_1(\bar{z}) \cap I_2(\bar{z})$.

Problem	BA		LF		MLF		NCP	REG		
	$\bar{u}$	$\bar{p}$	$\bar{\mu}_L$	$\bar{\mu}_U$	$\bar{u}_L$	$\bar{u}_U$	$\bar{u}_{\text{NCP}}$	$\bar{\nu}_G$	$\bar{\nu}_H$	$\bar{\nu}_{GH}$
ex9.2.2	5.74	0	0	2.20e5	0	5.74	6.67	0	0	1.83e3
	0	1.25e-6	0	0	0	0	0	0	0	0
qpec2	4.00	0	0	1.04e5	0	4.00	4.00	0	-	1.26e3
	4.00	0	0	1.04e5	0	4.00	4.00	0	-	1.26e3
	4.00	0	0	1.04e5	0	4.00	4.00	0	-	1.26e3
	4.00	0	0	1.04e5	0	4.00	4.00	0	-	1.26e3
	4.00	0	0	1.04e5	0	4.00	4.00	0	-	1.26e3
	4.00	0	0	1.04e5	0	4.00	4.00	0	-	1.26e3
	4.00	0	0	1.04e5	0	4.00	4.00	0	-	1.26e3
	4.00	0	0	1.04e5	0	4.00	4.00	0	-	1.26e3
	4.00	0	0	1.04e5	0	4.00	4.00	0	-	1.26e3
	4.00	0	0	1.04e5	0	4.00	4.00	0	-	1.26e3
scholtes4	2.00	0	0	8.98e4	0	2.00	2.00	0	0	6.32e2

Table 6: Multipliers of biactive elements in C-stationary cases. REG does not have $\bar{\nu}_H$ for qpec2, in that G_j and H_j have identical expressions in this example.

Performance degeneracy of REG can also be attributed to numerical failure in satisfying LICQ. For REG (16), consider the equation

$$\begin{aligned} 0 \; = \textstyle\sum_{i=1}^{n_c} \lambda_i \nabla c_i(z) \; & - \sum_{j \in I_G(z)} \nu_{G,j} \nabla G_j(z) - \sum_{j \in I_H(z)} \nu_{H,j} \nabla H_j(z) \\ & + \sum_{j \in I_{GH}(z,t)} \nu_{GH,j} \left[H_j(z) \nabla G_j(z) + G_j(z) \nabla H_j(z) \right], \end{aligned} \tag{77}$$

where $I_G(z) = \{j \,|\, G_j(z) = 0\}, I_H(z) = \{j \,|\, H_j(z) = 0\}$, and $I_{GH}(z,t) = \{j \,|\, G_j(z)H_j(z) = t\}$ are active sets of the inequality constraints. The nonzero multipliers $\bar{\nu}_{GH}$ in Table 6 indicate that $I_{GH}(\bar{z}, \bar{t}) \cap I_1(\bar{z}) \cap I_2(\bar{z}) \neq \emptyset$. In such circumstance, $G_j(z)$ and $H_j(z)$ in (77) both tend to zero as $t \to 0$, although from (16) they are nonzero for any $t > 0$. As a consequence, even if z is close to a feasible point of the MPCC where MPCC-LICQ holds, the multipliers $\nu_{GH,j}$ may not be zero to satisfy (77) with $t > 0$ sufficiently small. This degeneracy can be recognized from large REG multipliers and inaccuracy in satisfying the complementarity conditions.

Now we discuss stationarity of $\bar{z}$ for the MPCCs. Table 6 shows that ex9.2.2 has one of the biactive pairs that has multipliers obviously meeting the requirements for strong stationarity; in all the other cases we have $\bar{u}_j > 0$ (i.e., $\bar{p}_j = 0$) (by BA), and $\bar{u}_{U,j} > 0$ (by MLF). We know from (52) and (56) that $\bar{z}$ is at least C-stationary. More properties about these examples are analyzed below.

Example ex9.2.2:

$$\begin{array}{lrl} \min & & x^2 + (y-10)^2 \\ \text{s.t.} & & x \le 15 \\ & \lambda_1: & -x + y \le 0 \\ & & -x \le 0 \\ & \lambda_2: & x + y + s_1 = 20 \\ & \lambda_3: & -y + s_2 = 0 \\ & \lambda_4: & y + s_3 = 20 \\ & \lambda_5: & 2x + 4y + l_1 - l_2 + l_3 = 60 \\ & u_{li}, u_{si}: & 0 \le l_i \perp s_i \ge 0 \; i = 1, \ldots, 4, \end{array}$$

where the multipliers λ_i correspond to ordinary active constraints at $\bar{z}$. The KKT conditions at $\bar{z}$ for the RNLP requires that

$$\begin{array}{lll} u_{l1} = \lambda_5, & u_{s1} = -3\lambda_5 - 10, & 0 \le \lambda_1, \\ u_{l2} = -\lambda_5, & u_{s2} = 0, & 0 = \lambda_1 + \lambda_5 - 10, \\ u_{l3} = \lambda_5, & u_{s3} = 0, & 0 = \lambda_2 + 3\lambda_5 + 10, \\ u_{l4} = 0, & u_{s4} = 0, & 0 = \lambda_3 = \lambda_4. \end{array} \tag{78}$$

This implies that u_{l1}, u_{s1} cannot both be nonnegative, hence $\bar{z}$ cannot be strongly stationary. On the other hand, since LICQ does not hold at $\bar{z}$, we can choose $\lambda_5 = 0$ or $\lambda_5 = -10/3$, so that $\bar{z}$ is M-stationary.

