Solving Mathematical Programs with Equilibrium Constraints

Lei Guo, Gui-Hua Lin and Jane J. Ye

January 2014, Revised September and October in 2014

Communicated by Xiaoqi Yang

Abstract. This paper aims at developing effective numerical methods for solving mathematical programs with equilibrium constraints. Due to the existence of complementarity constraints, the usual constraint qualifications do not hold at any feasible point, and there are various stationarity concepts such as Clarke, Mordukhovich, and strong stationarities that are specially defined for mathematical programs with equilibrium constraints. However, since these stationarity systems contain some unknown index sets, there has been no numerical method for solving them directly. In this paper, we remove the unknown index sets from these stationarity systems successfully and reformulate them as smooth equations with box constraints. We further present a modified Levenberg-Marquardt method for solving these constrained equations. We show that, under some weak local error bound conditions, the method is locally and superlinearly convergent. Furthermore, we give some sufficient conditions for local error bounds and show that these conditions are not very stringent by a number of examples.

Key Words. Mathematical program with equilibrium constraints, Clarke/Mordukhovich/strong stationarity, Levenberg-Marquardt method, error bound.

2010 Mathematics Subject Classification. 90C26, 90C30, 90C33.

1 Introduction

Mathematical program with equilibrium constraints (MPEC) is a constrained optimization problem, in which the essential constraints are defined by some parametric variational inequalities or parametric complementarity systems. MPEC is a class of very important problems since they arise frequently in applications; see [1, 2] for references. One main source of MPEC comes from bilevel programming problems, which have numerous applications in practice. The challenge in theoretical and numerical treatment of MPEC arises from the fact that the Mangasarian-Fromovitz constraint qualification (MFCQ) is violated at every feasible point; see [3]. Nevertheless, there have been great progresses made on

Lei Guo, Sino-US Global Logistics Institute, Shanghai Jiao Tong University, Shanghai 200030, China, and Department of Mathematics and Statistics, University of Victoria, Victoria, BC, V8W 2Y2 Canada. E-mail: guolayne@gmail.com.

Gui-Hua Lin (corresponding author), School of Management, Shanghai University, Shanghai 200444, China. E-mail: guihualin@shu.edu.cn.

Jane J. Ye, Department of Mathematics and Statistics, University of Victoria, Victoria, BC, V8W 2Y2 Canada. E-mail: janeye@uvic.ca

theoretical issues including various necessary and sufficient optimality conditions, constraint qualifications, stability analysis, and sensitivity analysis; see, e.g., [4–8]. In particular, various stationarity concepts such as Clarke (or C-) stationarity, Mordukhovich (or M-) stationarity, strong (or S-) stationarity, and various constraint qualifications that ensure a local minimizer of MPEC is C-/M-/S-stationary have been studied; see [6, 7] for more discussions. Moreover, many numerical methods have been proposed to solve MPEC; see, e.g., [9] and the references therein.

One way to solve a standard nonlinear programming problem is to solve its Karush-Kuhn-Tucker (KKT) system by using some numerical methods such as Newton-type methods. It is known that solving an S-stationarity system is equivalent to solving the KKT system for the original MPEC as a nonlinear programming problem with equality and inequality constraints; see Theorem 3.1 given below. However, since the MFCQ fails to hold at every feasible point when the MPEC is treated as a standard nonlinear programming problem, a local minimizer of MPEC may not be a solution of the classical KKT system. Moreover, to guarantee the quadratic convergence of the Newton-type methods for solving the classical KKT system, the Jacobian of the classical KKT system is usually required to be nonsingular, which is implied by the linear independent constraint qualification (LICQ) and the second order sufficient condition; see, e.g., [10, page 441]. Since the LICQ fails to hold for MPEC, the classical KKT system may be degenerate, i.e., the Jacobians of the resulting system may be singular, and hence the Newton-type methods may not be stable. On the other hand, since the C-/M-/S-stationarity systems for MPEC contain some unknown index sets, they are all uncertain systems, so that we cannot solve them directly.

We present a novel approach in this paper: By removing the unknown index sets from the C-/M-/S-stationarity systems, we reformulate them as constrained equations. We further propose a modified Levenber-Marguardt (LM) method to solve the constrained equations, and show that the method is locally and superlinearly convergent under some local error bound conditions.

2 MPEC Stationarities

We consider the MPEC in the form

$$\min \ f(x) \quad \text{s.t.} \quad g(x) \leq 0, \ h(x) = 0, \ 0 \leq G(x) \ \perp \ H(x) \geq 0, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}, g : \mathbb{R}^n \to \mathbb{R}^p, h : \mathbb{R}^n \to \mathbb{R}^q$, and $G, H : \mathbb{R}^n \to \mathbb{R}^m$ are all twice differentiable, and their second order derivatives are locally Lipschitzian, whereas $a \perp b$ means that a is perpendicular to b.

For a nonconvex optimization problem, stationary points are good candidates for local minimizers, and most existing numerical algorithms aim at finding stationary points. Unlike the classical nonlinear programming problems, which have only one kind of KKT conditions, there are various kinds of KKT-type conditions for MPEC. In this section, we first review the definitions of the popular stationarity conditions, and then we give some examples to show that it is important to study various stationary points. We refer the reader to [6, 7] for more discussions of these stationarity conditions.

Let $\mathcal{F}$ be the feasible region of problem (1). For a given point $x^* \in \mathcal{F}$, let $I_g^* := \{ i \mid g_i(x^*) = 0\}$, $\mathcal{I}^* := \{ i \mid G_i(x^*) = 0 < H_i(x^*)\}$, $\mathcal{J}^* := \{ i \mid G_i(x^*) = 0 = H_i(x^*)\}$, and $\mathcal{K}^* := \{ i \mid G_i(x^*) > 0 = H_i(x^*)\}$. Obviously, $\{\mathcal{I}^*, \mathcal{J}^*, \mathcal{K}^*\}$ is a partition of $\{1, 2, \cdots, m\}$. Given a mapping $F : \mathbb{R}^n \to \mathbb{R}^l$ and a vector $x \in \mathbb{R}^n$, $\nabla F(x)$ stands for the transposed Jacobian of F at x.

Definition 2.1. (1) We call $x^* \in \mathcal{F}$ a weakly stationary point of problem (1) iff there exist multipliers $(\lambda, \mu, u, v) \in \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^m \times \mathbb{R}^m$ satisfying

$$\nabla f(x^*) + \nabla g(x^*)\lambda + \nabla h(x^*)\mu - \nabla G(x^*)u - \nabla H(x^*)v = 0, \tag{2}$$

$$\min(\lambda, -g(x^*)) = 0, \tag{3}$$

$$u_i = 0 \ (i \in \mathcal{K}^*), \quad v_i = 0 \ (i \in \mathcal{I}^*). \tag{4}$$

(2) We call $x^* \in \mathcal{F}$ a Clarke stationary point or a C-stationary point of problem (1) iff there exist multipliers $(\lambda, \mu, u, v) \in \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^m \times \mathbb{R}^m$ satisfying (2)–(4) and

$$u_i v_i \geq 0 \ \text{ for each } \ i \in \mathcal{J}^*. \tag{5}$$

(3) We call $x^* \in \mathcal{F}$ a Mordukhovich stationary point or an M-stationary point of problem (1) iff there exist multipliers $(\lambda, \mu, u, v) \in \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^m \times \mathbb{R}^m$ satisfying (2)–(4) and

$$\text{either } \ u_i v_i = 0 \ \text{ or } \ u_i > 0, v_i > 0 \ \text{ for each } \ i \in \mathcal{J}^*. \tag{6}$$

(4) We call $x^* \in \mathcal{F}$ a strongly stationary point or an S-stationary point of problem (1) iff there exist multipliers $(\lambda, \mu, u, v) \in \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^m \times \mathbb{R}^m$ satisfying (2)–(4) and

$$u_i \geq 0, \ v_i \geq 0 \ \text{ for each } \ i \in \mathcal{J}^*. \tag{7}$$

The relations among the above stationarities can be stated as follows:

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

In what follows, we use some examples to illustrate the importance of studying these stationarities.

Example 2.1. Consider the problem

$$\min \; x_1 - 2x_2 \quad \text{s.t.} \;\; x_1 - x_2 \geq 0, \;\; 0 \leq x_1 \perp x_2 \geq 0. \tag{8}$$

Since all constraint functions are affine, any local minimizer must be M-stationary [7]. By solving the weak stationarity conditions

$$\left(\begin{smallmatrix} 1 \\ -2 \end{smallmatrix}\right) - \left(\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right) \lambda - \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right) u - \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right) v = 0, \; \min(\lambda, x_1 - x_2) = 0, \; \min(x_1, x_2) = 0, \; x_1 u = x_2 v = 0,$$

we know that problem (8) has a unique weakly stationary point $x^* = (0,0)$ with multipliers $u = 1 - \lambda$, $v = \lambda - 2$, and $\lambda \geq 0$. The only nonempty index set is $\mathcal{J}^*$. Since u and v cannot be both non-negative, x^* is not an S-stationary point. By taking $\lambda = 1$ or $\lambda = 2$, we have $uv = 0$. Therefore, the unique minimizer $(0,0)$ is an M-stationary point, but not an S-stationary point.

Example 2.2. Consider the problem

$$\min \; x_1 + x_2 - x_3 - \frac{1}{2}x_4 \quad \text{s.t.} \quad -6x_1 + x_3 + x_4 \leq 0, \; -6x_2 + x_3 \leq 0, \; x_4^2 \leq 0, \; 0 \leq x_1 \perp x_2 \geq 0.$$

Similarly to Example 2.1, by solving the weak stationarity conditions, we can obtain the unique weakly stationary point $x^* = (0,0,0,0)$ with multipliers $u = v = -2$, $\lambda_1 = \lambda_2 = \frac{1}{2}$, and $\lambda_3 \geq 0$. The only nonempty index set is $\mathcal{J}^*$. Since $u < 0$ and $v < 0$, the unique minimizer $(0,0,0,0)$ is a C-stationary point, but not an M-stationary point.

