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 then employ 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. With appropriate assumptions, the resulting solution of the proposed formulations is proved to be C- and M-stationary for the MPCC problems. Numerical performance of the proposed formulations, and the formulations by Lin & Fukushima and 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:

\[ c(x, y, q) = 0, \quad y_j = \max(0, x_j) \quad (j = 1, \ldots, n_x), \]

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

\[
\begin{align*}
\min & \quad f(x, y, q) \\
\text{s.t.} & \quad c(x, y, q) = 0 \\
& \quad y_j - \max(0, x_j) = 0 \quad (j = 1, \ldots, n_x),
\end{align*}
\]

where \( f : \mathbb{R}^{2n_x+n_q} \rightarrow \mathbb{R} \) and \( c : \mathbb{R}^{2n_x+n_q} \rightarrow \mathbb{R}^{n_c} \) are twice continuously differentiable functions. 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, 27, 40, 43] have developed a powerful conceptual framework for the solution of engineering models with nonsmooth elements. In addition to providing algorithmic differentiation tools [27] 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 nonlinear programming (NLP) algorithms [34] 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 and Walther [15] 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:
\[ c(z) = 0, \quad 0 \leq G(z) \perp H(z) \geq 0. \]  
(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
\[ (0 \leq y \perp y - x \geq 0) \equiv (y_j = \max(0, x_j), \ j = 1, \ldots, n_x), \]

shows the equivalence of (3) to (1), and also (2) to the mathematical program with complementarity constraints:
\[ \text{MPCC} : \min f(x, y, q) \]
\[ \text{s.t.} \quad c(x, y, q) = 0 \]
\[ 0 \leq y \perp y - x \geq 0. \]

The purpose of this paper is to develop NLP-based frameworks for the solution of large-scale, nonsmooth optimization problems. The approach is to substitute nonsmooth constraints by complementarity constraints and then apply constraint relaxations to develop NLP formulations, which are solved with NLP strategies. The resulting solution is analyzed for its stationarity to the MPCC problem, which is also a characterization of its optimality to the original nonsmooth problem.

The remainder of this section presents preliminary information on NLP properties as well as a review of NLP-based MPCC and smoothing approaches. Section 2 develops solution
strategies synthesized from constraint relaxation, NCP functions and bounding properties. This leads to a new bounding algorithm that derives from NCP-based MPCC formulations. Convergence properties are analyzed in Section 3. Numerical studies in Section 4 demonstrate the effectiveness of these approaches.

1.1 NLP Preliminaries

Consider the general NLP given by:

\[
\min_z f(z) \quad \text{s.t. } c(z,p) = 0, \quad g(z,p) \leq 0,
\]

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. At a feasible point \( z \), denote the set \( I_g(z) = \{j|g_j(z) = 0\} \). To characterize the solution of (5) we define its KKT point.

**Definition 1.1.** (KKT, [34]) Karush–Kuhn–Tucker (KKT) conditions for Problem (5) are given by:

\[
\nabla f(\bar{z}) + \nabla c(\bar{z})\bar{\lambda} + \nabla g(\bar{z})\bar{\mu} = 0
\]
\[
c(\bar{z}) = 0, \quad 0 \leq \bar{\mu} \perp g(\bar{z}) \leq 0
\]

for some multipliers \((\bar{\lambda}, \bar{\mu})\), where \( \bar{z} \) is a KKT point. We also define \( \mathcal{L} = f(z) + c(z)^T\lambda + g(z)^T\mu \) 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, [34]) The linear independence constraint qualification (LICQ) holds at \( \bar{z} \) when the gradient vectors

\[
\nabla c_i(\bar{z}), i = 1, \ldots, n_c \quad \text{and} \quad \nabla g_j(\bar{z}), \forall j \in I_g(\bar{z})
\]

are linearly independent. LICQ also implies that the multipliers \( \bar{\lambda}, \bar{\mu} \) are unique.

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

\[
d^T \nabla_{zz} \mathcal{L}(\bar{z}, \bar{\lambda}, \bar{\mu})d > 0
\]

for all \( d \neq 0 \), such that

\[
\nabla c_i(\bar{z})^Td = 0, \quad i = 1, \ldots, n_c
\]
\[
\nabla g_j(\bar{z})^Td = 0, \quad \text{for all } \bar{\mu}_j > 0 \text{ and } j \in I_g(\bar{z})
\]
\[
\nabla g_j(\bar{z})^Td \leq 0, \quad \text{for all } \bar{\mu}_j = 0 \text{ and } j \in I_g(\bar{z}).
\]

**Definition 1.4.** (Strict Complementarity, [13]) At a KKT point \((\bar{z}, \bar{\lambda}, \bar{\mu})\) of (5), the strict complementarity condition (SC) is defined by \( \bar{\mu}_j - g_j(\bar{z}) > 0 \) for each \( j \in I_g(\bar{z}) \).
In particular, a feasible point $z$ conditions for a (local) minimizer of an MPCC are described by the concept of the following linear program with complementarity constraints (LPCC):

Theorem 1.5. ([13]) Let $f, c, g$ in (5) be at least $\ell + 1$ times differentiable in $z$ and $\ell$ times differentiable in $p$. Then

- $\bar{z}(p_0)$ is an isolated minimizer, and the associated multipliers $\bar{\lambda}$ and $\bar{\mu}$ 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 $\ell$ times differentiable function $s(p) = (\bar{z}(p), \bar{\lambda}(p), \bar{\mu}(p))$, that corresponds to a locally unique minimum for (5);
- there exist finite Lipschitz constants $L_s, L_f > 0$, such that

\[
\|s(p) - s(p_0)\| \leq L_s\|p - p_0\| \quad \text{and} \quad |f(\bar{z}(p)) - f(\bar{z}(p_0))| \leq L_f\|p - p_0\|.
\]

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

\[
\begin{align*}
\min_d & \quad \nabla f(z^*)^T d \\
\text{s.t.} & \quad c(z^*) + \nabla c(z^*)^T d = 0 \\
& \quad 0 \leq y^* - d_y \perp (y^* + d_y) - (x^* + d_x) \geq 0.
\end{align*}
\]

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

\[
I_1(z^*) = \{j | y_j^* = 0\},
I_2(z^*) = \{j | y_j^* - x_j^* = 0\}.
\]

On the other hand, a feasible point $z^*$ with the existence of multipliers $\lambda^*, \sigma^*_1, \sigma^*_2$ that satisfy

\[
\nabla f(z^*) + \sum_{i=1}^{n_c} \lambda_i^* \nabla c_i(z^*) - \sum_{j \in I_1(z^*)} \sigma^*_1 \begin{bmatrix} 0 \\ e_j \\ 0 \end{bmatrix} - \sum_{j \in I_2(z^*)} \sigma^*_2 \begin{bmatrix} -e_j \\ e_j \\ 0 \end{bmatrix} = 0,
\]

where $e_j$ is a vector with the $j$th element being 1 and other elements being 0, only guarantees weak stationarity.

In addition to weak and $B$-stationarity, three stationarity concepts are considered. Provided that the weak stationarity (13) holds at $z^*$, then $z^*$ satisfies

- C-stationarity, if $\sigma^*_1 \sigma^*_2 \geq 0$ for all $j \in I_1(z^*) \cap I_2(z^*)$;
- M-stationarity, if $\sigma^*_1 \sigma^*_2 > 0$ or $\sigma^*_1 \sigma^*_2 = 0$ for all $j \in I_1(z^*) \cap I_2(z^*)$;
- strong stationarity (S-stationarity), if $\sigma^*_1, \sigma^*_2 \geq 0$ for all $j \in I_1(z^*) \cap I_2(z^*)$.

If MPCC-LICQ holds at $z^*$, namely, the following set of gradients is linearly independent:

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

4
then B-stationarity of $z^*$ can be established [37] based on its equivalence to strong stationarity, which is defined by $z^*$ that solves the relaxed NLP given by:

$$\text{RNLP} : \min f(x, y, q) \quad \text{s.t.} \quad c(x, y, q) = 0$$

$$y_j = 0, \ y_j - x_j > 0, \ j \in I_1(z^*) \setminus I_2(z^*)$$

$$y_j > 0, \ y_j - x_j = 0, \ j \in I_2(z^*) \setminus I_1(z^*)$$

$$y_j \geq 0, \ y_j \geq x_j, \ j \in I_1(z^*) \cap I_2(z^*)$$

(15)

1.2 Previous Work on NLP-based MPCC Solution

It is well-known that MPCC problems, for example (4), pose challenges for ordinary NLP algorithms to find a solution. The complementarity conditions (4c) introduce a combinatorial structure to the problem, and violate standard regularity assumptions on NLP constraints at every feasible point. Starting from seminal monographs on MPCCs [11, 33], a rich, comprehensive framework has been developed to characterize MPCC solutions and algorithmic strategies.

Fukushima and Pang [14] 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 [37] 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 [38] considers a sequence of stationary points of parametric NLPs which relax MPCC solutions, i.e.,

$$\text{REG}(t) : \min f(x, y, q) \quad \text{s.t.} \quad c(x, y, q) = 0$$

$$y \geq 0, \ y - x \geq 0$$

$$y_j(y_j - x_j) \leq t, \ j = 1, \ldots, n_x$$

with a vanishing positive $t$. 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 [14]. 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 [32] present a bounding scheme for MPCCs using two relaxations of complementarity constraints. This double relaxation 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
double relaxation MPCC approaches have been developed including a nonsmooth relaxation [23], a local-support approach [41] and the double-sided relaxation [10]. As with [32], these approaches also converge to C-stationary points, under reasonable assumptions.

Moreover, Guo et al. [16] 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 [36] 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 [20] 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 [30]. Kanzow and Schwartz [26] 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. [21] and Schwartz [39], generalize the definition of stationarity conditions for MPCCs, through relaxed constraint qualifications (i.e., MPCC-LICQ $\Rightarrow$ MPCC-MFCQ $\Rightarrow$ MPCC-GCQ) to define appropriate stationarity conditions. In particular, they showed that a local minimum that satisfies MPCC-LICQ also satisfies strong stationarity, while a local minimum that satisfies MPCC-GCQ also satisfies M-stationarity. Weaker necessary conditions with MPCC-GCQ include C-, A- and weak stationarity. Also, they show that the C-stationarity results of [38] and [32] hold under the weaker MPCC-MFCQ condition.

More recently, Hegerhorst-Schultchen and coworkers [17–19] considered abs-normal NLPs, developed in [15], which are given by:

$$\begin{align*}
\min f(x, |z|) \quad \text{s.t.} \quad c_E(x, |z|) &= 0, \quad c_I(x, |z|) \geq 0, \quad c_Z(x, |z|) - z = 0.
\end{align*}$$

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

$$\begin{align*}
\min f(x, u+v) \quad \text{s.t.} \quad c_E(x, u+v) &= 0, \quad c_I(x, u+v) \geq 0, \quad c_Z(x, u+v) - u - v = 0, \quad 0 \leq u \perp v \geq 0,
\end{align*}$$

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 [19], 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.

This study develops an approach for complementarity problems based on relaxation of so-called NCP-functions $\varphi$ for constraints (4c), which satisfy

$$\varphi(y_j, y_j - x_j) = 0 \quad \text{if and only if} \quad y_j \geq 0, y_j - x_j \geq 0, y_j(y_j - x_j) = 0. \quad (17)$$
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. Typical examples include the smoothed square root function

\[
\varphi_{i}^{\min}(y_j, y_j - x_j) = y_j + (y_j - x_j) - \sqrt{(y_j - (y_j - x_j))^2 + t^2},
\]

and the perturbed Fischer-Burmeister function

\[
\varphi_{i}^{\text{FB}}(y_j, y_j - x_j) = y_j + (y_j - x_j) - \sqrt{y_j^2 + (y_j - x_j)^2 + t^2},
\]

where the smoothing factor \(t \geq 0\), and \(t = 0\) recovers the property (17). NCP-functions are developed and applied to reformulate MPCCs and approximate the solution by solving a sequence of NLPs with \(t\) tending to zero \([12, 14, 22, 26, 31]\). The method proposed in this research is closely related to the regularization by Lin and Fukushima \([32]\), which is discussed in more detail in Section 2.2.

### 1.3 Smoothing the Max Operator

The max operator in (1) can be approximated by a variety of smoothed equations based on distribution density functions \(\delta(\xi), \xi \in \mathbb{R}\) \([8]\), which approximate a Dirac delta function. We 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

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

The shape of the density function can be parametrized by \(\epsilon\) and we define the parametrized density as:

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

the smoothed step function as:

\[
s(\xi, \epsilon) = \int_{-\infty}^{\xi} d(\xi, \epsilon) d\xi, \tag{20}
\]

and the smoothed max function as:

\[
m(\xi, \epsilon) = \int_{-\infty}^{\xi} s(\xi, \epsilon) d\xi. \tag{21}
\]

Function (21) satisfies the following properties.