Example qpec2:

$$\begin{array}{lrl} \min & & \sum_{i=1}^{10}(x_i - 1)^2 + \sum_{i=1}^{10}(y_{1i} - 2)^2 + \sum_{j=1}^{10}(y_{2j} - 2)^2 \\ \text{s.t.} & \text{c1-c10:} & 0 \le y_{1i} - x_i \perp y_{1i} \ge 0, \; i = 1, \ldots, 10 \\ & \text{c11-c20:} & 0 \le y_{2j} \perp y_{2j} \ge 0, \; j = 1, \ldots, 10. \end{array}$$

The biactive complementarity conditions c11-c20 are actually ordinary equality constraints $y_{2j} = 0, j = 1, \ldots, 10$, and there is no restriction on the sign of their multipliers. Therefore, $\bar{z} = (\bar{x}, \bar{y}_1, \bar{y}_2)$ is a strongly stationary point if conditions c11-c20 are viewed as ordinary constraints; otherwise, $\bar{z}$ is C-stationary and qpec2 has no solution with stronger stationarity properties. Specifically, weak stationarity conditions for every $z = (x, y_1, y_2)$ feasible to the MPCC require

$$\begin{bmatrix} 2(x-1) \\ 2(y_1 - 2) \\ 2(y_2 - 2) \end{bmatrix} - \sum_{i \in I_1} \sigma_1 \begin{bmatrix} -e_i \\ e_i \\ 0 \end{bmatrix} - \sum_{i \in I_{y1}} \sigma_{y1} \begin{bmatrix} 0 \\ e_i \\ 0 \end{bmatrix} - \sum_{j \in I_{y2}} \sigma_{y2} \begin{bmatrix} 0 \\ 0 \\ e_j \end{bmatrix} = 0, \tag{79}$$

where I_1, I_{y1}, I_{y2} are the active sets for conditions $y_1 - x \geq 0, y_1 \geq 0, y_2 \geq 0$, respectively. Since $y_2 \equiv 0$, the multipliers σ_{y2} (of the artificially biactive elements) can only be $\sigma_{y2,j} = -2,\ j = 1, \ldots, 10$.

Example scholtes4:

$$\begin{array}{lrl} \min & & z_1 + z_2 - z_3 \\ \text{s.t.} & \lambda_1 : & -4z_1 + z_3 \leq 0 \\ & \lambda_2 : & -4z_2 + z_3 \leq 0 \\ & u_1, u_2 : & 0 \leq z_1 \perp z_2 \geq 0. \end{array}$$

The point $\bar{z} = (\bar{z}_1, \bar{z}_2, \bar{z}_3) = (0, 0, 0)$ is not strongly stationary, because it is not a KKT point of the RNLP. In fact, there do not exist multipliers $u_1, u_2 \geq 0$ satisfying the KKT conditions for the RNLP, because:

$$\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} + \lambda_1 \begin{bmatrix} -4 \\ 0 \\ 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 \\ -4 \\ 1 \end{bmatrix} - u_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - u_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = 0. \tag{80}$$

However, by noting that multipliers for this system are not unique, we can choose one of u_1 and u_2 be 0, and the other one be -2, such that $\bar{z}$ is M-stationary. This example was also discussed in [40] to demonstrate that MPCC-MFCQ ensures C-stationarity of $\bar{z}$, while strong stationarity requires more restrictive constraint qualification at the local minimizer of the MPCC, for example MPCC-SMFCQ, which fails at $\bar{z}$.

4.1.3 Discussion based on Asymptotic Weak Nondegeneracy

The two preceding subsections have shown examples for which we can judge the solutions to be strongly stationary or not, from the parameter p_j or multipliers $\bar{u}_j$ (by BA) or from the multipliers $\bar{u}_{U,j}$ (by MLF), for all $j \in I_1(\bar{z}) \cap I_2(\bar{z})$. Now we discuss a case where we do not have such strong evidence.

Example ralph2 is described as

$$\begin{array}{ll} \min & x^2 + y^2 - 4xy \\ \text{s.t.} & 0 \leq x \perp y \geq 0. \end{array}$$

Starting from $(x, y) = (1, 1)$, all the formulations converge to $(\bar{x}, \bar{y}) = (0, 0)$, which is the global optimum and satisfies the MPCC-LICQ. Numerical results are summarized in Table 7. At the solution, the lower bound of MLF is inactive such that $\bar{u}_L = 0$, while the upper bound is not strongly active such that $\bar{u}_U$ is nearly zero. Correspondingly, BA has the parameter $\bar{p} = 0$, and at the same time the multiplier $\bar{u} > 0$ is close to zero. These observations are not compliant with the immediate evidence for strong stationarity. Note that the multiplier $\bar{\nu}_{GH}$ is positive but bounded; Scholtes [41] has proved that with bounded multiplier ν_{GH} as $t \to 0$, the solution is strongly stationary.