From the above two examples, one may tend to think that the reason that a minimizer is not an S-stationary point (or not an M-stationary point) is the nonexistence of an S-stationary point (or an M-stationary point). The following two examples show that, even S- or M-stationary points exist, the problem may attain its optimum at M- or C-stationary points.

Example 2.3. Consider the problem

$$\min \; (x_1 - 1)^2 + (x_2 - \frac{1}{2})^2 \quad \text{s.t.} \;\; x_1 \leq 1, \; x_2 \geq 0, \; 0 \leq 2x_1 + x_2 \perp 2 - (x_1 - 1)^2 - (x_2 - 1)^2 \geq 0.$$

The weak stationarity conditions are

$$\left(\begin{smallmatrix} 2x_1-2 \\ 2x_2-1 \end{smallmatrix}\right) + \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right) \lambda_1 - \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right) \lambda_2 - \left(\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right) u - \left(\begin{smallmatrix} 2-2x_1 \\ 2-2x_2 \end{smallmatrix}\right) v = 0, \ \min(\lambda_1, 1-x_1) = 0, \ \ \min(\lambda_2, x_2) = 0,$$
$$\min(2x_1 + x_2, 2-(x_1-1)^2-(x_2-1)^2) = 0, \ (2x_1+x_2)u = (2-(x_1-1)^2-(x_2-1)^2)v = 0,$$

which yield three weakly stationary points: $(1, \sqrt{2}+1)$ with multipliers $u = 0, v = -1-\frac{\sqrt{2}}{4}$, and $\lambda_1 = \lambda_2 = 0$; $(0,0)$ with multipliers $u = \lambda_2 - 1, v = -\lambda_2, \lambda_1 = 0$, and $\lambda_2 \geq 0$; and $(-\frac{2}{5}, \frac{4}{5})$ with multipliers $u = \frac{7}{5}, v = -2$, and $\lambda_1 = \lambda_2 = 0$. At $(1, \sqrt{2}+1)$, since $\mathcal{J}^* = \emptyset$, it is also S-stationary. At $(0,0)$, the only nonempty index set is $\mathcal{J}^*$, and the point is not S-stationary since u and v cannot be both non-negative. Taking either $\lambda_2 = 0$ or $\lambda_2 = 1$, we have $uv = 0$, and hence $(0,0)$ is M-stationary but not S-stationary. At $(-\frac{2}{5}, \frac{4}{5})$, since the only nonempty index set is $\mathcal{J}^*$ and $u > 0, v < 0$, it is only weakly stationary. From graphing, it is easy to find that the M-stationary point $(0,0)$ is the unique global minimizer, while the unique S-stationary point $(1, \sqrt{2}+1)$ is not a local minimizer (in fact it is the unique global maximizer).

Example 2.3 gives us some hint. If we only solve the S-stationarity system, we might have missed the true solution. We next give an example whose local minimizers are C-stationary, but not M-stationary. Similarly, in this example, if we only solve S- or/and M-stationarity system, we might have missed the true solution.

Example 2.4. Consider the problem

$$\begin{aligned} \min \quad & (x_1-1)^2 + (x_2 - \frac{1}{2})^2 + \frac{1}{2} x_3 (x_1 - 1) \\ \text{s.t.} \quad & x_1 \leq 1, \ x_2 + x_3(x_1-1) \geq 0, \ x_3^2 \leq 0, \ 0 \leq 2x_1 + x_2 \perp 2-(x_1-1)^2-(x_2-1)^2 \geq 0. \end{aligned} \tag{9}$$

The weak stationarity conditions are

$$\begin{pmatrix} 2x_1-2+\frac{1}{2}x_3 \\ 2x_2-1 \\ \frac{1}{2}(x_1-1) \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \lambda_1 - \begin{pmatrix} x_3 \\ 1 \\ x_1-1 \end{pmatrix} \lambda_2 + \begin{pmatrix} 0 \\ 0 \\ 2x_3 \end{pmatrix} \lambda_3 - \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} u - \begin{pmatrix} 2-2x_1 \\ 2-2x_2 \\ 0 \end{pmatrix} v = 0,$$
$$\min(\lambda_1, 1-x_1) = 0, \quad \min(\lambda_2, x_2 + x_3(x_1-1)) = 0, \quad \min(\lambda_3, -x_3^2) = 0,$$
$$\min(2x_1 + x_2, 2-(x_1-1)^2-(x_2-1)^2) = 0, \ (2x_1+x_2)u = (2-(x_1-1)^2-(x_2-1)^2)v = 0,$$

which yield two weakly stationary points: $(1, \sqrt{2}+1, 0)$ with multipliers $u = 0, v = -1-\frac{\sqrt{2}}{4}$, $\lambda_1 = \lambda_2 = 0$, and $\lambda_3 \geq 0$; and $(0,0,0)$ with multipliers $u = v = -\frac{1}{2}, \lambda_1 = 0, \lambda_2 = \frac{1}{2}$, and $\lambda_3 \geq 0$. It is obvious that the unique minimizer $(0,0,0)$ is C-stationary but not M-stationary, and the unique maximizer $(1, \sqrt{2}+1, 0)$ is the unique S-stationary point.

3 Reformulations for Stationarity Conditions

Note that, unlike the KKT systems in the nonlinear programming theory, the C-/M-/S-stationarity systems contain the conditions (5), (6), and (7), which are all uncertain because the index sets $\mathcal{I}^*, \mathcal{J}^*$, and $\mathcal{K}^*$ are generally unknown before the solution x^* is found. There has been no numerical method proposed to solve the C-/M-/S-stationarity systems directly. In this section, by removing the unknown index sets from the systems, we reformulate them as equations with simple constraints, which are all certain systems and can be solved directly.

Theorem 3.1. *For any $x^* \in \mathcal{F}$, we have the following statements:*

(i) *Conditions* (4) *and* (5) *are equivalent to the equations*

$$u_i G_i(x^*) = v_i H_i(x^*) = 0, \ u_i v_i \geq 0 \ (i = 1, \cdots, m). \tag{10}$$

(ii) *Conditions* (4) *and* (6) *are equivalent to the equations*

$$u_i G_i(x^*) = v_i H_i(x^*) = 0, \ u_i v_i \geq 0, \ \max\{u_i, v_i\} \geq 0 \ (i = 1, \cdots, m). \tag{11}$$

(iii) *Conditions* (4) *and* (7) *are equivalent to the equations*

$$\alpha_i G_i(x^*) = \beta_i H_i(x^*) = 0, \alpha_i \geq 0, \beta_i \geq 0, u_i = \alpha_i - \zeta H_i(x^*), v_i = \beta_i - \zeta G_i(x^*) (i = 1, \cdots, m). \tag{12}$$

Proof (i) Condition (10) implies (4) and (5) evidently. We next show the converse part. Suppose that (4) and (5) hold. It follows that $u_i G_i(x^*) = 0$ and $v_i H_i(x^*) = 0$ for $i = 1, \cdots, m$. We next show $u_i v_i \geq 0$ for each i. Without any loss of generality, we may assume $i \notin \mathcal{J}^*$. Note that $x^* \in \mathcal{F}$ implies either $i \in \mathcal{I}^*$ or $i \in \mathcal{K}^*$. This means that either $u_i = 0$ or $v_i = 0$ by (4), and hence we must have $u_i v_i = 0$.

(ii) Condition (11) implies (4) and (6) evidently. Suppose that (4) and (6) hold. Since (6) implies (5), we have (10) from (i) immediately. It suffices to show the last inequality in (11) for each i.

- If $i \notin \mathcal{J}^*$, as shown in (i), then there must hold $u_i v_i = 0$ and hence $\max\{u_i, v_i\} \geq 0$.
- If $i \in \mathcal{J}^*$, then it follows from (6) that either $u_i > 0$ and $v_i > 0$, or $u_i v_i = 0$. In any case, we always have $\max\{u_i, v_i\} \geq 0$.

(iii) If there exist $\{\alpha, \beta, \zeta\}$ satisfying (12), conditions (4) and (7) hold obviously. Conversely, suppose that u and v satisfy conditions (4) and (7). Define $\zeta \in \mathbb{R}$ as

$$\zeta := \begin{cases} 0, & \mathcal{I}^* = \mathcal{K}^* = \emptyset, \\ \max\{-\frac{u_i}{H_i(x^*)} \mid i \in \mathcal{I}^*\}, & \mathcal{I}^* \neq \emptyset, \mathcal{K}^* = \emptyset, \\ \max\{-\frac{v_j}{G_j(x^*)} \mid j \in \mathcal{K}^*\}, & \mathcal{I}^* = \emptyset, \mathcal{K}^* \neq \emptyset, \\ \max\{-\frac{u_i}{H_i(x^*)}, -\frac{v_j}{G_j(x^*)} \mid i \in \mathcal{I}^*,\ j \in \mathcal{K}^*\}, & \mathcal{I}^* \neq \emptyset, \mathcal{K}^* \neq \emptyset. \end{cases}$$

Let $\alpha := u + \zeta H(x^*)$ and $\beta := v + \zeta G(x^*)$. It is easy to see that (12) hold. This completes the proof of the theorem. □

In Theorem 3.1, the unknown index sets $\mathcal{I}^*, \mathcal{J}^*$, and $\mathcal{K}^*$ have been removed successfully from conditions (5)–(7). As a result, by introducing some slack and auxiliary variables, the equivalent C-/M-/S-stationarity systems can be reformulated as constrained equations in the form

$$F(w) = 0, \quad w \in W, \tag{13}$$

where the constraint set $W := \{w \in \mathbb{R}^l \mid w_i \geq 0, i \in \mathcal{I}\}$ and $\mathcal{I}$ is a fixed index set. There may be more than one kind of equivalent reformulations. In this paper, we suggest the following smooth formulations for the C-/M-/S-stationarities.

