A Class of Smoothing Functions for Nonlinear and Mixed Complementarity Problems

Dedicated to Richard W. Cottle, friend and colleague,

on the occasion of his sixtieth birthday

Chunhui Chen* & O. L. Mangasarian*

Mathematical Programming Technical Report 94-11
August 1994 / Revised February 1995

Abstract

We propose a class of parametric smooth functions that approximate the fundamental plus function, $(x)_+ = \max\{0, x\}$, by twice integrating a probability density function. This leads to classes of smooth parametric nonlinear equation approximations of nonlinear and mixed complementarity problems (NCPs and MCPs). For any solvable NCP or MCP, existence of an arbitrarily accurate solution to the smooth nonlinear equation as well as the NCP or MCP, is established for sufficiently large value of a smoothing parameter α. Newton-based algorithms are proposed for the smooth problem. For strongly monotone NCPs, global convergence and local quadratic convergence are established. For solvable monotone NCPs, each accumulation point of the proposed algorithms solves the smooth problem. Exact solutions of our smooth nonlinear equation for various values of the parameter α, generate an interior path, which is different from the central path for interior point method. Computational results for 52 test problems compare favorably with those for another Newton-based method. The smooth technique is capable of solving efficiently the test problems solved by Dirkse & Ferris [8], Harker & Xiao [13] and Pang & Gabriel [30].

1 Introduction

The complementarity condition

$$0 \leq x \perp y \geq 0,$$

where x and y are vectors in R^n and the symbol $\perp$ denotes orthogonality, plays a fundamental role in mathematical programming. Many problems can be formulated by using this complementarity condition. For example, most optimality conditions of mathematical programming [26] as well as variational inequalities [6] and extended complementarity problems [23, 11, 40] can be so formulated. It is obvious that the vectors x and y satisfy complementarity condition if and only if

$$x = (x - y)_+,$$

*Computer Sciences Department, University of Wisconsin, 1210 West Dayton Street, Madison, WI 53706, email: *chunhui@cs.wisc.edu, olvi@cs.wisc.edu.* This material is based on research supported by Air Force Office of Scientific Research Grant F49620-94-1-0036 and National Science Foundation Grant CCR-9322479.

where the plus function $(\cdot)_+$ is defined as

$$(\xi)_+ = \max\{\xi, 0\},$$

for a real number ξ. For a vector x, the vector $(x)_+$ denotes the plus function applied to each component of x. In this sense, the plus function plays an important role in mathematical programming. But one big disadvantage of the plus function is that it is not smooth because it is not differentiable. Thus numerical methods that use gradients cannot be directly applied to solve a problem involving a plus function. The basic idea of this paper is to use a smooth function approximation to the plus function. With this approximation, many efficient algorithms, such as the Newton method, can be easily employed.

There are many Newton-based algorithms for solving nonlinear complementarity problems, variational inequalities and mixed complementarity problems. In [12] a good summary and references up to 1988 are given. Generalizations of the Newton method to nonsmooth equations can be found in [34], [35] and [36]. Since then, several approaches based on B-differentiable equations were investigated in [13], [28] and [29]. In addition, an algorithm based on nonsmooth equations and successive quadratic programming was given [30], as well as a Newton method with a path following technique [32, 8], and a trust region Newton method for solving a nonlinear least squares reformulation of the NCP [24]. With the exception of [24], a feature common to all these methods is that the subproblem at each Newton iteration is still a combinatorial problem. In contrast, by using the smooth technique proposed here, we avoid this combinatorial difficulty by approximately reformulating the nonlinear or mixed complementarity problem as a smooth nonlinear equation. Consequently, at each Newton step, we only need to solve a linear equation. This is much simpler than solving a mixed linear complementarity problem or a quadratic program.

Smoothing techniques have already been applied to different problems, such as, l_1-minimization problems [21], multi-commodity flow problems [31], nonsmooth programming [41, 20], linear and convex inequalities [5], and linear complementarity problems [4], [5] and [17]. These successful techniques motivate a systematic study of the smoothing approach. Questions we wish to address include the following. How to generate new smoothing functions? What is a common property of smoothing functions?

In Section 2, we relate the plus function through a parametric smoothing procedure, to a probability density function with a parameter β. As the parameter β approaches zero, the smooth plus function approaches the nonsmooth plus function $(\cdot)_+$. This gives us a tool for generating a class of smooth plus functions and a systematic way to develop properties of these functions. In Section 3, we approximate the NCP by a smooth parametric nonlinear equation. For the strongly monotone case, we establish existence of a solution for the nonlinear equation and estimate the distance between its solution and the solution of original NCP. For a general solvable NCP, existence of an arbitrarily accurate solution to the nonlinear equation, and hence to the NCP, is established. For a fixed value of the smoothing parameter $\alpha = \frac{1}{\beta}$, we give a Newton-Armijo type algorithm and establish its convergence. In Section 4, we treat the MCP, the mixed complementarity problem (21). For the case of a solvable monotone MCP with finite bounds $l, u \in R^n$, we prove that if the smoothing parameter β is sufficiently small, then the smooth system has a solution. An efficient smooth algorithm based on the Newton-Armijo approach with an adjusted smoothing parameter is also given and convergence is established. In Section 5 we show that exact solutions of our smooth nonlinear equation, for various values of the smoothing parameter β generate an interior path to the feasible region, different from the central path of the interior point method [19]. We compare the two paths on a simple example and show that our path gives a smaller error for the same value of the smoothing parameter β. In Section 6, encouraging numerical testing results are given for

52 problems from the MCPLIB [9] which includes all the problems attempted in [13], [30] and [8]. These problems range in size of up to 8192 variables. These examples include the difficult von Thünen NCP model [30, 39] which is solved here to an accuracy of 1.0e-7.