In fact, strong stationarity of $\bar{z}$ for this simple example can be verified conveniently by the RNLP, namely, by that $\bar{z}$ is a KKT point of the RNLP. However, strict complementarity does not hold at $\bar{z}$ for the RNLP, and hence the upper level strict complementarity (ULSC) for the MPCC does not hold either. Indeed, $x^* = y^* = \sigma_1^* = \sigma_2^* = 0$ in the limit, which explains the nearly zero NLP multipliers at $\bar{z}$.

ralph2	BA		LF		MLF		NCP	REG		
Multiplier	$\bar{u}$	$\bar{p}$	$\bar{\mu}_L$	$\bar{\mu}_U$	$\bar{u}_L$	$\bar{u}_U$	$\bar{u}_{\text{NCP}}$	$\bar{\nu}_G$	$\bar{\nu}_H$	$\bar{\nu}_{GH}$
	5.01e-6	0	0	2.00	0	5.00e-6	5.01e-6	0	0	2.00
Total iteration	57		44		40		57	40		
Complementarity residual	1.25e-6		9.5e-7		1.25e-6		1.25e-6	1.58e-3		

Table 7: Results of Example ralph2.

Another way to analyze the stationarity with the NCP-based formulations takes advantage of the concept of *asymptotic weak nondegeneracy* [16]. We will see more usage of this approach in the sequel for larger cases. Recall that $z^k \to z^*$ as $\epsilon^k \to 0$. For every $\epsilon^k > 0$ sufficiently small, we only need to consider the cases $p_j^k \equiv 0$ ($j \in I_1(z^*) \cap I_2(z^*)$) (for BA) and $j \in I_U(z^k, \epsilon^k) \cap I_1(z^*) \cap I_2(z^*)$ (for MLF). For any accumulation point r^* of $\left\{\nabla \Phi_j^\epsilon(z^k)\right\}$, asymptotic weak nondegeneracy requires that

$$\xi_j^* > 0,\ \eta_j^* > 0. \tag{81}$$

According to the expression (44) of $\nabla \Phi_j^\epsilon$, this means that there exist $\rho_1, \rho_2 > 0$ such that $0 < \rho_1 \le G_j(z^k)/H_j(z^k) \le \rho_2 < +\infty$; that is, $G_j(z^k)$ and $H_j(z^k)$ approach to zero in the same order of magnitude.

For problem ralph2, we have $G(z^k) = x^k, H(z^k) = y^k, \Phi^\epsilon(z^k) = x^k - \frac{1}{2}[x^k - y^k + \sqrt{(x^k - y^k)^2 + (\epsilon^k)^2}]$. Obviously, every z^k is asymptotically weakly nondegenerate. Suppose that the point $\bar{z} = (\bar{x}, \bar{y}) = (0, 0)$ is not strongly stationary. Then $\bar{u} > 0$ by (52), $\bar{u}_U > 0$ by (56). With the asymptotic condition and the null space matrix $\left[\begin{smallmatrix} -\bar{x}/\bar{y} \\ 1 \end{smallmatrix}\right]$, the reduced Hessian of $\Phi^\epsilon(\bar{z})$ tends to $-\infty$, i.e.,

$$\left[\begin{array}{cc} -\bar{x}/\bar{y} & 1 \end{array}\right] \cdot \frac{\bar{u} \cdot 2\bar{x}\bar{y}}{(\bar{x}+\bar{y})^3} \left[\begin{array}{cc} -1 & 1 \\ 1 & -1 \end{array}\right] \cdot \left[\begin{array}{c} -\bar{x}/\bar{y} \\ 1 \end{array}\right] = -2\bar{u} \cdot \frac{\bar{x}/\bar{y}}{\bar{y}(\bar{x}/\bar{y}+1)} \to -\infty. \tag{82}$$

The same result can be derived for MLF by substituting $\bar{u}_U$ for $\bar{u}$. Because (82) contradicts the optimality of $\bar{z}$ for the NLP, the hypothesis is false and $\bar{z}$ must be strongly stationary. (Note that strong stationarity instead of M-stationarity is concluded because of the presence of asymptotic weak nondegeneracy.) This analysis is validated by the numerical results. As shown in Table 7, we have $\bar{u}$ and $\bar{u}_U$ very close to zero, and $\bar{z}$ is declared to be (locally) optimal

4.2 Distillation Case Studies

Phase behavior of a vapor liquid system is determined by minimization of the Gibbs free energy. Embedded within distillation optimization, or other process optimization problems, these conditions lead to two-level optimization problems, which can be modeled through complementarity constraints. These conditions allow phase disappearance to be described in distillation systems for optimization of both steady state and dynamic tray columns [31,38]. The distillation column models consisting of MESH (Mass Balance, Equilibrium, Summation, and Heat Balance) equations are incorporated within an MPCC to optimize the feed tray location and total tray count.

The MPCC model is developed in [3]. This distillation MPCC formulation uses distribution functions that direct all feed, reflux and intermediate product streams to the column trays. As shown in Figure 2, streams for the feed and the reflux are fed to all trays as dictated by two discretized Gaussian distribution functions. Note that the grayed area in Figure 2 consists only of vapor traffic, and consequently, each tray model (i) includes a relaxed phase equilibrium model and the following complementarities that allow for disappearance of the liquid phase.