(1) C-stationarity:

$$F(x, y, z_1, z_2, z_3, \lambda, \mu, u, v) := \begin{pmatrix} \nabla f(x) + \nabla g(x)\lambda + \nabla h(x)\mu - \nabla G(x)u - \nabla H(x)v \\ \lambda^T z_1 \\ z_1 + g(x) \\ h(x) \\ z_2^T z_3 \\ z_2 - G(x) \\ z_3 - H(x) \\ u \circ z_2 \\ v \circ z_3 \\ y - u \circ v \end{pmatrix}, \tag{14}$$

$$W := \left\{(x, y, z_1, z_2, z_3, \lambda, \mu, u, v) \,\middle|\, y \geq 0; z_i \geq 0 (1 \leq i \leq 3); \lambda \geq 0\right\}, \tag{15}$$

where $\circ$ means the Hadamard product, i.e., $a \circ b := (a_1 b_1, \cdots, a_n b_n)$ for $a, b \in \mathbb{R}^n$.

(2) M-stationarity:

$$F(x,y_1,y_2,y_3,y_4,z_1,z_2,z_3,\lambda,\mu,u,v) := \begin{pmatrix} \nabla f(x)+\nabla g(x)\lambda+\nabla h(x)\mu-\nabla G(x)u-\nabla H(x)v \\ \lambda^T z_1 \\ z_1+g(x) \\ h(x) \\ z_2^T z_3 \\ z_2-G(x) \\ z_3-H(x) \\ u\circ z_2 \\ v\circ z_3 \\ y_1-u\circ v \\ y_3^T y_4 \\ y_2-y_3-u \\ y_2-y_4-v \end{pmatrix}, \tag{16}$$

$$W := \Big\{(x,y_1,y_2,y_3,y_4,z_1,z_2,z_3,\lambda,\mu,u,v) \,\Big|\, y_i \geq 0 (1\leq i\leq 4); z_i \geq 0 (1\leq i\leq 3); \lambda\geq 0\Big\}. \tag{17}$$

(3) S-stationarity:

$$F(x,z_1,z_2,z_3,\lambda,\mu,\alpha,\beta,\zeta) := \begin{pmatrix} \nabla f(x)+\nabla g(x)\lambda+\nabla h(x)\mu-\nabla G(x)(\alpha-\zeta H(x))-\nabla H(x)(\beta-\zeta G(x)) \\ \lambda^T z_1 \\ z_1+g(x) \\ h(x) \\ z_2^T z_3 \\ z_2-G(x) \\ z_3-H(x) \\ \alpha^T z_2 \\ \beta^T z_3 \end{pmatrix}, \tag{18}$$

$$W := \Big\{(x,z_1,z_2,z_3,\lambda,\mu,\alpha,\beta,\zeta)\Big| z_i\geq 0 (1\leq i\leq 3); \lambda\geq 0; \alpha\geq 0; \beta\geq 0\Big\}. \tag{19}$$

Note that, if any function in $\{g_i, 1\leq i\leq p; -G_j, -H_j, 1\leq j\leq m\}$ is convex, then it is not necessary to add it in function F by introducing a slack variable, i.e., we may keep it in the abstract set W. Note also that, if we treat (1) as a standard nonlinear programming problem, the multipliers $\{\alpha,\beta,\zeta\}$ in (12) or (18)–(19) are just the usual Lagrange multipliers corresponding to the constraints $\{G(x)\geq 0, H(x)\geq 0, G(x)^T H(x)=0\}$, respectively. Moreover, we add slack variables in the above systems, so that the constraint sets become a polyhedron. In fact, if any function in $\{g_i, 1\leq i\leq p; G_j, H_j, 1\leq j\leq m\}$ is affine, then we may move it to the constraint set W, that is, we may not use a slack variable for it.

4 Modified LM Method for Constrained Equations

Consider the constrained equation (13), in which W is a nonempty, closed and convex subset of $\mathbb{R}^l$ and $F:\mathbb{R}^l\to\mathbb{R}^\nu$ is a differentiable function. The results given in this section are of independent interest. Throughout this section, we suppose that the solution set W^* of (13) is nonempty. Kanzow et al. [11] propose an LM-type method for solving constrained equations with a locally quadratic rate of convergence. Applying their results to (13) directly would require the local error bound condition $c\,\mathrm{dist}(w,W^*)\leq\|F(w)\|$ for $w\in\mathcal{B}_\delta(w^*)\cap W$, where $\mathrm{dist}(w,W^*)$ denotes the distance from w to W^*, $c>0$, $\delta>0$, $w^*\in W^*$, and $\mathcal{B}_\delta(w^*) := \{w\in\mathbb{R}^l \mid \|w-w^*\|<\delta\}$. It is well known that this error bound condition is equivalent to the calmness of the perturbed constrained equations as a set-valued mapping $S(p) := \{w \in W \mid F(w) = p\}$ around $(0, w^*)$. Hence, the above local error bound condition is weaker than the pseudo-Lipschitz continuity of the set-valued mapping around $(0, w^*)$, which is equivalent to the classical nondegeneracy condition, i.e., the Jacobian of F at w^* has maximum row rank, and there exists a vector d in the interior of the tangent cone of W at w^* such that $\nabla F_i(w^*)^T d = 0, i = 1, \ldots, \nu$; see, e.g., [8] for discussions on this topic.

Instead of using the regularization parameter in terms of $\|F(w)\|^2$ as in [11], we suggest to use a regularization parameter in terms of $\|F(w)\|^\sigma$ with $\sigma \in [1, 2]$, and our superlinear convergence result holds under the error bound condition

$$c \,\mathrm{dist}^{1/\gamma}(w, W^*) \le \|F(w)\|, \quad w \in \mathcal{B}_\delta(w^*) \cap W, \tag{20}$$

where γ is a suitable constant. Since our method reduces to the one in [11] when $\sigma = 2$ and $\gamma = 1$, and our assumptions are more general than the ones in [11], our results include the ones in [11] as a special case. In addition, the formula we derive below for the convergence rate also indicates that the parameter σ may be used to adjust the convergence rate since the bigger is the parameter σ, the smaller is the number τ.

Obviously, solving (13) is equivalent to solving the optimization problem

$$\min \quad \theta(w) := \frac{1}{2}\|F(w)\|^2 \quad \text{s.t.} \quad w \in W. \tag{21}$$

The LM method proposed for solving constrained equations in [11] determines the iterations by solving

$$\min \;\theta_k(w) := \frac{1}{2}\|F(w^k) + \nabla F(w^k)^T(w - w^k)\|^2 + \frac{\eta_k}{2}\|w - w^k\|^2 \quad \text{s.t.} \;\; w \in W, \tag{22}$$

where w^k is the current point, and $\eta_k > 0$ is a positive parameter. Since (22) is a strongly convex program, the iteration is well-defined. We now describe our modified method.

Algorithm 4.1.

Step 1: Choose $w^0 \in W$, $\eta > 0$, $\sigma \in [1, 2]$, and set $k := 0$.

Step 2: If $F(w^k) = 0$, stop. Otherwise, set $\eta_k := \eta\|F(w^k)\|^\sigma$, and solve problem (22) to get w^{k+1}.

Step 3: If $d^k := w^{k+1} - w^k = 0$, stop. Otherwise, let $k := k + 1$, and go to Step 2.

The regularization parameter η_k plays an important role in convergence analysis. Note that, in [11], this parameter is defined by $\eta\|F(w^k)\|^2$. However, as pointed out by Fan and Yuan in [12], the choice $\eta_k := \eta\|F(w^k)\|^2$ may have some unsatisfactory properties: When w^k is close to the solution set of (13), the parameter η_k gets very small, and hence it may lose its role in (22). On the contrary, if w^k is far from the solution set of (13), then η_k gets very large, so that the step d^k will be very small. In our new algorithm, the regularization parameter is chosen as $\eta_k := \eta\|F(w^k)\|^\sigma$ for $\sigma \in [1,2]$.

We next establish the convergence theory for the modified method. Our proof techniques are inspired by the recent work of Rehling and Fisher [13], in which an inexact version of a constrained LM method with similar regularization parameter is analyzed under the classical error bound condition with $\gamma = 1$.

Note that every point in W^* satisfies the necessary optimality condition $0 \in \nabla F(w)F(w) + \mathcal{N}_W(w)$, where $\mathcal{N}_W$ denotes the normal cone to W. Consider the perturbed generalized equation

$$s \in \nabla F(w)F(w) + \mathcal{N}_W(w) \tag{23}$$

with parameter $s \in \mathbb{R}^l$. Denote by W_s^* the solution set of (23). It is obvious that $W^* \subseteq W_0^*$.

In what follows, we assume that Algorithm 4.1 generates an infinite sequence of iterations. In order to establish the convergence theory of the algorithm, we make the following assumption throughout this section.

Assumption 4.1. There exist $w^* \in W^*$ and positive constants $\{c, \delta, \gamma\}$ with $\delta \in]0,1]$ and $\gamma \in [\frac{1}{2}, 1]$ such that (i) (20) holds; (ii) both F and ∇F are Lipschitz continuous in $\mathcal{B}_{2\delta}(w^*)$ with Lipschitz constant L.

The following lemma indicates that the set-valued mapping $s \rightrightarrows W_s^*$ is calm with exponent $\gamma/(2-\gamma)$ at w^*. It extends the recent result of Rehling and Fisher [13, Lemma 1] to the case where $\gamma \neq 1$.