A few words about our notation. For $f : R \to R$ and $x \in R^n$, the vector $f(x)$ in R^n is defined by the components $(f(x))_i = f(x_i), i = 1, \cdots, n$. The support set of $f(x)$, which is the set of points such that $f(x) \neq 0$, will be denoted by supp { f(x) }. The set of m-by-n real matrices will be denoted by $R^{m\times n}$. The notation $\mathbf{0}$ and $\mathbf{1}$ will represent vectors with all components 0 and 1 respectively, of appropriate dimension. The infinity, l_1 and l_2 norms will be denoted by $\|\cdot\|_\infty$, $\|\cdot\|_1$ and $\|\cdot\|_2$ respectively. The identity matrix of arbitrary dimension will be denoted by I. For a differentiable function $f\colon R^n \to R^m$, ∇f will denote the $m \times n$ Jacobian matrix of partial derivatives. If $F(x)$ has Lipschitz continuous first partial derivatives on R^n with constant $K > 0$, that is

$$\|\nabla F(x) - \nabla F(y)\| \le K\|x - y\|, \quad \forall x, y \in R^n,$$

we write $F(x) \in LC^1_K(R^n)$.

2 A Class of Smoothing Functions

We consider a class of smooth approximations to the fundamental function $(x)_+ = \max\{x, 0\}$. Notice first that $(x)_+ = \int_{-\infty}^{x} \sigma(y)dy$, where $\sigma(x)$ is the step function:

$$\sigma(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \le 0 \end{cases}$$

The step function $\sigma(x)$ can in turn be written as, $\sigma(x) = \int_{-\infty}^{x} \delta(y)dy$, where $\delta(x)$ is the Dirac delta function which satisfies the following properties

$$\delta(x) \ge 0, \quad \int_{-\infty}^{+\infty} \delta(y)dy = 1.$$

Figures 1 to 3 depict the above functions. The fact that the plus function is obtained by twice integrating the Dirac delta function, prompts us to propose probability density functions as a means of smoothing the Dirac delta function and its integrals. Hence we consider the piecewise continuous function $d(x)$ with finite number of pieces which is a density function, that is it satisfies

$$d(x) \ge 0 \ \text{ and } \ \int_{-\infty}^{\infty} d(x)\, dx = 1. \tag{1}$$

To parametrize the density function we define

$$\hat{t}(x, \beta) = \frac{1}{\beta} d(\frac{x}{\beta}) \tag{2}$$

where β is a positive parameter. When β goes to 0, the limit of $\hat{t}(x, \beta)$ is the Dirac delta function $\delta(x)$. This motivates a class of smooth approximations as follows:

$$\hat{s}(x, \beta) = \int_{-\infty}^{x} \hat{t}(t, \beta)dt \approx \sigma(x)$$

and

$$\hat{p}(x, \beta) = \int_{-\infty}^{x} \hat{s}(t, \beta)dt \approx (x)_+ \tag{3}$$

Therefore, we can get an approximate plus function by twice integrating a density function. In fact, this is the same as defining

$$\hat{p}(x,\beta) = \int_{-\infty}^{+\infty} (x-t)_+ \hat{t}(t,\beta)dt = \int_{-\infty}^{x} (x-t)\hat{t}(t,\beta)dt. \tag{4}$$

This formulation was given in [20] and [15, p.12] for a density(kernel) function with finite support. We will give our results in terms of a density function with arbitrary support. This includes the finite support density function as a special case.

Proposition 2.1 *Let $d(x)$ be a probability density function and $\hat{t}(x,\beta) = \frac{1}{\beta}d(\frac{x}{\beta})$, where β is a positive parameter. Let $d(x)$ satisfy the following assumptions:*

(A1) *$d(x)$ is piecewise continuous with finite number of pieces and satisfies (1).*

(A2) $E[|x|]_{d(x)} = \int_{-\infty}^{+\infty} |x| \; d(x) \; dx < +\infty$.

Then the definitions of $\hat{p}(x,\beta)$ given by (3) and (4) are consistent.

Proof By the definition (2) and assumption (A2), we have that $\hat{p}(x,\beta)$ defined by (4) satisfies

$$\hat{p}(x,\beta) = x\int_{-\infty}^{\frac{x}{\beta}} d(t)dt - \beta \int_{-\infty}^{\frac{x}{\beta}} td(t)dt \tag{5}$$

By direct computation,

$$\hat{p}'(x,\beta) = \int_{-\infty}^{\frac{x}{\beta}} d(t)dt \tag{6}$$

$$= \int_{-\infty}^{x} \hat{t}(t,\beta)dt = \hat{s}(x,\beta)$$

Hence the derivatives of $\hat{p}(x,\beta)$ defined by (3) and (4) are the same and the difference between the two representations of $\hat{p}(x,\beta)$ is a constant, say c. If we let x approach $-\infty$ in both (3) and (4), then $\hat{p}(x,\beta)$ approaches 0 in both, and hence $c = 0$. Therefore the definitions of $\hat{p}(x,\beta)$ given by (3) and (4) are consistent. ◻

Now we give properties of $\hat{p}(x,\beta)$ that show that it is an accurate approximation of the plus function $(x)_+$ as β approaches zero.