**Proposition 1.7.** \([8, \text{Proposition 2.2}]\) Let Assumptions 1.6 hold, then:

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

\[ 0 \leq m(\xi, \epsilon) - \max(0, \xi) \leq \kappa \epsilon / 2, \quad \text{where} \quad \kappa = \int_{-\infty}^{0} |\xi| \delta(\xi) d\xi; \]
\[ m(\xi, \epsilon) > \xi; \]  
\[ 0 < m'(\xi, \epsilon) < 1. \]  

3. \( m(\xi, \epsilon) \) is strictly increasing with \( \xi \) 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 \( \phi_+'(\xi) < 0 \) and \( \phi_-'(\xi) > 0 \). 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 \( m(\xi) - \max(0, \xi) \) is absolutely continuous, it has its maximum value at \( \xi = 0 \). \( \square \)

In this study we consider two popular examples of \( m(\xi, \epsilon) \) that satisfy Proposition 1.7:

- The smoothed square root function [1,6] with
  \[ m(\xi, \epsilon) = \frac{1}{2}(\xi + \sqrt{\xi^2 + \epsilon^2}), \]  
  where \( \kappa = 1 \) and \( m'(0, \epsilon) = s(0, \epsilon) = 1/2 \).
- The neural network function [7] with
  \[ m(\xi, \epsilon) = \xi + \frac{\epsilon}{2} \log(1 + e^{-2\xi/\epsilon}), \]  
  where \( \kappa = \log 2 \) and \( m'(0, \epsilon) = s(0, \epsilon) = 1/2 \).

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

\[
\begin{align*}
\min & \quad f(x, y, q) \\
\text{s.t.} & \quad c(x, y, q) = 0 \\
& \quad y - h^\epsilon(x) = 0.
\end{align*}
\]

2 Solution Strategies

To solve (2) we develop a solution strategy that considers the errors of the smoothed function \( h^\epsilon(x) \), which is embedded within a modification of the smoothed NLP (25). This modification can be adapted to provide a local upper bound to (2). A relaxed smooth problem is also formulated, which provides a local lower bound to (2). These two smoothed problems are the basis for a strategy that solves sequence of problems with \( \epsilon \to 0 \). This section provides background concepts and properties for the smoothed NLPs and the resulting algorithm.
2.1 Smooth NLP Formulations for Problem (2)

The KKT conditions for (25) can be written as:

\[\nabla_x f(x, y, q) + \nabla_x c(x, y, q)\lambda - \nabla_x h'(x) u = 0\] (26a)

\[\nabla_y f(x, y, q) + \nabla_y c(x, y, q)\lambda + u = 0\] (26b)

\[\nabla_q f(x, y, q) + \nabla_q c(x, y, q)\lambda = 0\] (26c)

\[c(x, y, q) = 0, \quad y - h'(x) = 0\] (26d)

Assume that NLP (25) satisfies LICQ and SOSC at its (local) solution \(\bar{z} = (\bar{x}, \bar{y}, \bar{q})\). Then from Theorem 1.5, there exist solutions \(z(\epsilon)\) within an \(\epsilon\)-ball of \(\bar{z}\).

The conditions (26) also apply to the KKT conditions of the parametric program

\[
\begin{align*}
\text{min} & \quad f(x, y, q) \\
\text{s.t.} & \quad c(x, y, q) = 0 \\
& \quad y - h'(x) + p = 0,
\end{align*}
\] (27a)

with the addition of parameter \(p \in \mathbb{R}^{n_x}\) in the last equation. With \(\epsilon\) sufficiently small, then for all \(p_j \in [0, \kappa \epsilon/2]\) such that \(p_j = h'_j(x) - \max(0, x_j)\), Theorem 1.5 can be used to obtain sensitivity information with respect to \(p_j\), directly from the KKT conditions for (27).

In addition, consider the relaxed NLP given by:

\[
\begin{align*}
\text{min} & \quad f(x, y, q) \\
\text{s.t.} & \quad c(x, y, q) = 0 \\
& \quad -(\kappa \epsilon/2)e \leq y - h'(x) \leq 0,
\end{align*}
\] (28a)

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 provides a lower bound to (2). The KKT conditions for (28) are given by:

\[
\begin{align*}
\nabla_x f(x, y, q) + \nabla_x c(x, y, q)\lambda - \nabla_x h'(x)(u_U - u_L) = 0 \quad & \text{(29a)} \\
\nabla_y f(x, y, q) + \nabla_y c(x, y, q)\lambda + (u_U - u_L) = 0 \quad & \text{(29b)} \\
\nabla_q f(x, y, q) + \nabla_q c(x, y, q)\lambda = 0 \quad & \text{(29c)} \\
0 \leq u_L \perp y - h'(x) + \kappa \epsilon/2 \geq 0 \quad & \text{(29d)} \\
0 \leq u_U \perp y - h'(x) \leq 0, \quad & \text{(29e)}
\end{align*}
\]

where \(u\) 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 \(0 \leq u_L \perp u_U \geq 0\).

To derive \(\epsilon\)-bounds for the solution of (2), note that the solution of (27) with \(p_j = h'_j(x) - \max(0, x_j)\) is also feasible to (2), and hence forms an upper bound to (2). On the other hand, the solution of (2) is feasible to Problem (28), and hence (28) is a lower bound to (2). Moreover, problems (27) and (28) are closely related such that for \(\epsilon\) sufficiently small, we can obtain sensitivity corrections with respect to \(p\) at the solution of (27) that can provide approximate solutions to (28). In fact, for \(\epsilon\) sufficiently small, a (local)
solution of (27), with a sensitivity correction for $p$, will differ from the solution of (28) by $O(\epsilon^2)$ and bound the solution of (2). These observations can be summarized by the following proposition.

**Proposition 2.1.** Let LICQ and SOSC hold at the solutions of (25) and (27), and SC, LICQ and SOSC hold at the solution of (28). Then for $p_j \in [0, \kappa \epsilon/2]$ and $\epsilon$ sufficiently small, the following statements hold.

1. The solution of (25) along with a correction based on linearization of the KKT conditions provides an $O(\epsilon^2)$-approximate solution to (27) for $\epsilon$ suitably small.
2. Using the sensitivity corrections of (25) (or (27) with any $p_j \in [0, \kappa \epsilon/2]$), provides an $\epsilon^2$-approximate solution to (28).
3. Let $z(p)$ be a KKT point for (27), then the sensitivity of $f(z(p))$ to $p$, i.e., $\frac{df(z(p))}{dp}$ for (27), is given directly by $u(p)$.
4. The solution $z^+$ of (27) with $p_j^+ = h'_j(x^+)-\max(0, x^+_j)$ ($j = 1, \ldots, n_x$) satisfies $f(z^+) \geq f(z^*)$, where $z^*$ is the solution of (2).
5. The solution $z^-$ of (28) satisfies $f(z^+) \geq f(z^*) \geq f(z^-)$, where $z^+$ and $z^*$ are the solutions of (27) and (2), respectively.

**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 $\mathcal{L} = f(z) + c(z)^T\lambda + (y - h(x) + p)^T u$, 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:

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

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

$$\lim_{\|p\| \to 0} \left( \nabla_{sp} \mathcal{L}(s(0), 0)^T + \nabla_{ss} \mathcal{L}(s(0), 0) \frac{ds}{dp} \right) \frac{p}{\|p\|} = 0 \quad \Rightarrow \quad \frac{ds}{dp} = -\nabla_{ss} \mathcal{L}(s(0), 0)^{-1} \nabla_{sp} \mathcal{L}(s(0), 0), \quad (30)$$

and for all $p$,

$$\frac{ds}{dp} p = -\nabla_{ss} \mathcal{L}(s(0), 0)^{-1} \nabla_{sp} \mathcal{L}(s(0), 0) p = -\nabla_{ss} \mathcal{L}(s(0), 0)^{-1} \nabla_s \mathcal{L}(s(0), p). \quad (31)$$

Also since LICQ and SOSC hold, $\nabla_{ss} \mathcal{L}$ is nonsingular and bounded in $\epsilon$ neighborhood of solutions of (25) and (27). Applying Taylor’s theorem and using $\nabla_s \mathcal{L}(s(p), p) = 0$
leads to:
\[
s(0) + \frac{ds}{dp}^T p - s(p) = s(0) - s(p) - \nabla_{ss} \mathcal{L}(s(0), 0)^{-1} \nabla_s \mathcal{L}(s(0), p) \\
= \nabla_{ss} \mathcal{L}(s(0), 0)^{-1} [\nabla_{ss} \mathcal{L}(s(0), 0)(s(0) - s(p)) - (\nabla_s \mathcal{L}(s(0), p) - \nabla_s \mathcal{L}(s(p), p))] \\
= \nabla_{ss} \mathcal{L}(s(0), 0)^{-1} \int_0^1 (\nabla_{ss} \mathcal{L}(s(0), 0) - \nabla_{ss} \mathcal{L}(s(\tau p), p)(s(0) - s(p)))d\tau \\
= L_{inv} L\|s(0) - s(p)\|^2 = O(\|p\|^2) = O(\epsilon^2), \quad (32)
\]
where \( L \) is the Lipschitz constant for \( \nabla_{ss} \mathcal{L} \), while \( L_{inv} \) is the bound on the norm of its inverse; and the last line follows from the fourth item of Theorem 1.5 and from \( p_j \in [0, \kappa \epsilon/2] \).

2. From the solution of Problem (28) given by \((z^-, \lambda, u_L, u_U)\), define \( p^- = h'(x^-) - y^- \). To show that the solutions of (25) and (27) with sensitivity corrections are \( \epsilon^2 \)-approximate to (28), apply the result from Item 1 and note that \( \|s(0) + \frac{ds}{dp}^T p^- - s(p^-)\|^2 = O(\|p^-\|^2) = O(\epsilon^2) \) and \( \|s(p) + \frac{ds}{dp}^T (p^- - p) - s(p^-)\|^2 = O(\|p-p^-\|^2) = O(\epsilon^2) \).

3. Consider the Lagrange function for (27) at two parameter values \( p \neq p' \). Applying Taylor’s theorem with \( p(\tau) = p + \tau(p' - p) \) leads to:
\[
f(z(p')) - f(z(p)) = \mathcal{L}(s(p'), p') - \mathcal{L}(s(p), p) \\
= \int_0^1 \left( \nabla_p \mathcal{L}(s(p(\tau)), p(\tau))^T + \nabla_s \mathcal{L}(s(p(\tau)), p(\tau))^T \frac{ds}{dp}^T \right) (p' - p)d\tau \\
= \int_0^1 u(p(\tau))^T (p' - p)d\tau, \quad (33)
\]
where we used \( \nabla_s \mathcal{L}(s(p(\tau))) = 0 \) to deduce the last equality. As \( p' \to p \), we obtain that \( df(z(p))/dp = u(p) \).

4. Using \( p_j^+ = h_j'(x_j^+) - \max(0, x_j^+) \) in (27), \( z^+ \) is also a feasible point for (2). Moreover, since \( z^* \) is the optimal solution of (2), we have \( f(z^+) \geq f(z^*) \).

5. At the solution \( z^* \) of (2), \( y_j^* - \max(0, x_j^*) = y_j^* - h_j'(x_j^*) + p_j^* = 0 \) \( (j = 1, \ldots, n_x) \). Since \( y_j^* - h_j'(x_j^*) = -p_j^* \in [-\kappa \epsilon/2, 0] \), \( z^* \) is a feasible point for (28). Moreover, since \( z^- \) is the optimal solution of (28), we have \( f(z^+) \geq f(z^-) \). As a result, \( f(z^+) \geq f(z^*) \geq f(z^-) \).

\[
\square
\]

### 2.2 Relation to Lin-Fukushima Regularization Scheme

This section investigates the relation between (28) and the well-studied regularization scheme proposed by Lin and Fukushima in [32].

According to the method of [32], MPCC (4) is approximated by
\[
\text{LF}(t) : \min_{x, y, q} f(x, y, q) \quad \text{s.t.} \quad c(x, y, q) = 0, \quad (34a) \\
\Psi_{L,j}(z) = (y_j + t)(y_j - x_j + t) \geq t^2, \quad j = 1, \ldots, n_x, \quad (34b) \\
\Psi_{U,j}(z) = y_j(y_j - x_j) \leq t^2, \quad j = 1, \ldots, n_x, \quad (34c)
\]

\[
\]
where $t$ is a positive parameter. Suppose MPCC-LICQ (14) holds at a feasible point $z^*$ of MPCC (4). It has been proved that in a neighborhood $N$ of $z^*$, standard LICQ holds at every feasible point $z \in N$ of NLP (34) for $t$ sufficiently small.

With this constraint qualification, convergence results have been established for MPCC (4). In particular, for a sequence $\{t^k\}$ with $\lim_{k \to \infty} t^k = 0$, suppose the stationary points $\{z^k\}$ of (34) have an accumulation point $z^*$ where MPCC-LICQ holds, then $z^*$ is a C-stationary point of MPCC (4). Furthermore, if every $z^k$ meets additional second-order conditions

$$d^T \nabla z_k L_\Psi(z^k, \lambda^k, \mu^k_U, \mu^k_L)d \geq -\alpha \|d\|^2,$$

for the Lagrangian $L_\Psi = f(z^k) + c(z^k)^T \lambda^k - (\Psi_L(z^k) - (t^k)^2)^T \mu^k_L + (\Psi_U(z^k) - (t^k)^2)^T \mu^k_U$, some finite constant $\alpha > 0$, and all the directions $d$ in the set:

$$D(z^k) = \left\{ d \bigg| \begin{array}{l}
\nabla c_i(z^k)^T d = 0, \quad i = 1, \ldots, n_c \\
\n\nabla \Psi_{L,j}(z^k)^T d = 0, \quad \forall j \in I_{\Psi_L}(z^k, t^k) = \{ j \mid \Psi_{L,j}(z^k) = (t^k)^2 \} \\
\n\nabla \Psi_{U,j}(z^k)^T d = 0, \quad \forall j \in I_{\Psi_U}(z^k, t^k) = \{ j \mid \Psi_{U,j}(z^k) = (t^k)^2 \} 
\end{array} \right\},$$

then $z^*$ is a M-stationary point of MPCC (4).