Lemma 4.1. *There are $\hat{L} > 0$ and $\delta_1 > 0$ such that $W_s^* \cap \mathcal{B}_{\delta_1}(w^*) \subseteq W^* + \hat{L}\|s\|^{\gamma/(2-\gamma)}\mathcal{B}_1(0)$ for $s \in \mathbb{R}^l$.*

Proof Let $\delta_1 := \min\{\delta, (\frac{2c^2}{L^2})^{\gamma/(4\gamma-2)}\}$ and $s \in \mathbb{R}^l$. Let $W_s^* \cap \mathcal{B}_{\delta_1}(w^*) \neq \emptyset$ and $\hat{w} \in W_s^* \cap \mathcal{B}_{\delta_1}(w^*)$ be given. This means that there exists $\hat{y} \in \mathcal{N}_W(\hat{w})$ such that $s = \nabla F(\hat{w})F(\hat{w}) + \hat{y}$. Since W^* is nonempty and closed, there exists $\hat{w}^* \in W^*$ such that $\|\hat{w} - \hat{w}^*\| = \text{dist}(\hat{w}, W^*)$. It is easy to see that $\hat{w}^* \in \mathcal{B}_{2\delta}(w^*)$, and hence, by the mean-value theorem and Assumption 4.1, we have

$$\|F(\hat{w}) + \nabla F(\hat{w})^T(\hat{w}^* - \hat{w})\| \leq \|\hat{w}^* - \hat{w}\| \int_0^1 \|\nabla F(\hat{w} + t(\hat{w}^* - \hat{w})) - \nabla F(\hat{w})\|\, dt \leq \frac{L}{2}\,\text{dist}^2(\hat{w}, W^*). \tag{24}$$

Since $\hat{w}^* \in W, \hat{w} \in \mathcal{B}_\delta(w^*) \cap W$, and $\hat{y} \in \mathcal{N}_W(\hat{w})$, we have $(\hat{w}^* - \hat{w})^T\hat{y} \leq 0$, and hence, by (20)–(24),

$$(\hat{w}^* - \hat{w})^T s \leq (\hat{w}^* - \hat{w})^T \nabla F(\hat{w})F(\hat{w}) \leq \frac{L^2}{8}\,\text{dist}^4(\hat{w}, W^*) - \frac{c^2}{2}\,\text{dist}^{2/\gamma}(\hat{w}, W^*). \tag{25}$$

Noting that $\gamma \geq \frac{1}{2}$, $\hat{w} \in \mathcal{B}_{\delta_1}(w^*)$, and $\delta_1 \leq \big(\frac{2c^2}{L^2}\big)^{\gamma/(4\gamma-2)}$, we have

$$\frac{L^2}{8}\operatorname{dist}^{4-2/\gamma}(\hat{w}, W^*) \leq \frac{L^2}{8}\|\hat{w} - w^*\|^{4-2/\gamma} \leq \frac{L^2}{8}\delta_1^{4-2/\gamma} \leq \frac{1}{4}c^2,$$

and hence, by (25),

$$\frac{c^2}{4}\operatorname{dist}^{2/\gamma}(\hat{w}, W^*) \leq \frac{c^2}{2}\operatorname{dist}^{2/\gamma}(\hat{w}, W^*) - \frac{L^2}{8}\operatorname{dist}^{4}(\hat{w}, W^*) \leq -(\hat{w}^* - \hat{w})^T s \leq \|s\|\operatorname{dist}(\hat{w}, W^*),$$

i.e., $\operatorname{dist}(\hat{w}, W^*) \leq \Big(\frac{4}{c^2}\|s\|\Big)^{\gamma/(2-\gamma)}$. This indicates that the conclusion holds with $\hat{L} := (\frac{4}{c^2})^{\gamma/(2-\gamma)}$. □

Lemma 4.2. *Let $w^k \in \mathcal{B}_\delta(w^*) \cap W$. There exists $\kappa > 0$ such that $\|w^{k+1} - w^k\| \leq \kappa \operatorname{dist}^{\gamma_1}(w^k, W^*)$, where $\gamma_1 := \min\{2 - \frac{\sigma}{2\gamma}, 1\}$.*

Proof Let $w^k \in \mathcal{B}_\delta(w^*) \cap W$. Since W^* is nonempty and closed, there exists $\hat{w}^k \in W^*$ such that $\|w^k - \hat{w}^k\| = \operatorname{dist}(w^k, W^*)$. It is easy to see that $\hat{w}^k \in \mathcal{B}_{2\delta}(w^*)$. Thus, by Assumption 4.1, we have

$$\eta c^\sigma \|w^k - \hat{w}^k\|^{\sigma/\gamma} = \eta c^\sigma \operatorname{dist}^{\sigma/\gamma}(w^k, W^*) \leq \eta_k = \eta\|F(w^k)\|^\sigma. \tag{26}$$

Since w^{k+1} is the unique minimizer of problem (22), we have

$$\frac{\eta_k}{2}\|w^{k+1} - w^k\|^2 \leq \theta_k(w^{k+1}) \leq \theta_k(\hat{w}_k). \tag{27}$$

As a result, we have from (24) and (26)–(27) that

$$\begin{aligned} \|w^{k+1} - w^k\|^2 &\leq \frac{2}{\eta_k}\theta_k(\hat{w}^k) \leq \frac{1}{\eta_k}\|F(w^k) + \nabla F(w^k)^T(\hat{w}^k - w^k)\|^2 + \|\hat{w}^k - w^k\|^2 \\ &\leq \frac{L^2}{4\eta_k}\|\hat{w}^k - w^k\|^4 + \|\hat{w}^k - w^k\|^2 \leq \frac{L^2}{4\eta c^\sigma}\|\hat{w}^k - w^k\|^{4-\sigma/\gamma} + \|\hat{w}^k - w^k\|^2. \end{aligned}$$

Letting $\kappa := \sqrt{1 + \frac{L^2}{4\eta c^\sigma}}$, by $\|w^k - \hat{w}^k\| \leq \|w^k - w^*\| \leq \delta \leq 1$, we get the conclusion immediately. □

Given $w^k \in \mathbb{R}^l$, we define

$$\Theta_k(w) := \nabla F(w)F(w) - \nabla F(w^k)(F(w^k) + \nabla F(w^k)^T(w - w^k)) - \eta_k(w - w^k). \tag{28}$$

Lemma 4.3. *There exist $C > 0$ and $\delta_2 > 0$ such that $\|\Theta_k(w^{k+1})\| \leq C \operatorname{dist}^{\gamma_2}(w^k, W^*)$ when the iteration $w^k \in \mathcal{B}_{\delta_2}(w^*) \cap W$, where $\gamma_2 := \min\{\sigma + \gamma_1, 2\gamma_1\} = \min\{\frac{4\gamma-\sigma}{\gamma}, 2\}$.*

Proof Let $\delta_2 := \min\left\{\delta, \left(\frac{\delta}{\kappa}\right)^{1/\gamma_1}\right\}$ and $w^k \in \mathcal{B}_{\delta_2}(w^*) \cap W$. Let $\hat{w}^k$ be defined as in the proof of Lemma 4.2. It is easy to see that $\hat{w}^k \in \mathcal{B}_{2\delta}(w^*)$, and, by Lemma 4.2,

$$\|w^{k+1} - w^*\| \leq \|w^{k+1} - w^k\| + \|w^k - w^*\| \leq \kappa \operatorname{dist}^{\gamma_1}(w^k, W^*) + \delta_2 \leq 2\delta,$$

namely, $w^{k+1} \in \mathcal{B}_{2\delta}(w^*)$. We then have from Assumption 4.1 and Lemma 4.2 that

$$\begin{aligned} \|F(w^{k+1})\| &\leq L\|w^{k+1} - \hat{w}^k\| \leq L(\|w^{k+1} - w^k\| + \|w^k - \hat{w}^k\|) \\ &\leq L(\kappa \operatorname{dist}^{\gamma_1}(w^k, W^*) + \operatorname{dist}(w^k, W^*)) \leq L(\kappa + 1) \operatorname{dist}^{\gamma_1}(w^k, W^*), \end{aligned} \tag{29}$$

where the last inequality follows from the fact that $\operatorname{dist}(w^k, W^*) \leq \|w^k - w^*\| \leq \delta_2 \leq \delta \leq 1$. Note that $\eta_k = \eta\|F(w^k) - F(\hat{w}^k)\|^\sigma \leq \eta L^\sigma \|w^k - \hat{w}^k\|^\sigma$, and Assumption 4.1 implies the boundedness of $\|\nabla F\|$ on $\mathcal{B}_{2\delta}(w^*)$, i.e., there exists $L' > 0$ such that $\|\nabla F(w)\| \leq L'$ for $w \in \mathcal{B}_{2\delta}(w^*)$. It then follows from (28)–(29), Assumption 4.1, Lemma 4.2, and the mean-value theorem that

$$\begin{aligned} \|\Theta_k(w^{k+1})\| &\leq \|\nabla F(w^{k+1}) - \nabla F(w^k)\|\|F(w^{k+1})\| \\ &\quad + \|\nabla F(w^k)\|\|F(w^{k+1}) - F(w^k) - \nabla F(w^k)^T(w^{k+1} - w^k)\| + \eta_k\|w^{k+1} - w^k\| \\ &\leq L\|w^{k+1} - w^k\|\|F(w^{k+1})\| + L'L\|w^{k+1} - w^k\|^2 + \eta L^\sigma\|w^k - \hat{w}^k\|^\sigma\|w^{k+1} - w^k\| \\ &\leq C \operatorname{dist}^{\gamma_2}(w^k, W^*), \end{aligned}$$

where $\gamma_2 := \min\{\gamma_1 + \sigma, 2\gamma_1\}$ and $C := L^2\kappa(\kappa + 1) + L'L\kappa^2 + \eta L^\sigma \kappa$. This completes the proof. □

Lemma 4.4. *There exist* $C' > 0$ *and* $\delta_3 > 0$ *such that* $\operatorname{dist}(w^{k+1}, W^*) \leq C' \operatorname{dist}^\tau(w^k, W^*)$ *when the iteration* $w^k \in \mathcal{B}_{\delta_3}(w^*) \cap W$*, where* $\tau := \frac{\gamma_2\gamma}{2-\gamma} = \min\{\frac{4\gamma-\sigma}{2-\gamma}, \frac{2\gamma}{2-\gamma}\}$.

