Numer. Math. Theor. Meth. Appl.
doi: 10.4208/nmtma.2010.33.6

Vol. **3**, No. 3, pp. 367-386
August 2010

Superlinear Convergence of a Smooth Approximation Method for Mathematical Programs with Nonlinear Complementarity Constraints

Fujian Duan* and Lin Fan

College of Mathematics and Computational Science, Guilin University of Electronic Technology, Guilin, 541004, China.

Received 24 November 2009; Accepted (in revised version) 24 November 2009

Available online 2 July 2010

Abstract. Mathematical programs with complementarity constraints (MPCC) is an important subclass of MPEC. It is a natural way to solve MPCC by constructing a suitable approximation of the primal problem. In this paper, we propose a new smoothing method for MPCC by using the aggregation technique. A new SQP algorithm for solving the MPCC problem is presented. At each iteration, the master direction is computed by solving a quadratic program, and the revised direction for avoiding the Maratos effect is generated by an explicit formula. As the non-degeneracy condition holds and the smoothing parameter tends to zero, the proposed SQP algorithm converges globally to an S-stationary point of the MPEC problem, its convergence rate is superlinear. Some preliminary numerical results are reported.

AMS subject classifications: 90C30, 65K05

Key words: Mathematical programs with complementarity constraints, nonlinear complementarity constraints, aggregation technique, S-stationary point, global convergence, super-linear convergence.

1. Introduction

Mathematical programs with equilibrium constraints (MPEC) is an optimization problem whose constraints include variational inequalities or complementarity system. It forms a relatively new and interesting class of optimization problems, which have found many applications in engineering and economics, we refer to [10] for an extensive bibliography on this topic and its applications. In this paper, we consider an important subclass of the MPEC problem, which is called mathematical programs with complementarity constraints (MPCC):

$$\begin{aligned} \min \quad & f(x,y) \\ \text{s.t.} \quad & 0 \le F(x,y) \perp y \ge 0, \end{aligned} \tag{1.1}$$

*Corresponding author. *Email addresses:* `duanfj@guet.edu.cn` (F. Duan), `fanlin0803@163.com` (L. Fan)

where $f : R^{n+m} \to R, F : R^{n+m} \to R^m$ are continuously differentiable functions, and $w \perp y$ indicates orthogonality of any vectors $w, y \in R^m$. The system $0 \le F(x, y) \perp y \ge 0$ is said to be the lower-level or equilibrium constraints.

The major difficulty in solving problem (1.1) is that its constraint fails to satisfy the standard Mangasarian-Fromovitz constraints qualification at any feasible point because of the existence of the complementarity constraint, see [1] and [13]. So the theory for nonlinear programming can not be directly applied to problem (1.1), hence the standard methods are not guaranteed to solve such problem. In recent years, optimal conditions and various stationarity concepts were deeply studied by some authors, see [10, 13, 15]. For example, Luo et al. [10] provided a comprehensive study on the MPEC, such as the exact penalization theory, stationarity conditions. Scheel and Scholtes [13] made an excellent clarification on these concepts and elucidated their connections. More recently, Qi et al. [12] investigated the differentiable properties of the aggregation function, and used it to propose a smoothing method for nonlinear complementarity problems. Jiang and Ralph [8] proposed two smooth SQP methods for MPEC. Global convergence of the methods depend on the lower level strict complementarity condition amongst some conditions, such as the LICQ or MFCQ. Fukushima and Tseng [5] proposed an ε-active set algorithm, they used a sequence of SSNPs based on an ε-active set to approach the discussed MPCC. Under a uniform LICQ on the ε-feasible set, this algorithm generates iterates whose cluster points are B-stationary points of the problem. However, the proof has a gap and shows only that each cluster point is an M-stationary point. Subsequently, Fukushima and Tseng [6] discussed this gap and a modified algorithm that achieves B-stationarity under an additional error bound condition. Tao [14] proposed a class of smoothing methods for MPCC, they used an available probability density function to obtain a corresponding approximation of the original problem. Under some conditions, the MPCC-LICQ holds for the class of smooth methods. However, the methods of [5, 6, 8, 12, 14] do not adopt any technique to avoid the Maratos effect, they only possess global convergence.

Motivated by the ideas of [8, 12, 14], we present a new smoothing SQP algorithm for the problem (1.1). Some notable features of the new algorithm are as follows:at each iteration, the master direction is computed by solving a quadratic program (QP) which only includes equality constraints, the form of QP is different from in [8]. The revised direction for avoiding the Maratos effect is obtained by an explicit formula. The proposed algorithm possesses not only global convergence, but also super-linear convergence.

The structure of this paper is organized as follows: In Section 2, some known results are restated. In Section 3, the algorithm is proposed. In Section 4, we show that the algorithm is globally convergent. Super-linear convergence rate is proved in Section 5, and finally some preliminary numerical results are reported in Section 6.

Throughout this paper, we use the following notations:

$$z = (x, y, w), \quad p = (x, y), \quad q = (y, w),$$
$$dp = (dx, dy), \quad dq = (dy, dw), \quad dz = (dx, dy, dw),$$
$$p^k = (x^k, y^k), \quad q^k = (y^k, w^k), \quad dz^k = (dx^k, dy^k, dw^k).$$

2. Preliminaries

In this section, we recall some concepts about the MPCC (1.1).

For the sake of simplicity, we denote the feasible set of the problem (1.1) as follows:

$$\mathscr{F}=\Big\{(x,y):0\le F(x,y)\perp y\ge 0\Big\},$$

and the tangent cone of $\mathscr{F}$ at a vector $p^*=(x^*,y^*)\in\mathscr{F}$ as follows:

$$\mathscr{T}(p^*,\mathscr{F})=\Big\{d=\lim_{k\to\infty}\frac{p^k-p^*}{\tau_k}:p^k\in\mathscr{F},\lim_{k\to\infty}p^k=p^*,\tau_k\downarrow 0\Big\}.$$

For any $p=(x,y)$, we decompose the index set $L=\{1,2,\cdots,m\}$ into three disjoint subsets,

$$\begin{aligned}\alpha(p)&=\{1\le i\le m:F_i(p)<y_i\},\\ \beta(p)&=\{1\le i\le m:F_i(p)=y_i\},\\ \gamma(p)&=\{1\le i\le m:F_i(p)>y_i\}.\end{aligned}$$

And we define the decomposition index set $\mathscr{A}(p)$ at p by

$$\mathscr{A}(p)=\Big\{(\mathscr{J},\mathscr{K}):\mathscr{J}\supseteq\alpha(p),\mathscr{K}\supseteq\gamma(p),\mathscr{J}\cup\mathscr{K}=1,2,\cdots,m,\mathscr{J}\cap\mathscr{K}=\emptyset\Big\}.$$

If $p^*=(x^*,y^*)\in\mathscr{F}$, then

$$\begin{aligned}\alpha(p^*)&=\{1\le i\le m:F_i(p^*)=0<y_i^*\},\\ \beta(p^*)&=\{1\le i\le m:F_i(p^*)=0=y_i^*\},\\ \gamma(p^*)&=\{1\le i\le m:F_i(p^*)>0=y_i^*\}.\end{aligned}$$

Now, we give some main definitions about the optimal conditions of the MPCC (1.1).