In later studies, these convergence properties have been extended further. C-stationarity convergence is improved to hold under a weaker MPCC-MFCQ assumption on the limit point $z^*$ [21]. In addition, if the sequence of NLPs (34) are only solved approximately, C-stationarity of $z^*$ is also validated for such cases [24,25].

To investigate the relation between NLPs (28) and (34), consider $h_j^*(x)$ defined by the square root function (23). From the upper bound of (28c) we have

$$y_j - \frac{x_j + \sqrt{x_j^2 + \epsilon^2}}{2} \leq 0$$

$$\Leftrightarrow y_j + (y_j - x_j) \leq \sqrt{x_j^2 + \epsilon^2} = \sqrt{(y_j - (y_j - x_j))^2 + \epsilon^2}$$

$$\Leftrightarrow y_j(y_j - x_j) \leq \epsilon^2/4;$$

and from the lower bound we have

$$y_j - \frac{x_j + \sqrt{x_j^2 + \epsilon^2}}{2} \geq -\epsilon/2$$

$$\Leftrightarrow y_j + (y_j - x_j) + \epsilon \geq \sqrt{x_j^2 + \epsilon^2} = \sqrt{(y_j - (y_j - x_j))^2 + \epsilon^2}$$

$$\Leftrightarrow y_j(y_j - x_j) + \frac{\epsilon}{2}(y_j + (y_j - x_j)) \geq 0$$

$$\Leftrightarrow (y_j + \epsilon/2)(y_j - x_j + \epsilon/2) \geq \epsilon^2/4.$$  

With $\epsilon = 2t$, inequalities (36) and (37) are identical to $\Psi_{U,j}(z)$ and $\Psi_{L,j}(z)$, respectively, and leads to the same relaxation.

The equivalence between the relaxation by (28c) and (34c)(34d) means that the schemes for MPCCs derived from NLP (28) converge to solutions that are C- or M-stationary, under the same circumstances as for the schemes based on NLP (34). On the other hand, because the complementary elements $y_j$ and $y_j - x_j$ are not explicit in (28c), unlike in the regularization functions (34c)(34d), we cannot compare Lagrange gradients and Hessians of these two reformulations directly; these comparisons are needed for the test on M-stationarity.
2.3 Bounding Algorithm for Problem (2)

According to the properties in Proposition 2.1, for all parameters $0 \leq p_j, p'_j \leq (\kappa \epsilon / 2)(j = 1, \ldots, n_x)$ in (27) and the corresponding solutions $z(p)$ and $z(p')$, it is straightforward to show that:

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

(38)

Since the solution of (27) is equivalent to a solution of (28) with $p' = p^-$ (proof of Item 2, Proposition 2.1) and an upper bound to (2) when $p' = p^+$ (proof of Item 4, Proposition 2.1), the inequalities above can be generalized for any $p_j \in [0, \kappa \epsilon / 2]$ and $\epsilon$ sufficiently small, as:

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

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

(39)

To isolate the solution of (2), we apply the following procedure for the sequence $\epsilon^k \to 0$, and apply the lower and upper bounds in (39). By solving (27) with the generated values $p^k$, this approximates the solutions of (28), and finds an $\epsilon$-approximate solution of (2).

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

Specify $\epsilon_{tol} > 0, \gamma \in (0, 1)$; initialize $\epsilon^0 > 0, p^0 = 0$ and $z^0 = (x^0, y^0, p^0)$; set $k \leftarrow 0$.

(Optional) For some $\hat{\epsilon} < \epsilon^0$, solve a sequence of problems (25) with $\epsilon \to \hat{\epsilon}$. Then calculate $\hat{p}_j = \max(0, x_j) - y_j$; set $\epsilon^0 \leftarrow \hat{\epsilon}$ and $p^0 \leftarrow \hat{p}$.

For $\epsilon^k \geq \epsilon_{tol}$,

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

2. Since $p_j^k \in [0, \kappa \epsilon^k / 2]$, the upper bound of (28) is

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

3. Calculate an $O(\epsilon^2)$ approximation and lower bound to Problem (28). From the solution of (27),

   - let $J_0 = \{j | p_j^k = 0 \text{ and } u_j^k < 0\}$, setting $p_j^k$ to $\kappa \epsilon^k / 2$ for $j \in J_0$ would reduce the objective function of (27);
   - let $J_\epsilon = \{j | p_j^k = \kappa \epsilon^k / 2 \text{ and } u_j^k > 0\}$, setting $p_j^k = 0$ for $j \in J_\epsilon$ would reduce the objective function of (27);
   - The lower bound $f_{low} = f(z^k) - (\kappa \epsilon^k) \sum_{j \in J_0 \cup J_\epsilon} |u_j^k|$.

4. Set $\epsilon^{k+1} = \gamma \epsilon^k$, correct the parameters as $p_j^{k+1} = \begin{cases} \kappa \epsilon^{k+1} / 2, & j \in J_0; \\ 0, & j \in J_\epsilon; \\ \gamma p_j^k, & \text{otherwise}. \end{cases}$

Set $k \leftarrow k + 1$ and go to Step 1.
3 Convergence Analysis

The KKT conditions for NLPs (25), (27), and (28) have the same structure. This section analyzes convergence properties of the latter two. Note that all the analyses for (27) are directly applicable to (25), which can be viewed as a special case with $p^k$ invariably being zero.

To develop our convergence results, denote $\Phi_j^f(z) = y_j - h_j^f(x)$. At a feasible point $z$ of NLP (27), $\Phi_j^f(z) + p_j = 0$, where the value of $p_j$ is determined by the Bounding Algorithm. On the other hand, at the solution of NLP (28), we have $\Phi_j^f(z) + p_j = 0$, where $p_j = 0$ if $z$ locates on the upper bound of the feasible region, $p_j = \kappa \epsilon / 2$ if $z$ on the lower bound, and $p_j \in (0, \kappa \epsilon / 2)$ if $z$ in the interior.

Assume an infinite sequence of stationary points $z^k$ of NLP (27) or (28) is generated with $\epsilon^k$ tending to zero. We consider stationarity of the limit point $z^*$ for the MPCC (4), following the convergence analysis in [14] (which is related to Problem (25)). In Sections 3.1 and 3.2 we develop convergence results for the relaxation based on the smoothed square root function (23). In Section 3.3 we extend these results to the smoothed neural network function (24).

3.1 Stationary Conditions of Subproblems

For the square root function (23) with $\kappa = 1$, it follows from $\Phi_j^f(z) + p_j = 0$ that:

\[
2(y_j + p_j) - x_j = (x_j^2 + \epsilon^2)^{1/2} > 0 \quad (40a)
\]
\[
\implies (2(y_j + p_j) - x_j)^2 = x_j^2 + \epsilon^2 \quad (40b)
\]
\[
\implies (y_j + p_j)(y_j - x_j + p_j) = \epsilon^2 / 4. \quad (40c)
\]

This leads to the equivalence:

\[
\Phi_j^0(z) = 0 \iff y_j = 0 \text{ or } y_j - x_j = 0, \quad y_j(y_j - x_j) = 0,
\]
\[
\Phi_j^f(z) = 0 \iff y_j + p_j > 0, \quad y_j - x_j + p_j > 0, \quad (41)
\]

which follows by noting from (40c) 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 (40a) is violated.

To facilitate and generalize the analysis, we apply the MPCC notation in (3) and denote $G_j(z) = y_j, H_j(z) = y_j - x_j$. Then we have:

\[
\Phi_j^f(z) + p_j = \frac{1}{2} \left( G_j(z) + H_j(z) - \sqrt{(G_j(z) - H_j(z))^2 + \epsilon^2 + 2p_j} \right),
\]

\[
\nabla^G \Phi_j^f(z) = \frac{1}{2} - \frac{G_j(z) - H_j(z)}{2 \sqrt{(G_j(z) - H_j(z))^2 + \epsilon^2}},
\]

\[
\nabla^H \Phi_j^f(z) = \frac{1}{2} + \frac{G_j(z) - H_j(z)}{2 \sqrt{(G_j(z) - H_j(z))^2 + \epsilon^2}},
\]

\[
\nabla^{GG} \Phi_j^f(z) = \nabla^{HH} \Phi_j^f(z) = -\frac{2(G_j(z) + p_j)(H_j(z) + p_j)}{((G_j(z) - H_j(z))^2 + \epsilon^2)^{3/2}},
\]

\[
\nabla^{GH} \Phi_j^f(z) = \nabla^{HG} \Phi_j^f(z) = \frac{2(G_j(z) + p_j)(H_j(z) + p_j)}{((G_j(z) - H_j(z))^2 + \epsilon^2)^{3/2}}.
\]
At a point \( z \) such that \( \Phi_j^\epsilon(z) + p_j = 0 \), it follows from (40c) and (41) that
\[
\sqrt{(G_j(z) - H_j(z))^2 + \epsilon^2} = \sqrt{((G_j(z) + p_j) - (H_j(z) + p_j))^2 + \epsilon^2}
\]
\[
= \sqrt{(G_j(z) + p_j)^2 + (H_j(z) + p_j)^2 + 2(G_j(z) + p_j)(H_j(z) + p_j)}
\]
\[
= |G_j(z) + H_j(z) + 2p_j| = G_j(z) + H_j(z) + 2p_j.
\]
Thus, the above derivatives can be simplified to
\[
\nabla \Phi_j^\epsilon(z) = \begin{cases} \frac{H_j(z) + p_j}{G_j(z) + H_j(z) + 2p_j}, & G_j(z) + H_j(z) + 2p_j \\ \frac{G_j(z) + p_j}{G_j(z) + H_j(z) + 2p_j}, & 0 \leq G_j(z) + H_j(z) + 2p_j \end{cases}
\]
\[
\nabla \Phi_j^\epsilon(z) = \frac{-2(G_j(z) + p_j)(H_j(z) + p_j)}{(G_j(z) + H_j(z) + 2p_j)^3},
\]
\[
\nabla \Phi_j^\epsilon(z) = \frac{2(G_j(z) + p_j)(H_j(z) + p_j)}{(G_j(z) + H_j(z) + 2p_j)^3},
\]
and the gradient of \( \Phi_j^\epsilon(z) \) at a point \( \Phi_j^\epsilon(z) + p_j = 0 \) is given by
\[
\nabla \Phi_j^\epsilon(z) = \frac{H_j(z) + p_j}{G_j(z) + H_j(z) + 2p_j} \nabla G_j(z) + \frac{G_j(z) + p_j}{G_j(z) + H_j(z) + 2p_j} \nabla H_j(z)
\]
\[
= \frac{y_j - x_j + p_j}{2y_j - x_j + 2p_j} \begin{bmatrix} 0 \\ e_j \\ 0 \end{bmatrix} + \frac{y_j + p_j}{2y_j - x_j + 2p_j} \begin{bmatrix} -e_j \\ e_j \\ 0 \end{bmatrix}.
\]
For \( \epsilon = 0 \), at a feasible point \( z^* \) of the MPCC, with \( x_j^* = y_j^* = 0 \) \((j \in I_1(z^*) \cap I_2(z^*))\), the Clarke generalized gradient
\[
\partial \Phi_j^0(z^*) = \left\{ r \mid r = \lim_{k \to \infty} \nabla \Phi_j^0(z^k), \text{ with } z^k \to z^* \text{ and } \nabla \Phi_j^0(z^k) \text{ exist} \right\}
\]
is contained in the set
\[
\partial C \Phi_j^0(z^*) = \left\{ r \mid r = \xi_j \begin{bmatrix} 0 \\ e_j \\ 0 \end{bmatrix} + \eta_j \begin{bmatrix} -e_j \\ e_j \\ 0 \end{bmatrix}, \ (\xi_j, \eta_j) \in B \right\},
\]
where \( B = \left\{ (\xi_j, \eta_j) | (1 - \xi_j^2) + (1 - \eta_j^2) \leq 1 \right\} \). In other words, any accumulation point \( r^* \)
of \( \nabla \Phi_j^\epsilon(z) \) for \( j \in I_1(z^*) \cap I_2(z^*) \) can be represented by
\[
r^* = \xi_j \nabla G_j(z^*) + \eta_j \nabla H_j(z^*) = \xi_j \begin{bmatrix} 0 \\ e_j \\ 0 \end{bmatrix} + \eta_j \begin{bmatrix} -e_j \\ e_j \\ 0 \end{bmatrix},
\]
for some \( (\xi_j, \eta_j) \) satisfying \((1 - \xi_j^2) + (1 - \eta_j^2) \leq 1 \). This leads to the following result.
Theorem 3.1. Let $z$ be a feasible point of NLP (27) or (28) with smoothing factor $\epsilon$. Suppose $z \to z^*$ as $\epsilon \to 0$. Then $z^*$ is feasible for MPCC (4). Moreover, if MPCC-LICQ holds at $z^*$, then LICQ holds at every feasible point $z$ of NLP (27) or (28) for all $\epsilon > 0$ sufficiently small.