Proof Let $\delta_3 := \min\left\{\delta_2, \frac{\delta_1}{2}, \left(\frac{\delta_1}{2\kappa}\right)^{1/\gamma_1}\right\}$ and $w^k \in \mathcal{B}_{\delta_3}(w^*) \cap W$. Note that problem (22) is equivalent to the generalized equation $\Theta_k(w) \in \nabla F(w)F(w) + \mathcal{N}_W(w)$. This means that w^{k+1} is a unique solution of the perturbed generalized equation (23) with $s := \Theta_k(w^{k+1})$. Then, we have $w^{k+1} \in W^*_{\Theta_k(w^{k+1})}$. Since

$$\|w^{k+1} - w^*\| \leq \|w^{k+1} - w^k\| + \|w^k - w^*\| \leq \kappa \operatorname{dist}^{\gamma_1}(w^k, W^*) + \delta_3 \leq \delta_1$$

by Lemma 4.2, we have $w^{k+1} \in W^*_{\Theta_k(w^{k+1})} \cap \mathcal{B}_{\delta_1}(w^*)$. It follows from Lemmas 4.1 and 4.3 that

$$\operatorname{dist}(w^{k+1}, W^*) \leq \hat{L}\|\Theta_k(w^{k+1})\|^{\gamma/(2-\gamma)} \leq \hat{L}C^{\gamma/(2-\gamma)} \operatorname{dist}^{\gamma_2\gamma/(2-\gamma)}(w^k, W^*).$$

We obtain the conclusion immediately by letting $C' := \hat{L}C^{\gamma/(2-\gamma)}$. □

Lemma 4.5. *Assume that the number τ given in Lemma 4.4 is larger than one, that is, $\gamma > \max\{\frac{2}{3}, \frac{2+\sigma}{5}\}$. Assume also that, in Algorithm 4.1, the starting point $w^0 \in \mathcal{B}_{\delta_0}(w^*)$, where*

$$\delta_0 := \min\left\{\tfrac{1}{2}{C'}^{1/(1-\tau)}, \tfrac{\delta_3}{2}, \left(\tfrac{\delta_3}{2a\kappa}\right)^{1/\gamma_1}\right\}, \tag{30}$$

and $a := \sum_{i=0}^{\infty} \frac{1}{2^{\gamma_1(\tau^i-1)}} < +\infty$. Then, the sequence $\{w^k\}$ generated by Algorithm 4.1 is contained in $\mathcal{B}_{\delta_3}(w^*)$.

Proof It is obvious that $w^0 \in \mathcal{B}_{\delta_3}(w^*)$. Therefore, by mathematical induction, it is sufficient to show that, for any k, $\{w^0, w^1, \cdots, w^k\} \subset \mathcal{B}_{\delta_3}(w^*)$ implies $w^{k+1} \in \mathcal{B}_{\delta_3}(w^*)$. In fact,

$$\begin{aligned} \|w^{k+1} - w^*\| &\leq \|w^k - w^*\| + \|d^k\| \leq \|w^{k-1} - w^*\| + \|d^{k-1}\| + \|d^k\| \\ &\leq \cdots\cdots \leq \|w^0 - w^*\| + \sum_{i=0}^{k}\|d^i\| \leq \delta_0 + \kappa\sum_{i=0}^{k}\operatorname{dist}^{\gamma_1}(w^i, W^*), \end{aligned} \tag{31}$$

where the last inequality follows from Lemma 4.2. Since $\{w^0, w^1, \cdots, w^k\} \subset \mathcal{B}_{\delta_3}(w^*)$, we have from Lemma 4.4 that $\operatorname{dist}(w^i, W^*) \leq C' \operatorname{dist}^{\tau}(w^{i-1}, W^*)$ for each i. It follows that, for each $i = 1, \cdots, k$,

$$\operatorname{dist}(w^i, W^*) \leq {C'}^{1+\tau}\operatorname{dist}^{\tau^2}(w^{i-2}, W^*) \leq \cdots\cdots \leq {C'}^{1+\tau+\cdots+\tau^{i-1}}\operatorname{dist}^{\tau^i}(w^0, W^*) \leq {C'}^{\frac{\tau^i-1}{\tau-1}}\delta_0^{\tau^i}. \tag{32}$$

From the definition of δ_0 and (31)–(32), we have

$$\begin{aligned} \|w^{k+1} - w^*\| &\leq \delta_0 + \kappa\sum_{i=0}^{k}\operatorname{dist}^{\gamma_1}(w^i, W^*) \leq \delta_0 + \kappa\sum_{i=0}^{k}\left({C'}^{\frac{\tau^i-1}{\tau-1}}\delta_0^{\tau^i}\right)^{\gamma_1} \\ &\leq \delta_0 + \kappa\delta_0^{\gamma_1}\sum_{i=0}^{k}\left({C'}^{\frac{\gamma_1}{\tau-1}}\delta_0^{\gamma_1}\right)^{\tau^i-1} \leq \delta_0 + \kappa\delta_0^{\gamma_1}a \leq \delta_3. \end{aligned}$$

This completes the proof. □

Now, we give the first convergence result.

Theorem 4.1. *Assume that $\gamma > \max\{\frac{2}{3}, \frac{2+\sigma}{5}\}$ and $\{w^k\}$ is a sequence generated by Algorithm 4.1 with starting point $w^0 \in \mathcal{B}_{\delta_0}(w^*)$, where δ_0 is defined as in (30). Then, the sequence $\{\operatorname{dist}(w^k, W^*)\}$ converges superlinearly to zero with order no less than $\tau = \min\{\frac{4\gamma-\sigma}{2-\gamma}, \frac{2\gamma}{2-\gamma}\}$.*

Proof It follows from Lemma 4.5 that $\{w^k\}$ is contained in the ball $\mathcal{B}_{\delta_3}(w^*)$. Noting that $\delta_0 \leq \delta_3 \in (0, 1)$ and $\tau > 1$, we have from (32) that $\operatorname{dist}(w^k, W^*) \to 0$ as $k \to \infty$. Furthermore, by Lemma 4.4, there holds

$\limsup_{k\to\infty} \frac{\mathrm{dist}(w^{k+1},W^*)}{\mathrm{dist}^\tau(w^k,W^*)} \le C'$. In consequence, $\{\mathrm{dist}(w^k, W^*)\}$ converges superlinearly to zero with order no less than $\tau > 1$. This completes the proof. □

We next investigate the convergence of the sequence $\{w^k\}$ generated by Algorithm 4.1.

Theorem 4.2. *Assume that $\gamma > \max\{\frac{2}{3}, \frac{2+\sigma}{5}\}$ and $\gamma \ge \frac{\sigma}{2}$. Let $\{w^k\}$ be a sequence generated by Algorithm 4.1 with starting point $w^0 \in \mathcal{B}_{\delta_0}(w^*)$, where δ_0 is defined as in (30). Then, $\{w^k\}$ converges superlinearly to a solution of problem (13) with order no less than $\tau = \min\{\frac{4\gamma-\sigma}{2-\gamma}, \frac{2\gamma}{2-\gamma}\}$.*

Proof Note that $\gamma \ge \frac{\sigma}{2}$ implies $\gamma_1 = \min\{2 - \frac{\sigma}{2\gamma}, 1\} = 1$. It follows from Lemma 4.2 and Theorem 4.1 that, for any k sufficiently large and any i,

$$\begin{aligned} \|w^{k+i} - w^k\| &\le \|w^{k+i-1} - w^k\| + \|d^{k+i-1}\| \le \cdots\cdots \le \sum_{\iota=1}^{i} \|d^{k+\iota-1}\| \\ &\le \kappa \sum_{\iota=1}^{i} \mathrm{dist}(w^{k+\iota-1}, W^*) \le \kappa \sum_{\iota=0}^{i-1} \frac{1}{2^\iota} \mathrm{dist}(w^k, W^*) \le 2\kappa\, \mathrm{dist}(w^k, W^*). \end{aligned} \tag{33}$$

This means that $\{w^k\}$ is a Cauchy sequence, and hence $\{w^k\}$ converges to a solution, say $\bar{w}$, of (13).

It remains to show that the convergence of $\{w^k\}$ is superlinear. In fact, letting $i \to \infty$ in (33), we get $\|w^k - \bar{w}\| \le 2\kappa\, \mathrm{dist}(w^k, W^*)$. Combining this with $\|w^{k-1} - \bar{w}\| \ge \mathrm{dist}(w^{k-1}, W^*)$, we have from the proof of Theorem 4.1 that

$$\limsup_{k\to\infty} \frac{\|w^k - \bar{w}\|}{\|w^{k-1} - \bar{w}\|^\tau} \le \limsup_{k\to\infty} \frac{2\kappa\, \mathrm{dist}(w^k, W^*)}{\mathrm{dist}^\tau(w^{k-1}, W^*)} \le 2\kappa C'.$$

That is, $\{w^k\}$ converges to $\bar{w} \in W^*$ superlinearly with order no less than $\tau > 1$. □

As shown above, under the assumptions of Theorem 4.1 or 4.2, $\{\mathrm{dist}(w^k, W^*)\}$ or $\{w^k\}$ converges superlinearly with order no less than $\tau = \min\{\frac{4\gamma-\sigma}{2-\gamma}, \frac{2\gamma}{2-\gamma}\}$. In particular, provided that the local error bound condition (20) holds with $\gamma = 1$, the rate of convergence is at least quadratic for any $\sigma \in [1, 2]$.