Definition 2.1. (S-stationary point) *A point $p^*=(x^*,y^*)\in\mathscr{F}$ is an S-stationary point of the* (1.1)*, if each $\forall(\mathscr{J},\mathscr{K})\in\mathscr{A}(p^*)$, there exist K-T multipliers $\eta^*\in R^m,\pi^*\in R^m$, such that*

$$\begin{aligned}&\nabla f(x^*,y^*)-\nabla F(x^*,y^*)\eta^*-\begin{pmatrix}0_{n\times m}\\ I_{m\times m}\end{pmatrix}\pi^*=0,\\ &F_i(x^*,y^*)=0,\quad 0\le y_i^*\perp\pi_i^*\ge 0,\quad\forall i\in\mathscr{J},\\ &0\le F_i(x^*,y^*)\perp\eta_i^*\ge 0,\quad y_i^*=0,\quad\forall i\in\mathscr{K}.\end{aligned}\tag{2.1}$$

Definition 2.2. (Lower-level nondegenerate condition) *A point $(x,y)\in R^{n+m}$ is said to be lower-level non-degenerate for the MPCC* (1.1)*, if $y_i\neq F_i(x,y),\forall i=1,2,\cdots,m$.*

3. Equivalent reformulation of MPCC and algorithm

In this section, we first give an equivalent reformulation of the MPCC (1.1), and then propose a new SQP algorithm for solving problem (1.1).

Let $F(x,y)=w$, by using the aggregation technique, we define

$$\Phi(y,w,\mu)=\begin{pmatrix} \phi(y_1,w_1,\mu) \\ \vdots \\ \phi(y_m,w_m,\mu) \end{pmatrix},$$

where

$$\phi(y_i,w_i,\mu)=-\mu\ln\Big[\exp\left(-y_i/\mu\right)+\exp\left(-w_i/\mu\right)\Big],\quad i=1,\cdots,m.$$

For $\mu>0$, it is easy to know that $\Phi(y,w,\mu)$ is continuously differentiable with respect to y,w, let

$$\lim_{\mu\to 0}\Phi(y,w,\mu)=\Phi(y,w).$$

As such, it is natural to define $\Phi(y,w)=\Phi(y,w,0)$,

$$\phi(y_i,w_i,0)=\lim_{\mu\to 0}-\mu\ln\Big[\exp\left(-y_i/\mu\right)+\exp\left(-w_i/\mu\right)\Big],\quad i=1,\cdots,m.$$

Then we construct the following parametric nonlinear programming problems (P_μ):

$$\begin{aligned} \min\quad & f(x,y) \\ \text{s.t.}\quad & F(x,y)=w, \\ & \Phi(y,w,\mu)=0. \end{aligned} \tag{3.1}$$

For $\mu>0$, (3.1) is said to be the smoothing approximation of MPCC (1.1). Obviously, as $\mu\to 0$, the solution of (P_μ) tends to the solution of MPCC (1.1).

In order to analyze the LICQ for the MPCC (3.1), we consider the following mapping $H: R^{n+2m}\times(0,+\infty)\to R^{2m}$

$$H(z,\mu)=\begin{pmatrix} F(x,y)-w \\ \Phi(y,w,\mu) \end{pmatrix}. \tag{3.2}$$

It is clear that the gradient of $H(z,\mu)$ can be described as

$$\nabla H(z,\mu)=\begin{pmatrix} \nabla_x F(x,y) & 0 \\ \nabla_y F(x,y) & \nabla_y\Phi(y,w,\mu) \\ -I_m & \nabla_w\Phi(y,w,\mu) \end{pmatrix}. \tag{3.3}$$

Some basic assumptions are given as follows:

Assumption 3.1. *The function $f(x,y), F(x,y)$ are two-times continuously differentiable.*

Assumption 3.2. *For any*

$$z \in X = \{(x,y,w) \in R^{n+m+m} : (x,y) \in \mathscr{F}, w = F(x,y)\},$$

the gradient matrix $\nabla H(z,\mu)$ has full column rank.

Assumption 3.3. *The feasible set of* (1.1) *is nonempty, i.e., $\mathscr{F} \neq \emptyset$.*

Let $z^k \in X$ and $\mu_k > 0$, we consider the following quadratic program $QP(z^k,\mu_k)$:

$$\begin{aligned} \min \quad & \nabla f(x^k,y^k)^T \begin{pmatrix} dx \\ dy \end{pmatrix} + \frac{1}{2}(dx^T, dy^T, dw^T) B_k \begin{pmatrix} dx \\ dy \\ dw \end{pmatrix} \\ \text{s.t.} \quad & F(x^k,y^k) - w^k + \nabla F(x^k,y^k)^T \begin{pmatrix} dx \\ dy \end{pmatrix} - dw = 0, \\ & \Phi(y^k,w^k,\mu_k) + \nabla\Phi(y^k,w^k,\mu_k)^T \begin{pmatrix} dy \\ dw \end{pmatrix} = 0. \end{aligned} \tag{3.4}$$

where the gradient $\nabla\Phi(y^k,w^k,\mu_k)$ can be computed by

$$\nabla\Phi(y^k,w^k,\mu_k) = (D_a^k \quad D_b^k)^T, \tag{3.5}$$

$$\begin{aligned} D_a^k &= \operatorname{diag}\big(\partial_{y_i}\phi(y_i^k,w_i^k,\mu_k), i = 1 \sim m\big), \\ D_b^k &= \operatorname{diag}\big(\partial_{w_i}\phi(y_i^k,w_i^k,\mu_k), i = 1 \sim m\big). \end{aligned} \tag{3.6}$$

In this paper, we define the ℓ_1 penalty function $\theta : R^{n+2m} \times (0,+\infty) \times (0,+\infty) \to R$ for problem (3.1) as a merit function:

$$\theta(z,c,\mu) = f(x,y) + c\big(\|F(x,y) - w\|_1 + \|\Phi(y,w,\mu)\|_1\big). \tag{3.7}$$

Obviously, the function $\theta(.,c,\mu)$ is directionally differentiable. From the equality constraints of (3.4), we can obtain direction derivative $\theta'(z^k,c_k,\mu_k;dz^k)$ of $\theta(.,c,\mu)$ at z^k along dz^k by

$$\begin{aligned} & \theta'(z^k,c_k,\mu_k;dz^k) \\ =& \nabla f(x^k,y^k)^T \begin{pmatrix} dx^k \\ dy^k \end{pmatrix} - c_k\|\Phi(y^k,w^k,\mu_k)\|_1 - c_k\|F(x^k,y^k) - w^k\|_1. \end{aligned} \tag{3.8}$$

The SQP algorithm for solving problem (1.1) is given in Algorithm 3.1.

Algorithm 3.1:

Step 0 Initialization:

Given a initial point $z^0 = (x^0, y^0, w^0) \in X$, scalars $\delta \in (0, \infty), c_{-1} > 0$, $\gamma \in [1, 2), \alpha \in (0, \frac{1}{2}), \beta \in (0, 1)$, and a sequence $\{\mu_k\}_{k=0}^{\infty}$ such that

$$\mu_k > 0, \quad \mu_{k+1} < \mu_k, \quad \lim_{k\to\infty} \mu_k = 0, \quad \lim_{k\to\infty} \frac{\mu_{k+1}}{\mu_k^\gamma} = \eta \in (0, 1),$$

a symmetric positive definite matrix $B_0 \in R^{(n+2m)\times(n+2m)}$. Set $k := 0$.

Step 1 Computation of the main search direction dz^k:

Solving the quadratic programming $QP(z^k, \mu_k)$(3.4) to obtain the K-T point dz^k and its corresponding K-T multipliers $\lambda^k = (u^k, v^k) \in R^{m+m}$. Let $dp^k = (dx^k, dy^k)$, $dq^k = (dy^k, dw^k)$.

Step 2 Termination check:

If $dz^k = 0, \mu_k \le \varepsilon$, STOP! If $dz^k = 0, \mu_k > \varepsilon$, then choose a new parameter $\mu_k' \in (\mu_{k+1}, \mu_k)$, and let $\mu_k = \mu_k'$, go to step 1. If $dz^k \ne 0$, select a new parameter $\overline{\mu}_k \in (\mu_{k+1}, \mu_k)$, and set $\mu_k = \overline{\mu}_k$, go to step 3.

