A Smoothing Method for a Mathematical Program with P-Matrix Linear Complementarity Constraints[1]

Xiaojun Chen[2] and Masao Fukushima[3]

June 5, 2001

Abstract

We consider a mathematical program whose constraints involve a parametric P-matrix linear complementarity problem with the design (upper level) variables as parameters. Solutions of this complementarity problem define a piecewise linear function of the parameters. We study a smoothing function of this function for solving the mathematical program. We investigate the limiting behaviour of optimal solutions, KKT points and B-stationary points of the smoothing problem. We show that a class of mathematical programs with P-matrix linear complementarity constraints can be reformulated as a piecewise convex program and solved through a sequence of continuously differentiable convex programs.

Key words. mathematical program with equilibrium constraints, P-matrix linear complementarity problem, reformulation, smoothing approximation.

AMS subject classifications. 90C30, 90C33.

---

[1]This work was supported in part by the Grant-in-Aid for Scientific Research (B) from the Japan Society for the Promotion of Science.

[2]Department of Mathematics and Computer Science, Shimane University, Matsue 690-8504, Japan (chen@math.shimane-u.ac.jp).

[3]Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan (fuku@i.kyoto-u.ac.jp).

1 Introduction

We consider the following mathematical program with linear complementarity constraints:

$$\begin{aligned} \text{minimize} \quad & f(x,y) \\ \text{subject to} \quad & g(x) \le 0,\ \ h(x) = 0 \\ & Nx + My + q \ge 0,\ \ y \ge 0 \\ & (Nx + My + q)^T y = 0, \end{aligned} \tag{1.1}$$

where $f : R^n \times R^m \to R$, $g : R^n \to R^\ell$, $h : R^n \to R^{\ell'}$ are continuously differentiable functions, $M \in R^{m\times m}$, $N \in R^{m\times n}$ and $q \in R^m$.

Throughout this paper, we assume that M is a P-matrix [6, 15], that is, all principal minors of M are positive, or equivalently

$$\max_{1\le i\le m} y_i(My)_i > 0 \qquad \text{for all } y(\neq 0) \in R^m.$$

Then for any fixed $x \in R^n$, the linear complementarity problem

$$Nx + My + q \ge 0,\ \ y \ge 0,\ \ (Nx + My + q)^T y = 0 \tag{1.2}$$

has a unique solution [6, Theorem 3.3.7]. We denote the solution of (1.2) by $y(x)$ and define the function $\theta : R^n \to R$ by $\theta(x) := f(x, y(x))$. Then it is not difficult to show that problem (1.1) is equivalent to the following problem, which we call the *reduced problem*:

$$\begin{aligned} \text{minimize} \quad & \theta(x) \\ \text{subject to} \quad & g(x) \le 0,\ \ h(x) = 0, \end{aligned} \tag{1.3}$$

in the following sense:

- If (x^*, y^*) is a solution of (1.1), then x^* is a solution of (1.3).
- If x^* is a solution of (1.3), then $(x^*, y(x^*))$ is a solution of (1.1).

The major difficulty to solve (1.1) is not only that the feasible set of (1.1) is generally nonsmooth or even disconnected, but also that it fails to satisfy a standard constraint qualification in nonlinear programming [4].

It is well known that the complementarity problem (1.2) can be reformulated as a system of nonsmooth equations. For example, (x, y) satisfies (1.2) if and only if it solves the system of equations

$$\Phi(x,y) := \begin{pmatrix} \phi(y_1, (My + Nx + q)_1) \\ \vdots \\ \phi(y_m, (My + Nx + q)_m) \end{pmatrix} = 0,$$

where

$$\phi(a,b) = a + b - \sqrt{a^2 + b^2}.$$

The function $\phi : R^2 \to R$ is called the Fischer-Burmeister function [8]. Thus, problem (1.1) can be rewritten as

$$\begin{array}{ll} \text{minimize} & f(x,y) \\ \text{subject to} & g(x) \le 0, \;\; h(x) = 0, \\ & \Phi(x,y) = 0. \end{array} \tag{1.4}$$

Smoothing methods for solving (1.4) have been studied extensively in recent years [7, 9, 10]. The feature of smoothing methods for solving (1.4) is to approximate the nonsmooth function $\Phi : R^n \times R^m \to R^m$ by a continuously differentiable function $\Phi_\mu : R^n \times R^m \to R^m$ and to solve the smoothing problem

$$\begin{array}{ll} \text{minimize} & f(x,y) \\ \text{subject to} & g(x) \le 0, \;\; h(x) = 0, \\ & \Phi_\mu(x,y) = 0, \end{array} \tag{1.5}$$

where $\mu > 0$ and $\Phi_\mu(x,y) \to \Phi(x,y)$ as $\mu \to 0$. For such smoothing algorithms developed so far, global convergence to a B-stationary point of problem (1.1) has only been established under the (lower level) nondegeneracy assumption or the so-called MPEC linear independence constraint qualification at a limit point [10]. However, it is not easy to verify these requirements in practice.

In this paper, we introduce a smoothing method for solving the reduced problem (1.3). We use the solution $y_\mu(x)$ of $\Phi_\mu(x,y) = 0$ as a smoothing function of $y(x)$, and approximate the objective function $\theta(x)$ of (1.3) by $\theta_\mu(x) := f(x, y_\mu(x))$.

The plan of the paper is as follows: In Section 2, we study some properties of the smoothing function $y_\mu(\cdot)$. In particular, we show that the component functions of $y_\mu(\cdot)$ as well as $y(\cdot)$ turn out to be convex under certain conditions on M. This result is of particular interest because it reveals that there is a class of MPECs, though somewhat special, that may be reformulated as a convex program (Section 4). To the authors' knowledge, this is a first result that gives a condition under which MPEC can be reduced to convex programs.

In Section 3, we investigate the limiting behaviour of sequences of optimal solutions, KKT points and B-stationary points of the (reduced) smoothing problem

$$\begin{array}{ll} \text{minimize} & \theta_\mu(x) \\ \text{subject to} & g(x) \le 0, \;\; h(x) = 0 \end{array} \tag{1.6}$$

as $\mu \to 0$. Throughout, we simply call (1.6) the *smoothing problem*.

In Section 4, as an application, we mention that a class of mathematical programs with P-matrix linear complementarity constraints can be reformulated as a piecewise convex program and solved through a sequence of continuously differentiable convex programs.

A few words about notations. We let $\|\cdot\|$ denote the Euclidean norm. For any matrix $B \in R^{s\times t}$, let $B_{\mathcal{K}\mathcal{J}}$ be the submatrix of B whose entries lie in the rows of B indexed by $\mathcal{K}$ and the columns indexed by $\mathcal{J}$. If $\mathcal{J} = \{1, 2, \ldots, t\}$, we simply denote $B_{\mathcal{K}\mathcal{J}}$ by $B_{\mathcal{K}}$. For any function $c : R^s \to R^t$, we denote its Jacobian and the transposed Jacobian at $x \in R^s$ by $c'(x) \in R^{t\times s}$ and $\nabla c(x) = c'(x)^T \in R^{s\times t}$, respectively. For a nonsingular matrix A, we denote $A^{-T} = (A^T)^{-1} = (A^{-1})^T$.