As shown in Section 3, the C-/M-/S-stationarity conditions can be reformulated as constrained equations. Now, we apply the results given above to solve these equations. Note that, since all the functions $\{f, g, h, G, H\}$ are assumed to be twice differentiable, and their second order derivatives are locally Lipschitzian, condition (ii) in Assumption 4.1 is satisfied by the constrained equation (13) with the mapping F defined by (14), (16), (18), and the set W defined by (15), (17), (19), respectively. As a result, the convergence of Algorithm 4.1 only needs the local error bound condition (20). We have the following convergence results from Theorems 4.1 and 4.2 immediately.

Theorem 4.3. *Consider the C-stationarity system, M-stationarity system, and S-stationarity system as constrained equations defined by* (14)–(15), (16)–(17), *and* (18)–(19), *respectively. Suppose that there exist some* $w^* \in W^*$ *and constants* $c > 0$, $\delta \in]0,1]$, $\gamma > \max\{\frac{2}{3}, \frac{2+\sigma}{5}\}$ *such that the error bound condition* (20) *holds, and let* $\{w^k\}$ *be a sequence generated by Algorithm* 4.1 *with starting point* $w^0 \in \mathcal{B}_{\delta_0}(w^*)$, *where* δ_0 *is defined as in Section* 4. *Then,* $\{\text{dist}(w^k, W^*)\}$ *converges to zero superlinearly, and, if* $\gamma \geq \frac{\sigma}{2}$ *holds,* $\{w^k\}$ *converges to a solution superlinearly with order no less than* $\tau = \min\{\frac{4\gamma-\sigma}{2-\gamma}, \frac{2\gamma}{2-\gamma}\}$.

5 Sufficient Conditions for Error Bounds

Since the success of Algorithm 4.1 depends on the existence of an error bound with exponent greater than the number $\max\{\frac{2}{3}, \frac{2+\sigma}{5}\}$, we devote this section to the study of existence of the error bounds with exponent greater than $\max\{\frac{2}{3}, \frac{2+\sigma}{5}\}$. Since $\|F(w)\|$ is close to zero when w is close to the solution w^*, it is obvious that $\|F(w)\| \leq \|F(w)\|^\gamma$ when $\gamma \leq 1$ for such w. Therefore, the local error bound condition with exponent $\gamma = 1$ implies the local error bound condition with $\gamma < 1$, but not vice versa. In the literature, there are some sufficient conditions for existence of local error bounds with exponent not equal to 1 (see, e.g., [14]), but they are usually not easy to verify. We next give an example to show that, for the C-/M-/S-stationarity systems, it is possible to have an error bound with $\gamma < 1$.

Consider the MPEC

$$\min\ x_1 + x_2 + x_3^{\frac{11}{5}} \quad \text{s.t.}\ \ x_3^{\frac{12}{5}} = 0,\ 0 \leq x_1 \perp x_2 \geq 0.$$

The S-stationarity system (18)–(19) for the above problem is $F_S(w) = 0$ with $w \in W_S$, where

$$F_S(x_1, x_2, x_3, \mu, \alpha, \beta, \zeta) := \begin{pmatrix} 1-\alpha+\zeta x_2 \\ 1-\beta+\zeta x_1 \\ \frac{11}{5}x_3^{6/5}+\frac{12}{5}x_3^{7/5}\mu \\ x_3^{12/5} \\ x_1 x_2 \\ \alpha x_1 \\ \beta x_2 \end{pmatrix}$$

and $W_S := \{w = (x_1, x_2, x_3, \mu, \alpha, \beta, \zeta) \mid x_1 \geq 0, x_2 \geq 0, \alpha \geq 0, \beta \geq 0\}$. It is easy to see that the set of S-stationary points is $W_S^* := \{(0,0,0,\mu,1,1,\zeta) \mid \mu \in]-\infty,+\infty[, \zeta \in]-\infty,+\infty[\}$. We first show that $\|F_S(w)\|$ cannot provide a local error bound. In fact, for any $w^* \in W_S^*$ and any $w \notin W_S^*$, we have

$$\frac{\text{dist}^2(w, W_S^*)}{\|F_S(w)\|^2} = \frac{x_1^2 + x_2^2 + x_3^2 + (\alpha-1)^2 + (\beta-1)^2}{(1-\alpha+\zeta x_2)^2 + (1-\beta+\zeta x_1)^2 + x_3^{\frac{12}{5}}(\frac{11}{5} + \frac{12}{5}x_3^{\frac{1}{5}}\mu)^2 + x_3^{\frac{24}{5}} + x_1^2x_2^2 + \alpha^2x_1^2 + \beta^2x_2^2}.$$

Taking the limit route of $x_1 = x_2 = 0$, $\alpha = \beta = 1$, $\mu = \mu^*$, $\zeta = \zeta^*$, and $x_3 = t$ with $t \to 0^+$, we can get

$$\limsup_{w \to w^*} \frac{\mathrm{dist}^2(w, W_S^*)}{\|F_S(w)\|^2} = +\infty,$$

which indicates that $\|F_S(w)\|$ cannot provide a local error bound. We next show that $\|F_S(w)\|^{\frac{5}{6}}$ can provide a local error bound. In fact, noting that $(\alpha - 1)^2 \leq 2(1 - \alpha + \zeta x_2)^2 + 2\zeta^2 x_2^2$ and $(\beta - 1)^2 \leq 2(1 - \beta + \zeta x_1)^2 + 2\zeta^2 x_1^2$, we have

$$\begin{aligned}
\frac{\mathrm{dist}^2(w, W_S^*)}{\|F_S(w)\|^{\frac{5}{3}}} &\leq \frac{(1+2\zeta^2)x_1^2 + (1+2\zeta^2)x_2^2 + x_3^2 + 2(1-\alpha+\zeta x_2)^2 + 2(1-\beta+\zeta x_1)^2}{\left((1-\alpha+\zeta x_2)^2 + (1-\beta+\zeta x_1)^2 + x_3^{\frac{12}{5}}(\frac{11}{5} + \frac{12}{5}x_3^{\frac{1}{5}}\mu)^2 + x_3^{\frac{24}{5}} + x_1^2 x_2^2 + \alpha^2 x_1^2 + \beta^2 x_2^2\right)^{\frac{5}{6}}} \\
&\leq \frac{(1+2\zeta^2)x_1^2}{|\alpha x_1|^{\frac{5}{3}}} + \frac{(1+2\zeta^2)x_2^2}{|\beta x_2|^{\frac{5}{3}}} + \frac{x_3^2}{x_3^2|\frac{11}{5} + \frac{12}{5}x_3^{\frac{1}{5}}\mu|^{\frac{5}{3}}} + \frac{2(1-\alpha+\zeta x_2)^2}{|1-\alpha+\zeta x_2|^{\frac{5}{3}}} + \frac{2(1-\beta+\zeta x_1)^2}{|1-\beta+\zeta x_1|^{\frac{5}{3}}} \\
&\to \left(\frac{5}{11}\right)^{\frac{5}{3}} \quad \text{as} \quad w \to w^*.
\end{aligned}$$

This means that the error bound condition (20) holds with $\gamma = \frac{5}{6}$, $c = ((\frac{5}{11})^{\frac{5}{3}} + 1)^{-\frac{1}{2}}$, and some appropriate $\delta > 0$, that is, $\|F_S(w)\|^{\frac{5}{6}}$ can provide a local error bound for the S-stationarity system.

The following three conditions are well-known sufficient conditions for the local error bound with $\gamma = 1$:

(C1) F is affine, and W is a polyhedron (by the well-known result of Hoffman's bound [15]).

(C2) The classical nondegeneracy condition, or equivalently, the MFCQ for the constrained system at w^*, i.e., the gradients $\{\nabla F_1(w^*), \nabla F_2(w^*), \cdots, \nabla F_\nu(w^*)\}$ are linearly independent, and there exists a vector d in the interior of the tangent cone $\mathcal{T}_W(w^*)$ such that $\nabla F_i(w^*)^T d = 0$ for each $i = 1, \cdots, \nu$, or equivalently, there is no nonzero vector $\eta \in R^\nu$ such that $0 \in \sum_{i=1}^{\nu} \nabla F_i(w^*)\eta_i + \mathcal{N}_W(w^*)$; see, e.g., [16, page 546].

(C3) The LICQ for the constrained system holds at w^*, i.e., w^* is in the interior of W, and the gradient vectors $\{\nabla F_1(w^*), \nabla F_2(w^*), \cdots, \nabla F_\nu(w^*)\}$ are linearly independent, or equivalently, w^* is in the interior of W, and $\nabla F(w^*)$ has maximal column rank.

Except the first criterion, the other criteria are based on the point w^*. Moreover, these criteria are actually much stronger than the existence of local error bounds. Since Hoffman shows in [15] that linear systems always have global error bounds in 1952, many researchers have tried to find weaker sufficient conditions for the existence of error bounds. In particular, Minchenko and Stakhovski [17] show the existence of error bounds under the relax constant regularity condition that is weaker than the criteria (C1) and (C3). Guo et al. [18] obtain the existence of error bounds under the quasi-normality condition that is weaker than the criteria (C1)–(C3). Other criteria for local error bounds with exponent $\gamma \neq 1$ can be found in, e.g., [14].

Unfortunately, for the C-/M-/S-stationarity systems, due to the complementarity constraints, the criterion (C1) never holds, and the point-based criteria (C2)–(C3) and other weaker criteria in the literature are unlikely to hold. Nevertheless, since the criteria (C1)–(C3) are stronger than the existence of error bounds with $\gamma = 1$, it does not mean that the error bound condition is unlikely to hold with $\gamma = 1$. It turns out that, by eliminating certain components of the mapping F, we could still derive the existence of error bounds with $\gamma = 1$ by making use of the above criteria and other weaker criteria in the literature. For this purpose, we first introduce a lemma given in [19].