Step 3 Penalty update:

$s^k = \max_{1\le i\le m}\{|u_i^k|, |v_i^k|\}$,

$$c_k = \begin{cases} c_{k-1}, & \text{if } c_{k-1} \ge s^k + \delta, \\ \max\{s^k + \delta, c_{k-1} + 2\delta\}, & \text{otherwise.} \end{cases}$$

Step 4 Computation of the revised direction $\widetilde{d}z^k$:

$$\widetilde{d}z^k = -\nabla H(z^k, \mu_k)(\nabla H(z^k, \mu_k)^T \nabla H(z^k, \mu_k))^{-1} H(z^k + dz^k, \mu_k), \tag{3.9}$$

where

$$H(z^k + dz^k, \mu_k) = \begin{pmatrix} F(p^k + dp^k) - w^k - dw^k \\ \Phi(q^k + dq^k, \mu_k) \end{pmatrix}. \tag{3.10}$$

If $\|\widetilde{d}z^k\| > \|dz^k\|$, let $\widetilde{d}z^k = 0$, go to step 5.

Step 5 Do line search:

Compute the step size t_k, which is the first number t of the sequence $\{1, \beta, \beta^2, \cdots\}$ satisfying

$$\theta(z^k + t dz^k + t^2 \widetilde{d}z^k, c_k, \mu_k) \le \theta(z^k, c_k, \mu_k) + \alpha t \theta'(z^k, dz^k, c_k, \mu_k). \tag{3.11}$$

Step 6 Update:

Generate a new iteration point by $z^{k+1} = z^k + t_k dz^k + t_k^2 \widetilde{d}z^k$ and a new symmetric positive definite matrix $B_{k+1} \in R^{(n+2m)\times(n+2m)}$ by BFGS formula. Set $k := k+1$ and go back to step 1.

4. Global convergence of algorithm

In this section, we firstly show that Algorithm 3.1 is well-defined, then prove that the algorithm converges globally to an S-stationary point of problem (1.1).

Proposition 4.1. *Let $\mu_k > 0$ be given and B_k be symmetric positive definite, then the quadratic programming $QP(z^k, \mu_k)$ (3.4) problem has a unique optimal solution.*

Proposition 4.2. *For any $(y_i, w_i, \mu) \in [0, +\infty) \times [0, +\infty) \times (0, +\infty)$, it holds that*

$$\left[\partial_{y_i}\phi(y_i, w_i, \mu)\right]^2 + \left[\partial_{w_i}\phi(y_i, w_i, \mu)\right]^2 \geq \frac{1}{2} > 0.$$

Proof. Since $0 < \exp\left(-y_i/\mu\right) \leq 1, 0 < \exp\left(-w_i/\mu\right) \leq 1$, we have that

$$\begin{aligned}&\frac{\left[\exp\left(-y_i/\mu\right)\right]^2 + \left[\exp\left(-w_i/\mu\right)\right]^2}{\left[\exp\left(-y_i/\mu\right) + \exp\left(-w_i/\mu\right)\right]^2} - \frac{1}{2}\\ =&\frac{\left[\exp\left(-y_i/\mu\right) - \exp\left(-w_i/\mu\right)\right]^2}{2\left[\exp\left(-y_i/\mu\right) + \exp\left(-w_i/\mu\right)\right]^2} \geq 0.\end{aligned}$$

Thereby, we conclude that

$$\begin{aligned}&\left[\partial_{y_i}\phi(y_i, w_i, \mu)\right]^2 + \left[\partial_{w_i}\phi(y_i, w_i, \mu)\right]^2\\ =&\frac{\left[\exp\left(-y_i/\mu\right)\right]^2 + \left[\exp\left(-w_i/\mu\right)\right]^2}{\left[\exp\left(-y_i/\mu\right) + \exp\left(-w_i/\mu\right)\right]^2} \geq \frac{1}{2} > 0.\end{aligned}$$

The proof is completed. □

Lemma 4.1. *If $dz^k \neq 0$, then it holds that*

$$\theta'(z^k, dz^k, c_k, \mu_k) \leq -(dz^k)^T B_k dz^k < 0. \tag{4.1}$$

Proof. Since (dz^k, u^k, v^k) is a K-T point pair of the $QP(z^k, \mu_k)$(3.4), we have

$$\begin{aligned}&\begin{pmatrix} \nabla f(x^k, y^k) \\ 0 \end{pmatrix} + B_k \begin{pmatrix} dx^k \\ dy^k \\ dw^k \end{pmatrix} + \begin{pmatrix} \nabla F(x^k, y^k) \\ -I \end{pmatrix} u^k\\ &\qquad + \begin{pmatrix} 0 \\ \nabla\Phi(y^k, w^k, \mu^k) \end{pmatrix} v^k = 0,\end{aligned} \tag{4.2}$$

$$\begin{aligned}&F(x^k, y^k) - w^k + \nabla F(x^k, y^k)^T \begin{pmatrix} dx^k \\ dy^k \end{pmatrix} - dw^k = 0,\\ &\Phi(y^k, w^k, \mu^k) + \nabla\Phi(y^k, w^k, \mu^k)^T \begin{pmatrix} dy^k \\ dw^k \end{pmatrix} = 0.\end{aligned} \tag{4.3}$$

It follows from (4.2) that

$$\left(\nabla f(x^k,y^k)^T,0\right)\begin{pmatrix} dx^k \\ dy^k \\ dw^k \end{pmatrix}+(dz^k)^TB_kdz^k+(u^k)^T\left(\nabla F(x^k,y^k)^T,-I\right)dz^k$$
$$+(v^k)^T\left(0,\nabla\Phi(y^k,w^k,\mu_k)^T\right)dz^k=0. \tag{4.4}$$

By combining (4.3) with (4.4), we obtain

$$\nabla f(x^k,y^k)^T\begin{pmatrix} dx^k \\ dy^k \end{pmatrix}+(dz^k)^TB_kdz^k-(u^k)^T(F(x^k,y^k)-w^k)$$
$$-(v^k)^T\Phi(y^k,w^k,\mu_k)=0. \tag{4.5}$$

In view of (3.8) and (4.5), we conclude that

$$\begin{aligned}\theta'(z^k,dz^k,c_k,\mu_k)=&-(dz^k)^TB_kdz^k+(u^k)^T\Big(F(x^k,y^k)-w^k\Big)+(v^k)^T\Phi(y^k,w^k,\mu_k)\\ &-c_k\big(\|F(x^k,y^k)-w^k\|_1+\|\Phi(y^k,w^k,\mu_k)\|_1\Big)\\ \le&-(dz^k)^TB_kdz^k+\sum_{i=1}^m\big(|u_i^k|-c_k\big)\big|F_i(x^k,y^k)-w_i^k\big|\\ &+\sum_{i=1}^m\big(|v_i^k|-c_k\big)\big|\phi(y_i^k,w_i^k,\mu_k)\big|.\end{aligned}$$

Taking into account $c_k>s^k=\max_{1\le i\le m}\left\{|u_i^k|,|v_i^k|\right\}$, we get that

$$\theta'(z^k,dz^k,c_k,\mu_k)\le-(dz^k)^TB_kdz^k<0.$$

The proof is completed. □

Lemma 4.2. *The line search in step 5 is well defined, i.e., step 5 yields a step-size $t_k=\beta^{j_k}$ for some finite j_k.*