$$y_{ij} = \beta_i \frac{P_j(T_i)}{P_i} x_{ij},\ \beta_i = 1 - s_i^l + s_i^v, \tag{83a}$$

$$0 \le L_i \perp s_i^l \ge 0,\ (s_i^l - \max(0, s_i^l - L_i) = 0, \tag{83b}$$

$$0 \le V_i \perp s_i^v \ge 0,\ (s_i^v - \max(0, s_i^v - V_i) = 0, i \in \mathcal{S}, \tag{83c}$$

where i is tray index numbered from reboiler$(= 1)$, j is components index, β_i is the relaxation parameter, P is column pressure, T_i is temperature of tray i, $L_i/V_i \ge 0$ is flow rates of liquid/vapor, $x_{ij}/y_{ij} \in [0, 1]$ is fraction of component j in liquid/vapor leaving tray i, s_i^l and s_i^v are slack variables, the tray set $\mathcal{S} = \{2, \ldots, N-1\}$ with N as total number of trays. We choose two additional continuous optimization variables N_f, the feed location, and N_t, the number of trays, with $N_t \ge N_f$. We also specify feed and reflux flowrates for $i \in \mathcal{S}$ based on the value of the distributions at tray i, given by

$$F_i = F \frac{\exp(\frac{-(i-N_f)^2}{\sigma_f})}{\sum\limits_{j \in \mathcal{S}} \exp(\frac{-(j-N_f)^2}{\sigma_f})}, \quad R_i = R \frac{\exp(\frac{-(i-N_t)^2}{\sigma_t})}{\sum\limits_{j \in \mathcal{S}} \exp(\frac{-(j-N_t)^2}{\sigma_t})} \ \ i \in \mathcal{S}, \tag{84}$$

where $\sigma_f, \sigma_t = 0.5$ are parameters in the distribution. Note that feed and reflux flowrates are allowed on all trays $i \in \mathcal{S}$.

The resulting MPCC model is used to determine the optimal number of trays, reflux ratio and feedtray location for a benzene/toluene separation. The five MPCC formulations are considered for the distillation case study. All of these solution strategies were solved in a sequence of 13 NLPs with $\epsilon^k = 10^{k/2}, k = 0, \ldots, 12$. These formulations are modeled in GAMS and solved with CONOPT4, using default options.

Benzene-Toluene Separation

The MPCC model from [3], consisting of the mass, summation, energy balances, equlibrium complementarities (83) for each tray and tray distributions (84), is applied to a binary column with a maximum of $N = 10$ trays; its feed is 100 mol/s of a 70%/30% mixture of benzene/toluene and distillate flow is specified to be 50% of the feed. The objective function for the benzene-toluene separation minimizes

$$objective = wt \cdot D \cdot x_{D,Toluene} + wr \cdot r + wn \cdot N_t, \tag{85}$$

where N_t is the number of trays, $r = R/D$ is the reflux ratio, D is distillate flow, $x_{D,Toluene}$ is the toluene mole fraction and weighting parameters are set to $wt = 1$, $wr = 0.01$, and $wn = 0.45$; these weights allow the optimization to trade off product purity, energy cost and

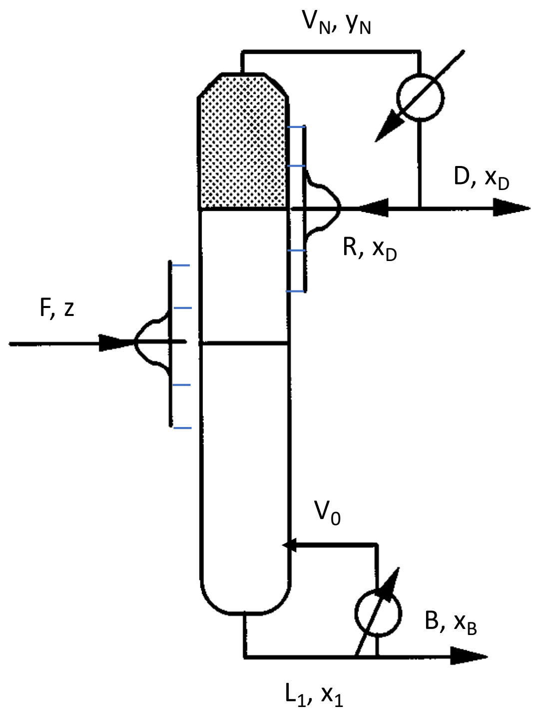

Figure 2: Distillation column showing feed and reflux flows distributed according to (84). The grayed column section is above the reflux location and has negligible liquid flows.

capital cost. The reflux ratio r is allowed to vary between 1 and 20, the feedtray location N_f varies between 2 and 20, and the total tray number N_t varies between $N_f + 1$ and $N - 1$. For all cases, there were $2N - 2$ complementarity constraints.

With $N = 10$ the resulting GAMS models consists of 142 (inequality and equality) constraints and 148 variables for BA and NCP formulations, 144 constraints and 149 variables for the REG formulation, 160 constraints and 148 variables for the LF and MLF formulations. Three cases were considered with $x_{D,Toluene} \leq \zeta$ with $\zeta = 0.05, 0.01, 0.005$. In addition, a fourth case was considered with $\zeta = 0.001$ and with $N = 25$ trays. This larger case leads to GAMS models with 352 constraints and 358 variables for BA and NCP formulations, 354 constraints and 359 variables for REG formulation, and 400 constraints and 358 variables for LF and modified LF formulations.