Proof. Feasibility of $z^*$ for MPCC (4) can be deduced from the continuity of $\Phi^\epsilon(z) + p$ for sufficiently small $\epsilon$ and its properties (41). Then it follows from the continuity of function $c$, and the gradients of $\Phi^\epsilon(z)$ given by (44) and (46), that

$$\lim_{\epsilon \to 0} \nabla c_i(z) = \nabla c_i(z^*), \quad i = 1, \ldots, n_c,$$

$$\lim_{\epsilon \to 0} \nabla \Phi^\epsilon_j(z) = \nabla G_j(z^*) = \begin{bmatrix} 0 \\ e_j \\ 0 \end{bmatrix}, \quad j \notin I_2(z^*),$$

$$\lim_{\epsilon \to 0} \nabla \Phi^\epsilon_j(z) = \nabla H_j(z^*) = \begin{bmatrix} -e_j \\ e_j \\ 0 \end{bmatrix}, \quad j \notin I_1(z^*),$$

$$\lim_{\epsilon \to 0} \nabla \Phi^\epsilon_j(z) = \xi_j^* \nabla G_j(z^*) + \eta_j^* \nabla H_j(z^*) = \xi_j^* \begin{bmatrix} 0 \\ e_j \\ 0 \end{bmatrix} + \eta_j^* \begin{bmatrix} -e_j \\ e_j \\ 0 \end{bmatrix}, \quad j \in I_1(z^*) \cap I_2(z^*),$$

(47)

where $(\xi_j^*, \eta_j^*) \in \mathcal{B}$. From MPCC-LICQ (14) at $z^*$, and the relations $I_1(z) \subseteq I_1(z^*)$ and $I_2(z) \subseteq I_2(z^*)$ for $z$ in the neighborhood of $z^*$, we obtain for NLP (27) that the gradients

$$\{\nabla c_i(z) | i = 1, \ldots, n_c\} \cup \{\nabla \Phi^\epsilon_j(z) | j = 1, \ldots, n_x\}$$

are linearly independent for $\epsilon$ small enough, and that

$$\sum_{i=1}^{n_c} \lambda_i \nabla c_i(z) + \sum_{j} u_j \nabla \Phi^\epsilon_j(z) = 0$$

(48)

can only be satisfied by $\lambda = 0, u = 0$.

To show this also holds for NLP (28) we consider

$$\sum_{i=1}^{n_c} \lambda_i \nabla c_i(z) - \sum_{j \in I_L(z,\epsilon)} u_{L,j} \nabla \Phi^\epsilon_j(z) + \sum_{j \in I_U(z,\epsilon)} u_{U,j} \nabla \Phi^\epsilon_j(z) = 0$$

(49)

and show that (49) can only be satisfied by $\lambda = 0, u_L = 0, u_U = 0$, where:

$$I_L(z,\epsilon) = \{j | \Phi^\epsilon_j(z) = -\epsilon/2\},$$

$$I_U(z,\epsilon) = \{j | \Phi^\epsilon_j(z) = 0\}$$
are the active sets of the inequality constraints. Equation (49) can be rewritten as:

\[
0 = \sum_{i=1}^{n_c} \lambda_i \nabla c_i(z) - \sum_{\substack{j \in I_L(z,\epsilon) \\ j \notin I_1(z^*) \cap I_2(z^*)}} u_{L,j} \left[ \frac{H_j(z) + p_j}{G_j(z) + H_j(z) + 2p_j} \right] \nabla G_j(z) + \frac{G_j(z) + p_j}{G_j(z) + H_j(z) + 2p_j} \nabla H_j(z) + \sum_{\substack{j \in I_U(z,\epsilon) \\ j \notin I_1(z^*) \cap I_2(z^*)}} u_{U,j} \left[ \frac{H_j(z)}{G_j(z) + H_j(z)} \right] \nabla G_j(z) + \frac{G_j(z)}{G_j(z) + H_j(z)} \nabla H_j(z) - \sum_{\substack{j \in I_L(z,\epsilon) \\ j \in I_1(z^*) \cap I_2(z^*)}} u_{L,j} \nabla \Phi_j^*(z) + \sum_{\substack{j \in I_U(z,\epsilon) \\ j \in I_1(z^*) \cap I_2(z^*)}} u_{U,j} \nabla \Phi_j^*(z),
\]

(50)

where \( \nabla \Phi_j^* \) in the last line approaches set (45) as \( \epsilon \to 0 \). Because MPCC-LICQ holds for (27), we compare \((\lambda, u)\) in (48) and \((\lambda, u_L, u_U)\) (49) for \( z \) in the neighborhood of \( z^* \). Noting that \( I_L(z,\epsilon) \cup I_U(z,\epsilon) \subseteq I_1(z^*) \cup I_2(z^*) \) we have:

\[
u_{L,j} \left[ \frac{H_j(z) + p_j}{G_j(z) + H_j(z) + 2p_j} \right] = 0, \quad u_{L,j} \left[ \frac{G_j(z) + p_j}{G_j(z) + H_j(z) + 2p_j} \right] = 0, \quad \lambda_i = 0, \quad i = 1, \ldots, n_c \]

\[
u_{U,j} \left[ \frac{H_j(z)}{G_j(z) + H_j(z)} \right] = 0, \quad u_{U,j} \left[ \frac{G_j(z)}{G_j(z) + H_j(z)} \right] = 0, \quad j \in I_U(z,\epsilon) \quad \text{and} \quad j \notin I_1(z^*) \cap I_2(z^*)
\]

\[
u_{L,j} \xi_j^* = 0, \quad u_{L,j} \eta_j^* = 0, \quad j \in I_L(z,\epsilon) \quad \text{and} \quad j \notin I_1(z^*) \cap I_2(z^*)
\]

\[
u_{U,j} \xi_j^* = 0, \quad u_{U,j} \eta_j^* = 0, \quad j \in I_U(z,\epsilon) \quad \text{and} \quad j \notin I_1(z^*) \cap I_2(z^*)
\]

Also, we note that (41) indicates for \( \epsilon > 0 \) that

\[
\frac{H_j(z) + p_j}{G_j(z) + H_j(z) + 2p_j} > 0, \quad \frac{G_j(z) + p_j}{G_j(z) + H_j(z) + 2p_j} > 0, \quad \text{for} \quad j \in I_L(z,\epsilon),
\]
\[
\frac{H_j(z)}{G_j(z) + H_j(z)} > 0, \quad \frac{G_j(z)}{G_j(z) + H_j(z)} > 0, \quad \text{for} \quad j \in I_U(z,\epsilon);
\]

and (45) indicates that \( \xi_j^* \) and \( \eta_j^* \) cannot both be zero as \( \epsilon \to 0 \), for \( j \in I_1(z^*) \cap I_2(z^*) \). Therefore, we can only have \( \lambda_i = u_{L,j} = u_{U,j} = 0 \) as the solution of (49). Hence, the gradients for NLP (28),

\[
\{ \nabla c_i(z) \mid i = 1, \ldots, n_c \} \cup \{ \nabla \Phi_j^*(z) \mid j \in I_L(z,\epsilon) \} \cup \{ \nabla \Phi_j^*(z) \mid j \in I_U(z,\epsilon) \}
\]

are linearly independent for \( \epsilon \) sufficiently small. \( \square \)

### 3.2 Stationarity Properties

This section builds on the relation between the MPCC multipliers and the multipliers at the NLP solutions. Based on this, the limit point of the stationary points of the NLP sequence is analyzed for its stationarity for MPCC (4).

**Theorem 3.2.** Let \( z^k \) be a stationary point of NLP (27) obtained with \( \epsilon = \epsilon^k \). Let \( z^k \to z^* \) as \( \epsilon^k \to 0 \) and suppose MPCC-LICQ holds at \( z^* \). Then \( z^* \) is a C-stationary point of MPCC (4).
Proof. Because $z^k$ is a stationary point of NLP (27), there exist multipliers $\lambda^k, u^k$ such that the KKT conditions (26) can be represented as follows:

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

Using (47), this system at $z^*$ becomes

$$
\nabla f(z^*) + \nabla c(z^*)\lambda^* + \sum_{j \notin I_2(z^*)} u_j^* \nabla G_j(z^*) + \sum_{j \notin I_1(z^*)} u_j^* \nabla H_j(z^*)
$$

$$
+ \sum_{j \in I_1(z^*) \cap I_2(z^*)} \left( u_j^* \xi_j \nabla G_j(z^*) + u_j^* \eta_j \nabla H_j(z^*) \right) = 0,
$$

where the multipliers are unique because LICQ holds at every $z^k$ with sufficiently small $\epsilon^k$ (Theorem 3.1). By setting

$$
\begin{align*}
\sigma^*_1 &= -u_j^*, & j \notin I_2(z^*), \\
\sigma^*_2 &= -u_j^*, & j \notin I_1(z^*), \\
\sigma^*_1 &= -u_j^* \xi_j, & j \in I_1(z^*) \cap I_2(z^*), \\
\sigma^*_2 &= -u_j^* \eta_j, & j \in I_1(z^*) \cap I_2(z^*),
\end{align*}
$$

(52)

the point $z^*$ satisfies the weak stationarity conditions (13). For the biactive set, since $(\xi^*_j, \eta^*_j) \in B$ (i.e., $\xi^*_j, \eta^*_j \geq 0$), then $\sigma^*_1 \sigma^*_2 = (u_j^*)^2 \xi^*_j \eta^*_j \geq 0$ and C-stationarity of $z^*$ can be deduced. □

We also note from (52) that if $u_j^* \leq 0$ for all $j \in I_1(z^*) \cap I_2(z^*)$ then $z^*$ is B-stationary, or equivalently, strongly stationary since MPCC-LICQ is assumed. Recall that in the Bounding Algorithm $u_j = u_{U,j} - u_{L,j}$ and $0 \leq u_{L,j} \perp u_{U,j} \geq 0$ as we discussed for (29). Therefore, $u_j^* \leq 0$ implies for the biactive set that the lower bounds of the relaxed NLP (28) are active, or, in other words, the parameters of NLP (27) have values $p_j = \epsilon/2$. This conclusion motivates Step 3 of the Bounding Algorithm, and indicates that in such circumstance the objective function has correct sensitivities with respect to the parameters associated with the biactive set.

On the other hand, proving C-stationarity of the limit point $z^*$ of $z^k$ generated by NLP (28) is not so direct as for NLP (27), because there may exist active $G_j(z^*)$ or $H_j(z^*)$ that lies in the interior of the feasible region of (28), for which the stationary conditions of the NLP do not reflect their influence on linear independence.

**Theorem 3.3.** Let $z^k$ be a stationary point of NLP (28) obtained with $\epsilon = \epsilon^k$. Let $z^k \to z^*$ as $\epsilon^k \to 0$ and suppose MPCC-LICQ holds at $z^*$. Then $z^*$ is a C-stationary point of MPCC (4).
\textbf{Proof.} It follows from the stationary conditions for NLP (28) at \( z^k \) that

\[
0 = \nabla f(z^k) + \nabla c(z^k) \lambda^k - \sum_{j \in I_L(z^k, e^k)} u_{L,j}^k \nabla \Phi_j^s(z^k) + \sum_{j \in I_U(z^k, e^k)} u_{U,j}^k \nabla \Phi_j^f(z^k) \\
= \nabla f(z^k) + \nabla c(z^k) \lambda^k \\
- \sum_{j \in I_L(z^k, e^k)} u_{L,j}^k \left[ \frac{H_j(z^k) + p_j^k}{G_j(z^k) + H_j(z^k) + 2p_j^k} \nabla G_j(z^k) + \frac{G_j(z^k) + p_j^k}{G_j(z^k) + H_j(z^k) + 2p_j^k} \nabla H_j(z^k) \right] \\
+ \sum_{j \in I_U(z^k, e^k)} u_{U,j}^k \left[ \frac{H_j(z^k)}{G_j(z^k) + H_j(z^k)} \nabla G_j(z^k) + \frac{G_j(z^k)}{G_j(z^k) + H_j(z^k)} \nabla H_j(z^k) \right].
\]

Defining the multipliers as:

\[
\sigma_{1j}^k = u_{L,j}^k \left[ \frac{H_j(z^k) + p_j^k}{G_j(z^k) + H_j(z^k) + 2p_j^k} \right], \quad j \in I_L(z^k, e^k) \text{ and } j \in I_1(z^*) \\
\sigma_{2j}^k = u_{U,j}^k \left[ \frac{G_j(z^k) + p_j^k}{G_j(z^k) + H_j(z^k) + 2p_j^k} \right], \quad j \in I_2(z^*) \\
\sigma_{1j}^k = -u_{L,j}^k \left[ \frac{H_j(z^k)}{G_j(z^k) + H_j(z^k)} \right], \quad j \in I_1(z^k, e^k) \text{ and } j \in I_1(z^*) \quad (53) \\
\sigma_{2j}^k = -u_{U,j}^k \left[ \frac{G_j(z^k)}{G_j(z^k) + H_j(z^k)} \right], \quad j \in I_2(z^k, e^k) \text{ and } j \in I_2(z^*).
\]

we rewrite the above conditions as

\[
0 = \nabla f(z^k) + \sum_{i=1}^{n_c} \lambda_i^k \nabla c_i(z^k) \\
- \sum_{j \in I_L(z^k, e^k)} \sigma_{1j}^k \left[ \nabla G_j(z^k) + \frac{G_j(z^k) + p_j^k}{H_j(z^k) + H_j(z^k) + 2p_j^k} \nabla H_j(z^k) \right] - \sum_{j \in I_L(z^k, e^k)} \sigma_{2j}^k \left[ \nabla H_j(z^k) + \frac{H_j(z^k) + p_j^k}{G_j(z^k) + H_j(z^k) + 2p_j^k} \nabla G_j(z^k) \right] \\
- \sum_{j \in I_L(z^k, e^k)} \sigma_{1j}^k \nabla G_j(z^k) + \sum_{j \in I_1(z^*) \cap I_2(z^*)} \sigma_{1j}^k \nabla G_j(z^k) + \sigma_{2j}^k \nabla H_j(z^k) \\
- \sum_{j \in I_U(z^k, e^k)} \sigma_{1j}^k \left[ \nabla G_j(z^k) + \frac{G_j(z^k)}{H_j(z^k)} \nabla H_j(z^k) \right] - \sum_{j \in I_U(z^k, e^k)} \sigma_{2j}^k \left[ \nabla H_j(z^k) + \frac{H_j(z^k)}{G_j(z^k)} \nabla G_j(z^k) \right] \\
- \sum_{j \in I_U(z^k, e^k)} \sigma_{1j}^k \nabla G_j(z^k) + \sum_{j \in I_1(z^*) \cap I_2(z^*)} \sigma_{1j}^k \nabla G_j(z^k) + \sigma_{2j}^k \nabla H_j(z^k), \\
(54)
\]