2 A smoothing function of $y(x)$

It can be shown that the solution function $y(\cdot) : R^n \to R^m$ of the parametric P-matrix linear complementarity problem (1.2) is piecewise linear [14, page 171] and, in particular, continuous. For a given $x \in R^n$, we define the index sets

$$\begin{aligned}
\mathcal{J}(x) &= \{\, i \,|\, y_i(x) > 0 = (My(x) + Nx + q)_i\},\\
\mathcal{K}(x) &= \{\, i \,|\, y_i(x) = 0 = (My(x) + Nx + q)_i\},\\
\mathcal{L}(x) &= \{\, i \,|\, y_i(x) = 0 < (My(x) + Nx + q)_i\}.
\end{aligned}$$

Note that $\mathcal{J}(x), \mathcal{K}(x)$ and $\mathcal{L}(x)$ are mutually disjoint, and $\mathcal{J}(x) \cup \mathcal{K}(x) \cup \mathcal{L}(x) = \{1, 2, \ldots, m\}$.

Since M is a P-matrix, the principal minor $M_{\mathcal{J}\mathcal{J}}$ is nonsingular for any index set $\mathcal{J} \subseteq \{1, 2, \ldots, m\}$. Applying Theorem 7.4.2 in [6] to (1.2), we have the following lemma.

Lemma 2.1 *The function $y(\cdot)$ is differentiable at x if and only if $y(x)$ is a strictly complementary solution of (1.2) or*

$$-M_{\mathcal{K}(x)\mathcal{J}(x)} M^{-1}_{\mathcal{J}(x)\mathcal{J}(x)} N_{\mathcal{J}(x)} + N_{\mathcal{K}(x)} = 0. \tag{2.1}$$

In the case $y(\cdot)$ is differentiable at x, the derivatives of $y(\cdot)$ are given by

$$\begin{aligned}
\nabla y_{\mathcal{J}(x)}(x) &= -N^T_{\mathcal{J}(x)} M^{-T}_{\mathcal{J}(x)\mathcal{J}(x)},\\
\nabla y_{\mathcal{K}(x)}(x) &= 0,\\
\nabla y_{\mathcal{L}(x)}(x) &= 0.
\end{aligned} \tag{2.2}$$

Since $y(\cdot)$ is piecewise linear, the Clarke generalized Jacobian [5] of $y(\cdot)$ exists everywhere and is given by

$$\partial y(x) = \mathrm{co}\Big\{ \lim_{\substack{x_k \to x \\ x_k \in D_y}} y'(x_k) \Big\},$$

where co stands for the convex hull and D_y denotes the set of points where $y(\cdot)$ is differentiable. The mean-value theorem [5, Proposition 2.6.5] states that

$$y(x) - y(\bar{x}) \in \mathrm{co}\{\partial y([x,\bar{x}])(x-\bar{x})\}, \quad \text{for all } x, \bar{x} \in R^n.$$

The right-hand side above denotes the convex hull of all points of the form $V(x-\bar{x})$ where $V \in \partial y(u)$ for some point u in the line segment $[x, \bar{x}]$. By Lemma 2.1, $\partial y(u)$ is a convex hull of a finite number of matrices, and hence it is bounded. As a consequence, we have the following lemma.

Lemma 2.2 *There is a positive number γ such that*

$$\|y(x) - y(\bar{x})\| \le \gamma \|x - \bar{x}\| \quad \textit{for all} \;\; x, \bar{x} \in R^n.$$

To simplify our discussion, we introduce a new variable $w \in R^m$ and define the function

$$F(x,y,w) := \begin{pmatrix} My + Nx + q - w \\ \phi(y_1, w_1) \\ \vdots \\ \phi(y_m, w_m) \end{pmatrix}.$$

For any $x \in R^n$, the unique solution $y(x)$ of $\Phi(x,y) = 0$ along with

$$w(x) = My(x) + Nx + q$$

constitutes the unique solution of $F(x,y,w) = 0$. Conversely, for any $x \in R^n$, the first m components of the unique solution $(y(x), w(x))$ of $F(x,y,w) = 0$ is the unique solution of $\Phi(x,y) = 0$. Note that we have

$$\begin{aligned} \mathcal{J}(x) &= \{\, i \,|\, y_i(x) > 0, w_i(x) = 0\}, \\ \mathcal{K}(x) &= \{\, i \,|\, y_i(x) = w_i(x) = 0\}, \\ \mathcal{L}(x) &= \{\, i \,|\, y_i(x) = 0, w_i(x) > 0\}. \end{aligned}$$

Although smoothing functions of F have been studied extensively [1, 2, 3, 13], little attention has been paid on smoothing functions of $y(\cdot)$. To start our study on smoothing functions of $y(\cdot)$, we consider the following smoothing function of F:

$$F_\mu(x,y,w) = \begin{pmatrix} My + Nx + q - w \\ \phi_\mu(y_1, w_1) \\ \vdots \\ \phi_\mu(y_m, w_m) \end{pmatrix},$$

where

$$\phi_\mu(a,b) = a + b - \sqrt{a^2 + b^2 + 2\mu^2},$$

and $\mu > 0$ is a smoothing parameter. The function $\phi_\mu : R^2 \to R$ is called a smoothing Fischer-Burmeister function of ϕ. Kanzow [13] first introduced a smoothing parameter μ in the Fischer-Burmeister function. The smoothing function $F_\mu : R^n \times R^m \times R^m \to R^m \times R^m$ has many desirable properties. In particular, due to the P-matrix property of M, we have the following proposition [1, 2, 3, 13].

Proposition 2.1 *Let $x \in R^n$ be fixed.*

1. *For any $\mu > 0$, the equation*

$$F_\mu(x,y,w) = 0 \tag{2.3}$$

   has a unique solution $(y_\mu(x), w_\mu(x))$, which satisfies

$$y_\mu(x) > 0, \quad w_\mu(x) > 0$$

   and

$$(y_\mu(x))_i (w_\mu(x))_i = \mu^2, \quad i = 1, 2, \ldots m.$$

2. *For any fixed $\mu > 0$, F_μ is continuously differentiable with respect to $(y,w) \in R^m \times R^m$ and $F'_\mu(x,y,w)$ is nonsingular for any (y,w), where $F'_\mu(x,y,w)$ denotes the Jacobian of F_μ with respect to (y,w).*

By Proposition 2.1, the function $(y_\mu(\cdot), w_\mu(\cdot)) : R^n \to R^m \times R^m$ is well-defined since the equation $F_\mu(x,y,w) = 0$ always has a unique solution $(y_\mu(x), w_\mu(x))$ for a given x. Now we study some properties of the function $y_\mu(\cdot)$.