All 20 problems were initialized far away from the optimum with $N_t = 21, N_f = 7, R = 2.2$. Temperature and mole fraction profiles were initialized with linear interpolations based on the top and bottom product properties. While the model is nonconvex and admits multiple optima, all five MPCC formulations converged to tolerance with CONOPT4 (10^{-7}) with essentially the same optimal solutions. Optimal objective function and design variable values are presented in Table 8, along with the cardinality of the biactive complementarity set. Iteration counts for the five formulations are presented in Table 9 along with an estimated $\epsilon-$convergence rate of the 13 NLPs for each MPCC. From this table we observe that NCP approach requires the lowest computational cost, followed by the BA approach. On the other hand, the LF and MLF approaches require the most effort to solve, while the REG approach takes intermediate effort. These trends can be especially observed in the last case, which is larger and overall takes more effort to solve.

ζ	N	Objective	N_t	N_f	r	$\lvert I_1 \cap I_2 \rvert$
0.05	10	3.9916	7.6804	3.85	0.72	0
0.01	10	2.9078	8.84	2.94	1.52	0
0.005	10	3.2336	8.93	2.86	2.09	0
0.001	25	3.0755	12.83	2.93	1.74	10

Table 8: Solutions of Benzene/Toluene cases.

ζ	N	BA	LF	MLF	NCP	REG
0.05	10	127/1.08	154/1.07	154/1.08	92/1.91	184/ 0.90
0.01	10	120/1.08	167/1.08	140/1.08	77/1.64	125/1.01
0.005	10	125/1.06	145/1.08	131/1.07	66/2.02	107/1.03
0.001	25	208/1.06	461/1.08	287/1.07	189/1.08	266/0.55

Table 9: Total iterations/convergence rate (ψ) of MPCC formulations for Benzene/Toluene cases. The iteration count is the sum of 13 consecutive CONOPT4 solutions with $\epsilon_i = 10^{-i/2}, i = 0, \ldots, 12$. The estimated convergence rate ψ is calculated from $|f(z(\epsilon_i) - f(\bar{z})|/|f(z(\epsilon_{i'}) - f(\bar{z})| = (\epsilon_i/\epsilon_{i'})^{\psi}$.

At the solution of all cases, CONOPT reports no negative curvature directions nor Jacobian degeneracies, which indicates satisfaction of LICQ and second-order necessary conditions. For the first three cases, $I_1(\bar{z}) \cap I_2(\bar{z}) = \emptyset$ at the last NLP solution $\bar{z}$, hence $\bar{z}$ is strongly stationary for the MPCC. For the last case, $I_1(\bar{z}) \cap I_2(\bar{z}) \neq \emptyset$, and there is no evidence supporting easy discrimination between S- and M-stationarity. Specifically, for all $j \in I_1(\bar{z}) \cap I_2(\bar{z})$, we have $\bar{p}_j = 0$ and small positive $\bar{u}_j$ (with BA), $\bar{u}_{L,j} = 0$ and small positive $\bar{u}_{U,j}$ (with MLF), and positive and bounded $\bar{\nu}_{GH,j}$ (with REG).

We employ again the concept of asymptotic weak nondegeneracy to determine the stationarity properties at the solution. We solve the last case to smaller tolerances, and all the formulations converge to the same point $\bar{z}$ as before. When decreasing ϵ^k to 10^{-12}, at the result of BA, for $j \in I_1(\bar{z}) \cap I_2(\bar{z})$, we still have all $\bar{p}_j = 0$; however, except for three small multipliers ($0 < \bar{u}_j < 0.3$), the remaining multipliers are essentially zero ($0 \leq \bar{u}_j < 10^{-5}$). Recall that u_j is the sensitivity $\mathrm{d}f(z)/\mathrm{d}p_j$. These vanishing multipliers suggest that small changes of the $\bar{p}_j$ (say, increasing $\bar{p}_j$ to $\bar{\epsilon}/2$) would not influence the objective value; on the other hand, it would not matter for MLF to have $\bar{u}_{L,j} > 0$ or $\bar{u}_{U,j} > 0$ for these constraints. In the course of decreasing ϵ^k from 10^{-12} to 10^{-15}, the multipliers u_j^k are still approaching zero. In addition, we observe that the biactive elements corresponding to the nonzero multipliers satisfy $O(10^{-2}) \leq G_j(z^k)/H_j(z^k) \leq O(10^2)$, which meets the asymptotic weak nondegeneracy requirement. Again, CONOPT reports no negative curvature directions nor Jacobian degeneracies. Therefore we conclude with all these observations that the solution $\bar{z}$ should be strongly stationary. Finally, it is worth mentioning that the largest REG multiplier at $\epsilon^k = 10^{-12}$ and $\epsilon^k = 10^{-15}$ is $\nu_{GH,j}^k \approx 2297$, which also suggests that we are probably approaching a strongly stationary solution.

Argon Column Optimization

As the second distillation MPCC we consider a larger, ternary model, which deals with the separation of argon from air. The argon separation column has $N = 63$ trays, its feed is 6546.54 lbmol/h of a 0.005%/9.753%/90.24% mixture of nitrogen/argon/oxygen, and distillate flow is specified to be 202.4576 lbmol/h with less than 1 mol % oxygen. The reflux ratio r is allowed to vary between 20 and 100, the feedtray location N_f varies between 1 and 10, and the total tray number N_t varies between 31 and 63. The objective function for the argon problem minimizes

$$objective = r + N_t, \tag{86}$$

where r is the reflux ratio. For all cases, there were $2N - 2$ complementarity constraints.