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

\[
\nabla f(z^k) + \bar{A}(z^k) \bar{\alpha}^k = \nabla f(z^k) + \begin{bmatrix} \frac{\lambda_i^k}{j \in I_L(z^k, e^k) \cup I_U(z^k, e^k)} \frac{\nabla G_j(z^k)}{j \in I_1(z^*)} \frac{-\nabla H_j(z^k)}{j \in I_2(z^*)} \end{bmatrix} \begin{bmatrix} \alpha^k \\ 0 \\ 0 \end{bmatrix} = 0.
\]
In the limit, the matrix $\bar{A}(z^k)$ converges to matrix $\bar{A}(z^*)$ with columns
\[
\{\nabla c_i(z^*) \mid i = 1, \ldots, n_c\} \cup \{-\nabla G_j(z^*) \mid j \in I_1(z^*)\} \cup \{-\nabla H_j(z^*) \mid j \in I_2(z^*)\},
\] (55)
and $\bar{\alpha}^k$ converges to vector $\bar{\alpha}^*$, whose elements are the multipliers (refer to (53))
\[
\lambda_{i}^{*} = \lim_{k \to 0} \lambda_{i}^{k}, \quad i = 1, \ldots, n_x,
\]
\[
\sigma_{1j}^{*} = \begin{cases} 
 u_{L,j}^{*} \geq 0, & j \in I_{L}^{0} \text{ and } j \in I_{1}(z^{*}) \setminus I_{2}(z^{*}) \\
 u_{L,j}^{*} \xi_{j}^{*} \geq 0, & j \in I_{L}^{0} \text{ and } j \in I_{1}(z^{*}) \cap I_{2}(z^{*}) \\
 -u_{U,j}^{*} \leq 0, & j \in I_{U}^{0} \text{ and } j \in I_{1}(z^{*}) \setminus I_{2}(z^{*}) \\
 -u_{U,j}^{*} \xi_{j}^{*} \leq 0, & j \in I_{U}^{0} \text{ and } j \in I_{1}(z^{*}) \cap I_{2}(z^{*}) \\
 0, & j \notin (I_{L}^{0} \cup I_{U}^{0}) \text{ and } j \in I_{1}(z^{*}),
\end{cases}
\]
\[
\sigma_{2j}^{*} = \begin{cases} 
 u_{L,j}^{*} \geq 0, & j \in I_{L}^{0} \text{ and } j \in I_{2}(z^{*}) \setminus I_{1}(z^{*}) \\
 u_{L,j}^{*} \eta_{j}^{*} \geq 0, & j \in I_{L}^{0} \text{ and } j \in I_{1}(z^{*}) \cap I_{2}(z^{*}) \\
 -u_{U,j}^{*} \leq 0, & j \in I_{U}^{0} \text{ and } j \in I_{2}(z^{*}) \setminus I_{1}(z^{*}) \\
 -u_{U,j}^{*} \eta_{j}^{*} \leq 0, & j \in I_{U}^{0} \text{ and } j \in I_{1}(z^{*}) \cap I_{2}(z^{*}) \\
 0, & j \notin (I_{L}^{0} \cup I_{U}^{0}) \text{ and } j \in I_{2}(z^{*}),
\end{cases}
\]
where $(\xi_{j}^{*}, \eta_{j}^{*}) \in \mathcal{B}$, and
\[
I_{L}^{0} = \{j \mid j \in I_{L}(z^{k}, e^{k}) \text{ for infinitely many } k\},
\]
\[
I_{U}^{0} = \{j \mid j \in I_{U}(z^{k}, e^{k}) \text{ for infinitely many } k\}.
\]
The multipliers in (56) are unique because $\bar{A}(z^*)$ has full column rank by assumption. Then $z^*$ is C-stationary because
\[
\sigma_{1j}^{*} \sigma_{2j}^{*} = \begin{cases} 
 (u_{L,j}^{*})^{2} \xi_{j}^{*} \eta_{j}^{*} \geq 0, & j \in I_{L}^{0} \text{ and } j \in I_{1}(z^{*}) \cap I_{2}(z^{*}) \\
 (u_{U,j}^{*})^{2} \xi_{j}^{*} \eta_{j}^{*} \geq 0, & j \in I_{U}^{0} \text{ and } j \in I_{1}(z^{*}) \cap I_{2}(z^{*}) \\
 0, & j \notin (I_{L}^{0} \cup I_{U}^{0}) \text{ and } j \in I_{1}(z^{*}) \cap I_{2}(z^{*}).
\end{cases}
\]
\[
\square
\]
From (56) we observe that if the solutions of (28) have no nonzero multipliers $u_{U,j}^{*}$ with $j \in I_{U}^{0}(z) \cap I_{1}(z^{*}) \cap I_{2}(z^{*})$, then $z^*$ is strongly stationary, or equivalently, B-stationary. This observation is consistent with that of Theorem 3.2.

To explore stronger results beyond C-stationarity, we consider the second-order conditions at $z^k$. Define the Lagrangian for NLP (27) as $\mathcal{L}_{\Phi}(z, \lambda, u) = f(z) + c(z)^{T}\lambda + (\Phi^{c}(z) + p)^{T}u$. The Hessian of $\mathcal{L}_{\Phi}$ is given by
\[
\nabla_{zz} \mathcal{L}_{\Phi}(z, \lambda, u) = \nabla_{zz} f(z) + \sum_{i=1}^{n_c} \lambda_i \nabla_{zz} c_i(z) + \sum_{j=1}^{n_x} u_j \nabla_{zz} \Phi^c_j(z).
\]
(57)
We say that the reduced Hessian of the Lagrangian at $z^\alpha$ is bounded below, if
\[ d^T \nabla_{zz} \mathcal{L}_\alpha(z^\alpha, \lambda^k, u^k) d \geq -\alpha \|d\|^2, \quad \forall d \in \mathcal{T}(z^k), \tag{58} \]
where the constant $\alpha > 0$ is bounded and $d$ lies in the tangent space of the constraints, i.e.,
\[ \mathcal{T}(z^k) = \left\{ d \mid \nabla c_i(z^k)^T d = 0, \quad i = 1, \ldots, n_c \right\}. \tag{59} \]
Note that condition (58) is a little weaker than the standard second-order necessary conditions (8), because of the negative right-hand side.

**Theorem 3.4.** Let $z^k$ be a stationary point of NLP (27) tending to $z^\ast$. In addition to the assumptions for Theorem 3.2, suppose the reduced Hessian of the Lagrangian at each $z^k$ is bounded below in the sense of (58). Then $z^\ast$ is an M-stationary point of MPCC (4).

**Proof.** The proof is similar to the proof of Theorem 3.4 in [32]. For the purpose of deriving a contradiction, suppose $z^\ast$ is not M-stationary. It follows from Theorem 3.2 that there exists some index $j_0 \in I_1(z^\ast) \cap I_2(z^\ast)$ such that
\[ \sigma_{1j_0}^* = -u^*_{j_0} \xi_{j_0}^* < 0, \]
\[ \sigma_{2j_0}^* = -u^*_{j_0} \eta_{j_0}^* < 0. \tag{60} \]
This implies $u^*_{j_0}, \xi_{j_0}^*, \eta_{j_0}^* > 0$, since $(\xi_{j_0}^*, \eta_{j_0}^*) \in \mathcal{B}$. We can choose a direction $d^k$ such that
\[ \nabla c_i(z^k)^T d^k = 0, \quad i = 1, \ldots, n_c, \]
\[ \nabla G_j(z^k)^T d^k = 0, \quad j \in I_1(z^\ast), j \neq j_0, \]
\[ \nabla H_j(z^k)^T d^k = 0, \quad j \in I_2(z^\ast), j \neq j_0, \]
\[ \nabla G_{j_0}(z^k)^T d^k = d_G = \nabla H \Phi^e_{j_0}(z^k), \]
\[ \nabla H_{j_0}(z^k)^T d^k = d_H = -\nabla G \Phi^e_{j_0}(z^k). \tag{61} \]

The direction $d^k$ is well-defined because the MPCC-LICQ at $z^\ast$ guarantees the coefficient vectors of $d^k$ are linearly independent for $\epsilon^k > 0$ sufficiently small, and the right-hand sides of the last two equations in (61) tend to $\eta^*_{j_0}$ and $-\xi^*_{j_0}$ in the limit. Note that $d^k \in \mathcal{T}(z^k)$.

The contribution from function $\Phi^e_{j_0}(z^k) + p^e_j$ to the reduced Hessian of the Lagrangian is
\[ u^e_{j_0}(d^k)^T \nabla_{zz} \Phi^e_{j_0}(z^k) d^k = \]
\[ u^e_{j_0}(d^k)^T \left[ \nabla_G \Phi^e_{j_0}(z^k) \nabla_{zz} G_j(z^k) + \nabla_H \Phi^e_{j_0}(z^k) \nabla_{zz} H_j(z^k) \right. \]
\[ + \nabla_G G \Phi^e_{j_0}(z^k) \nabla G_j(z^k) \nabla G_j(z^k)^T + \nabla_G H \Phi^e_{j_0}(z^k) \nabla G_j(z^k)^T \]
\[ + \nabla_H G \Phi^e_{j_0}(z^k) \nabla H_j(z^k) + \nabla_H H \Phi^e_{j_0}(z^k) \nabla H_j(z^k)^T d^k. \tag{62} \]

For MPCC (4), $\nabla_{zz} G_j(z^k) = \nabla_{zz} H_j(z^k) = 0$. According to the definition of $d^k$, the remaining terms of (62) for $j \neq j_0$ are all zeros, while for the index $j_0$ the following can be derived from
(43):

\[
u_{j_0}(d^k)^T \nabla_{zz} \Phi_{j_0}^\epsilon(z^k) d^k = u_{j_0}^k \left[ \nabla_{GG} \Phi_{j_0}^\epsilon(z^k) d_G^2 + 2 \nabla_{GH} \Phi_{j_0}^\epsilon(z^k) d_G d_H + \nabla_{HH} \Phi_{j_0}^\epsilon(z^k) d_H^2 \right] = -2u_{j_0}^k (G_{j_0}(z^k) + p^k_{j_0})(H_{j_0}(z^k) + p^k_{j_0}) \left( (G_{j_0}(z^k) + H_{j_0}(z^k) + 2p^k_{j_0})^2 \right) = -2u_{j_0}^k (G_{j_0}(z^k) + p^k_{j_0})(H_{j_0}(z^k) + p^k_{j_0}) = -2u_{j_0}^k \nabla_{G} \Phi_{j_0}^\epsilon(z^k) \nabla_{H} \Phi_{j_0}^\epsilon(z^k).
\]

As \( \epsilon_k \to 0 \),

\[
u_{j_0}(d^k)^T \nabla_{zz} \Phi_{j_0}^\epsilon(z^k) d^k \to -\infty,
\]

because \( u_{j_0}^k, \nabla_{G} \Phi_{j_0}^\epsilon(z^k), \nabla_{H} \Phi_{j_0}^\epsilon(z^k) \) converge to \( u_{j_0}^*, \xi_{j_0}^*, \eta_{j_0}^* \), which are positive and bounded, while \( G_{j_0}(z^k), H_{j_0}(z^k) \) and \( p^k_{j_0} \) tend to zero. Since all other terms in

\[ (d^k)^T \nabla_{zz} \mathcal{L}_\Phi(z^k, \lambda^k, u^k) d^k \]

are bounded, (64) leads the reduced Hessian of the Lagrangian to \(-\infty\), which contradicts (58). Therefore \( z^* \) is M-stationary.