Lemma 5.1. *Let δ be a positive constant, and $\Omega \subseteq \mathbb{R}^l$ be a nonempty and closed set. If $w \in \Omega$ and $y \in \mathcal{B}_{\delta/2}(w)$, then* $\operatorname{dist}(y, \Omega) = \operatorname{dist}(y, \Omega \cap \mathcal{B}_\delta(w))$.

We now present the error bound results by elimination.

Theorem 5.1. *Let W^* be the solution set of* (13)*, and $w^* \in W^*$. Suppose that $\bar{F}$ is constituted by some components of F, and $\bar{W}^* := \{w \,|\, \bar{F}(w) = 0\}$. If there exist $\delta > 0$, $c > 0$, and $\gamma > 0$ such that*

$$c \operatorname{dist}^{1/\gamma}(w, \bar{W}^*) \leq \|\bar{F}(w)\|, \qquad w \in \mathcal{B}_{\delta/2}(w^*), \tag{34}$$

$$\bar{W}^* \cap \mathcal{B}_\delta(w^*) = W^* \cap \mathcal{B}_\delta(w^*), \tag{35}$$

then there holds $c \operatorname{dist}^{1/\gamma}(w, W^*) \leq \|F(w)\|$ for $w \in \mathcal{B}_{\delta/2}(w^*) \cap W$.

Proof By Lemma 5.1, (34) implies $c \operatorname{dist}^{1/\gamma}(w, \bar{W}^* \cap \mathcal{B}_\delta(w^*)) \leq \|\bar{F}(w)\|$ for $w \in \mathcal{B}_{\delta/2}(w^*)$. By (35), it implies that $c \operatorname{dist}^{1/\gamma}(w, W^* \cap \mathcal{B}_\delta(w^*)) \leq \|\bar{F}(w)\| \leq \|F(w)\|$ for $w \in \mathcal{B}_{\delta/2}(w^*) \cap W$. Applying Lemma 5.1 again, we obtain the conclusion. □

In a similar way, we have the following theorem.

Theorem 5.2. *Let $w^* \in W^*$, $\hat{F}$ be constituted by some components of F, and $\hat{W}^* := \{w \in W \,|\, \hat{F}(w) = 0\}$. If there exist $\delta > 0$, $c > 0$, and $\gamma > 0$ such that $c \operatorname{dist}^{1/\gamma}(w, \hat{W}^*) \leq \|\hat{F}(w)\|$ for $w \in \mathcal{B}_{\delta/2}(w^*) \cap W$, and $\hat{W}^* \cap \mathcal{B}_\delta(w^*) = W^* \cap \mathcal{B}_\delta(w^*)$, then we have $c \operatorname{dist}^{1/\gamma}(w, W^*) \leq \|F(w)\|$ for $w \in \mathcal{B}_{\delta/2}(w^*) \cap W$.*

We now illustrate the applications of Theorem 5.1 by Example 2.1. The C-stationarity system (14) and (15) are $F_C(w) = 0$ with $w \in W_C$, where

$$F_C(x_1, x_2, x_3, y, \lambda, u, v) := \begin{pmatrix} 1-\lambda-u \\ -2+\lambda-v \\ x_1-x_2-x_3 \\ x_1x_2 \\ \lambda x_3 \\ x_1u \\ x_2v \\ y-uv \end{pmatrix}$$

and $W_C := \{(x_1, x_2, x_3, y, \lambda, u, v) \mid x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, y \geq 0, \lambda \geq 0\}$. It is not difficult to see that the set of C-stationary points is $W_C^* := \{(0, 0, 0, (1-\lambda)(\lambda-2), \lambda, 1-\lambda, \lambda-2) \mid 1 \leq \lambda \leq 2\}$. Let w^* be the C-stationary point corresponding to $\lambda = \frac{3}{2}$. By analyzing the transposed Jacobian $\nabla F_C(w^*)$, we eliminate the third and fourth components of F_C, and let

$$\bar{F}_C(x_1, x_2, x_3, y, \lambda, u, v) := \begin{pmatrix} 1-\lambda-u \\ -2+\lambda-v \\ \lambda x_3 \\ x_1 u \\ x_2 v \\ y-uv \end{pmatrix},$$

whose Jacobian has full row rank at w^*. It follows that (34) holds with exponent $\gamma = 1$. Furthermore, when $\delta > 0$ is small sufficiently, we can get by straightforward calculation that the solution set after elimination does not change, that is, $\{w \in \mathcal{B}_\delta(w^*) \mid \bar{F}_C(w) = 0\} = W_C^* \cap \mathcal{B}_\delta(w^*)$. Consequently, by Theorem 5.1, the local error bound with $\gamma = 1$ holds for the equivalent C-stationarity system in Example 2.1.

We next consider some examples from the MPEC literature, and try to verify conditions (34)–(35) with $\gamma = 1$ using the elimination method. We omit the verification process, and state the results only.

Example 5.1. [20] Consider the problem

$$\min\ (x_1 + x_2) \quad \text{s.t.} \ \ x_2^2 \geq 1,\ 0 \leq x_1 \perp x_2 \geq 0.$$

The weak stationarity conditions are

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 2x_2 \end{pmatrix} \lambda - \begin{pmatrix} 1 \\ 0 \end{pmatrix} u - \begin{pmatrix} 0 \\ 1 \end{pmatrix} v = 0,\ \min(\lambda, x_2^2 - 1) = \min(x_1, x_2) = x_1 u = x_2 v = 0,$$

which yield a unique weakly stationary point $(0, 1)$ with multipliers $u = 1, v = 0$, and $\lambda = \frac{1}{2}$. It is obviously an S-stationary point and the unique global minimizer.

Example 5.2. [6, 20] Consider the problem

$$\min\ x_1 + x_2 - x_3 \quad \text{s.t.} \quad -4x_1 + x_3 \leq 0,\ -4x_2 + x_3 \leq 0,\ 0 \leq x_1 \perp x_2 \geq 0.$$

The weak stationarity conditions are

$$\begin{gathered} \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} + \begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix} \lambda_1 + \begin{pmatrix} 0 \\ -4 \\ 1 \end{pmatrix} \lambda_2 - \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} u - \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} v = 0, \\ \min(\lambda_1, 4x_1 - x_3) = \min(\lambda_2, 4x_2 - x_3) = \min(x_1, x_2) = x_1 u = x_2 v = 0, \end{gathered}$$

which yield a unique weakly stationary point $(0, 0, 0)$ with multipliers $u = 4\lambda_2 - 3, v = 1 - 4\lambda_2,\ \lambda_1 = 1 - \lambda_2,$

and $\lambda_2 \in [0,1]$. It is obvious that $(0,0,0)$ is M-stationary, but not S-stationary. Moreover, $(0,0,0)$ is the unique global minimizer.

Example 5.3. [20] Consider the problem

$$\min\ -x_1 - \frac{1}{2}x_2 \quad \text{s.t.}\ \ x_1 + x_2 \le 2,\ 0 \le x_1^2 - x_1 \perp x_2 \ge 0.$$

The weak stationarity conditions are

$$\begin{pmatrix} -1 \\ -\frac{1}{2} \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix}\lambda - \begin{pmatrix} 2x_1-1 \\ 0 \end{pmatrix}u - \begin{pmatrix} 0 \\ 1 \end{pmatrix}v = 0, \min(\lambda, 2-x_1-x_2) = \min(x_1^2-x_1, x_2) = (x_1^2-x_1)u = x_2 v = 0,$$

which yield five weakly stationary points: $(2,0)$ with multipliers $u = 0, v = \frac{1}{2}$, and $\lambda = 1$; $(1,0)$ with multipliers $u = -1, v = -\frac{1}{2}$, and $\lambda = 0$; $(0,0)$ with multipliers $u = 1, v = -\frac{1}{2}$, and $\lambda = 0$; $(0,2)$ with multipliers $u = \frac{1}{2}, v = 0$, and $\lambda = \frac{1}{2}$; $(1,1)$ with multipliers $u = -\frac{1}{2}, v = 0$, and $\lambda = \frac{1}{2}$. It is not difficult to see that $\{(2,0),(0,2),(1,1)\}$ are S-stationary, and $(1,0)$ is C-stationary, while $(0,0)$ is only weakly stationary. In addition, $(0,2)$ is a local minimizer, while $(2,0)$ is the unique global minimizer.

Tab. 1: Verification results for Examples 5.1–5.3[a]

	C-system	M-system	S-system
Example 5.1	Yes	Yes	Yes
Example 5.2	Yes	No	$\emptyset$
Example 5.3	Yes	No	Yes

[a] "Yes" means that one can find some $\bar{F}$ such that condition (35) holds, and $\nabla \bar{F}(w^*)$ has full column rank. "No" means the converse, while $\emptyset$ means that the system has no solution.

The verification results given in Tab. 1 reveal that conditions (34)–(35) with $\gamma = 1$ may be satisfied in many cases. Note that, even in the M-systems for Examples 5.2 and 5.3, conditions (34)–(35) may still hold since the nonsingularity of Jabobians is only a sufficient condition for the existence of error bounds.