The first case is Reflux Constrained and limits the reflux ratio to $r \leq 35$, while the second case has a reflux ratio with $r \leq 100$ which is not constrained at the solution. These two cases were initialized with $N_t = 25, N_f = 5, R = 25$. The resulting GAMS models for these cases consist of 1260 (inequality and equality) constraints and 1264 variables for the BA and NCP formulations, 1260 constraints and 1264 variables for the REG formulation, 1382 constraints and 1264 variables for the LF and MLF formulations.

In addition, a third case was considered. This case is Tray Constrained with $N_t \leq N = 30$, and it was initialized with $N_t = 25, N_f = 5, R = 35$. Temperature and mole fraction profiles were initialized with linear interpolations based on the top and bottom product properties. This GAMS model consists of 598 constraints and 604 variables for BA, NCP and REG formulations, and 656 constraints and 604 variables for LF and modified LF formulations.

While the models are nonconvex and admit multiple optima, all five MPCC formulations converged to KKT tolerance (10^{-7}) with CONOPT4 with the same optimal solutions for each case. Optimal objective function and design variable values are presented in Table 10, along with the cardinality of biactive complementarity set. Iteration counts for the five formulations are presented in Table 11 along with an estimated $\epsilon-$convergence rate of the 13 NLPs for each MPCC. From the table we again observe that the NCP approach requires the lowest computational cost, followed by the BA approach. Also, the LF and MLF approaches require the most effort while the REG approach takes intermediate effort. These trends can be especially observed in the first and second cases, which are larger and overall take more effort to solve.

Case	Objective	N_t	N_f	r	$\lvert I_1 \cap I_2 \rvert$
Reflux constrained	72.47	37.47	5.20	35.0	22
Unconstrained	72.20	36.46	4.89	35.79	23
Tray constrained	84.85	29.17	2.73	56.67	0

Table 10: Solutions of Argon column cases

At the solution of all cases, CONOPT reports no negative curvature directions nor Jacobian degeneracies, indicating satisfaction of LICQ and second-order necessary conditions. The Tray Constrained case converges to a strongly stationary point $\bar{z}$, where $I_1(\bar{z}) \cap I_2(\bar{z}) = \emptyset$.

Case	BA	LF	MLF	NCP	REG
Reflux Constrained	293/1.08	1193/1.08	1100/1.08	259/1.17	552/0.85
Unconstrained	285/1.04	993/1.06	455/1.07	247/1.43	517/0.88
Tray Constrained	148/1.08	182/1.08	242/1.08	92/2.0	229/1.02

Table 11: Total iterations/convergence rate (ψ) of MPCC formulations for Argon column cases. The iteration count is the sum of 13 consecutive CONOPT4 solutions with $\epsilon_i = 10^{-i/2}, i = 0, \dots 12$. The estimated convergence rate ψ is calculated from $|f(z(\epsilon_i) - f(\bar{z})|/|f(z(\epsilon_{i'}) - f(\bar{z})| = (\epsilon_i/\epsilon_{i'})^\psi$.

In the other two cases, $I_1(\bar{z}) \cap I_2(\bar{z}) \neq \emptyset$ and we need the condition of asymptotic weak nondegeneracy to estimate the stationarity properties. The problems are further solved to $\bar{\epsilon} = 10^{-15}$. In the Reflux Constrained case, for $j \in I_1(\bar{z}) \cap I_2(\bar{z})$, BA has four nonzero multipliers, the largest of which is $0 < \bar{u}_j < 0.008$, whose corresponding biactive elements show $G_j(\bar{z})/H_j(\bar{z}) < 3.59 \times 10^7$. In the Unconstrained case, BA has three nonzero multipliers, and the largest one is $0 < \bar{u}_j < 0.0008$, whose corresponding biactive elements show $G_j(\bar{z})/H_j(\bar{z}) < 10$. Again, while ϵ decreasing, no negative curvature directions nor Jacobian degeneracies are reported. Based on these observations, we prefer to be slightly conservative and conclude that the solutions of the Reflux Constrained and Unconstrained cases are M-stationary and strongly stationary, respectively. For comparison, the largest REG multiplier is $\bar{\nu}_{GH,j} \approx 4415$ (Relux Constrained case) and $\bar{\nu}_{GH,j} \approx 6.66$ (Unconstrained case).

To conclude, the LF and MLF formulations require the most effort for both distillation systems. This is likely because determining the active sets becomes more difficult as $\epsilon \to 0$, which is also observed in [29]. With the BA approach, the underlying optimality conditions relate closely to LF and MLF, while the NLP structure is similar to the faster NCP formulation. As a result, BA provides a good compromise among the five MPCC formulations.

5 Conclusions

Nonlinear programs involving nonsmooth systems occur frequently in practice. This study deals with nonsmoothness arising from max operators, by expressing them equivalently in complementarity form. Strategies have been widely investigated to converge the resulting MPCC problems to a meaningful solution, by employing nonlinear programming formulations and algorithms. We put forward two NLP formulations (BA and MLF) based on NCP-functions generated from ϵ-smoothed square root and neural network functions. In particular, BA operates together with a sensitivity directed bounding strategy to isolate the minimizer of the MPCCs. It has been proved that with sensitivity corrections, the solution of the two formulations differs by $O(\epsilon^2)$ from the solution of the MPCCs (Proposition 2.1).