Theorem 2.1 *There is a positive constant κ such that for any $x \in R^n$ and $\mu > \mu' \geq 0$, we have*

$$\|y_\mu(x) - y_{\mu'}(x)\| \leq \kappa(\mu - \mu').$$

Proof: By straightforward calculus, we obtain

$$\begin{aligned} 0 &= F_\mu(x, y_\mu(x), w_\mu(x)) - F_{\mu'}(x, y_{\mu'}(x), w_{\mu'}(x)) \\ &= \begin{pmatrix} M & -I \\ I - D_1(x) & I - D_2(x) \end{pmatrix} \begin{pmatrix} y_\mu(x) - y_{\mu'}(x) \\ w_\mu(x) - w_{\mu'}(x) \end{pmatrix} \\ &\quad - (\mu - \mu') \begin{pmatrix} 0 & 0 \\ 0 & D_3(x) \end{pmatrix} \begin{pmatrix} 0 \\ e \end{pmatrix}, \end{aligned}$$

where $D_1(x), D_2(x), D_3(x) \in R^{m\times m}$ are diagonal matrices whose diagonal elements are

$$(D_1(x))_{ii} = \frac{1}{s_i}(y_\mu(x) + y_{\mu'}(x))_i$$

$$(D_2(x))_{ii} = \frac{1}{s_i}(w_\mu(x) + w_{\mu'}(x))_i$$

$$(D_3(x))_{ii} = \frac{2}{s_i}(\mu + \mu')$$

and

$$e = (1, 1, \dots, 1)^T \in R^m,$$

$$\begin{aligned} s_i &= \sqrt{(y_\mu(x))_i^2 + (w_\mu(x))_i^2 + 2\mu^2} + \sqrt{(y_{\mu'}(x))_i^2 + (w_{\mu'}(x))_i^2 + 2(\mu')^2} \\ &= (y_\mu(x))_i + (w_\mu(x))_i + (y_{\mu'}(x))_i + (w_{\mu'}(x))_i. \end{aligned}$$

The last equality follows from Proposition 2.1 (1), i.e.,

$$(y_\mu(x))_i > 0,\ (w_\mu(x))_i > 0,\ (y_\mu(x))_i(w_\mu(x))_i = \mu^2, \quad i = 1, 2, \dots m.$$

Since M is a P-matrix and all diagonal elements of $D_1(x)$ and $D_2(x) = I - D_1(x)$ lie in the interval $[0, 1]$, by [11, Theorem 4.4], the matrix

$$H(x) = \begin{pmatrix} M & -I \\ I - D_1(x) & I - D_2(x) \end{pmatrix} = \begin{pmatrix} M & -I \\ I - D_1(x) & D_1(x) \end{pmatrix}$$

is nonsingular for all $x \in R^n$. Moreover

$$\sup_{x\in R^n} \|H(x)^{-1}\| \le \max_{0\le d\le 1} \left\| \begin{pmatrix} M & -I \\ I - D & D \end{pmatrix}^{-1} \right\| < \infty,$$

where $d \in R^m$ and $D = \operatorname{diag}(d_i)$, that is, $\|H(x)^{-1}\|$ is uniformly bounded. Futhermore all diagonal elements of $D_3(x)$ lie in the interval $[0, 2]$. Hence there is a constant $\kappa \geq \|H^{-1}(x)\|\|D_3(x)\|\|e\|$ such that

$$\begin{aligned} \|y_\mu(x) - y_{\mu'}(x)\| &\leq \left\| \begin{pmatrix} y_\mu(x) - y_{\mu'}(x) \\ w_\mu(x) - w_{\mu'}(x) \end{pmatrix} \right\| \\ &\leq \|H^{-1}(x)\|\|D_3(x)\|\|e\|(\mu - \mu') \\ &\leq \kappa(\mu - \mu'), \end{aligned}$$

where κ is independent of μ, μ' and x. ∎

Recall that, when $\mu = 0$, ϕ_μ is the ordinary Fischer-Burmeister function. From Theorem 2.1, we immediately obtain the following corollary.

Corollary 2.1 *There exists a positive constant κ such that for any $x \in R^n$ and $\mu > 0$, we have*

$$\|y_\mu(x) - y(x)\| \leq \kappa\mu.$$

Now we study the differentiability of the function $y_\mu(\cdot)$. Let $G : R^n \to R^m \times R^m$ be defined by $G(x) = F(x, y(x), w(x))$. Then $G(x) \equiv 0$ and $G'(x) \equiv 0$ in R^n.

Lemma 2.3 *If the function $y(\cdot)$ is differentiable at x, then $w(\cdot)$ is also differentiable at x. Moreover, we have*

$$\nabla w_{\mathcal{K}(x)}(x) = \nabla y_{\mathcal{K}(x)}(x) = 0$$

and, for $i = 1, \ldots, m$,

$$G'_{m+i}(x) = 0 = \begin{cases} w_i'(x) & i \in \mathcal{J}(x) \\ y_i'(x) & i \in \mathcal{L}(x) \\ \alpha y_i'(x) + \beta w_i'(x) & i \in \mathcal{K}(x), \end{cases} \tag{2.4}$$

where α and β are arbitrary real numbers.

Proof: Since $w(x) = My(x) + Nx + q$, $w(\cdot)$ is differentiable at x if $y(\cdot)$ is differentiable at x. By continuity, there is a neighborhood $\mathcal{N}_x$ of x such that for all $z \in \mathcal{N}_x$,

$$G_{m+i}(z) = w_i(z) = 0, \quad i \in \mathcal{J}(x)$$

and

$$G_{m+i}(z) = y_i(z) = 0, \quad i \in \mathcal{L}(x).$$

Hence this lemma holds when $y(x)$ is a strictly complementarity solution of (1.2), i.e., $\mathcal{K}(x) = \emptyset$. Now we suppose $\mathcal{K}(x) \neq \emptyset$. By Lemma 2.1, we have

$$\nabla y_{\mathcal{K}(x)}(x) = 0$$

and

$$\begin{aligned}
\nabla w_{\mathcal{K}(x)}(x) &= \nabla y(x) M_{\mathcal{K}(x)}^T + N_{\mathcal{K}(x)}^T \\
&= \nabla y_{\mathcal{J}(x)}(x) M_{\mathcal{K}(x)\mathcal{J}(x)}^T + N_{\mathcal{K}(x)}^T \\
&= -N_{\mathcal{J}(x)}^T M_{\mathcal{J}(x)\mathcal{J}(x)}^{-T} M_{\mathcal{K}(x)\mathcal{J}(x)}^T + N_{\mathcal{K}(x)}^T \\
&= 0.
\end{aligned}$$

Hence we may write

$$G'_{m+i}(x) = 0 = \alpha y'_i(x) + \beta w'_i(x)$$

for any $\alpha, \beta \in R$. Consequently, the derivatives of G_{m+i} at x satisfy (2.4) if $y(\cdot)$ is differentiable at x. ∎

Theorem 2.2 *For any $\mu > 0$, the function $y_\mu(\cdot)$ is continuously differentiable in R^n, and there exists a bounded set $\Omega \subset R^{m\times n}$, which is independent of μ, such that $y'_\mu(x) \in \Omega$ for all $x \in R^n$. Moreover, for any $\bar{x} \in R^n$, we have*

$$\Big\{ \lim_{\substack{\mu\downarrow 0 \\ x\to\bar{x}}} y'_\mu(x) \Big\} \subseteq \partial y(\bar{x}).$$