6 Numerical Results

In this section, we compare the performance of Algorithm 4.1 with the methods presented in [21, 22] on Examples 2.1–2.4 and 5.1–5.3. In our experiments, we chose all the starting points to be $(5,5,\cdots,5)$, the parameters in the partial augmented Lagrangian method were the same as in [21], and the parameters in the $\ell_{1/2}$ penalty method were almost the same as in [22], except that the initial penalty parameter was chosen to be 10 instead of 1[1]. In addition, for Algorithm 4.1, we set the parameter $\eta = 0.1$, and terminated the iteration

[1] In fact, when we chose the initial penalty parameter to be 1, the numerical results obtained are not satisfactory.

if $\|F(w^k)\| \leq 10^{-6}$ or $\|d^k\| \leq 10^{-6}$. The numerical results were reported in Tabs. 2–4, respectively. In the tables, the values of variables and multipliers denote the values obtained within 100 iterations, *Iter* denotes the number of iterations by solving the corresponding approximation problems, and (u^k, v^k) is defined by $u^k := \alpha^k - \zeta^k H(x^k), v^k := \beta^k - \zeta^k G(x^k)$ for the S-systems. In particular,

- in Examples 2.1 and 5.2, since all constraint functions are affine, all local minimizers must be M-stationary, and hence, we only solved the M-systems;
- in Example 5.1, since the MPEC-LICQ holds, any local minimizer must be S-stationary, and hence, we only solved the S-system;
- in Example 2.2, since both S- and M-stationary points do not exist, we only solved the C-system;
- in Examples 2.3–2.4 and 5.3, since we cannot make sure which kinds of stationarity points the minimizers are, we solved all of the three systems.

The results show that, in some cases, we may not find the minimizers by solving the S-systems only.

Tab. 2: Numerical results for Examples 2.1–2.4 and 5.1–5.3 by Algorithm 4.1 with $\sigma = 1$

	Systems	Iter	x^k	(u^k, v^k)	$\|F(w^k)\|$	Time
Example 2.1	M	16	(0,0.0000)	(0,3.0000)	2.0683e-07	0.4305
Example 2.2	C [a]	78	(0.0429,0,-0.0000,0)	(-0.0051,-3.9966)	0.4204	3.5504
Example 2.3	C	11	(0.0000,0)	(0.0000,-1.0000)	3.6890e-10	0.3430
	M [a]	15	(-0.0076,0.0157)	(-1.0070,-0.0004)	0.0013	1.3910
	S	100	(-0.4594,0.7073)	(-0.8477,-0.8319)	1.1766	3.1200
Example 2.4	C [a]	22	(0.0720,-0.0671,-0.0721)	(-0.0001,0.0000)	0.0052	1.0187
	M [a]	18	(-0.0768,0.1526,0.1188)	(-1.0691,-0.0005)	0.0166	1.1543
	S	100	(0.3736,2.2679,0.0067)	(9.2337,-7.5256)	13.4476	11.0329
Example 5.1	S	6	(0,1.0000)	(1.0000,0.0000)	1.2986e-06	0.1134
Example 5.2	M	12	(0.0000,0.0000,0.0000)	(0.0000,-2.0000)	8.9677e-08	0.4540
Example 5.3	C	21	(2.0000,0.0000)	(0.0000,0.5000)	1.4344e-07	0.3893
	M	26	(2.0000,0.0000)	(0.0000,0.5000)	5.4236e-06	0.7313
	S	100	(0.8857,1.1333)	(0.1045,0.5728)	0.1037	2.2114

[a] The algorithm stopped since the magnitude of search direction became too small.

Tab. 3: Numerical results for Examples 2.1–2.4 and 5.1–5.3 by the method in [21]

	x^k	Iter	$f(x^k)$	$G(x^k)^T H(x^k)$	Time
Example 2.1	(0.7937,0.7937)	100	-0.7937	0.6300	3.7640e-04
Example 2.2	(0.0001,0.0015,-0.0016,-0.0000)	1	0.0032	1.8072e-07	1.2074e-06
Example 2.3	(0.4061,0.0000)	100	0.4061	0.5257	0.0015
Example 2.4	(-0.0000,0.0000,0.0000)	1	-2.4982e-08	4.0747e-19	1.5092e-06
Example 5.1	(0.0000,1.0000)	1	1.0000	8.0066e-08	9.0553e-07
Example 5.2	(1.0000,1.0000,4.0000)	100	-2.0000	1.0000	0.0010
Example 5.3	(2.0000,0.0000)	1	-2.0000	3.2470e-07	1.2074e-06

Tab. 4: Numerical results for Examples 2.1–2.4 and 5.1–5.3 by the $\ell_{1/2}$ penalty algorithm in [22]

	x^k	Iter	$f(x^k)$	$\min(G(x^k), H(x^k))$	Time
Example 2.1	(-0.0000,-0.0000)	16	1.9158e-07	-1.8035e-07	0.2974
Example 2.2	1.0e+19*(-4.7764,-4.7359,2.9522,-0.1547)	100	-1.2387e+20	-4.7764e+19	1.1181
Example 2.3	(0.0000, 0.0000)	18	1.2500	5.4404e-11	0.3184
Example 2.4	(-0.0003, 0.0008, 0.0003)	24	1.2502	4.4384e-07	0.4752
Example 5.1	(-0.0000, 1.0000)	28	1.0000	-1.4205e-07	0.6031
Example 5.2	(0.0000,0.0000,0.0000)	13	-2.3305e-05	1.3595e-10	0.1985
Example 5.3	(2.0000,-0.0000)	14	-2.0000	2.0874e-05	0.2465

The results shown in Tab. 2 reveal that Algorithm 4.1 was able to obtain global minimizers by solving the stationarity systems for all examples except Examples 2.2 and 2.4. Even for Examples 2.2 and 2.4, although the algorithm stopped at only approximate solutions, one can expect that the solutions will be closer and closer to the true solutions by increasing the tolerance.

From Tab. 3, we see that the partial augmented Lagrangian method in [21] found global minimizers for Examples 2.2, 2.4, 5.1, and 5.3. However, for Examples 2.1, 2.3, and 5.2, the algorithm stopped at infeasible points, and thus failed to find the solutions. But this is not unexpected since an accumulation point of an augmented Lagrangian method is generally not guaranteed to be feasible. From Tab. 4, we can see that the $\ell_{1/2}$ penalty method found global minimizers for all examples except Examples 2.2 and 2.4. In particular, for Example 2.2, the iteration sequence moves away from the feasible region.

7 Conclusions

We have reformulated the popular stationarity conditions for MPEC as systems of equations with box constraints, and presented a modified LM algorithm for solving these constrained equations. Since the success of proposed algorithm depends greatly on the existence of local error bounds, we have developed some sufficient conditions for local error bounds. Note that, since Algorithm 4.1 is only locally convergent, how to choose starting points is very important. As in [11], in order to achieve global convergence, some kinds of line search techniques may need to be used. We leave this issue as a future work.

Acknowledgements. The first author's work was supported by the NSFC Grant (No. 11401379) and the China Postdoctoral Science Foundation (No. 2014M550237). The second author's work was supported in part by the NSFC Grant (No. 11431004) and the Innovation Program of Shanghai Municipal Education Commission. The third author's work was supported in part by the NSERC. The authors are grateful to the anonymous referees for their helpful comments and suggestions.

References

1 Luo, Z.Q., Pang, J.S., Ralph, D.: Mathematical Programs with Equilibrium Constraints, Cambridge University Press. Cambridge, United Kingdom (1996)

2 Outrata, J.V., Kocvara, M., Zowe, J.: Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory, Applications and Numerical Results. Kluwer Academic Publishers, Boston (1998)

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

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

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

6 Scheel, H.S., Scholtes, S.: Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. Math. Oper. Res. 25, 1–22 (2000)

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

8 Ye, J.J., Ye, X.Y.: Necessary optimality conditions for optimization problems with variational inequality constraints. Math. Oper. Res. 22, 977–997 (1997)

9 Fukushima, M., Lin, G.H.: Smoothing methods for mathematical programs with equilibrium constraints. In: Proceedings of the ICKS'04, pp. 206–213. IEEE Computer Society Press, Los Alamitos (2004)

10 Bazaraa, M.S., Sherali, H.D., Shetty, C.M.: Nonlinear programming Theory and Algorithms. John Wiley & Sons, New York (1993)

11 Kanzow, C., Yamashita, N., Fukushima, M.:, Levenberg-Marquardt methods with strong local convergence properties for solving nonlinear equations with convex constraints. J. Comput. Appl. Math. 172, 375–397 (2004)

12 Fan, J.Y., Yuan, Y.X.: On the quadratic convergence of the Levenberg-Marquardt method without nonsingularity assumption. Computing, 74, 23–39 (2005)

13 Behling, R., Fischer, A.: A unified local convergence analysis of inexact constrained Levenberg-Marquardt methods. Optim. Lett. 6, 927–940 (2012)

14 Wu, Z., Ye, J.J.: First-order and second-order conditions for error bounds. SIAM J. Optim. 14, 621–645 (2003)

15 Hoffman, A.J.: On approximate solutions of systems of linear equalities. J. Res. Nat. Bur. Standards 49, 263–265 (1952)

16 Jourani, A.: Constraint qualifications and Lagrange multipliers in nondifferentiable programming problems. J. Optim. Theory Appl. 81, 533–548 (1994)

17 Minchenko, L., Stakhovski, S.: Parametric nonlinear programming problems under the relaxed constant rank condition. SIAM J. Optim. 21, 314–332 (2011)

18 Guo, L., Ye, J.J., Zhang, J.: Mathematical programs with geometric constraints in Banach spaces: enhanced optimality, exact penalty, and sensitivity. SIAM J. Optim. 23, 2295–2319 (2013)

19 Scholtes, S., Stöhr, M.: Exact penalization of mathematical programs with complementarity constraints. SIAM J. Control Optim. 37, 617–652 (1999)

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

21 Huang, X.X., Yang, X.Q., Teo., K.L.: Partial augmented Lagrangian method and mathematical programs with complementarity constraints. J. Global Optim. 35, 235–254 (2006)

22 Yang, X.Q., Huang, X.X.: Lower-order penalty methods for mathematical programs with complementarity constraints. Optim. Methods Softw. 19, 693–720 (2004)