Stationarity at the limit of the stationary points of the proposed formulations are investigated. In the presence of MPCC-LICQ, LICQ holds at every feasible point of the proposed NLPs in a neighborhood of a point feasible to the MPCC (Theorems 3.1 and 3.2); in the limit, the stationary point of the NLP is guaranteed to be C-stationary (Theorems 3.3 and 3.4); furthermore, M-stationarity is established with additional second-order conditions (Theorems 3.5 and 3.6).

The proposed formulations, together with the closely related Lin-Fukushima formulation (LF), well-studied Scholtes' formulation (REG), and the ordinary NCP formulation (non-parametric), are applied to selected MacMPEC examples and two large-scale distillation cases. It turns out that the NCP-based approaches (BA/MLF/NCP) have advantages in dealing with biactive elements, because of the robustness in satisfying LICQ in the presence of nonempty biactive set; this is a potential benefit from the accumulation point of the derivatives of the NCP-functions. On the other hand, regularization methods LF and REG may have LICQ failure when biactive elements arise, with the phenomena of very large NLP multipliers and inaccurate solutions. Numerical studies of the large-scale cases also demonstrate that the two-side bounded formulations LF and MLF need the most iterations to converge; a possible cause is the challenge in determining the active set with vanishing ϵ. Instead, the BA and ordinary NCP formulations are the most efficient alternatives in these cases.

In this research, all the theoretical results on stationarities are developed from the MPCC-LICQ and NLP LICQ assumptions. Inspecting convergence properties of the NLP-based strategies with weaker constraint qualifications will be considered for future work. In addition, the asymptotic weak nondegeneracy condition is hard to enforce in practice; more practical conditions to characterize B-stationarity would be beneficial.

Acknowledgments. Support for this study was provided from an unrestricted grant from Carrier Corporation, and is gratefully acknowledged. We are also grateful to Dr. Alexandra Schwartz for her great help in MacMPEC/MATLAB interface, and to Dr. Helena Sofia Rodrigues and Dr. M. Teresa T. Monteiro for their kind help in AMPL/MATLAB interface.

References

1. Balakrishna, S., Biegler, L.T. Targetting Strategies for the Synthesis and Heat Integration of Nonisothermal Reactor Networks. *I & EC Research*, 31(9): 2152-2164 (1992)

2. Barton, P.I., Khan, K.A., Stechlinski, P., Watson, H.A.J. Computationally relevant generalized derivatives: Theory, evaluation and applications. *Optimization Methods and Software*, 33(4-6): 1030-1072 (2018)

3. Biegler, L.T. *Nonlinear Programming: Concepts, Algorithms and Applications to Chemical Processes*. SIAM, Philadelphia (2010)

4. Bonnans, J.F., Shapiro, A. *Perturbation Analysis of Optimization Problems*, Springer, New York (2000)

5. Byrd, R.H., Liu, G., Nocedal, J. On the local behavior of an interior point method for nonlinear programming, *In: Numerical Analysis 1997, Griffths D.F., Higham D.J. (eds)*. Addison-Wesley Longman, Harlow, UK, 37-56 (1997)

6. Chen, B., Harker, P.T. A non-interior-point continuation method for linear complementarity problems. *SIAM J. Matrix Analysis & Applics.*, 14(4): 1168-1190 (1993)

7. Chen, C., Mangasarian, O.L. Smoothing methods for convex inequalities and linear complementarity problems. *Mathematical Programming*, 71: 51-69 (1995)

8. Chen, C., Mangasarian, O.L. A class of smoothing functions for nonlinear and mixed complementarity problems. *Computational Optimization and Applications*, 5: 97-138 (1996)

9. Clarke, F.H. *Optimization and Nonsmooth Analysis*. Wiley, New York (1983)

10. DeMiguel, A.V., Friedlander, M.P., Nogales, F.J., Scholtes, S. A two-sided relaxation scheme for mathematical programs with equilibrium constraints. *SIAM J. Opt.*, 16(2): 587-609 (2005)

11. Dempe, S. *Foundations of Bilevel Programming, Nonconvex Optimization and Its Applications*. vol. 61. Kluwer, Dordrecht (2002)

12. Facchinei, F., Jiang, H., Qi, L. A smoothing method for mathematical programs with equilibrium constraints. *Mathematical Programming*, 85, 107-134 (1999)

13. Fiacco, A. *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming.* Academic Press, New York (1983)

14. Flegel, M.L., Kanzow, C. A direct proof for M-stationarity under MPEC-GCQ for mathematical programs with equilibrium constraints, *In: Optimization with Multivalued Mappings. Springer Optimization and Its Applications, vol 2, Dempe S., Kalashnikov V. (eds)*. Springer, Boston, 111-122 (2006)

15. Flegel, M.L., Kanzow, C. A Fritz John approach to first order optimality conditions for mathematical programs with equilibrium constraints. *Optimization*, 52(3): 277-286 (2003)

16. Fukushima, M., Pang, J-S. Convergence of a smoothing continuation method for mathematical programs with complementarity constraints, *In: Ill-posed Variational Problems and Regularization Techniques. Lecture Notes in Economics and Mathematical Systems, vol 477, Théra M., Tichatschke R. (eds)*. Springer, Berlin, Heidelberg, 99-110 (1999)

17. Griewank, A., Walther, A. First and second order optimality conditions for piecewise smooth objective functions. *Optimization Methods and Software*, 31(5): 904-930 (2016)