Proof. For the sake of simplicity, we denote that

$$\begin{aligned}T_1&=f(p^k+tdp^k+t^2\widetilde{d}p^k)-f(p^k),\\ T_2&=\sum_{i=1}^m\Big[|F_i(p^k+tdp^k+t^2\widetilde{d}p^k)-w_i^k-tdw_i^k-t^2\widetilde{d}w_i{}^k|-|F_i(p^k)-w_i^k|\Big],\\ T_3&=\sum_{i=1}^m\Big[|\phi(q_i^k+tdq_i^k+t^2\widetilde{d}q_i{}^k,\mu_k)|-|\phi(q_i^k,\mu_k)|\Big].\end{aligned}$$

Then

$$\theta(z^{k+1},c_k,\mu_k)-\theta(z^k,c_k,\mu_k)=T_1+c_kT_2+c_kT_3. \tag{4.6}$$

Using Taylor expansion, we have

$$T_1 = t\nabla f(p^k)^T dp^k + t^2 f(p^k)^T \widetilde{d}p^k + o(t) = t\nabla f(p^k)^T dp^k + o(t).$$

Observe that

$$\begin{aligned}
&F_i(p^k + tdp^k + t^2\widetilde{d}p^k) - w_i^k - tdw_i^k - t^2\widetilde{d}w_i{}^k \\
=&F_i(p^k) + \nabla F_i(p^k)^T(tdp^k + t^2\widetilde{d}p^k) - w_i^k - tdw_i^k - t^2\widetilde{d}w_i{}^k + o(t) \\
=&F_i(p^k) + t\nabla F_i(p^k)^T dp^k - w_i^k - tdw_i^k + o(t) \\
=&F_i(p^k) - w_i^k - t\big(F_i(p^k) - w_i^k\big) + o(t) \\
=&(1-t)\big(F_i(p^k) - w_i^k\big) + o(t),
\end{aligned}$$

and

$$\begin{aligned}
&\phi(q_i^k + tdq_i^k + t^2\widetilde{d}q_i{}^k, \mu_k) \\
=&\phi(q_i^k, \mu_k) + \nabla\phi(q_i^k, \mu_k)^T(tdq_i^k + t^2\widetilde{d}q_i{}^k) + o(t) \\
=&\phi(q_i^k, \mu^k) + t\nabla\phi(q_i^k, \mu_k)^T dq_i^k + o(t) \\
=&(1-t)(q_i^k, \mu_k) + o(t).
\end{aligned}$$

So, we conclude that

$$\begin{aligned}
T_2 =& \sum_{i=1}^m \Big[(1-t)|F_i(p^k) - w_i^k| - |F_i(p^k) - w_i^k|\Big] + o(t) \\
=& \sum_{i=1}^m \Big[(-t)|F_i(p^k) - w_i^k|\Big] + o(t) \\
=& -t\|F(p^k) - w^k\|_1 + o(t).
\end{aligned}$$

Similarly, we have

$$T_3 = -t\|\Phi(q^k, \mu_k)\|_1 + o(t).$$

Consequently,

$$\begin{aligned}
&T_1 + c_k T_2 + c_k T_3 \\
=&t\Big[\nabla f(p^k)^T dp^k - c_k\|F(p^k) - w^k\|_1 - c_k\|\Phi(q^k, \mu_k)\|_1\Big] + o(t).
\end{aligned}$$

This, along with (4.6), shows that

$$\theta(z^{k+1}, c_k, \mu_k) - \theta(z^k, c_k, \mu_k) = t\theta'(z^k, dz^k, c_k, \mu_k) + o(t).$$

In view of the fact $\theta'(z^k, dz^k, c_k, \mu_k) < 0$, there exists a constant $t_k > 0$ such that

$$\theta(z^k + t_k dz^k + t_k^2\widetilde{d}z^k, c_k, \mu_k) \le \theta(z^k, c_k, \mu_k) + \alpha t_k \theta'(z^k, dz^k, c_k, \mu_k).$$

The proof is finished. □

Before showing the algorithm is globally convergent, we make some assumptions as follows:

Assumption 4.1. *The sequence* $\{z^k\}$ *is bounded.*

Assumption 4.2. *There exist constants* $b \geq a > 0$,*such that*

$$a\|z\|^2 \leq z^T B_k z \leq b\|z\|^2, \quad \forall z \in R^{n+2m}, \quad \forall k = 0, 1, 2, \ldots$$

Assumption 4.3. *For any limit point* $z^* = (x^*, y^*, w^*)$ *of* $\{z^k\}$*, it is satisfied the lower-level non-degeneracy:* $(y_i^*, w_i^*) = (y_i^*, F_i(x^*, y^*)) \neq (0,0), i = 1 \sim m$*, i.e.,* $\beta(p^*) = \emptyset$.

The following lemma summarizes several important properties of the sequence $\{z^k, k \in K\}$:

Lemma 4.3. ([4]) *If Assumptions* 3.1, 4.1, 4.2 *hold and* $\lim_{k \in K} z^k = z^*$*, then*

(a) The sequence $\{dz^k : k \in K\}, \{\widetilde{d}z^k : k \in K\}$ *and the multiplier sequences* $\{(u^k, v^k) : k \in K\}$ *are bounded.*

(b) There exists a positive integer k_0 *such that* $c_k \equiv c_{k_0} = c$*, for all* $k \geq k_0$.

According to Lemma 4.3, Assumption 4.2, we might as well assume that there exists a subsequence K, such that

$$\begin{aligned} & dz^k \to dz^*, \quad \widetilde{d}z^k \to \widetilde{d}z^*, \quad B_k \to B_*, x^k \to x^*, \\ & u^k \to u^*, \quad v^k \to v^*, \quad c_k \equiv c, \quad k \in K. \end{aligned} \tag{4.7}$$

Proposition 4.3. *For any* $\mu > t > 0$ *and* $(y, w) \in [0, +\infty) \times [0, +\infty)$*, we have*

$$|\phi(y, w, t)| \leq |\phi(y, w, \mu)| + \left(\ln 2 + 2M/\sqrt{t}\right)\mu. \tag{4.8}$$

Proof. Using Mean Value Theorem, we know that there exists $\sigma \in [t, \mu]$ such that

$$\begin{aligned} |\phi(y, w, t)| = & |\phi(y, w, \mu) + \phi'_\mu(y, w, \sigma)(t - \mu)| \\ \leq & |\phi(y, w, \mu)| + |\phi'_\mu(y, w, \sigma)|(\mu - t) \\ \leq & |\phi(y, w, \mu)| + |\phi'_\mu(y, w, \sigma)|\mu. \end{aligned} \tag{4.9}$$

Set

$$t_1 = \frac{\exp\left(-y/\sigma\right)}{\exp\left(-y/\sigma\right) + \exp\left(-w/\sigma\right)}, \quad t_2 = \frac{\exp\left(-w/\sigma\right)}{\exp\left(-y/\sigma\right) + \exp\left(-w/\sigma\right)}.$$

Then it can be verified that

$$\begin{aligned} & t_1 + t_2 = 1, \\ & \phi'_\mu(y, w, \sigma) \\ = & -\ln\left[\exp\left(-y/\sigma\right) + \exp\left(-w/\sigma\right)\right] - \frac{1}{\sigma} \frac{y \exp\left(-y/\sigma\right) + w \exp\left(-w/\sigma\right)}{\exp\left(-y/\sigma\right) + \exp\left(-w/\sigma\right)} \\ = & -\ln\left[\exp\left(-y/\sigma\right) + \exp\left(-w/\sigma\right)\right] - \frac{1}{\sigma}(y t_1 + w t_2). \end{aligned}$$