Proposition 2.2 Properties of $\hat{\mathbf{p}}(\mathbf{x},\beta)$, $\beta > \mathbf{0}$
Let $d(x)$ and $\hat{t}(x,\beta)$ be as in Proposition 2.1, and let $d(x)$ satisfy (A1) and (A2). Then $\hat{p}(x,\beta)$ has the following properties:

(1) $\hat{p}(x,\beta)$ is continuously differentiable. If, in addition, $d(x)$ is k-times continuously differentiable, $\hat{p}(x,\beta)$ is (k+2)-times continuously differentiable.

(2) $-D_2\beta \leq \hat{p}(x,\beta) - (x)_+ \leq D_1\beta$, where

$$D_1 = \int_{-\infty}^{0} |x|d(x)dx \tag{7}$$

and

$$D_2 = \max\{\int_{-\infty}^{+\infty} xd(x)dx, 0\} \tag{8}$$

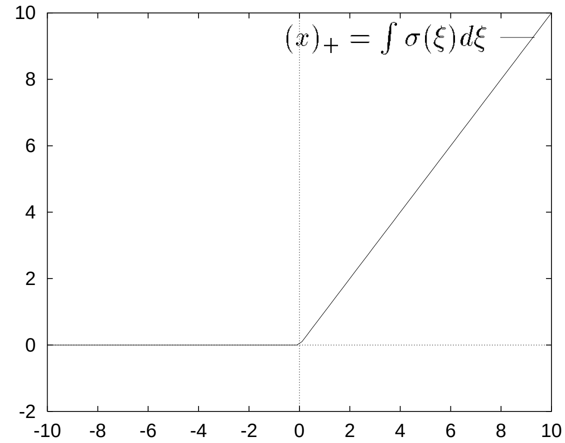

Figure 1: **The plus function** $(\mathbf{x})_+ = \max\{\mathbf{x}, \mathbf{0}\}$

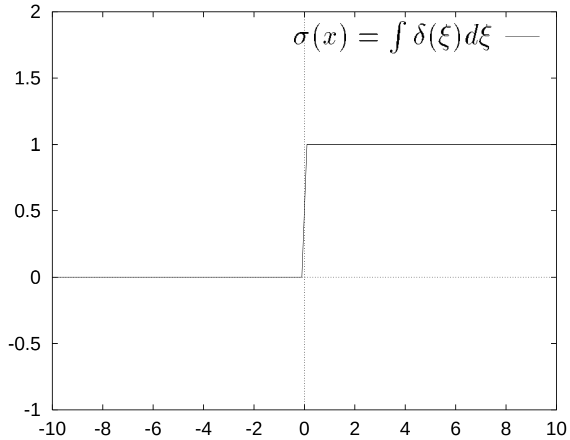

Figure 2: **The step function** $\sigma(\mathbf{x}) = \mathbf{1}$ **if** $\mathbf{x} > \mathbf{0}$**,** $\mathbf{0}$ **if** $\mathbf{x} \leq \mathbf{0}$

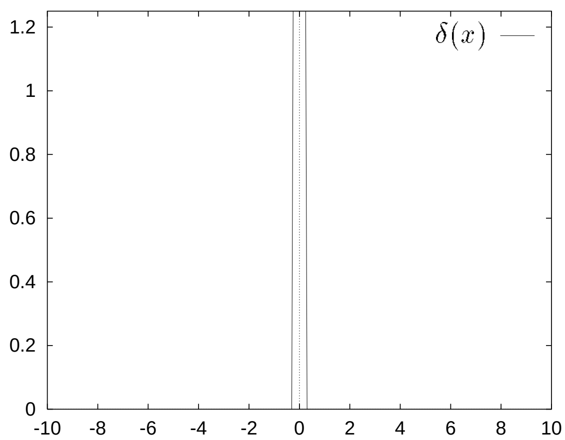

Figure 3: **The Dirac delta function** $\delta(\mathbf{x})$

(3) $\hat{p}(x,\beta)$ *is nondecreasing and convex. In addition, let* $d(x)$ *satisfy*

(A3) *supp* $\{\ d(x)\ \} = R$.

Then $\hat{p}(x,\beta)$ *is strictly increasing and strictly convex in* x *for a fixed* $\beta > 0$.

(4) $0 \le \hat{p}'(x,\beta) \le 1$ *and if (A3) is satisfied, then* $0 < \hat{p}'(x,\beta) < 1$.

(5) Let (A3) be satisfied and $D_2 = 0$, *then*

$$\hat{p}(x,\beta) > x.$$

Proof (1) By equation (6) in the proof of last proposition, the conclusion follows.

(2) If $x \ge 0$, by using (5), we have that

$$\hat{p}(x,\beta) - (x)_+ = x\int_{-\infty}^{\frac{x}{\beta}} d(t)dt - \beta\int_{-\infty}^{\frac{x}{\beta}} td(t)dt - x$$

$$= -x\int_{\frac{x}{\beta}}^{\infty} d(t)dt + \beta\int_{\frac{x}{\beta}}^{\infty} td(t)dt - \beta\int_{-\infty}^{\infty} td(t)dt$$

$$= \beta\int_{\frac{x}{\beta}}^{\infty} (t - \frac{x}{\beta})d(t)dt - \beta\int_{-\infty}^{\infty} td(t)dt$$