To further investigate stationarity for the stationary points of NLP (28), we redefine the Lagrangian as \( \mathcal{L}_\Phi(z, \lambda, u_L, u_U) = f(z) + c(z)^T \lambda - \sum_{j \in I_L(z, \epsilon)} \Phi_j^\epsilon(z) u_{L,j} + \sum_{j \in I_U(z, \epsilon)} \Phi_j^\epsilon(z) u_{U,j} \). Thus the Hessian of \( \mathcal{L}_\Phi \) is given by

\[
\nabla_{zz} \mathcal{L}_\Phi(z, \lambda, u_L, u_U) = \nabla_{zz} f(z) + \sum_{i=1}^{n_c} \lambda_i \nabla_{zz} c_i(z) - \sum_{j \in I_L(z, \epsilon)} u_{L,j} \nabla_{zz} \Phi_j^\epsilon(z) + \sum_{j \in I_U(z, \epsilon)} u_{U,j} \nabla_{zz} \Phi_j^\epsilon(z).
\]

Similar to (35), the bounded below condition for the reduced Hessian of the Lagrangian at \( z^k \) is restated as

\[
w^T \nabla_{zz} \mathcal{L}_\Phi(z^k, \lambda^k, u^k_L, u^k_U) w \geq -\alpha \|w\|^2, \ \forall w \in \mathcal{C}(z^k),
\]

where the constant \( \alpha > 0 \) is bounded, and the set \( \mathcal{C} \) is defined as

\[
\mathcal{C}(z^k) = \left\{ w \mid \nabla c_i(z^k)^T w = 0, \quad i = 1, \ldots, n_c \right\} \cup \left\{ w \mid \nabla \Phi_j^\epsilon(z^k)^T w = 0, \quad \forall j \in I^0_L \cup I^0_U \right\}.
\]

Note that set \( \mathcal{C}(z^k) \) does not involve those \( w \) that direct to the interior of the feasible region, i.e. \( \nabla \Phi_j^\epsilon(z^k)^T w > 0 \) for \( j \in I^0_L \) with \( u_{L,j}^k = 0 \), or \( \nabla \Phi_j^\epsilon(z^k)^T w < 0 \) for \( j \in I^0_U \) with \( u_{U,j}^k = 0 \). Instead, only those \( w \) that adhere to the active inequality constraints for sufficiently small \( \epsilon \) need to be considered; otherwise, inactive constraint terms in the Lagrangian vanish as \( \epsilon \to 0 \).

**Theorem 3.5.** Let \( z^k \) be a stationary point of NLP (28) tending to \( z^* \). In addition to the assumptions for Theorem 3.3, suppose the reduced Hessian of the Lagrangian at each \( z^k \) is bounded below in the sense of (65). Then \( z^* \) is an M-stationary point of MPCC (4).
Proof. The following shows $z^*$ is M-stationary, by adapting the proof of Theorem 3.4 for (28). Suppose $z^*$ is not M-stationary. It follows from Theorem 3.3 that there exists some index $j_0 \in I_1(\mathbf{z}^*) \cap I_2(\mathbf{z}^*)$ such that $\sigma^*_{j_0} < 0$ and $\sigma^*_{2j_0} < 0$. As with (56), we choose $j_0 \in I^0_U \cap I_1(\mathbf{z}^*) \cap I_2(\mathbf{z}^*)$ and a direction $w^k$ such that

\[
\nabla c_i(z^k)T w^k = 0, \ i = 1, \ldots, n_c, \\
\nabla G_j(z^k)T w^k = 0, \ j \in I_1(\mathbf{z}^*), j \neq j_0, \\
\nabla H_j(z^k)T w^k = 0, \ j \in I_2(\mathbf{z}^*), j \neq j_0, \\
\n\nabla G_{j_0}(z^k)T w^k = w_G = \nabla H_{j_0}(z^k), \\
\nabla H_{j_0}(z^k)T w^k = w_H = -\nabla c_i\Phi_{j_0}(z^k).
\]

The direction $w^k$ is well-defined due to the MPCC-LICQ assumption. Also note that $w^k \in C(z^k)$ since $I^0_L \cup I^0_U \subseteq I_1(\mathbf{z}^*) \cup I_2(\mathbf{z}^*)$. It is easy to see that $(w^k)^T \nabla_{zz} \Phi_{j_0}(z^k)w^k = 0$ for all $j \in I^0_L$ and $j \in I^0_U \setminus \{j_0\}$. For index $j_0$, we have that

\[
u_{U,j_0}^k (w^k)^T \nabla_{zz} \Phi_{j_0}(z^k)w^k \\
= u_{U,j_0}^k \left[ \nabla_G \Phi_{j_0}^e (z^k)w^2_G + 2\nabla_{GH} \Phi_{j_0}^e (z^k)w_G w_H + \nabla_{HH} \Phi_{j_0}^e (z^k)w^2_H \right] \\
= -2u_{U,j_0}^k \frac{G_{j_0}(z^k)w_G}{(G_{j_0}(z^k) + H_{j_0}(z^k))^3} (w_G - w_H)^2 \\
= -2u_{U,j_0}^k \frac{G_{j_0}(z^k)H_{j_0}(z^k)}{(G_{j_0}(z^k) + H_{j_0}(z^k))^3} \\
= -2u_{U,j_0}^k \frac{\nabla_G \Phi_{j_0}^e (z^k) \nabla_H \Phi_{j_0}^e (z^k)}{G_{j_0}(z^k) + H_{j_0}(z^k)} \rightarrow -\infty,
\]

because $u_{U,j_0}^k, \nabla_G \Phi_{j_0}^e (z^k), \nabla_H \Phi_{j_0}^e (z^k)$ 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. Then we have

$(w^k)^T \nabla_{zz} \mathcal{L}_{\Phi}(z^k, \lambda^k, u_L^k, u_U^k)w^k \rightarrow -\infty$ as well, since all other terms except (68) in the reduced Hessian are bounded. This contradicts (65), and therefore $z^*$ is M-stationary. \hfill \Box

### 3.3 Extension to Neural Network Functions

This section extends the convergence results for the reformulation with the square root function to that with the neural network function. For simplicity, the following only presents outlines of the extension based on NLP (27); the results for NLP (28) can be extended similarly. For the neural network function (24), the smoothed function and its gradients are
given by
\[
\Phi_j^\epsilon(z) + p_j = H_j(z) - \frac{\epsilon}{2} \log(1 + e^{-2(G_j(z) - H_j(z))/\epsilon}) + p_j,
\]
\[
\nabla_G \Phi_j^\epsilon(z) = \frac{e^{-2(G_j(z) - H_j(z))/\epsilon}}{1 + e^{-2(G_j(z) - H_j(z))/\epsilon}},
\]
\[
\nabla_H \Phi_j^\epsilon(z) = \frac{1}{1 + e^{-2(G_j(z) - H_j(z))/\epsilon}},
\]
\[
\nabla_{GG} \Phi_j^\epsilon(z) = \nabla_{HH} \Phi_j^\epsilon(z) = \frac{-2e^{-2(G_j(z) - H_j(z))/\epsilon}}{\epsilon(1 + e^{-2(G_j(z) - H_j(z))/\epsilon})^2},
\]
\[
\nabla_{GH} \Phi_j^\epsilon(z) = \nabla_{HG} \Phi_j^\epsilon(z) = \frac{2e^{-2(G_j(z) - H_j(z))/\epsilon}}{\epsilon(1 + e^{-2(G_j(z) - H_j(z))/\epsilon})^2}.
\]

It follows from Item 2, Proposition 1.7 that the smoothed NCP-function \( \Phi_j^\epsilon \) satisfies [8, Proposition 3.3]:
\[
\max(0, -G_j(z)) \leq \epsilon \log 2/2, \quad \max(0, -H_j(z)) \leq \epsilon \log 2/2, \quad \max(0, G_j(z)H_j(z)) \leq \epsilon^2/2.
\]

At a point \( z \) such that \( \Phi_j^\epsilon(z) + p_j = 0 \), we have
\[
e^{2(H_j(z)+p_j)/\epsilon} = 1 + e^{-2(G_j(z)-H_j(z))/\epsilon}.
\]

Substituting (71) into (69), we can simplify the derivatives to
\[
\nabla_G \Phi_j^\epsilon(z) = e^{-2(G_j(z)+p_j)/\epsilon},
\]
\[
\nabla_H \Phi_j^\epsilon(z) = e^{-2(H_j(z)+p_j)/\epsilon},
\]
\[
\nabla_{GG} \Phi_j^\epsilon(z) = \nabla_{HH} \Phi_j^\epsilon(z) = \frac{-2e^{-2(G_j(z)+H_j(z)+2p_j)/\epsilon}}{\epsilon},
\]
\[
\nabla_{GH} \Phi_j^\epsilon(z) = \nabla_{HG} \Phi_j^\epsilon(z) = \frac{2e^{-2(G_j(z)+H_j(z)+2p_j)/\epsilon}}{\epsilon}.
\]

The proof for constraint qualification (Theorem 3.1) can be easily applied to a feasible point \( z \) of the NLP (27) generated from the smoothed neural network function. Here, we have
\[
\nabla \Phi_j(z) = \nabla_G \Phi_j^\epsilon \nabla_G(z) + \nabla_H \Phi_j^\epsilon \nabla_H(z) = e^{-2(G_j(z)+p_j)/\epsilon} \begin{bmatrix} 0 \\ e_j \\ 0 \end{bmatrix} + e^{-2(H_j(z)+p_j)/\epsilon} \begin{bmatrix} -e_j \\ e_j \\ 0 \end{bmatrix}.
\]

Multiplying both sides of (71) with \( e^{-2(H_j(z)+p_j)/\epsilon} \) leads to
\[
e^{-2(G_j(z)+p_j)/\epsilon} + e^{-2(H_j(z)+p_j)/\epsilon} = 1,
\]
which implies with (72) that \( \lim_{\epsilon \to 0} \nabla_G \Phi_j^\epsilon(z) = 0 \) and \( \lim_{\epsilon \to 0} \nabla_H \Phi_j^\epsilon(z) = 1 \) for all \( j \notin I_1(z^*) \), and \( \lim_{\epsilon \to 0} \nabla_G \Phi_j^\epsilon(z) = 1 \) and \( \lim_{\epsilon \to 0} \nabla_H \Phi_j^\epsilon(z) = 0 \) for all \( j \notin I_2(z^*) \), where \( z^* \) is the limit point of \( z \). Hence (47) still holds and the previous proof for LICQ at \( z \) applies.
C-stationarity of \( z^* \) for the MPCC immediately follows from Theorem 3.2. The key point in obtaining the M-stationarity (Theorems 3.4 and 3.5) is to establish a contradiction between the negative multipliers (60) associated with the biactive sets at \( z^* \), and the bounded below condition (58) on the reduced Hessian at each stationary point \( z^k \) of the NLP, with \( \epsilon^k \) sufficiently small. Here we again assume (60) holds for some index \( j_0 \in I_1(z^*) \cap I_2(z^*) \), and choose the direction \( d^k \in T(z^k) \) as defined in (61). Then the contribution from function \( \Phi^e_{\epsilon_0}(z^k) + p^e_{j_0} \) to the reduced Hessian of the Lagrangian of NLP (27) is

\[
\begin{align*}
    u^k_{\epsilon_0}(d^k)^T \nabla_{zz} \Phi^e_{\epsilon_0}(z^k) d^k \\
    = u^k_{\epsilon_0} \left[ \nabla_{GG} \Phi^e_{\epsilon_0}(z^k) d^2_G + 2 \nabla_{GH} \Phi^e_{\epsilon_0}(z^k) d_G d_H + \nabla_{HH} \Phi^e_{\epsilon_0}(z^k) d^2_H \right] \\
    = -\frac{2u^k_{\epsilon_0}}{\epsilon} \nabla_{G} \Phi^e_{\epsilon_0}(z^k) \nabla_{H} \Phi^e_{\epsilon_0}(z^k) \\
    = \frac{-2u^k_{\epsilon_0}}{\epsilon} \nabla_{G} \Phi^e_{\epsilon_0}(z^k) \nabla_{H} \Phi^e_{\epsilon_0}(z^k) \rightarrow -\infty \text{ as } \epsilon \rightarrow 0,
\end{align*}
\]

where we have used (69) and (74) to derive the last equality, and the limit is obtained by noting that \( u^k_{\epsilon_0}, \nabla_{G} \Phi^e_{\epsilon_0}(z^k), \nabla_{H} \Phi^e_{\epsilon_0}(z^k) \) converge to \( u^*_{\epsilon_0}, \xi^*_{j_0}, \eta^*_{j_0} \), which are positive due to the assumption (60) and that \( (\xi^*_{j_0}, \eta^*_{j_0}) \in B \), and bounded due to the MPCC-LICQ at \( z^* \). The reduced Hessian of the Lagrangian \( (d^k)^T \nabla_{zz} \mathcal{L}_\Phi d^k \) tends to \(-\infty \) too, in that, except for (75), all other terms in the reduced Hessian are bounded. This contradicts the condition (58) that the reduced Hessian is bounded below. As a consequence, the assumption (60) cannot hold, and the M-stationarity of \( z^* \) can be proved as before.

## 4 Numerical Results

To demonstrate the performance of the above methods, this section provides numerical results of MPCCs drawn from two sources. The first set is selected from the MacMPEC test set, while the second is a set of distillation case studies. These optimization problems have nonsmooth relations in the form of (1) and are solved as reformulated MPCCs. Five MPCC formulations, analyzed in Section 3, are considered for the numerical study.