According to Assumption 4.1, we may assume that there exists a constant $M\sqrt{\sigma} > 0$, such that

$$\|z^k\| \le M\sqrt{\sigma}.$$

Thereby, we have

$$\begin{aligned}|\phi'_\mu(y,w,\sigma)| &\le \ln\left[\exp\left(-y/\sigma\right) + \exp(-w/\sigma)\right] + \frac{1}{\sigma}(yt_1 + wt_2)\\ &\le \ln 2 + \frac{1}{\sigma}M\sqrt{\sigma} = \ln 2 + \frac{M}{\sqrt{\sigma}} + \frac{M}{\sqrt{\sigma}}\\ &\le \ln 2 + \frac{M}{\sqrt{t}} + \frac{M}{\sqrt{t}} = \ln 2 + 2\frac{M}{\sqrt{t}}.\end{aligned}$$

This along with (4.9) yields the desired inequality (4.8). □

Proposition 4.4. *Suppose that Assumptions* 3.1-4.3 *hold. Then*

$$\lim_{k\to\infty}\theta(z^k,c,\mu_k) = \lim_{k\to\infty}\theta(z^{k+1},c,\mu_k) = \theta(z^*,c,0). \tag{4.10}$$

Proof. For k large enough, from Proposition 4.3, we have

$$|\phi(q_i^{k+1},\mu_{k+1})| \le |\phi(q_i^{k+1},\mu_k)| + \left(\ln 2 + 2\frac{M}{\sqrt{\mu_{k+1}}}\right)\mu_k,$$

which together with (3.7) shows that

$$\begin{aligned}\theta(z^{k+1},c,\mu_{k+1}) &= f(p^{k+1}) + c\|F(p^{k+1}) - w^{k+1}\| + c\sum_{i=1}^{m}\left|\phi(q_i^{k+1},\mu_{k+1})\right|\\ &\le \theta(z^{k+1},c,\mu_k) + c_m\left(\ln 2 + 2\frac{M}{\sqrt{\mu_{k+1}}}\right)\mu_k.\end{aligned} \tag{4.11}$$

In view of (3.11) and (4.1), we have

$$\theta(z^{k+1},c,\mu_k) \le \theta(z^k,c,\mu_k). \tag{4.12}$$

Consequently,

$$\theta(z^{k+1},c,\mu_{k+1}) \le \theta(z^k,c,\mu_k) + c_m\left(\ln 2 + 2\frac{M}{\sqrt{\mu_{k+1}}}\right)\mu_k.$$

Taking into account

$$\lim_{k\to\infty}\frac{\mu_{k+1}}{\mu_k^\gamma} = \eta\in(0,1), \quad \gamma\in[1,2), \quad k\in K, \quad \mu_k\to\mu_*,$$

we have $\|\mu_k\| \le \varrho$. Then

$$\sum_{k\in K} \frac{\mu_k}{\sqrt{\mu_{k+1}}} \le \sum_{k\in K} \sqrt{\frac{\pi}{\eta}} \mu_k^{1-\gamma/2} = \sum_{k\in K} \sqrt{\frac{\pi}{\eta}} \mu_k^{1-\gamma/2} < +\infty, \quad \pi > 1,$$

while

$$\sum_{k=0}^{\infty} \mu_k < +\infty, \quad c_m > 0.$$

Hence, it holds that

$$\sum_{k=0}^{\infty} c_m \left(\ln 2 + 2\frac{M}{\sqrt{\mu_{k+1}}} \right) \mu_k = c_m \ln 2 \sum_{k=0}^{\infty} \mu_k + 2c_m M \sum_{k=0}^{\infty} \frac{\mu_k}{\sqrt{\mu_{k+1}}} < +\infty. \tag{4.13}$$

From (4.13) and Proposition 4.3 [7], we know the entire sequence $\{\theta(z^k, c, \mu_k)\}$ is convergent. So, we have

$$\lim_{k\to\infty} \theta(z^k, c, \mu_k) = \theta(z^*, c, 0).$$

From (4.11) and (4.12), it holds that

$$\theta(z^{k+1}, c, \mu_{k+1}) - c_m \left(\ln 2 + 2\frac{M}{\sqrt{\mu_{k+1}}} \right) \mu_k \le \theta(z^{k+1}, c, \mu_k) \le \theta(z^k, c, \mu_k).$$

Passing to the limit $k \to \infty$ in the above inequality, we conclude that

$$\lim_{k\to\infty} \theta(z^{k+1}, c, \mu_k) = \theta(z^*, c, 0).$$

The proof is completed. □

Lemma 4.4. *Both sequences* $\{dz^k : k \in K\}$ *and* $\{\widetilde{d}z^k : k \in K\}$ *converge to zero, i.e.,* $\widetilde{d}z^* = dz^* = 0$.

Proof. Since $\|\widetilde{d}z^k\| < \|dz^k\|$, it holds that $\|\widetilde{d}z^*\| \le \|dz^*\|$. We only prove that $dz^* = 0$. From (4.1), (3.11) and Assumption 4.2, we have that

$$\theta(z^{k+1}, c, \mu_k) \le \theta(z^k, c, \mu_k) - a\alpha t_k \|dz^k\|^2.$$

In view of Proposition 4.4 and Lemma 4.3(a), we obtain that

$$\lim_{k\in K, k\to\infty} t_k \|dz^k\| = 0.$$

If $\liminf_{k\in K, k\to\infty} t_k > 0$, then

$$\lim_{k\in K, k\to\infty} \|dz^k\| = 0.$$

Suppose that $\liminf_{k\in K, k\to\infty} t_k = 0$. Without loss of generality, we may assume that

$$\lim_{k\in K, k\to\infty} t_k = 0.$$

Set $\rho_k = \beta^{-1} t_k, k \in K$. Using (3.11), we have

$$\frac{\theta(z^k + \rho_k dz^k + \rho_k^2 \widetilde{d}z^k, c, \mu_k) - \theta(z^k, c, \mu_k)}{\rho_k} > \alpha \theta'(z^k, c, \mu_k; dz^k).$$

Passing to the limit $k \in K$ and $k \to \infty$, in view of (3.8) and Lemma 4.3 [7], we deduce

$$\theta'(z^*, c, 0; dz^*) \geq \alpha \theta'(z^*, c, 0; dz^*). \tag{4.14}$$

Combining (4.1) with Assumption 4.2, we conclude that

$$\theta'(z^k, c, \mu_k; dz^k) \leq -a \|dz^k\|^2.$$

Passing to the limit $k \in K$ and $k \to \infty$ in the above equality, we have

$$\theta'(z^*, c, 0; dz^*) \leq -a \|dz^*\|^2, \tag{4.15}$$

which together with (4.14), we obtain that $dz^* = 0$. □

Theorem 4.1. *Suppose that Assumptions* 3.1-4.3 *hold. If* $(x^*, y^*) \in \mathscr{F}$ *is a lower-level non-degenerate point of the MPCC* (1.1)*, and* $\lim_{k\to\infty} \mu_k = 0$*, then Algorithm* 3.1 *generates an infinite sequence* $\{z^k\}$ *whose any accumulation point* $z^* = (x^*, y^*, w^*)$ *is a KKT point of* (3.1)*. Furthermore,* (x^*, y^*) *is an S-stationary point of* (1.1)*.*