Therefore

$$\hat{p}(x,\beta) - (x)_+ \ge -\beta\int_{-\infty}^{\infty} td(t)dt \ge -\beta D_2$$

and

$$\hat{p}(x,\beta) - (x)_+ \le \beta\int_{\frac{x}{\beta}}^{\infty} td(t)dt - \beta\int_{-\infty}^{\infty} td(t)dt$$

$$\le \beta\int_{0}^{\infty} td(t)dt - \beta\int_{-\infty}^{\infty} td(t)dt$$

$$= -\beta\int_{-\infty}^{0} td(t)dt = \beta\int_{-\infty}^{0} |t|d(t)dt = \beta D_1.$$

Otherwise, $x < 0$, then

$$\hat{p}(x,\beta) - (x)_+ = x\int_{-\infty}^{\frac{x}{\beta}} d(t)dt + \beta\int_{-\infty}^{\frac{x}{\beta}} |t|\ d(t)dt$$

$$\le \beta\int_{-\infty}^{0} |t|\ d(t)dt = D_1\beta$$

and

$$\hat{p}(x,\beta) - (x)_+ = \beta\int_{-\infty}^{\frac{x}{\beta}} (\frac{x}{\beta} - t)d(t)dt \ge 0 \ge -\beta D_2.$$

(3) By equation (6) and the fact that $d(x) \geq 0$,

$$\hat{p}'(x,\beta) = \int_{-\infty}^{\frac{x}{\beta}} d(t)dt \geq 0$$

and

$$(\hat{p}'(x,\beta) - \hat{p}'(y,\beta))(x-y) = (x-y)\int_{y}^{x} d(t)dt \geq 0$$

Therefore $\hat{p}(x,\beta)$ is nondecreasing and convex. If, in addition, (A3) is true, then

$$\hat{p}'(x,\beta) > 0 \quad \text{and} \quad (\hat{p}'(x,\beta) - \hat{p}'(y,\beta))(x-y) > 0, \text{ for } x \neq y$$

So $\hat{p}(x,\beta)$ is strictly increasing and strictly convex.

(4) By formula (6), it is easy to see that $0 \leq \hat{p}'(x,\beta) \leq 1$. If (A3) is true, $d(x) > 0$ and $\hat{p}'(x,\beta)$ is strictly increasing. Therefore $0 < \hat{p}'(x,\beta) < 1$.

(5) By (A3), we have

$$\int_{\frac{x}{\beta}}^{\infty} (t - \frac{x}{\beta})d(t)dt > 0, x \geq 0$$

and

$$\int_{-\infty}^{\frac{x}{\beta}} (t - \frac{x}{\beta})d(t)dt > 0, x < 0$$

By the similar proof of (2), we have

$$\hat{p}(x,\beta) - (x)_+ > -D_2\beta = 0$$

Therefore

$$\hat{p}(x,\beta) > (x)_+ \geq x.$$

□

Now we have a systematic way for generating a class of smooth plus functions. Given any probability density function $d(x)$ satisfying (A1) and (A2) we define $\hat{t}(x,\beta)$ and $\hat{p}(x,\beta)$ as in (2) and (3) respectively. The smooth function $\hat{p}(x,\beta)$ approximates the plus function with increasing accuracy as the parameter β approaches 0. The properties of the function $\hat{p}(x,\beta)$ are given in Proposition 2.2 above.

We now give examples of smooth plus functions. The first example, which will be used throughout this paper, is based on the sigmoid function of neural networks [14, 22, 5] and defined as follows:

$$s(x,\alpha) = \frac{1}{1+e^{-\alpha x}}, \quad \alpha > 0$$

This function approximates the step function $\sigma(x)$ as α tends to infinity. Since the derivative with respect to x of this function tends to the Dirac delta function as α tends to infinity, it follows that α plays the role of $\frac{1}{\beta}$ and we shall therefore take

$$\alpha = \frac{1}{\beta}.$$

Example 2.1 Neural Networks Smooth Plus Function [5]
Let

$$d(x) = \frac{e^{-x}}{(1+e^{-x})^2}$$

Here $D_1 = \log 2,\quad D_2 = 0$ *and* $supp\{d(x)\} = R$*, where* D_1 *and* D_2 *are defined by (7) and (8). Integrating* $\frac{1}{\beta}d(\frac{x}{\beta})$ *twice gives*

$$\hat{p}(x,\beta) = x + \beta \log(1 + e^{-\frac{x}{\beta}})$$

Letting $\alpha = \frac{1}{\beta}$*, we have*

$$p(x,\alpha) = \hat{p}(x,\frac{1}{\alpha}) = \int s(\xi,\alpha)d\xi = x + \frac{1}{\alpha}\log(1+e^{-\alpha x})$$

$$s(x,\alpha) = \hat{s}(x,\frac{1}{\alpha}) = \frac{1}{1+e^{-\alpha x}} = \int t(\xi,\alpha)d\xi$$

$$t(x,\alpha) = \hat{t}(x,\frac{1}{\alpha}) = \frac{\alpha e^{-\alpha x}}{(1+e^{-\alpha x})^2} = \alpha s(x,\alpha)(1-s(x,\alpha))$$

The function $p(x,\alpha)$ above can also be obtained by using an appropriate recession function given in Example 2.2 of [1], or by an exponential penalty function as in [3, p. 313, equation (25)]. Figures 4 to 6 depict the functions $p(x,5)$, $s(x,5)$ and $t(x,5)$ respectively.

Following are several other smooth plus functions based on probability density functions proposed by other authors.