Proof: By Proposition 2.1 and the Implicit Function Theorem 5.2.4 in [15], the function $(y_\mu(\cdot), w_\mu(\cdot))$ is continuously differentiable and

$$\begin{aligned}
\begin{pmatrix} y'_\mu(x) \\ w'_\mu(x) \end{pmatrix} &= -\begin{pmatrix} \nabla_y F_\mu(x,y,w) \\ \nabla_w F_\mu(x,y,w) \end{pmatrix}^{-T} \nabla_x F_\mu(x,y,w)^T \\
&= \begin{pmatrix} M & -I \\ I - \operatorname{diag}((t_\mu(x))_i) & \operatorname{diag}((t_\mu(x))_i) \end{pmatrix}^{-1} \begin{pmatrix} -N \\ 0 \end{pmatrix}, \qquad (2.5)
\end{aligned}$$

where

$$0 < (t_\mu(x))_i := \frac{(y_\mu(x))_i}{\sqrt{(y_\mu(x))_i^2 + (w_\mu(x))_i^2 + 2\mu^2}} = \frac{(y_\mu(x))_i}{(y_\mu(x))_i + (w_\mu(x))_i} < 1,$$

for $i = 1, 2, \ldots, m$. Here we use

$$(y_\mu(x))_i > 0,\ (w_\mu(x))_i > 0,\ (y_\mu(x))_i (w_\mu(x))_i = \mu^2.$$

By Corollary 2.1 and Lemma 2.2,

$$\begin{aligned}
\|y_\mu(x) - y(\bar{x})\| &\le \|y_\mu(x) - y(x)\| + \|y(x) - y(\bar{x})\| \\
&\le \kappa\mu + \gamma\|x - \bar{x}\|
\end{aligned}$$

and

$$\begin{aligned}\|w_\mu(x) - w(\bar{x})\| &\leq \|M\|\|y_\mu(x) - y(\bar{x})\| + \|N\|\|x - \bar{x}\| \\ &\leq \kappa\mu\|M\| + (\gamma\|M\| + \|N\|)\|x - \bar{x}\|.\end{aligned}$$

Hence we have

$$y_\mu(x) \to y(\bar{x}), \quad w_\mu(x) \to w(\bar{x}), \quad \text{as } \mu \downarrow 0,\ x \to \bar{x}.$$

This implies that

$$\Big\{ \lim_{\substack{\mu\downarrow 0 \\ x\to\bar{x}}} (t_\mu(x))_i \Big\} \subseteq \begin{cases} \{1\} & i \in \mathcal{J}(\bar{x}) \\ \{0\} & i \in \mathcal{L}(\bar{x}) \\ [0,1] & i \in \mathcal{K}(\bar{x}). \end{cases}$$

Moreover, since M is a P-matrix, by [11, Theorem 4.4], the matrix

$$H(x) = \begin{pmatrix} M & -I \\ I - \mathrm{diag}((t_\mu(x))_i) & \mathrm{diag}((t_\mu(x))_i) \end{pmatrix}$$

is nonsingular and its inverse is uniformly bounded for all $0 \leq (t_\mu(x))_i \leq 1$. Hence, by (2.5), the sequence $\{(y'_\mu(x), w'_\mu(x))\}$ is bounded when $\mu \downarrow 0$ and $x \to \bar{x}$. Let $(y^\circ(\bar{x}), w^\circ(\bar{x}))$ be any accumulation point of $\{(y'_\mu(x), w'_\mu(x))\}$ as $\mu \downarrow 0$ and $x \to \bar{x}$. Then there exists a diagonal matrix $\mathrm{diag}(\bar{t}_i)$ whose diagonal elements are

$$\bar{t}_i \in \begin{cases} \{1\} & i \in \mathcal{J}(\bar{x}) \\ \{0\} & i \in \mathcal{L}(\bar{x}) \\ [0,1] & i \in \mathcal{K}(\bar{x}), \end{cases}$$

such that

$$\begin{pmatrix} y^\circ(\bar{x}) \\ w^\circ(\bar{x}) \end{pmatrix} = \begin{pmatrix} M & -I \\ I - \mathrm{diag}(\bar{t}_i) & \mathrm{diag}(\bar{t}_i) \end{pmatrix}^{-1} \begin{pmatrix} -N \\ 0 \end{pmatrix}.$$

On the other hand, the function $(y(\cdot), w(\cdot))$ is actually piecewise linear. Thus, by Rademacher's Theorem, $(y(\cdot), w(\cdot))$ is almost everywhere differentiable. Let $\{x_k\}$ be a sequence converging to $\bar{x}$ and $(y(\cdot), w(\cdot))$ be differentiable at x_k for each k. Recall that, by Lemma 2.3, the function G defined by $G(x) = F(x, y(x), w(x))$ vanishes everywhere and particularly satisfies (2.4).

Now we differentiate both sides of $G(x) = 0$ at x^k, and obtain from Lemma 2.3

$$\begin{pmatrix} My'(x_k) + N - w'(x_k) \\ (I - \mathrm{diag}((\tau_k)_i))y'(x_k) + \mathrm{diag}((\tau_k)_i)w'(x_k) \end{pmatrix} = 0,$$

where

$$(\tau_k)_i = \begin{cases} 1 & i \in \mathcal{J}(x_k) \\ 0 & i \in \mathcal{L}(x_k) \\ (\alpha_k)_i & i \in \mathcal{K}(x_k) \end{cases} \tag{2.6}$$

with $(\alpha_k)_i$ being an arbitrary number in $[0,1]$. Hence we get

$$\begin{pmatrix} y'(x_k) \\ w'(x_k) \end{pmatrix} = \begin{pmatrix} M & -I \\ I - \operatorname{diag}((\tau_k)_i) & \operatorname{diag}((\tau_k)_i) \end{pmatrix}^{-1} \begin{pmatrix} -N \\ 0 \end{pmatrix}.$$

Passing $\{x_k\}$ to $\bar{x}$, we obtain an element in $\partial(y,w)(\bar{x})$. Suppose that there is $i \in \mathcal{K}(\bar{x})$. If $i \in \mathcal{K}(x_k) \cap \mathcal{K}(\bar{x})$ for all sufficiently large k, then by (2.6), $(\tau_k)_i$ can be any number in $[0,1]$. Otherwise, we can conclude that $y_i(\cdot)$ is not differentiable at $\bar{x}$. In fact, if $y_i(\cdot)$ is differentiable at $\bar{x}$, then from the piecewise linearity and the nonnegativity of $y_i(\cdot)$ and $w_i(\cdot)$, the differentiability at $\bar{x}$ implies that there is a neighborhood $\mathcal{N}_{\bar{x}}$ of $\bar{x}$ such that for all $x \in \mathcal{N}_{\bar{x}}$, $y_i(x) = w_i(x) = 0$, so that $i \in \mathcal{K}(x) \cap \mathcal{K}(\bar{x})$. This is a contradiction. Therefore, from the nondifferentiability of $y_i(\cdot)$ at $\bar{x}$, we may find two sequences $\{x_{k_1}\}, \{x_{k_2}\} \subset D_y$ such that

$$\lim_{k_1 \to \infty} x_{k_1} = \bar{x}, \qquad i \in \mathcal{J}(x_{k_1}),$$

and

$$\lim_{k_2 \to \infty} x_{k_2} = \bar{x}, \qquad i \in \mathcal{L}(x_{k_2}) \cup \mathcal{K}(x_{k_2}).$$