Proof. Firstly, we solve $\nabla_{y_i} \phi(y_i^*, w_i^*, 0)$, $\nabla_{w_i} \phi(y_i^*, w_i^*, 0)$. Observe that

$$\begin{aligned}
\nabla_{y_i} \phi(y_i^*, w_i^*, 0) &= \lim_{\mu_k \to 0} \frac{\exp\left(-y_i^*/\mu_k\right)}{\exp\left(-y_i^*/\mu_k\right) + \exp\left(-w_i^*/\mu_k\right)} \\
&= \lim_{\mu_k \to 0} \frac{1}{1 + \exp\left((y_i^* - w_i^*)/\mu_k\right)}, \\
\nabla_{w_i} \phi(y_i^*, w_i^*, 0) &= \lim_{\mu_k \to 0} \frac{\exp\left(-w_i^*/\mu_k\right)}{\exp\left(-y_i^*/\mu_k\right) + \exp\left(-w_i^*/\mu_k\right)} \\
&= \lim_{\mu_k \to 0} \frac{1}{1 + \exp\left((w_i^* - y_i^*)/\mu_k\right)}.
\end{aligned}$$

Then

$$\nabla_{y_i} \phi(y_i^*, w_i^*, 0) = \begin{cases} 0, & i \in \alpha(p^*), \\ 1, & i \in \gamma(p^*). \end{cases} \tag{4.16}$$

$$\nabla_{w_i} \phi(y_i^*, w_i^*, 0) = \begin{cases} 1, & i \in \alpha(p^*), \\ 0, & i \in \gamma(p^*). \end{cases} \tag{4.17}$$

From (4.2), (4.3), (4.7) and Lemma 4.4, we get that

$$\begin{pmatrix} \nabla f(x^*,y^*) \\ 0 \end{pmatrix} + \begin{pmatrix} \nabla F(x^*,y^*) \\ -I \end{pmatrix} u^* + \begin{pmatrix} 0 \\ \nabla \Phi(y^*,w^*,0) \end{pmatrix} v^* = 0, \tag{4.18a}$$

$$F(x^*,y^*) = w^*, \qquad \Phi(y^*,w^*,0) = 0. \tag{4.18b}$$

In view of $(x^*,y^*) \in \mathscr{F}$, $\beta(p^*) = \emptyset$, (4.16) and (4.17), we conclude that

$$\begin{aligned} &\nabla_{y_i}\phi(y_i^*,w_i^*,0) = 0, \quad \nabla_{w_i}\phi(y_i^*,w_i^*,0) = 1, \quad i \in \alpha(p^*), \\ &\nabla_{y_i}\phi(y_i^*,w_i^*,0) = 1, \quad \nabla_{w_i}\phi(y_i^*,w_i^*,0) = 0, \quad i \in \gamma(p^*). \end{aligned}$$

From (4.18), we have $u^* = \nabla_w \Phi(y^*,w^*,0)v^*$. Let

$$\eta^* = -u^* = -\nabla_w \Phi(y^*,w^*,0)v^*, \quad \pi^* = -\nabla_y \Phi(y^*,w^*,0)v^*.$$

Then

$$\begin{aligned} &\begin{pmatrix} \nabla f(x^*,y^*) \\ 0 \end{pmatrix} - \begin{pmatrix} \nabla F(x^*,y^*) \\ -I \end{pmatrix} \eta^* \\ &\quad + \begin{pmatrix} 0 \\ 0 \\ \nabla_w \Phi(y^*,w^*,0) \end{pmatrix} v^* + \begin{pmatrix} 0 \\ \nabla_y \Phi(y^*,w^*,0) \\ 0 \end{pmatrix} v^* = 0. \end{aligned}$$

Thereby, we have

$$\nabla f(x^*,y^*) - \nabla F(x^*,y^*)\eta^* - \begin{pmatrix} 0_{n\times m} \\ I_{m\times m} \end{pmatrix} \pi^* = 0, \tag{4.19a}$$

$$w_i^* = F_i(x^*,y^*) = 0, \quad y_i^* > 0, \quad \nabla_{y_i}\phi(y_i^*,w_i^*,0) = 0, \quad \pi_i^* = 0, \quad \forall i \in \alpha(p^*), \tag{4.19b}$$

$$w_i^* = F_i(x^*,y^*) > 0, \quad y_i^* = 0, \quad \nabla_{w_i}\phi(y_i^*,w_i^*,0) = 0, \quad \eta_i^* = 0, \quad \forall i \in \gamma(p^*). \tag{4.19c}$$

From Definition 2.1, (4.19) shows that (x^*,y^*) is an S-stationary point of (1.1). □

5. Super-linear convergence

In this section, we first prove that the step-size t_k of Algorithm 3.1 always equals 1 for k sufficiently large. Then we show that Algorithm 3.1 is super-linearly convergent under additional hypotheses.

In order to obtain the superlinear convergence rate, we make the following additional assumptions:

Assumption 5.1. *Suppose that the strong second-order sufficiency conditions for* (1.1) *hold, i.e.,*

$$(dz)^T \nabla^2_{zz} L(z^*, u^*, v^*, 0) dz > 0, \quad \forall dz \in \Omega,$$

where

$$\Omega \stackrel{def}{=} \{dz \in R^{n+2m} : dz \neq 0, \nabla H(z^*)^T dz = 0\},$$
$$L(z,u,v,\mu) = f(x,y) + (F(x,y) - w)^T u + \Phi(y,w,\mu)^T v.$$

Assumption 5.2. $B_k \to B_*, k \to \infty$.

Assumption 5.3. *The sequence of matrices* $\{B_k\}$ *satisfies*

$$\|(\nabla^2_{zz} L(z^k, u^k, v^k, \mu_k) - B_k) dz^k\| = o(\|dz^k\|).$$

Lemma 5.1. *Assume that Assumptions* 3.1-5.2 *hold. Then*

(i) *The entire sequence* $\{z^k\}$ *converges to* z^*.

(ii) $\lim_{k\to\infty} dz^k = \lim_{k\to\infty} \widetilde{d}z^k = 0$.

Proof. (i) From (3.11), (4.1) and Assumption 4.2, we have that

$$\theta(z^{k+1}, c, \mu_k) \leq \theta(z^k, c, \mu_k) - a\alpha t_k \|dz^k\|^2,$$

which, together with Proposition 4.4 and Lemma 4.3(a), yield

$$\lim_{k\to\infty} t_k \|dz^k\| = 0.$$

Thereby, we conclude

$$\begin{aligned} \|z^{k+1} - z^k\| &\leq t_k \|dz^k\| + t_k^2 \|\widetilde{d}z^k\| \\ &\leq t_k \|dz^k\| + t_k \|\widetilde{d}z^k\| \leq 2t_k \|dz^k\| \to 0, \quad k \to \infty. \end{aligned}$$

According to Assumptions 5.1-5.2 and Proposition 4.1 in [11], we can get $z^k \to z^*, k \to \infty$.

(ii) From part (i) and Lemma 4.4, it holds that

$$\lim_{k\to\infty} dz^k = \lim_{k\to\infty} \widetilde{d}z^k = 0.$$

The proof is completed. □

Lemma 5.2. *For k sufficiently large, we have*

$$\|\widetilde{d}z^k\| = \mathcal{O}(\|dz^k\|^2). \tag{5.1}$$

Proof. By Taylor expansion, we have

$$\begin{aligned}&F(p^k+dp^k)-w^k-dw^k\\=&F(p^k)+\nabla F(p^k)^T dp^k-w^k-dw^k+\mathscr{O}(\|dp^k\|^2).\end{aligned}$$

From (4.3), we have

$$F(p^k+dp^k)-w^k-dw^k=\mathscr{O}(\|dp^k\|^2)=\mathscr{O}(\|dz^k\|^2).$$

With the same reason, we can conclude that

$$\Phi(q^k+dq^k,\mu_k)=\mathscr{O}(\|dz^k\|^2).$$

Thereby, we obtain (5.1). □

Theorem 5.1. *For k sufficiently large, we have $z^{k+1}=z^k+dz^k+\widetilde{d}z^k$, i.e., $t_k\equiv 1$.*