Example 2.2 Chen-Harker-Kanzow-Smale Smooth Plus Function [38], [17] and [4]
Let

$$d(x) = \frac{2}{(x^2+4)^{\frac{3}{2}}}$$

Here $D_1 = 1,\quad D_2 = 0,\quad supp\{d(x)\} = R$ *and*

$$\hat{p}(x,\beta) = \frac{x + \sqrt{x^2+4\beta^2}}{2}$$

Example 2.3 Pinar-Zenios Smooth Plus Function [31]
Let

$$d(x) = \begin{cases} 1 & \textit{if } 0 \le x \le 1 \\ 0 & \textit{otherwise} \end{cases}$$

Here $D_1 = 0,\quad D_2 = \frac{1}{2},\quad supp\{d(x)\} = [0,1]$ *and*

$$\hat{p}(x,\beta) = \begin{cases} 0 & \textit{if } x < 0 \\ \frac{x^2}{2\beta} & \textit{if } 0 \le x \le \beta \\ x - \frac{\beta}{2} & \textit{if } x > \beta \end{cases}$$

This function can also be obtained by applying the Moreau-Yosida regularization [15, p.13] to the plus function.

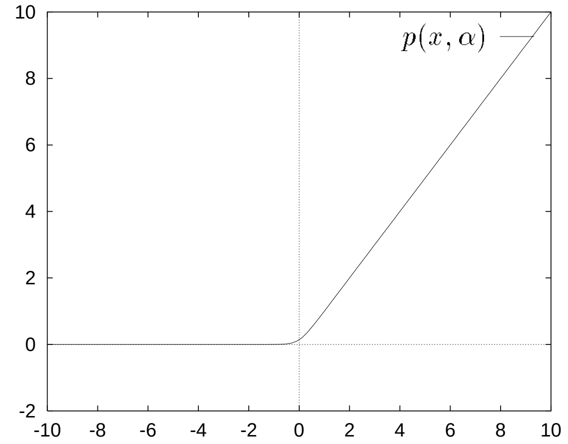

Figure 4: **The p function $\mathbf{p}(\mathbf{x},\alpha) = \mathbf{x} + \frac{1}{\alpha}\log(\mathbf{1}+\mathbf{e}^{-\alpha\mathbf{x}})$ with $\alpha = \mathbf{5}$**

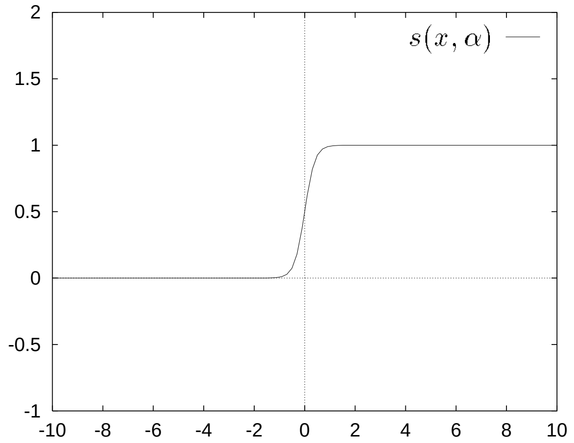

Figure 5: **The sigmoid function $\mathbf{s}(\mathbf{x},\alpha) = \frac{\mathbf{1}}{\mathbf{1}+\mathbf{e}^{-\alpha\mathbf{x}}}$ with $\alpha = \mathbf{5}$**

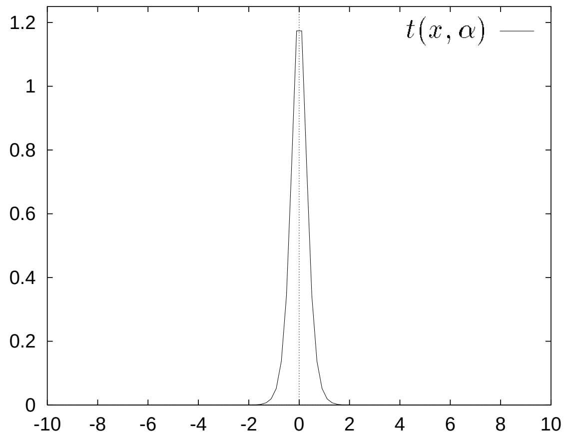

Figure 6: **The t function $\mathbf{t}(\mathbf{x},\alpha) = \frac{\alpha\mathbf{e}^{-\alpha\mathbf{x}}}{(\mathbf{1}+\mathbf{e}^{-\alpha\mathbf{x}})^{\mathbf{2}}}$ with $\alpha = \mathbf{5}$**

Example 2.4 Zang Smooth Plus Function [41]
Let

$$d(x) = \begin{cases} 1 & if -\frac{1}{2} \leq x \leq \frac{1}{2} \\ 0 & otherwise \end{cases}$$

Here $D_1 = \frac{1}{8}$, $D_2 = 0$, $supp\{d(x)\} = [-\frac{1}{2}, \frac{1}{2}]$ *and*

$$\hat{p}(x,\beta) = \begin{cases} 0 & if\ x < -\frac{\beta}{2} \\ \frac{1}{2\beta}(x+\frac{\beta}{2})^2 & if\ |x| \leq \frac{\beta}{2} \\ x & if\ x > \frac{\beta}{2} \end{cases}$$

Note that in Examples 2.3 and 2.4 above, the density function $d(x)$ has compact support while the smooth function $\hat{p}(x,\beta)$ is only once continuously differentiable. In Examples 2.1 and 2.2, d(x) has infinite support while the functions $p(x,\alpha)$ and $\hat{p}(x,\beta)$ are differentiable infinitely often.