Then, corresponding to these sequences, there are two sequences $\{(\tau_{k_1})_i\}$ and $\{(\tau_{k_2})_i\}$ such that $\lim\limits_{k_1 \to \infty} (\tau_{k_1})_i = 1$ and $\lim\limits_{k_2 \to \infty} (\tau_{k_2})_i = 0$. Note that the Clarke generalized Jacobian of $(y(\cdot), w(\cdot))$ at $\bar{x}$ is the convex hull of accumulation points of $\{(y'(x_k), w'(x_k))\}$ corresponding to all such sequences. Since the convex hull of all accumulation points of $\{(\tau_k)_i\}$ is the interval $[0,1]$ and, for sufficiently large k, $\mathcal{J}(\bar{x}) \subseteq \mathcal{J}(x_k)$ and $\mathcal{L}(\bar{x}) \subseteq \mathcal{L}(x_k)$, we can claim that

$$\begin{pmatrix} M & -I \\ I - \operatorname{diag}(\bar{\tau}_i) & \operatorname{diag}(\bar{\tau}_i) \end{pmatrix}^{-1} \begin{pmatrix} -N \\ 0 \end{pmatrix}$$

belongs to $\partial(y,w)(\bar{x})$, if

$$\bar{\tau}_i \in \begin{cases} \{1\} & i \in \mathcal{J}(\bar{x}) \\ \{0\} & i \in \mathcal{L}(\bar{x}) \\ [0,1] & i \in \mathcal{K}(\bar{x}). \end{cases}$$

Therefore, we have $(y^\circ(\bar{x}), w^\circ(\bar{x})) \in \partial(y,w)(\bar{x})$. Moreover, by [5, Proposition 2.6.2], we have $\partial(y,w)(\bar{x}) \subseteq \partial y(\bar{x}) \times \partial w(\bar{x})$, and hence $y^\circ(\bar{x}) \in \partial y(\bar{x})$. ∎

The next lemma gives a sufficient condition for the functions $(y_\mu(\cdot))_i$ to be convex. This result will then imply that the functions $y_i(\cdot)$ are convex under the same condition (Corollary 2.1).

Lemma 2.4 *Let M be a triangular P-matrix with nonpositive off-diagonal entries. Then, for any $\mu > 0$, each $(y_\mu(\cdot))_i$ is a convex function on R^n.*

Proof: Without loss of generality, we only consider the case where M is lower triangular. The P-matrix property gives that all diagonal elements m_{ii} of M are positive. Then it is not difficult to deduce from $(y_\mu(x))_i(w_\mu(x))_i = \mu^2$, $i = 1, 2, \ldots, m$, and $y_\mu(x) > 0$ that

$$(y_\mu(x))_i = \frac{1}{2m_{ii}}\left(-(r_\mu(x))_i + \sqrt{(r_\mu(x))_i^2 + 4m_{ii}\mu^2}\right),$$

where

$$(r_\mu(x))_i = \begin{cases} (Nx+q)_1 & i = 1 \\ \sum_{j=1}^{i-1} m_{ij}(y_\mu(x))_j + (Nx+q)_i & i = 2, \ldots, m. \end{cases}$$

By induction, we show that $(-r_\mu(x))_i$ and $(y_\mu(x))_i$ are convex for all $i = 1, 2, \ldots, m$. First, let us define a function $b_i : R \to R$ by

$$b_i(t) = \frac{1}{2m_{ii}}(-t + \sqrt{t^2 + 4m_{ii}\mu^2}).$$

The function b_i is strictly convex and decreasing as $b_i'(t) < 0$ and $b_i''(t) > 0$ for all $t \in R$. Using this function, we can write

$$(y_\mu(x))_i = b_i((r_\mu(x))_i).$$

Let x and z be two points in R^n. If $(-r_\mu(x))_i$ is convex, i.e, $(r_\mu(x))_i$ is concave, then $(y_\mu(x))_i$ is convex, which is shown as follows:

$$\begin{aligned} (y_\mu(\alpha x + (1-\alpha)z))_i &= b_i((r_\mu(\alpha x + (1-\alpha)z))_i) \\ &\le b_i(\alpha(r_\mu(x))_i + (1-\alpha)(r_\mu(z))_i) \\ &\le \alpha b_i((r_\mu(x))_i) + (1-\alpha)b_i((r_\mu(z))_i) \\ &= \alpha(y_\mu(x))_i + (1-\alpha)(y_\mu(z))_i. \end{aligned}$$

For $i = 1$, $(-r_\mu(x))_i$ is linear and hence it is convex, and so $(y_\mu(x))_1$ is convex. If $(y_\mu(x))_i$, $i = 1, 2, \ldots, k-1$, are convex, then since $m_{ij} \le 0$ for $i \ne j$ and

$$(-r_\mu(x))_k = -\sum_{j=1}^{k-1} m_{ij}(y_\mu(x))_j - (Nx+q)_k,$$

$(-r_\mu(x))_k$ is convex. Hence $(y_\mu(\cdot))_k$ is convex. ∎

Any P-matrix with nonpositive off-diagonal entries is an M-matrix [6, Definition 3.11.1]. Recall that a matrix $A \in R^{m\times m}$ is called an M-matrix [15] if A is invertible, $a_{ij} \le 0$ for $i \ne j$, and $A^{-1} \ge 0$, where the last condition means all entries of A^{-1} are nonnegative. An M-matrix is also called a K-matrix [6].

By Corollary 2.1, letting $\mu \to 0$ in

$$(y_\mu(\alpha x + (1-\alpha)z))_i \le \alpha(y_\mu(x))_i + (1-\alpha)(y_\mu(z))_i,$$

we obtain the following corollary.

Corollary 2.2 *Suppose that M is a triangular P-matrix with nonpositive off-diagonal entries, then each $y_i(\cdot)$ is a convex function on R^n.*

3 Convergence analysis

In this section, we investigate the limiting behaviour of sequences of optimal solutions, KKT points and B-stationary points of the smoothing problem (1.6) as $\mu \downarrow 0$. We denote the feasible set, the solution set and the optimal value of the reduced problem (1.3) by $S = \{x \in R^n \mid g(x) \le 0,\ h(x) = 0\}$, X^* and θ^*, respectively.