Proof. We only prove that

$$\triangle=\theta(z^k+dz^k+\widetilde{d}z^k,c,\mu_k)-\theta(z^k,c,\mu_k)-\alpha\theta'(z^k,dz^k,c,\mu_k)\le 0.$$

Note that

$$\begin{aligned}\triangle=&f(p^k+dp^k+\widetilde{d}p^k)-f(p^k)+c\|F(p^k+dp^k+\widetilde{d}p^k)-w^k-dw^k-\widetilde{d}w^k\|_1\\&+c\|\Phi(q^k+dq^k+\widetilde{d}q^k,\mu_k)\|_1-c\big(\|F(p^k)-w^k\|_1+\|\Phi(q^k,\mu_k)\|_1\big)\\&-\alpha\nabla f(p^k)^T dp^k+\alpha c\big(\|F(p^k)-w^k\|_1+\|\Phi(q^k,\mu_k)\|_1\big)\\=&f(p^k+dp^k+\widetilde{d}p^k)+c(\alpha-1)\big(\|F(p^k)-w^k\|_1+\|\Phi(q^k,\mu_k)\|_1\big)\\&+c\|\Phi(q^k+dq^k+\widetilde{d}q^k,\mu_k)\|_1-\alpha\nabla f(p^k)^T dp^k-f(p^k)\\&+c\|F(p^k+dp^k+\widetilde{d}p^k)-w^k-dw^k-\widetilde{d}w^k\|_1.\end{aligned}$$

From (3.3), (3.5), (3.9) and (3.10), we have

$$\nabla F(p^k)^T\widetilde{d}p^k-\widetilde{d}w^k=-\big[F(p^k+dp^k)-w^k-dw^k\big], \tag{5.2}$$

$$\nabla\Phi(q^k,\mu_k)^T\widetilde{d}q^k=-\Phi(q^k+dq^k,\mu_k). \tag{5.3}$$

In view of (5.1)-(5.3) and the Taylor expansion, we obtain that

$$\begin{aligned}&F(p^k+dp^k+\widetilde{d}p^k)-w^k-dw^k-\widetilde{d}w^k\\=&F(p^k+dp^k)+\nabla F(p^k+dp^k)^T\widetilde{d}p^k+o(\|\widetilde{d}p^k\|)-w^k-dw^k-\widetilde{d}w^k\\=&F(p^k+dp^k)+\nabla F(p^k)^T\widetilde{d}p^k-w^k-dw^k-\widetilde{d}w^k+o(\|\widetilde{d}p^k\|)+\mathscr{O}(\|dz^k\|^3)\\=&o(\|dz^k\|^2),\end{aligned} \tag{5.4}$$

$$\begin{aligned}
&\Phi(q^k + dq^k + \widetilde{d}q^k, \mu_k)\\
=&\Phi(q^k + dq^k, \mu_k) + \nabla\Phi(q^k + dq^k, \mu_k)^T \widetilde{d}q^k + o(\|\widetilde{d}q^k\|)\\
=&\Phi(q^k + dq^k, \mu_k) + \nabla\Phi(q^k, \mu_k)^T \widetilde{d}q^k + o(\|\widetilde{d}q^k\|) + \mathscr{O}(\|dz^k\|^3)\\
=&o(\|dz^k\|^2),
\end{aligned} \tag{5.5}$$

$$\begin{aligned}
&f(p^k + dp^k + \widetilde{d}p^k) - f(p^k)\\
=&\nabla f(p^k)^T(dp^k + \widetilde{d}p^k) + \frac{1}{2}(dp^k + \widetilde{d}p^k)^T \nabla^2 f(p^k)(dp^k + \widetilde{d}p^k) + o(\|dz^k\|^2).
\end{aligned} \tag{5.6}$$

So, we have

$$\begin{aligned}
\triangle = &\nabla f(p^k)^T(dp^k + \widetilde{d}p^k) + \frac{1}{2}(dp^k + \widetilde{d}p^k)^T \nabla^2 f(p^k)(dp^k + \widetilde{d}p^k)\\
&+ c(\alpha - 1)(\|F(p^k) - w^k\|_1 + \|\Phi(q^k, \mu_k)\|_1) - \alpha \nabla f(p^k)^T dp^k + o(\|dz^k\|^2).
\end{aligned} \tag{5.7}$$

We obtain from (4.2) that

$$\begin{aligned}
\nabla f(p^k)^T \widetilde{d}p^k =& -(dz^k)^T B_k \widetilde{d}z^k - (u^k)^T(\nabla F(p^k)^T, -I)\widetilde{d}z^k - (v^k)^T(0, \nabla\Phi(q^k, \mu_k)^T)\widetilde{d}z^k\\
=& -((u^k)^T, (v^k)^T)\nabla H(z^k, \mu_k)^T \widetilde{d}z^k + o(\|dz^k\|^2).
\end{aligned}$$

From (4.4), we conclude that

$$\nabla f(p^k)^T dp^k = -(dz^k)^T B_k dz^k - ((u^k)^T, (v^k)^T)\nabla H(z^k, \mu_k)^T dz^k.$$

Thereby, we have that

$$\begin{aligned}
&\nabla f(p^k)^T(dp^k + \widetilde{d}p^k)\\
=& -(dz^k)^T B_k dz^k - ((u^k)^T, (v^k)^T)\nabla H(z^k, \mu_k)^T(dz^k + \widetilde{d}z^k) + o(\|dz^k\|^2).
\end{aligned}$$

From (3.2), (3.3), (5.4), (5.5) and by Taylor expansion, we get that

$$\begin{aligned}
o(\|dz^k\|^2) =&H(z^k, \mu_k) + \nabla H(z^k, \mu_k)^T(dz^k + \widetilde{d}z^k)\\
&+ \frac{1}{2}\begin{pmatrix} (dz^k + \widetilde{d}z^k)^T \nabla^2_{zz} F_i(p^k)(dz^k + \widetilde{d}z^k), i = 1 \sim m.\\ (dz^k + \widetilde{d}z^k)^T \nabla^2_{zz} \phi(q_i^k, \mu_k)(dz^k + \widetilde{d}z^k), i = 1 \sim m. \end{pmatrix}.
\end{aligned}$$

So, we have that

$$\begin{aligned}
&\big((u^k)^T, (v^k)^T\big)\nabla H(z^k, \mu_k)^T(dz^k + \widetilde{d}z^k)\\
=& -\frac{1}{2}(dz^k + \widetilde{d}z^k)^T[\nabla^2_{zz} L(z^k, u^k, v^k, \mu_k) - \nabla^2_{zz} f(p^k)](dz^k + \widetilde{d}z^k)\\
&- ((u^k)^T, (v^k)^T)H(z^k, \mu_k) + o(\|dz^k\|^2).
\end{aligned}$$

Consequently, it holds that

$$\begin{aligned}
&\nabla f(p^k)^T(dp^k+\widetilde{d}p^k)\\
=&\frac{1}{2}(dz^k+\widetilde{d}z^k)^T[\nabla^2_{zz}L(z^k,u^k,v^k,\mu_k)-\nabla^2_{zz}f(p^k)](dz^k+\widetilde{d}z^k)\\
&-(dz^k)^TB_kdz^k+\big((u^k)^T,(v^k)^T\big)H(z^k,\mu_k)+o(\|dz^k\|^2\big). \qquad (5.8)
\end{aligned}$$

From (4.5), we have that

$$\nabla f(p^k)^T dp^k = -(dz^k)^T B_k dz^k + (u^k)^T(F(p^k)-w^k)+(v^k)^T\Phi(q^k,\mu_k). \qquad (5.9)$$