We summarize the various functions introduced as follows:

$$\begin{array}{ccccc} \hat{p}(x,\beta) & \overset{\int}{\longleftarrow} & \hat{s}(x,\beta) & \overset{\int}{\longleftarrow} & \hat{t}(x,\beta) = \frac{1}{\beta}d(\frac{x}{\beta}) \\ \downarrow \beta \longrightarrow 0 & & & & \\ (x)_+ & \overset{\int}{\longleftarrow} & \sigma(x) & \overset{\int}{\longleftarrow} & \delta(x) \\ \uparrow \alpha \longrightarrow \infty & & & & \\ p(x,\alpha) = \hat{p}(x,\frac{1}{\alpha}) & \overset{\int}{\longleftarrow} & s(x,\alpha) = \hat{s}(x,\frac{1}{\alpha}) & \overset{\int}{\longleftarrow} & t(x,\alpha) = \hat{t}(x,\frac{1}{\alpha}) \end{array}$$

Because of our favorable experience with the function $p(x,\alpha)$ [5] on linear complementarity problems and linear and convex inequalities, we chose it for our numerical experiments. Further comparisons using different approximations to the plus function are left for future work.

3 The Nonlinear Complementarity Problem

In this section we consider the nonlinear complementarity problem (NCP) of finding an x in R^n such that

$$0 \leq x \perp F(x) \geq 0 \qquad (9)$$

Here $F(x)$ is a differentiable function from R^n to R^n. By using the smooth function $\hat{p}(x,\beta)$ introduced in last section, we consider the smooth nonlinear equation

$$R(x) = x - \hat{p}(x - F(x), \beta) = 0 \qquad (10)$$

as an approximation to the following nonsmooth equivalent reformulation of the NCP

$$x = (x - F(x))_+ \qquad (11)$$

We first show that a natural residual for the NCP is easily bounded by a corresponding residual for the nonlinear equation (10).

Lemma 3.1

$$\|x-(x-F(x))_+\|_p \le \|x-\hat{p}(x-F(x),\beta)\|_p + \gamma_p \max\{D_1, D_2\}\beta, \quad p=1,2,\infty, \tag{12}$$

where $\gamma_1 = n$, $\gamma_2 = \sqrt{n}$ and $\gamma_\infty = 1$. The constants D_1 and D_2 depend on the density function used and are defined in (7) and (8).

Proof

$$\begin{aligned}\|x-(x-F(x))_+\|_p &\le \|x-\hat{p}(x-F(x),\beta)+\hat{p}(x-F(x),\beta)-(x-F(x))_+\|_p\\ &\le \|x-\hat{p}(x-F(x),\beta)\|_p + \|\hat{p}(x-F(x),\beta)-(x-F(x))_+\|_p\\ &\le \|x-\hat{p}(x-F(x),\beta)\|_p + \gamma_p \max\{D_1, D_2\}\beta. \qquad \Box\end{aligned}$$

The above result is also true for any monotone norm [27].

We first consider the strongly monotone NCP, that is there exists a $k>0$ such that for any $x, y \in R^n$

$$(F(x)-F(y))^T(x-y) \ge k\|x-y\|^2 \tag{13}$$

Since the NCP is strongly monotone, it has a unique solution [12]. The following error bound for the strongly monotone NCP is given as Theorem 3.2.1 in [33].

Lemma 3.2 *Let the NCP be strongly monotone and let $F(x)$ be Lipschitz continuous. Then for any $x \in R^n$*

$$\|x-\bar{x}\|_p \le C_p\|x-(x-F(x))_+\|_p, \quad p=1,2,\infty, \tag{14}$$

where $\bar{x}$ is the unique solution of the NCP and C_p is a condition constant of F independent of x.

Now, we give an error bound for the NCP by using the smooth function $\hat{p}(x,\beta)$. By Lemma 3.1 and Lemma 3.2, it is easy to get the following lemma.

Lemma 3.3 *Let the NCP be strongly monotone and let $F(x)$ be Lipschitz continuous. Then for any $x \in R^n$*

$$\|x-\bar{x}\|_p \le C_p(\|x-\hat{p}(x-F(x),\beta)\|_p + \gamma_p \max\{D_1, D_2\}\beta), \quad p=1,2,\infty, \tag{15}$$

where $\bar{x}$ and C_p are defined in Lemma 3.2, and γ_p and the constants D_1, D_2 are defined in Lemma 3.1.

Let the residual $f(x)$ of the nonlinear equation (10) be defined as follows

$$f(x) = \frac{1}{2}R(x)^T R(x) = \frac{1}{2}\|x-\hat{p}(x-F(x),\beta)\|_2^2 \tag{16}$$

We now prove that if x is a stationary point of $f(x)$ for a monotone $F(x)$, then x must be a solution of the nonlinear equation (10), and hence by (15), x is an approximate solution of the NCP.