Theorem 3.1 *Assume that the level set*

$$D = \{x \in S \mid \theta(x) \le \theta^* + \Gamma\}$$

is bounded for a positive constant Γ. Then there is $\mu_\Gamma > 0$ such that for every $\mu \in (0, \mu_\Gamma]$, the solution set X^*_μ of the smoothing problem (1.6) is nonempty and uniformly bounded. Moreover, we have

$$\Big\{ \lim_{\substack{\mu\downarrow 0 \\ x_\mu \in X^*_\mu}} x_\mu \Big\} \subseteq X^*.$$

Proof: The Lipschitz continuity of $y(\cdot)$ and the boundedness of D yield that $Z = \{(x, y(x)) \mid x \in D\}$ is bounded. By the mean-value theorem, we have

$$\begin{aligned}
\theta_\mu(x) &= f(x, y_\mu(x)) \\
&= f(x, y(x)) + (y_\mu(x) - y(x))^T \nabla_y f(x, y_\mu(x) + \alpha(y(x) - y_\mu(x))) \\
&= \theta(x) + (y_\mu(x) - y(x))^T \nabla_y f(x, y_\mu(x) + \alpha(y(x) - y_\mu(x))),
\end{aligned}$$

for some $\alpha \in [0,1]$. From Corollary 2.1 and the continuity of $\nabla_y f$, the boundedness of Z implies that there are $\mu_1 > 0$ and $\kappa_1 > 0$ such that $\|\nabla_y f(x, y_\mu(x) +$

$\alpha(y(x) - y_\mu(x)))\| \le \kappa_1$ for all $\mu \in (0, \mu_1]$ and $x \in D$. Hence, again by Corollary 2.1, there is $\kappa_0 > 0$ such that

$$|\theta(x) - \theta_\mu(x)| \le \kappa_0 \mu, \quad \text{for all } \ x \in D, \ \mu \in (0, \mu_1]. \tag{3.1}$$

Let

$$\mu_\Gamma = \min(\mu_1, \Gamma/(2\kappa_0)).$$

Then for any $x^* \in X^*$ and any $\mu \in (0, \mu_\Gamma]$, we obtain

$$\theta_\mu(x^*) \le \theta(x^*) + \kappa_0\mu \le \theta^* + \Gamma/2,$$

which implies that the level set

$$D_\mu = \{x \in S \mid \theta_\mu(x) \le \theta^* + \Gamma/2\}$$

is nonempty. Suppose that $x \in D_\mu$ and $\mu \in (0, \mu_\Gamma]$. Then, by (3.1), we have

$$\theta(x) \le \theta_\mu(x) + \kappa_0\mu \le \theta^* + \Gamma,$$

that is, $x \in D$, so $D_\mu \subseteq D$. Hence the solution set X_μ^* of (1.6) is nonempty and bounded. Furthermore, $X_\mu^* \subseteq D_\mu \subseteq D$.

Let $\{\mu_k\}$ be a sequence of positive numbers tending to 0, and choose $x_{\mu_k} \in X_{\mu_k}^*$ for each k. Then the sequence $\{x_{\mu_k}\} \subseteq D$ is bounded. Without loss of generality, we assume $x_{\mu_k} \to \bar{x}$. Then $\bar{x} \in D$, and so it is a feasible solution of the reduced problem (1.3).

Applying (3.1) to a global solution x^* of (1.3), we have

$$\theta_{\mu_k}(x_{\mu_k}) \le \theta_{\mu_k}(x^*) \le \theta(x^*) + \kappa_0\mu_k.$$

Therefore, we have

$$\theta(\bar{x}) = \lim_{\substack{\mu_k \downarrow 0 \\ x_{\mu_k} \to \bar{x}}} \theta_{\mu_k}(x_{\mu_k}) \le \theta(x^*) = \theta^*.$$

This, together with $\bar{x} \in S$, shows that $\theta(\bar{x}) = \theta^*$ and so $\bar{x} \in X^*$. ∎

Corollary 3.3 *Assume that the feasible set S of the reduced problem (1.3) is bounded. Then for every $\mu > 0$, the solution set X_μ^* of the smoothing problem (1.6) is nonempty and bounded. Moreover, we have*

$$\Big\{ \lim_{\substack{\mu \downarrow 0 \\ x_\mu \in X_\mu^*}} x_\mu \Big\} \subseteq X^*.$$

We say $(x_\mu, \lambda_\mu, \nu_\mu) \in R^n \times R^\ell \times R^{\ell'}$ is a KKT point of the smoothing problem (1.6) if it satisfies the following conditions:

$$\begin{array}{rcl} 0 &=& \nabla_x L_\mu(x, \lambda, \nu), \\ 0 &=& \min(\lambda, -g(x)), \\ 0 &=& h(x), \end{array} \tag{3.2}$$

where

$$L_\mu(x, \lambda, \nu) = f(x, y_\mu(x)) + \lambda^T g(x) + \nu^T h(x).$$

We say $(\bar{x}, \bar{\lambda}, \bar{\nu}) \in R^n \times R^\ell \times R^{\ell'}$ is a *weak KKT point* or a *Clarke stationary point* of the reduced problem (1.3) if it satisfies the following conditions:

$$\begin{array}{rcl} 0 &\in& \partial_x L(x, \lambda, \nu), \\ 0 &=& \min(\lambda, -g(x)), \\ 0 &=& h(x), \end{array} \tag{3.3}$$

where

$$L(x, \lambda, \nu) = f(x, y(x)) + \lambda^T g(x) + \nu^T h(x).$$

We say that the Mangasarian-Fromovitz constraint qualification (MFCQ) holds at $x \in S$ if $\{\nabla h_j(x) \,|\, j = 1, \ldots, \ell'\}$ is a linearly independent set and there exists a vector $d \in R^n$ such that

$$\nabla g_i(x)^T d < 0, \;\; i \in \mathcal{I}(x), \quad \nabla h_j(x)^T d = 0, \;\; j = 1, \ldots, \ell',$$

where $\mathcal{I}(x) = \{i \,|\, g_i(x) = 0\}$.

Theorem 3.2 *Suppose that the feasible set S of the reduced problem (1.3) is bounded and the MFCQ holds at any $x \in S$. Then for every $\mu > 0$, a KKT point $(x_\mu, \lambda_\mu, \nu_\mu)$ of the smoothing problem (1.6) exists and $\{(x_\mu, \lambda_\mu, \nu_\mu) \,|\, \mu > 0\}$ is contained in a bounded set. Moreover, every accumulation point $(\bar{x}, \bar{\lambda}, \bar{\nu})$ of $\{(x_\mu, \lambda_\mu, \nu_\mu)\}$ as $\mu \downarrow 0$ is a weak KKT point of the reduced problem (1.3).*

Proof: The existence of a KKT point $(x_\mu, \lambda_\mu, \nu_\mu)$ of (1.6) for every $\mu > 0$ follows from Corollary 3.3 and the MFCQ. The continuous differentiabilty of $f, g, h,$ and $y_\mu(\cdot)$ implies that $(x_\mu, \lambda_\mu, \nu_\mu)$ satisfies

$$\begin{array}{rcl} 0 &=& \nabla_x f(x_\mu, y_\mu(x_\mu)) + \nabla y_\mu(x_\mu) \nabla_y f(x_\mu, y_\mu(x_\mu)) \\ && + \nabla g(x_\mu)\lambda_\mu + \nabla h(x_\mu)\nu_\mu, \\ 0 &=& \min(\lambda_\mu, -g(x_\mu)), \\ 0 &=& h(x_\mu). \end{array} \tag{3.4}$$