18. Griewank, A. On stable piecewise linearization and generalized algorithmic differentiation. *Optimization Methods and Software*, 28(6): 1139–1178 (2013)

19. Guo, L., Lin, G-H, Ye, J.J. Solving mathematical programs with equilibrium constraints. *J. Optimization Theory and Applications*, 166: 234-256 (2015)

20. Hegerhorst-Schultchen, L.C., Steinbach, M.C. On first and second order optimality conditions for abs-Normal NLP. *Optimization*, 69(12): 2629-2656 (2020)

21. Hegerhorst-Schultchen, L.C., Kirches, C., Steinbach, M.C. On the relation between MPCCs and optimization problems in abs-normal form. *Optimization Methods and Software*, 35(3): 560-575 (2020)

22. Hegerhorst-Schultchen, L.C. Optimality Conditions for Abs-Normal NLPs. Doctoral dissertation, Fakultät Mathematik und Physik, Universität Hannover (2020)

23. Hoheisel, T., Kanzow, C., Schwartz A. Theoretical and numerical comparison of relaxation methods for mathematical programs with complementarity constraints. *Mathematical Programming*, 137: 257-288 (2013)

24. Hoheisel, T., Kanzow, C., Schwartz A. Improved convergence properties of the Lin-Fukushima-Regularization method for mathematical programs with complementarity constraints. *Numerical Algebra, Control and Optimization*, 1(1): 49-60 (2011)

25. Jiang, H., Ralph, D. Smooth SQP methods for mathematical programs with nonlinear complementarity constraints. *SIAM J. Opt.*, 10(3): 779-808 (2000)

26. Kadrani, A., Dussault, J.P., Benchakroun, A. A new regularization scheme for mathematical programs with complementarity constraints. *SIAM J. Opt.*, 20(1): 78-103 (2009)

27. Kanzow, C., Schwartz, A. The price of inexactness: Convergence properties of relaxation methods for mathematical programs with complementarity constraints revisited. *Mathematics of Operations Research*, 40(2): 253-275 (2015)

28. Kanzow, C., Schwartz, A. Convergence properties of the inexact Lin-Fukushima relaxation method for mathematical programs with complementarity constraints. *Computational Optimization and Applications*, 59: 249-262 (2014)

29. Kanzow, C., Schwartz, A. A new regularization method for mathematical programs with complementarity constraints with strong convergence properties. *SIAM J. Opt.*, 23(2): 770-798 (2013)

30. Khan, K.A., Barton, P.I. A vector forward mode of automatic differentiation for generalized derivative evaluation. *Optimization Methods and Software*, 30(6): 1185-1212 (2015)

31. Lang, Y-D, Biegler, L.T. A distributed stream method for tray optimization. *AIChE Journal*, 48(3): 582-595 (2002)

32. Leyffer, S., López-Calva, G., Nocedal, J. Interior methods for mathematical programs with complementarity constraints. *SIAM J. Opt.*, 17(1): 52-77 (2006)

33. Leyffer, S. MacMPEC: AMPL collection of MPECs, https://wiki.mcs.anl.gov/leyffer/index.php/MacMPEC (2000)

34. Lin, G., Fukushima, M. Hybrid approach with active set identification for mathematical programs with complementarity constraints. *J. Optimization Theory and Applications*, 128(1): 1-28 (2006)

35. Lin, G., Fukushima, M. A modified relaxation scheme for mathematical programs with complementarity constraints. *Annals of Operations Research*, 133: 63-84 (2005)

36. Luo, Z-Q, Pang, J-S, Ralph, D. *Mathematical Programs with Equilibrium Constraints.* Cambridge University Press, Cambridge (1996)

37. Nocedal, J., Wright, S.J, *Numerical Optimization*, Springer, New York (2006)

38. Raghunathan, A., Biegler, L.T. Mathematical programs with equilibrium constraints (MPECs) in process engineering. *Comp. and Chem. Eng.*, 27(10): 1381-1392 (2003)

39. Ralph, D., Wright, S.J. Some properties of regularization and penalization schemes for MPECs. *Optimization Methods and Software*, 19(5): 527-556 (2004)

40. Scheel, H., Scholtes, S. Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. *Math. Operations Research*, 25(1), 1-22 (2000)

41. Scholtes, S. Convergence properties of a regularization scheme for mathematical programs with complementary constraints. *SIAM J. Opt.*, 11(4): 918-936 (2001)

42. Schwartz, A. Mathematical programs with complementarity constraints and related problems. Course Notes, Graduate School CE, Technische Universitatät Darmstadt (2018) (https://github.com/alexandrabschwartz/ Winterschool2018/blob/master/LectureNotes.pdf).

43. Stechlinski, P., Khan, K.A., Barton, P.I. Generalized sensitivity analysis of nonlinear programs. *SIAM J. Opt.*, 28(1): 272-301 (2018)

44. Steffensen, S., Ulbrich, M. A new regularization scheme for mathematical programs with equilibrium constraints. *SIAM J. Opt.*, 20(5): 2504-2539 (2010)

45. Veelken, S. A New Relaxation Scheme for Mathematical Programs with Equilibrium Constraints: Theory and Numerical Experience. Ph.D. dissertation, TU Munich, Germany (2009)

46. Watson, H.A.J., Khan, K.A., Barton, P.I. Multistream heat exchanger modeling and design. *AIChE Journal*, 61(10): 3390–3403 (2015)