Proposition 3.1 *Suppose that $d(x)$ satisfies (A1) - (A3) and $\hat{p}(x,\beta)$ is defined by (4). For any monotone NCP, we have that $\nabla R(x)$ is positive definite. In addition, let x be a stationary point of $f(x)$, then x must be a solution of the nonlinear equation (10).*

Proof By definition,

$$\nabla R(x) = diag(\hat{p}'(x - F(x), \beta))(diag(\hat{p}'^{-1}(x - F(x), \beta) - \mathbf{1}) + \nabla \mathbf{F}(\mathbf{x}))$$

By (4) of Proposition 2.2, we have $0 < \hat{p}'(x, \beta) < 1$ and hence the diagonal matrices above are positive definite. Since $\nabla F(x)$ is positive semidefinite, it follows that $\nabla R(x)$ is positive definite. Let x be a stationary point of $f(x)$, then

$$\nabla f(x) = \nabla R(x)^T R(x) = 0.$$

Since $\nabla R(x)$ is nonsingular, $R(x) = 0$, it follows that x satisfies (10). □

When $F(x)$ is strongly monotone and Lipschitz continuous, then the level sets of $f(x)$ are compact. We state this result as the following proposition.

Proposition 3.2 *Consider the strongly monotone NCP with Lipschitz continuous $F(x)$. Then $f(x)$ defined by (16) has compact level sets.*

Proof Suppose not, then there exists a sequence $\{x_k\} \subset R^n$ and a positive number M such that $\|x_k\|_2 \to \infty$ as $k \to \infty$, and $\|x_k - \hat{p}(x_k - F(x_k), \beta)\|_2 \leq M$. Then by Lemma 3.3 ,

$$\|x_k - \bar{x}\|_2 \leq C_2(M + \gamma_2 \max\{D_1, D_2\}\beta),$$

where $\overline{x}$ is the unique solution of the NCP. Let $k \longrightarrow \infty$, the left hand side of the above inequality goes to ∞ and the right hand side stays finite. This is a contradiction. Hence the level sets of $f(x)$ are compact. □

We now show that, for the strongly monotone NCP with Lipschitz continuous function $F(x)$, the nonlinear equation (10) always has a unique solution.

Theorem 3.1 *Suppose that $d(x)$ satisfies (A1) - (A3), $\hat{p}(x, \beta)$ is defined by (4), $F(x)$ is strongly monotone and Lipschitz continuous. Then the nonlinear equation (10) has a unique solution.*

Proof By Proposition 3.2, the level sets of $f(x)$ are compact. So $\min_{x \in R^n} f(x)$ must have a solution x, which is a stationary point of $f(x)$. By Proposition 3.1 we get that x satisfies (10). If y is another solution of (10), then

$$0 = (x - y)(R(x) - R(y)) = (x - y) \int_{t=0}^{t=1} \nabla R(x + t(y - x)) dt (x - y)$$

for some $t \in [0, 1]$. Since ∇R is positive definite by Proposition 3.1, it follows that $x = y$. Therefore equation (10) has a unique solution. □

Let $x(\beta)$ be a solution of (10). Then $x(\beta) = \hat{p}(x(\beta) - F(x(\beta)), \beta)$. By Lemma 3.3, we have the following theorem which bounds the distance between the solution $x(\beta)$ of (10) and the solution point of the original NCP (9).

Theorem 3.2 *Consider a strongly monotone NCP with Lipschitz continuous $F(x)$. Let $x(\beta)$ be a solution of (10). Then, for the solution $\bar{x}$ of the NCP (9), we have that*

$$\|x(\beta) - \bar{x}\|_p \leq C_p \gamma_p \max\{D_1, D_2\}\beta, \ p = 1, 2, \infty.$$

Here C_p is the condition constant defined in Lemma 3.2, γ_p and D_1, D_2 are constants defined in Lemma 3.1.

By the above result, we know that if β is sufficiently small, a solution of (10) can approximate the solution of NCP to any desired accuracy. Hence we can solve (10) to get an approximate solution of the NCP.

For most part of this remaining section, we consider only the function $p(x, \alpha)$ defined in Example 2.1. We explore further the property of real numbers x and y that approximately satisfy the following equation

$$x = p(x - y, \alpha)$$

which is related to equation (10) that generates an approximate solution of the NCP. We claim that such x and y will approximately satisfy

$$0 \leq x \perp y \geq 0.$$

In order to prove this fundamental fact we establish the following lemma, the proof of which is relegated to the Appendix 1.

Lemma 3.4 *Let $h(x)$ be defined as follows:*

$$h(x) = -x \log(1 + \delta - e^{-x}), \quad \delta \geq 0$$

(i) If $0 < \delta < 1$, then

$$\max_{x \in [0, -\log \delta]} h(x) \leq 2.$$

(ii) If $\delta \geq 1$, then

$$\max_{x \in [-\log \delta, 0]} h(x) \leq \max\{\delta \log^2 \delta, \frac{1}{e}\}.$$

(iii) If $\delta = 0$, then

$$\max_{x \in [0, \infty)} h(x) \leq \frac{2}{e}.$$

We will now show that if an x and y satisfy

$$-\frac{\delta_1}{\alpha} \leq x - p(x - y, \alpha) \leq 0,$$

where $\delta_1 \geq 0$, then the complementarity condition $0 \leq x \perp y \geq 0$ is approximately satisfied for large α, in the following sense

$$(-x)_+ \leq \frac{\delta_1}{\alpha}, \qquad (-y)_+ \leq \frac{\delta_1}{\alpha}, \qquad (xy)_+ \leq \frac{C(\delta_1)}{\alpha^2},$$

where $C(\delta_1)$ is the constant defined in Proposition 3.3. Note that as $\alpha \longrightarrow \infty$, the complementarity condition $0 \leq x \perp y \geq 0$ is exactly satisfied.