- Regularized formulation (REG) (16) of Scholtes.
- The Lin-Fukushima (LF) formulation (34) 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. The reason is that CONOPT is a Newton-based, active set method that converges the sequence of problems with \( \epsilon^k \rightarrow 0 \) quickly, starting from the previous solution at \( \epsilon^{k-1} \). Also it checks for directions of negative curvature, in order to confirm whether second order necessary conditions are satisfied. Finally, it partitions and classifies the problem variables into basic, nonbasic and superbasic variables, and 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 CONOPT variable partition corresponds to satisfaction of the LICQ for the NLPs.
4.1 MacMPEC Results

Out of 133 problems from the MacMPEC collection [30], 15 problems are selected that have nonempty biactive sets at their solutions. They are of interest because we can distinguish different kinds of stationary points of the MPCCs by the biactive sets; in addition, biactive elements often pose numerical difficulties in solving the reformulated NLPs and thus can be employed to test performance of different NLP-based strategies. The above five MPCC formulations are implemented in GAMS and applied to these problems. The outer loop is controlled by \( \epsilon^0 = 0.25, \epsilon_{tol} = 10^{-6} \), and the reducing factor \( \gamma = 0.1 \); for BA, \( \epsilon = 0.01 \) 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 solution if both \( G_i(z^*) \) and \( H_i(z^*) \) are no more than \( 10^{-5} \).

The following demonstrates performance of the MPCC formulations in three parts, namely, for 11 problems of the selected set that converge to strongly stationary points, for 3 problems that converge to M-stationary points, and a special discussion for problem ralph2.

4.1.1 Examples with Strongly Stationary Solutions

Starting from the default initial points, formulations BA, MLF, and NCP solve all 11 problems to strongly stationary points. General information at the solutions is shown in Table 1. However, REG solves only 10 of them (except bilevel1), while LF solves 9 of them (except bilevel1 and outrata31) to the same solutions; for the exceptions, REG and LF converge to a different strongly stationary point with empty biactive set for bilevel1, and LF converges to a point infeasible to the original MPCC for outrata31.

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

Table 1: Strongly Stationary Points of Examples from MacMPEC

We first present detailed results of the 9 problems where the five formulations converge to the same points. Table 2 shows the iteration counts and complementary residuals at the solutions. For this group of problems, REG takes the fewest iterations to converge. Multipliers of the reformulated constraints corresponding to the biactive elements are given in Table 3. Strong stationarity of the solutions can be identified by BA with the nonzero
parameters $p_j^*$, or by MLF with the inactive upper bounds associated with $u_j^*$, for all $j \in I_1(z^*) \cap I_2(z^*)$. In addition, the REG multipliers $\nu_{GH,j}^* = 0$ for all $j \in I_1(z^*) \cap I_2(z^*)$ also validate strong stationarity of the points.

Note that LF often leads to very large multipliers for the biactive elements. In contrast to Eq. (50) for MLF, linear dependence for LF is written as:

$$0 = \sum_{i=1}^{n_c} \lambda_i \nabla c_i(z) - \sum_{j \in I_{gL}(z,t)} \mu_{L,j} [(H_j(z) + t_j) \nabla G_j(z) + (G_j(z) + t_j) \nabla H_j(z)]$$

$$+ \sum_{j \in I_{gU}(z,t)} \mu_{U,j} [H_j(z) \nabla G_j(z) + G_j(z) \nabla H_j(z)].$$

(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 for $t > 0$. As a consequence, with MPCC-LICQ at $z^*$ the LF multipliers $\mu_{L,j}$ and $\mu_{U,j}$ are not necessarily zero in order to satisfy (76) and, in practice, numerically dependent systems arise from the LF formulation for very small $t$.

<table>
<thead>
<tr>
<th>Problem</th>
<th>BA</th>
<th>LF</th>
<th>MLF</th>
<th>NCP</th>
<th>REG</th>
</tr>
</thead>
<tbody>
<tr>
<td>bard2m</td>
<td>61/1.25e-6</td>
<td>47/1.25e-6</td>
<td>59/1.25e-6</td>
<td>61/1.5e-7</td>
<td>9/8.88e-16</td>
</tr>
<tr>
<td>df1</td>
<td>9/9.16e-9</td>
<td>26/1.08e-5</td>
<td>25/2.6e-7</td>
<td>9/1.24e-6</td>
<td>7/0</td>
</tr>
<tr>
<td>ex9.2.3</td>
<td>58/1.256e-6</td>
<td>121/3.93e-16</td>
<td>69/1.25e-6</td>
<td>50/1.02e-6</td>
<td>17/1.7e-7</td>
</tr>
<tr>
<td>ex9.2.8</td>
<td>38/9.34e-16</td>
<td>21/1.56e-12</td>
<td>13/1.56e-10</td>
<td>85/1.73e-12</td>
<td>14/2.5e-6</td>
</tr>
<tr>
<td>ex9.2.9</td>
<td>28/1.25e-6</td>
<td>17/1.25e-6</td>
<td>12/1.25e-6</td>
<td>40/1.25e-6</td>
<td>7/0</td>
</tr>
<tr>
<td>kth1</td>
<td>40/3.37e-11</td>
<td>17/1.92e-10</td>
<td>12/6.38e-12</td>
<td>57/1.25e-6</td>
<td>7/0</td>
</tr>
<tr>
<td>qpec1</td>
<td>28/1.25e-6</td>
<td>124/1.25e-6</td>
<td>42/1.25e-6</td>
<td>28/1.25e-6</td>
<td>56/5.55e-17</td>
</tr>
<tr>
<td>scholtes2</td>
<td>43/2.3e-7</td>
<td>57/3.01e-6</td>
<td>59/2.3e-7</td>
<td>49/1.02e-6</td>
<td>10/0</td>
</tr>
<tr>
<td>sl1</td>
<td>33/1.25e-6</td>
<td>38/1.25e-6</td>
<td>31/1.25e-6</td>
<td>29/8.8e-7</td>
<td>6/2.22e-16</td>
</tr>
</tbody>
</table>

Table 2: Total Iterations/Complementary Residual ($\psi$) of MPCC Formulations (Strongly Stationary Case). The iteration count is the sum of consecutive CONOPT4 solutions till the outer loop converges. The complementary residual

$$\psi = \max_j (\min(G_j(z^*), H_j(z^*))), j \in I_1(z^*) \cap I_2(z^*).$$

If we solve these MPCC problems to a smaller $\epsilon_k$ (or $t_k$) limit, with $\epsilon_{tot} = 10^{-12}$, all the formulations except LF converge to the same solutions more accurately, with the complementary residuals vanishing to zero. On the other hand, LF converges more accurately only for ex9.2.9. For df1, qpec1, and sl1, the LF complementary residuals at the limit points cannot be decreased below 1.08e-5, 1.97e-5, and 2.7e-7, respectively. For the remaining problems, the LF multipliers associated with the biactive elements become significantly more ill-conditioned as $\epsilon$ decreases, the solutions diverge from the previous strongly stationary solutions, and LF stabilizes at new limit points, where there are no biactive elements, the complementary residuals are very small, and multipliers $\mu_{L}^*$ and $\mu_{U}^*$ are both well-conditioned (except for kth1). However, these points are infeasible to the original MPCC, with either $G_j$ or $H_j (j = 1, \ldots, n_x)$ negative and the other one close to zero. Performance profiles for these five problems (bard2m, ex9.2.3, ex9.2.8, kth1, scholtes2) are presented in Figure 1.
Table 3: Multipliers Corresponding to Biactive Elements (Strongly Stationary Case). REG does not have $\nu^*_H$ for qpec1, in that $G_j$ and $H_j$ have identical expressions in this example.

Moreover, with $\epsilon$ getting smaller, REG usually requires only a few iterations for convergence of the corresponding NLP, while for the NCP-based formulations, i.e. BA, MLF and NCP, the number of iterations in each NLP does not always decrease.

### 4.1.2 Examples with M-Stationary Solutions

Starting from the default initial points, all five formulations solve the 3 problems (ex9.2.2, qpec2, scholtes4) to their M-stationary solutions. General information at the solutions is given by Table 4. Iteration counts and complementary residuals at the solutions are shown in Table 5. For these problems, REG does not exhibit reduced iteration counts as before, and it cannot converge accurately, only to complementary residuals larger than $10^{-4}$. M-stationarity can be identified with $p^*$ (by BA) and $u^*_U$ (by MLF) in Table 6. Note that ex9.2.2 have two pairs of biactive elements, only one of which, however, has multipliers meeting the strongly stationary conditions.

When $\epsilon_{tol} = 10^{-12}$ is applied, all the formulations will converge to the same solutions as listed in Table 4. However, LF and REG still cannot converge accurately. For LF, their complementary residuals stay at their previous levels; for REG, the large multipliers $\nu^*_{GH}$ in Table 6 increase to $O(10^4)$ or $O(10^5)$, and the complementary residuals cannot decrease below $10^{-5}$.

In this test, NCP-based formulations outperform, in terms of iteration counts and complementary residuals. MLF is the only formulation that identifies the two biactive pairs of
Figure 1: Profiles of Objective (\(obj\)), Complementary Residual (\(comp\_red\)), Norm of Multipliers (\(mult\_norm\)), and Violation of Positivity (\(neg\_G/H\)) against \(\epsilon\) by LF. Here \(\epsilon\) decreases from 0.25 to \(10^{-15}\), \(comp\_red = \max_j(\min(G_j(z_k^*), H_j(z_k^*))\),
\[neg\_G/H = \max_j(0, -G_j(z_k^*), -H_j(z_k^*))\] \(j \in \{1, \ldots, n_x\}\), \(mult\_norm = \max(||\mu^*||_\infty, ||\mu^*||_\infty)\). LF stabilizes at a MPCC infeasible point as \(\epsilon^k \leq 10^{-12}\) except for kth1, which fails in NLP solution when \(\epsilon < 10^{-12}\).

<table>
<thead>
<tr>
<th>Problem</th>
<th>Obj</th>
<th>(z^*)</th>
<th>(I_1(z^<em>) \cap I_2(z^</em>))</th>
</tr>
</thead>
<tbody>
<tr>
<td>ex9.2.2</td>
<td>100</td>
<td>(10, 10, 0, 0, 0, 0, 0, 10, 10, 0)</td>
<td>{1,4}</td>
</tr>
<tr>
<td>qpec2</td>
<td>45</td>
<td>(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, 0)</td>
<td>{11, 12, 13, 14, 15, 16, 17, 18, 19, 20}</td>
</tr>
<tr>
<td>scholtes4</td>
<td>0</td>
<td>(0, 0, 0)</td>
<td>{1}</td>
</tr>
</tbody>
</table>

Table 4: M-Stationary Points of Examples from MacMPEC

ex9.2.2 precisely. The behavior of LF is about the same as for the strongly stationary examples. The degeneracy of REG can also be attributed to numerical failure to satisfy LICQ. For REG (16), linear dependence can be expressed as follows

\[
0 = \sum_{i=1}^{n_c} \lambda_i \nabla c_i(z) - \sum_{j \in I_G(z)} \nu_{G,j} \nabla G_j(z) - \sum_{j \in I_H(z)} \nu_{H,j} \nabla H_j(z) + \sum_{j \in I_{GH}(z,t)} \nu_{GH,j} [H_j(z) \nabla G_j(z) + G_j(z) \nabla H_j(z)] , \tag{77}
\]

where \(I_G(z) = \{j | G_j(z) = 0\}\), \(I_H(z) = \{j | H_j(z) = 0\}\), \(I_{GH}(z,t) = \{j | G_j(z)H_j(z) = t\}\) are active sets of the inequality constraints. For these M-stationary examples, the nonzero multipliers \(\nu_{GH,j}^*\) in Table 6 indicate \(I_{GH}(z,t) \cap I_1(z^*) \cap I_2(z^*) \neq \emptyset\) as \(t \to 0\). In such circumstances, \(G_j(z)\) and \(H_j(z)\) in (77) tend to zeros as \(t \to 0\), though they should be nonzero for \(t > 0\). As a consequence, even if MPCC-LICQ holds at \(z^*\), \(\nu_{GH,j}^*\) may not be zero to satisfy (77) when \(t\) is sufficiently small. The degenerate system results in large REG multipliers, and inaccuracy in satisfying the complementarity conditions.
Table 5: Total Iterations/Complementary Residual ($\psi$) of MPCC Formulations (M-Stationary Case). The iteration count is the sum of consecutive CONOPT4 solutions till the outer loop converges. The complementary residual 

$$\psi = \max_j (\min(G_j(z^*), H_j(z^*))), j \in I_1(z^*) \cap I_2(z^*).$$

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

Table 6: Multipliers Corresponding to Biactive Elements (M-Stationary Case). REG does not have $\nu^*_H$ for qpec2, in that $G_j$ and $H_j$ have identical expressions in this example.