Taking into account (5.7)-(5.9), we conclude that

$$\begin{aligned}
\triangle=&-(dz^k)^TB_kdz^k+(1-\alpha)((u^k)^T,(v^k)^T)\begin{pmatrix} F(p^k)-w^k\\ \Phi(q^k,\mu_k)\end{pmatrix}+\alpha(dz^k)^TB_kdz^k\\
&+\frac{1}{2}(dz^k+\widetilde{d}z^k)^T\nabla^2_{zz}L(z^k,u^k,v^k,\mu_k)(dz^k+\widetilde{d}z^k)+c(\alpha-1)(\|F(p^k)-w^k\|_1\\
&+\|\Phi(q^k,\mu_k)\|_1)+o(\|dz^k\|^2)\\
=&\Big(\alpha-\frac{1}{2}\Big)(dz^k)^TB_kdz^k+\frac{1}{2}(dz^k+\widetilde{d}z^k)^T[\nabla^2_{zz}L(z^k,u^k,v^k,\mu_k)-B_k](dz^k+\widetilde{d}z^k)\\
&+o(\|dz^k\|^2)+(1-\alpha)((u^k)^T,(v^k)^T)\begin{pmatrix} F(p^k)-w^k\\ \Phi(q^k,\mu_k)\end{pmatrix}\\
&+c(\alpha-1)(\|F(p^k)-w^k\|_1+\|\Phi(q^k,\mu_k)\|_1).
\end{aligned}$$

In view of Assumptions 4.2 and 5.3, $\alpha\in(0,\frac{1}{2})$ and $c>s^k=\max_{1\le i\le m}\{|u_i^k|,|v_i^k|\}$, we further get

$$\begin{aligned}
\triangle\le&a\Big(\alpha-\frac{1}{2}\Big)\|dz^k\|^2+c(1-\alpha)(\|F(p^k)-w^k\|_1+\|\Phi(q^k,\mu_k)\|_1)\\
&+c(\alpha-1)(\|F(p^k)-w^k\|_1+\|\Phi(q^k,\mu_k)\|_1)+o(\|dz^k\|^2)\\
=&a\Big(\alpha-\frac{1}{2}\Big)\|dz^k\|^2+o(\|dz^k\|^2)\le 0.
\end{aligned}$$

Hence (3.11) holds for $t_k=1$ and k large enough. □

Moreover, in view of Theorem 4.1, Assumption 5.3, Theorem 5.1 and the way of Theorem 5.2 in [3], it is easy to get the convergence theorem:

Theorem 5.2. *Suppose that Assumptions* 3.1-5.3 *hold. If* $\mu_k=o(\|dz^k\|)$, *then Algorithm* 3.1 *is super-linearly convergent, i.e., the sequence* $\{z^k\}$ *generated by Algorithm* 3.1 *satisfies that*

$$\|z^{k+1}-z^*\|=o(\|z^k-z^*\|).$$

6. Numerical experiments

In this section, we carry out numerical experiments based on the algorithm. The code of the proposed algorithm is written by using MATLAB 7.0.

In the implementation, we choose some parameters as follows: $\delta = 10, \alpha = 0.1, \beta = 0.5, \mu_0 = 1, \mu_{k+1} = 0.5\mu_k, c_{-1} = 10, B_0 = I_{n+2m}$. B_k is updated by the BFGS formula [2]. In the implementation, the stopping criterion of step 2 is changed to

$$\text{If } \|dz^k\| \le 10^{-6},\ \mu_k \le 10^{-6}, \text{ STOP!}$$

The test problems in Table 1 are selected from [9, 10]. Problems 1, 2, 4 are problem Scholtes 3, Jr 1, qpec 2 in [9], respectively, and Problem 3 is a three dimension example in [10]. A feasible initial point is provided for each problem. The results are summarized in Table 1. For each test problem, the Prob column lists the problem number; p and q are the number of variables and complementarity constraints, respectively; IP is the initial point; NT is the number of iterations. FV is the final value of the objective function.

Table 1:

Prob	p,q	IP	NT	(x^*, y^*)	FV	$\|dz^k\|$
1	(2,1)	(0,0)	6	$(0.70 \times 10^{-9}, 0.99)$	0.49, 0.50	1.02
2	(2,2)	(0,0)	29	(0.50, 0.49)	0.49, 0.5	9.42×10^{-7}
3	(3,2)	(-1,0,1)	52	$(1.2 \times 10^{-5}, 6.90 \times 10^{-5}, 1.28 \times 10^{-5})$	1.0×10^{-12},0	7.51×10^{-7}
4	(30,20)	(0,0, 0,0)	44	$(1.49, 1.49, 1.58 \times 10^{-7}, 1.58 \times 10^{-7})$	44.99, 45	9.35×10^{-7}

Acknowledgments This work was supported by the National Natural Science Foundation of China (No. 10861005), the Natural Science Foundation of Guangxi Province (No. 0728206), and the Innovation Project of Guangxi Graduate Education (No. 2009105950701M29).

References

[1] Y. Chen and M. Florian, *The nonlinear bilevel programming problem: formulations, regularity and optimality conditions*, Optimization, 32(1995), pp. 193-209.

[2] M.J.D. Powell, *A fast algorithm for nonlinearly constrained optimization calculations*, in: G.A.Waston (Ed.), Numerical Analysis, Springer, Berlin, 1978, pp. 144-157.

[3] F. Facchinei and S. Lucidi, *Quadraticily and superlinearly convergent algorithm for the solution of inequality constrainted optimization problems*, Journal of Optimizations Theory and Applications, 85(1995), pp. 265-289.

[4] M. Fukushima, Z.Q. Luo and J.S. Pang, *A globally convergent sequential quadratic programming algorithm for mathematical programming with linear complementarity constraints*, Computational Optimization and Applications, 10(1998), pp. 5-34.

[5] M. Fukushima and P. Tseng, *An implementable active-set algorithm for computing a B-stationary point of a mathematical program with linear complementarity constraints*, SIAM Journal on Optimization, 12(2002), pp. 724-739.

[6] M. Fukushima and P. Tseng, *An implementable active-set algorithm for computing a B-stationary point of a mathematical program with linear complementarity constraints:erratum*, SIAM Journal on Optimization, 17(2007), pp. 1253-1257.

[7] J.B. Jian, J.L. Li and X.D. Mo, *A strongly and superlinearly convergent SQP algorithm for optimization problems with linear complementarity constraints*, Applied Mathematics Optimization, 54(2006), pp. 17-46.

[8] H. Jiang and D. Ralph, *Smooth SQP method for mathematical programs with nonlinear complementarity constraints*, SIAM Journal on Optimization, 10(2000), pp. 779-808.

[9] S. Leyffer, *www-unix.mcs.anl.gov/ leyffer/MacMPEC/*, 2002.

[10] Z.Q. Luo, J.S. Pang, D. Ralph and S.Q. Wu, *Exact penalization and stationarity conditions of mathematical programs with equilibrium constraints*, Mathematical Programming, 75(1996), pp. 19-76.

[11] E.R. Painier and A.L. Tits, *A superlinearly convergent feasible method for the solution of enequality constrained optimization problems*, SIAM Journal Control and Optimization, 25(1987), pp. 934-950.

[12] L. Qi, D. Sun and G. Zhou, *A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities* , Mathematical Programming, 87(2000), pp. 1-35.

[13] H. Scheel and S. Scholtes, *Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity*, Mathematics of Operations Research, 25(2000), pp.1-22.

[14] Y. Tao, *A class of smoothing methods for mathematical programs with complementarity constraints*, Applied Mathematics and Computation, 186(2007), pp. 1-9.

[15] J.J. Ye, *Optimality conditions for optimization problems with complementarity constraints*, SIAM Journal on Optimization, 9(1999), pp. 374-387.