Since S is bounded, the sequence $\{x_\mu\}$ is contained in a bounded set. Let $\bar{x}$ be an accumulation point of $\{x_\mu\}$ as $\mu \downarrow 0$. Then by Theorem 2.2, we have

$$\Big\{ \lim_{\substack{\mu \downarrow 0 \\ x_\mu \to \bar{x}}} y'_\mu(x_\mu) \Big\} \subseteq \partial y(\bar{x}).$$

Moreover, by the continuity of $\nabla f, \nabla g, \nabla h$, and $y(\cdot)$, we have

$$\lim_{\substack{\mu \downarrow 0 \\ x_\mu \to \bar{x}}} \nabla_x f(x_\mu, y_\mu(x_\mu)) = \nabla_x f(\bar{x}, y(\bar{x})),$$

$$\lim_{\substack{\mu \downarrow 0 \\ x_\mu \to \bar{x}}} \nabla_y f(x_\mu, y_\mu(x_\mu)) = \nabla_y f(\bar{x}, y(\bar{x})),$$

$$\lim_{x_\mu \to \bar{x}} \nabla g(x_\mu) = \nabla g(\bar{x})$$

and

$$\lim_{x_\mu \to \bar{x}} \nabla h(x_\mu) = \nabla h(\bar{x}).$$

Hence from (3.4) and the MFCQ, we have the boundedness of $\{(\lambda_\mu, \nu_\mu)\}$. Assuming, without loss of generality, that $(x_\mu, \lambda_\mu, \nu_\mu) \to (\bar{x}, \bar{\lambda}, \bar{\nu})$, we obtain

$$\begin{array}{rcl} 0 & \in & \nabla_x f(\bar{x}, y(\bar{x})) + \partial y(\bar{x})^T \nabla_y f(\bar{x}, y(\bar{x})) + \nabla g(\bar{x})\bar{\lambda} + \nabla h(\bar{x})\bar{\nu}, \\ 0 & = & \min(\bar{\lambda}, -g(\bar{x})), \\ 0 & = & h(\bar{x}). \end{array}$$

By the Chain Rule [5, Theorem 2.6.6], the continuous differentiability of f yields

$$\nabla_x f(\bar{x}, y(\bar{x}))^T + \nabla_y f(\bar{x}, y(\bar{x}))^T \partial y(\bar{x}) = \partial f(\bar{x}, y(\bar{x})).$$

Using the continuous differentiability of g, h and Corollary 2 of [5, Proposition 2.3.3], we obtain

$$\nabla_x f(\bar{x}, y(\bar{x})) + \partial y(\bar{x})^T \nabla_y f(\bar{x}, y(\bar{x})) + \nabla g(\bar{x})\bar{\lambda} + \nabla h(\bar{x})\bar{\nu} = \partial_x L(\bar{x}, \bar{\lambda}, \bar{\nu}).$$

Hence $(\bar{x}, \bar{\lambda}, \bar{\nu})$ is a weak KKT point of (1.3). ∎

According to [6, Theorem 7.4.2], the function $y(\cdot)$ is directionally differentiable everywhere. The directional derivative of $y(\cdot)$ at $x \in R^n$ along a direction $h \in R^n$ is defined by

$$y'(x; h) = \lim_{t \downarrow 0} \frac{y(x + th) - y(x)}{t}.$$

We say $\bar{x} \in S$ is a *B-stationary point* of the reduced problem (1.3) if

$$\theta'(\bar{x}; h) = \nabla_x f(\bar{x}, y(\bar{x}))^T h + y'(\bar{x}; h)^T \nabla_y f(\bar{x}, y(\bar{x})) \geq 0, \quad \text{for all} \quad h \in \mathcal{T}(\bar{x}; S),$$

where $\mathcal{T}(\bar{x}; S)$ is the tangent cone of S at $\bar{x}$ that consists of all vectors h for which there exist a sequence $\{x_k\} \subseteq S$ converging to $\bar{x}$ and a sequence $\{\tau_k\}$ of positive numbers converging to zero such that $h = \lim_{k\to\infty}(x_k - \bar{x})/\tau_k$.

We say $x_\mu \in S$ is a B-stationary point of the smoothing problem (1.6) if

$$\begin{aligned} \theta'_\mu(x_\mu)^T h = \nabla_x f(x_\mu, y_\mu(x_\mu))^T h + \nabla_y f(x_\mu, y_\mu(x_\mu))^T y'_\mu(x_\mu) h \geq 0, \\ \text{for all} \quad h \in \mathcal{T}(x_\mu; S). \end{aligned} \tag{3.5}$$

For the smoothing problem (1.6), a KKT point is a B-stationary point and vise versa under a standard constraint qualification such as MFCQ. However, for the reduced problem (1.3), a weak KKT (Clarke stationary) point need not be a B-stationary point even if MFCQ is satisfied; thus the Clarke stationarity is a much weaker concept than the B-stationarity.

Lemma 3.1 *Let x_μ be a B-stationary point of the smoothing problem (1.6) and $\bar{x} \in S$ be an accumulation point of $\{x_\mu\}$ as $\mu \downarrow 0$. If MFCQ holds at $\bar{x}$, then there exists a $V \in \partial y(\bar{x})$ such that*

$$\nabla_x f(\bar{x}, y(\bar{x}))^T h + \nabla_y f(\bar{x}, y(\bar{x}))^T V h \geq 0, \quad \textit{for all} \quad h \in \mathcal{T}(\bar{x}; S).$$

Proof: By Theorem 2.2, any accumulation point of $\{y'_\mu(x_\mu)\}$ belongs to $\partial y(\bar{x})$. Taking a subsequence if necessary, we suppose that $y'_\mu(x_\mu) \to V \in \partial y(\bar{x})$. Choose $h \in \mathcal{T}(\bar{x}; S)$ arbitrarily. Then, under the MFCQ, it is not difficult to see that there exists $\{h_\mu\}$ such that $h_\mu \to h$ as $\mu \downarrow 0$ and $h_\mu \in \mathcal{T}(x_\mu; S)$ for each $\mu > 0$. Moreover, by taking the limit in (3.5), we can deduce that

$$\nabla_x f(\bar{x}, y(\bar{x}))^T h + \nabla_y f(\bar{x}, y(\bar{x}))^T V h \geq 0.$$

Since h is an arbitrary element in $\mathcal{T}(\bar{x}; S)$, we have the desired result. ∎

A function $\psi : R^n \to R$ is said to be *regular* at $x \in R^n$ if, for all $h \in R^n$, the directional derivative $\psi'(x; h)$ exists and

$$\psi'(x; h) = \limsup_{\substack{z \to x \\ t \downarrow 0}} \frac{\psi(z + th) - \psi(z)}{t}.$$

Theorem 3.3 *Suppose that $(y(\cdot))_i : R^n \to R$ is a convex function for each $i = 1, 2, \ldots, m$. If $\nabla_y f(x, y(x))$ has nonnegative components for all $x \in R^n$, S is bounded, and MFCQ holds at any $x \in S$, then the following statements are true.*

1. *θ is a regular function on R^n.*

2. *Let $\{\mu_k\}$ be a sequence of positive numbers converging to zero, and $\{x_{\mu_k}\}$ be a corresponding sequence of B-stationary points of the smoothing problems (1.6). Then every accumulation point of $\{x_{\mu_k}\}$ is a B-stationary point of the reduced problem (1.3).*

Proof: 1. It is shown in [5] that a convex function is regular. This statement then follows from the chain rule for regular functions [5, Theorem 2.3.9].