<table>
<thead>
<tr>
<th>Problem</th>
<th>BA</th>
<th>LF $\mu^*_L$</th>
<th>$\mu^*_U$</th>
<th>MLF $u^<em>_L$ $u^</em>_U$</th>
<th>NCP $u^*_ncp$</th>
<th>REG $\nu^<em>_G$ $\nu^</em>_H$ $\nu^*_GH$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ex9.2.2</td>
<td>5.74</td>
<td>0</td>
<td>2.20e5</td>
<td>0 5.74 6.67</td>
<td>0 0 1.83e3</td>
<td></td>
</tr>
<tr>
<td>qpec2</td>
<td>4.00</td>
<td>0</td>
<td>1.04e5</td>
<td>0 4.00 4.00</td>
<td>0 -1.26e3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>0</td>
<td>1.04e5</td>
<td>0 4.00 4.00</td>
<td>0 -1.26e3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>0</td>
<td>1.04e5</td>
<td>0 4.00 4.00</td>
<td>0 -1.26e3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>0</td>
<td>1.04e5</td>
<td>0 4.00 4.00</td>
<td>0 -1.26e3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>0</td>
<td>1.04e5</td>
<td>0 4.00 4.00</td>
<td>0 -1.26e3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>0</td>
<td>1.04e5</td>
<td>0 4.00 4.00</td>
<td>0 -1.26e3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>0</td>
<td>1.04e5</td>
<td>0 4.00 4.00</td>
<td>0 -1.26e3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>0</td>
<td>1.04e5</td>
<td>0 4.00 4.00</td>
<td>0 -1.26e3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>0</td>
<td>1.04e5</td>
<td>0 4.00 4.00</td>
<td>0 -1.26e3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>0</td>
<td>1.04e5</td>
<td>0 4.00 4.00</td>
<td>0 -1.26e3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>0</td>
<td>1.04e5</td>
<td>0 4.00 4.00</td>
<td>0 -1.26e3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>0</td>
<td>1.04e5</td>
<td>0 4.00 4.00</td>
<td>0 -1.26e3</td>
<td></td>
</tr>
<tr>
<td>scholtes4</td>
<td>2.00</td>
<td>0</td>
<td>8.98e4</td>
<td>0 2.00 2.00</td>
<td>0 0 6.32e2</td>
<td></td>
</tr>
</tbody>
</table>

4.1.3 Example of Strong Stationarity through Asymptotic Weak Degeneracy

From the two preceding subsections, we see examples for which the stationarity category of the solution can be decided by the bounding parameters $p^*$ or upper bound multipliers $u^*_U$. Now we discuss a case where the stationarity category is not evident from the numerical result.

Example ralph2 is described as

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

Starting from $(x, y) = (1, 1)$, all the formulations converge to $(x^*, y^*) = (0, 0)$, which is the global optimum and satisfies the MPCC-LICQ. Detailed information at the solution is summarized in Table 7. At the solution, the bounding parameter $p^* = 0$, the upper bound of MLF is active, and the REG multiplier is not zero but bounded. Scholtes [38] proved that with the bounded multiplier, the solution is strongly stationary. To analyze the stationarity with the NCP-based formulations, we take advantage of the concept of asymptotic weak nondegeneracy [14].
Consider the BA formulation with $p \equiv 0$ (i.e. ordinary NCP); MLF can be analyzed similarly. Recall that $z^k \to z^*$ as $\epsilon \to 0$. Asymptotic weak nondegeneracy requires that 

$$\xi_j^* > 0, \eta_j^* > 0, \forall j \in I_1(z^*) \cap I_2(z^*),$$

(78)

for any accumulation point of $\{\nabla \Phi_j(z^k)\}$. In view of (44), this implies that $0 < \rho_1 \leq G_j(z^k)/H_j(z^k) \leq \rho_2 < +\infty$, namely, neither $G_j(z^k)$ nor $H_j(z^k)$ vanishes to zero too quickly. For problem ralph2, with $G(z) = x, H(z) = y$, the asymptotic condition is obviously satisfied. Suppose the point $z^* = (x^*, y^*) = (0, 0)$ is not strongly stationary. Then by (52), the multiplier $u^* > 0$. Therefore, with the asymptotic condition and the null space matrix $[-x^k/y^k]$, the reduced Hessian of $\Phi^*(z^k)$ tends to $-\infty$ at $(z^*, u^*)$, i.e.,

$$\begin{bmatrix} -x^k/y^k & 1 \end{bmatrix} \cdot \frac{u^k \cdot 2 x^k y^k}{(x^k + y^k)^3} \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} -x^k/y^k \\ 1 \end{bmatrix} = -2u^k \cdot \frac{x^k/y^k}{y^k(x^k/y^k + 1)} \to -\infty.$$

Because this contradicts the optimality of $z^*$ for the NLP, the hypothesis is false and $z^*$ must be strongly stationary. The numerical experience of ralph2 validates the analysis; the multiplier $u^*$ is vanishing, the reduced Hessian of the Lagrangian is observed bounded below, and $z^*$ is declared to be (locally) optimal (where no direction of negative curvature is detected).

### 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 [28,35]. 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
where \( i \) is tray index numbered from reboiler (= 1), \( j \) is components index, \( \beta_i \) is the relaxation parameter, \( P \) is column pressure, \( T_i \) is temperature of tray \( i \), \( L_i \) and \( V_i \) ≥ 0 is flow rates of liquid/vapor, \( x_{ij} / y_{ij} \in [0, 1] \) is fraction of component \( j \) in liquid/vapor leaving tray \( i \), \( s_l^i \) and \( s_v^i \) are slack variables, the tray set \( S = \{2, \ldots, N - 1\} \) 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 \( N_t \geq N_f \). We also specify feed and reflux flowrates for \( i \in S \) based on the value of the distributions at tray \( i \), given by:

\[
F_i = F \frac{\exp\left(-\frac{(i-N_f)^2}{\sigma_f^2}\right)}{\sum_{j \in S} \exp\left(-\frac{(j-N_f)^2}{\sigma_f^2}\right)} , \quad R_i = R \frac{\exp\left(-\frac{(i-N_t)^2}{\sigma_t^2}\right)}{\sum_{j \in S} \exp\left(-\frac{(j-N_t)^2}{\sigma_t^2}\right)} \quad i \in S ,
\]

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

The resulting MPCC model is used to determine the optimal number of trays, reflux ratio and feed tray location for a benzene/toluene separation. Five MPCC formulations, analyzed in Section 3, 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, equilibrium complementarities (79) for each tray and tray distributions (79)–(80), 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:

\[
objective = wt \cdot D \cdot x_{D,Toluene} + wr \cdot r + wn \cdot N_t ,
\]

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 \( wt = 1, \ ws = 0.01, \) and \( wn = 0.45 \); these weights allow the optimization to trade off product purity, energy cost and capital cost. The reflux ratio \( r \) is allowed to vary between one and 20, the feed tray 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 \( x_{D,Tolune} \leq \zeta \) 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.

At the solution of all cases, CONOPT reports no negative curvature directions nor Jacobian degeneracies, which is consistent with LICQ and second-order necessary conditions satisfied. For the first three cases \( I_1 \cap I_2 = \emptyset \) and the solutions can be claimed to be strongly stationary. For the last case where \( I_1 \cap I_2 \neq \emptyset \), we cannot directly determine the station-
Table 8: Solutions of Benzene/Toluene Cases. $N = 10$ in the first three cases and $N = 25$ in the fourth case.

| $\zeta$ Cases | Objective | $N_t$ | $N_f$ | $r$ | $|I_1 \cap I_2|$ |
|---------------|-----------|-------|-------|-----|-----------------|
| 0.05          | 3.9916    | 7.6804| 3.85  | 0.72 | 0               |
| 0.01          | 2.9078    | 8.84  | 2.94  | 1.52 | 0               |
| 0.005         | 3.2336    | 8.93  | 2.86  | 2.09 | 0               |
| 0.001         | 3.0755    | 12.83 | 2.93  | 1.74 | 10              |

Table 9: Total Iterations/Convergence Rate ($\psi$) 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 $\psi$ is calculated from $|f(z(\epsilon_i) - f(z^*)|/|f(z(\epsilon_i) - f(z^*)| = (\epsilon_i/\epsilon_i')^\psi$.

<table>
<thead>
<tr>
<th>$\zeta$ Cases</th>
<th>BA</th>
<th>LF</th>
<th>MLF</th>
<th>NCP</th>
<th>REG</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>127/1.08</td>
<td>154/1.07</td>
<td>154/1.08</td>
<td>92/1.91</td>
<td>184/0.90</td>
</tr>
<tr>
<td>0.01</td>
<td>120/1.08</td>
<td>167/1.08</td>
<td>140/1.08</td>
<td>77/1.64</td>
<td>125/1.01</td>
</tr>
<tr>
<td>0.005</td>
<td>125/1.06</td>
<td>145/1.08</td>
<td>131/1.07</td>
<td>66/2.02</td>
<td>107/1.03</td>
</tr>
<tr>
<td>0.001</td>
<td>208/1.06</td>
<td>461/1.08</td>
<td>287/1.07</td>
<td>189/1.08</td>
<td>266/0.55</td>
</tr>
</tbody>
</table>

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: $\text{objective} = r + N_t$, 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 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 | $N_t$ | $N_f$ | $r$ | $|I_1 \cap I_2|$ |
|-----------------------|-----------|-------|-------|-----|----------------|
| Reflux Constrained    | 72.47     | 37.47 | 5.20  | 35  | 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

<table>
<thead>
<tr>
<th>Cases</th>
<th>BA</th>
<th>LF</th>
<th>MLF</th>
<th>NCP</th>
<th>REG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflux Constrained</td>
<td>293/1.08</td>
<td>1193/1.08</td>
<td>1100/1.08</td>
<td>259/1.17</td>
<td>552/0.85</td>
</tr>
<tr>
<td>Unconstrained</td>
<td>285/1.04</td>
<td>993/1.06</td>
<td>455/1.07</td>
<td>247/1.43</td>
<td>517/0.88</td>
</tr>
<tr>
<td>Tray Constrained</td>
<td>148/1.08</td>
<td>182/1.08</td>
<td>242/1.08</td>
<td>92/2.0</td>
<td>229/1.02</td>
</tr>
</tbody>
</table>

Table 11: Total Iterations/Convergence Rate ($\psi$) of MPCC Formulations for Argon Column Cases. The iteration count is the sum of 13 consecutive CONOPT4 solutions with $\epsilon_i = 10^{-i/2}, i = 0, \ldots, 12$. The estimated convergence rate $\psi$ is calculated from $|f(z(\epsilon_i) - f(z^*)|/|f(z(\epsilon_i) - f(z^*)| = (\epsilon_i/\epsilon_i')^{\psi}$.

At the solution of all cases, CONOPT reports no negative curvature directions nor Jacobian degeneracies, which is consistent with LICQ and second-order necessary conditions.
satisfied. The Tray Constrained case with $I_1 \cap I_2 = \emptyset$ has a solution that is strongly stationary and is easiest to solve. For the other two cases where $I_1 \cap I_2 \neq \emptyset$, evidence is insufficient to decide stationarity category of the solution. The problems are further solved to $\epsilon^k = 10^{-15}$, and all the formulations converge to the same solution as before. In both cases, most of the BA multipliers associated with the biactive elements have vanished to zero, except that four (Reflux Constrained case) and three (Unconstrained case) multipliers have values $|u^k_j| < 10^{-2}$ and $|u^k_j| < 10^{-3}$, respectively. As for the asymptotic condition (78), we observed $O(10^{-7}) \leq G^k_j / H^k_j \leq O(10^7)$ in the former case, and $O(10^{-1}) \leq G^k_j / H^k_j \leq O(10)$ in the latter. Again, as $\epsilon^k$ decreases, no negative curvature directions nor Jacobian degeneracies are reported, and the reduced Hessian of the Lagrangian is bounded below. Therefore, 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. In contrast, the nonzero REG multipliers associated with the biactive elements at $\epsilon^k = 10^{-15}$ have values $|\nu^k_{GH,j}| < 4414.57$ (Reflux Constrained case) and $|\nu^k_{GH,j}| < 6.67$ (Unconstrained case).

Finally, it is interesting to note that LF and MLF require the most effort for both distillation systems. This is likely because determining the active sets in (34) and (28c) becomes more difficult as $\epsilon \to 0$; this was also observed in [26]. We also note that the BA approach, with related optimality conditions to LF and MLF, shares convergence behavior to M-stationary points, but also has a similar structure similar to the faster NCP approach. 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 $\epsilon$-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 $O(\epsilon^2)$ from the solution of the MPCCs (Proposition 2.1). Stationarity of the approaching solution generated from the proposed NLP-based strategies are investigated. In the presence of MPCC-LICQ, the limit point of the stationary points of the NLP sequences is guaranteed to be C-stationary (Theorems 3.2 and 3.3). Furthermore, M-stationarity can be characterized with additional second-order conditions (Theorems 3.4 and 3.5). It is worth mentioning that taking advantage of the parameter (BA) or multiplier (MLF) information at the NLP solution can facilitate the discrimination of the stationarity categories in most cases, except for those, for example, where the condition of asymptotic weak nondegeneracy applies (78).

The proposed formulations, together with the closely related Lin-Fukushima formulation (LF), well-studied Scholtes’ formulation (REG), and the ordinary NCP formulation (without bounding), are applied to selected MacMPEC examples and two large-scale distillation cases, which have nonempty biactive sets at the solution. It turns out that the NCP-based formula-
tions have advantages in dealing with biactive elements, because LICQ of these formulations is robust to such circumstances, which is a potential benefit from the Clarke generalized gradient at the limit. 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 double bounded formulations LF and MLF need the most iterations to converge; a possible cause is the challenge in determining the active set with a vanishing $\epsilon$. Instead, the BA and ordinary NCP formulations are the most efficient alternatives in these cases.

In this research, all the theoretical results on stationarity are derived with the MPCC-LICQ and standard NLP LICQ assumptions. Inspecting convergence properties of the NLP-based strategies with weaker constraint qualifications will be considered for future work. In addition, more practical conditions to characterize B-stationarity would be beneficial, beyond the asymptotic weak nondegeneracy condition (78), which is hard to enforce.

**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