2. Since S is bounded, the sequence $\{x_{\mu_k}\}$ is contained in a compact set. Let $\bar{x}$ be an arbitrary accumulation point of $\{x_{\mu_k}\}$. Then by Lemma 3.1, there is a $V \in \partial y(\bar{x})$ such that

$$\nabla_x f(\bar{x}, y(\bar{x}))^T h + \nabla_y f(\bar{x}, y(\bar{x}))^T V h \geq 0, \quad \text{for all} \;\; h \in \mathcal{T}(\bar{x}; S). \tag{3.6}$$

By [5, Theorem 2.6.6],

$$\nabla_x f(\bar{x}, y(\bar{x}))^T + \nabla_y f(\bar{x}, y(\bar{x}))^T V \in \partial\theta(\bar{x}). \tag{3.7}$$

Since θ is regular, by [5, Proposition 2.1.2], we have

$$\theta'(\bar{x}; h) = \limsup_{\substack{z \to \bar{x} \\ t \downarrow 0}} \frac{\theta(z+th) - \theta(z)}{t} = \max\{\xi^T h \,:\, \xi \in \partial\theta(\bar{x})\}, \quad \text{for all} \;\; h \in R^n.$$

Therefore, it follows from (3.6) and (3.7) that

$$\theta'(\bar{x}; h) \geq \nabla_x f(\bar{x}, y(\bar{x}))^T h + \nabla_y f(\bar{x}, y(\bar{x}))^T V h \geq 0, \quad \text{for all} \;\; h \in \mathcal{T}(\bar{x}; S).$$

This shows that $\bar{x}$ is a B-stationary point of the reduced problem (1.3). ∎

4 An application

As an application, we consider the mathematical program (1.1) with

$$f(x, y) = f_1(x) + f_2(y),$$

or more generally

$$f(x, y) = \varphi(f_1(x), f_2(y)),$$

where $\varphi : R^2 \to R$ is a nondecreasing convex function, $f_1 : R^n \to R$ is a convex function, $f_2 : R^m \to R$ is a nondecreasing convex function. Recall that a function $\psi : R^t \to R$ is called nondecreasing if $z \leq z'$ implies $\psi(z) \leq \psi(z')$. Then we have the following lemma [12].

Lemma 4.2 *Let $\psi : R^t \to R$ be a nondecreasing convex function and $p_i : R^n \to R$ be convex functions for $i = 1, 2, \ldots, t$. Then the function $p_0 : R^n \to R$ defined by $p_0(x) = \psi(p_1(x), p_2(x), \ldots, p_t(x))$ is convex.*

Now suppose that the functions $y_i(\cdot) : R^n \to R$ are convex for $i = 1, 2, \ldots, m$, which is particularly the case when M is a triangular P-matrix with nonpositive off-diagonal elements. Then, by Corollary 2.2 and Lemma 4.2, it can be shown that $f_2(y(\cdot)) : R^n \to R$ is a convex function, and so the objective function $\theta(x) = \varphi(f_1(x), f_2(y(x)))$ of the reduced problem (1.3) is convex. Furthermore, by Lemmas 2.4 and 4.2, the function $f_2(y_\mu(\cdot))$ is also convex. Therefore, the objective function $\theta_\mu(x) = \varphi(f_1(x), f_2(y_\mu(x)))$ of the smoothing problem (1.6) is continuously differentiable and convex. Moreover, if $g_i, i = 1, 2, \ldots, \ell$, are convex and $h_j, j = 1, 2, \ldots, \ell'$, are affine, then the reduced problem (1.3) is a convex program and the smoothing problem (1.6) is a continuously differentiable convex program. In this case, a KKT point and a B-stationary point of the smoothing problem (1.6) are a global optimal solution of (1.6). Under the assumption that $D = \{x \in S \,|\, \theta(x) \le \theta^* + \Gamma\}$ is bounded, by solving the KKT system of (1.6) for $\mu_k > 0$ successively, we can find a solution x^* of (1.3) and then a solution $(x^*, y(x^*))$ of the original problem (1.1). Note that in this case, the feasible set of the original problem (1.1) can be disconnected and the feasible set of its smoothing problem (1.5) is not convex.

References

[1] B. Chen and P.T. Harker, Smooth approximations to nonlinear complementarity problems, SIAM J. Optim., 7 (1997), 403–420.

[2] C. Chen and O.L. Mangasarian, A class of smoothing functions for nonlinear and mixed complementarity problems, Comp. Optim. Appl., 5 (1996), 97–138.

[3] X. Chen, L. Qi and D. Sun, Global and superlinear convergence of the smoothing Newton method and its application to general box constrained variational inequalities, Math. Comp., 67 (1998), 519–540.

[4] Y. Chen and M. Florian, The nonlinear bilevel programming problem: Formulation, regularity and optimality conditions, Optimization, 32 (1995), 193–209.

[5] F.H. Clarke, Optimization and Nonsmooth Analysis, SIAM, Philadelphia, PA, 1990.

[6] R.W. Cottle, J.-S. Pang and R.E. Stone, The Linear Complementarity Problem, Academic Press, New York, NY, 1992.

[7] F. Facchinei, H. Jiang and L. Qi, A smoothing method for mathematical programs with equilibrium constraints, Math. Programming, 85 (1999), 107–134.

[8] A. Fischer, A special Newton-type optimization method, Optimization, 24 (1992), 269-284.

[9] M. Fukushima, Z.-Q. Luo and J.-S. Pang, A globally convergent sequential programming algorithm for mathematical programs with linear complementarity constraints, Comp. Optim. Appl., 10 (1998), 5–30.

[10] M. Fukushima and J.-S. Pang, Convergence of a smoothing continuation method for mathematical programs with complementarity constraints, Ill-posed Variational Problems and Regularization Techniques, Lecture Notes in Economics and Mathematical Systems, Vol. 477, M. Théra and R. Tichatschke (eds.), Springer-Verlag, Berlin/Heidelberg, 1999, pp. 99–110.

[11] S. A. Gabriel and J.J. Moré, Smoothing of mixed complementarity problems, Complementarity and Variational Problems- State of the Art, M.C. Ferris and J.-S. Pang (eds.), SIAM, Philadelphia, PA, 1997, pp. 105–116.

[12] J.-B. Hiriart-Urruty and C. Lemaréchal, Convex Analysis and Minimization Algorithms I, Springer-Verlag, Berlin/Heidelberg, 1993.

[13] C. Kanzow, Some noninterior continuation methods for linear complementarity problems, SIAM J. Matrix Analysis Appl., 17 (1996), 851–868.

[14] Z.Q. Luo, J.-S. Pang and D. Ralph, Mathematical Programs with Equilibrium Constraints, Cambridge University Press, Cambridge, UK, 1996.

[15] J.M. Ortega and W.C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York, NY, 1970.

[16] L. Qi and J. Sun, A nonsmooth version of Newton's method, Math. Programming, 58 (1993), 353–367.