Proposition 3.3 *Let $x, y \in R$ satisfy*

$$-\frac{\delta_1}{\alpha} \leq x - p(x - y, \alpha) \leq 0,$$

where $\delta_1 \geq 0$. Then

$$(-x)_+ \leq \frac{\delta_1}{\alpha}, \qquad (-y)_+ \leq \frac{\delta_1}{\alpha}, \qquad (xy)_+ \leq \frac{C(\delta_1)}{\alpha^2},$$

where

$$C(\delta_1) = \max\{2, (e^{\delta_1} - 1) \log^2(e^{\delta_1} - 1)\}.$$

Proof Let $\delta = e^{-\alpha x} + e^{-\alpha y} - 1$, since

$$-\frac{\delta_1}{\alpha} \le x - (x-y) - \frac{1}{\alpha}\log(1 + e^{-\alpha(x-y)}) \le 0,$$

we have

$$1 \le e^{-\alpha x} + e^{-\alpha y} \le e^{\delta_1}$$

Hence $0 \le \delta \le e^{\delta_1} - 1$. Since

$$e^{-\alpha x} \le e^{-\alpha x} + e^{-\alpha y} = 1 + \delta \le e^{\delta_1},$$

we have $x \ge -\frac{\delta_1}{\alpha}$. Hence

$$(-x)_+ \le \frac{\delta_1}{\alpha}.$$

Similarly,

$$(-y)_+ \le \frac{\delta_1}{\alpha}.$$

Now we consider the estimate of $(xy)_+$. Since $e^{-\alpha x} + e^{-\alpha y} = 1 + \delta$, we have

$$y = -\frac{1}{\alpha}\log(1 + \delta - e^{-\alpha x}),$$

Therefore

$$y \ge 0 \Longleftrightarrow 1 + \delta - e^{-\alpha x} \le 1 \Longleftrightarrow e^{-\alpha x} \ge \delta \tag{17}$$

and

$$y \le 0 \Longleftrightarrow 1 + \delta - e^{-\alpha x} \ge 1 \Longleftrightarrow e^{-\alpha x} \le \delta \tag{18}$$

Case 1 $0 < \delta < 1$.

If $y \ge 0$, then by (17),

$$x \le -\frac{1}{\alpha}\log\delta.$$

Hence

$$\begin{aligned}(xy)_+ &\le \max_{x\in[0,-\frac{\log\delta}{\alpha}]} -\frac{x}{\alpha}\log(1+\delta-e^{-\alpha x}) \\ &= \frac{1}{\alpha^2}\max_{y\in[0,-\log\delta]} h(y) \qquad (\text{Let } y = \alpha x) \\ &\le \frac{2}{\alpha^2} \qquad\qquad (\text{By (i) of Lemma 3.4})\end{aligned}$$

Otherwise $y \le 0$, then

$$x \ge -\frac{1}{\alpha}\log\delta \ge 0.$$

Hence $xy \le 0$, $(xy)_+ = 0$.

Case 2 $\delta \ge 1$.

If $y \ge 0$, then by (17),

$$x \le -\frac{1}{\alpha}\log\delta \le 0.$$

Hence $xy \leq 0$, $(xy)_+ = 0$. Otherwise $y \leq 0$, then $x \geq -\frac{1}{\alpha}\log\delta$. If $x \geq 0$, then $(xy)_+ = 0$. Now we only consider the case $x < 0$. Therefore

$$(xy)_+ \leq \max_{x\in[-\frac{\log\delta}{\alpha},0]} -\frac{x}{\alpha}\log(1+\delta-e^{-\alpha x}) = \frac{1}{\alpha^2}\max_{y\in[-\log\delta,0]} h(y) \qquad (\text{Let } y = \alpha x).$$

By (ii) of Lemma 3.4,

$$(xy)_+ \leq \frac{1}{\alpha^2}\max\{\frac{1}{e}, \delta\log^2\delta\} \leq \frac{1}{\alpha^2}\max\{\frac{1}{e}, (e^{\delta_1}-1)\log^2(e^{\delta_1}-1)\}.$$

Case 3 $\delta = 0$.

In this case, we have $x \geq 0$ and $y \geq 0$. By using (iv) of Lemma 3.4,

$$(xy)_+ \leq \frac{1}{\alpha^2}\max_{x\in[0\infty)} h(x) \leq \frac{1}{\alpha^2}\frac{2}{e}$$

Combining the above three cases, we get

$$(xy)_+ \leq \frac{C(\delta_1)}{\alpha^2}.$$

□

Even in the case of a solvable monotone nonlinear complementarity problem (e.g. $0 \leq x \perp F(x) \geq 0$, $F(x) := 0$), the nonlinear equation (10) may not necessarily have a solution. However, for all $\delta_1 \geq D_1$, and $\delta_2 \geq D_2$, the following system of inequalities

$$-\delta_1\beta\mathbf{1} \leq \mathbf{x} - \hat{\mathbf{p}}(\mathbf{x}-\mathbf{F}(\mathbf{x}),\beta) \leq \delta_2\beta\mathbf{1}, \tag{19}$$

always has a solution for $\beta > 0$. In particular, for the $p(x,\alpha)$ defined in Example 2.1 we have that for all $\delta_1 \geq \log 2$, the following system of inequalities

$$-\frac{\delta_1}{\alpha}\mathbf{1} \leq \mathbf{x} - \mathbf{p}(\mathbf{x}-\mathbf{F}(\mathbf{x}),\alpha) \leq \mathbf{0}, \tag{20}$$

always has a solution. Hence by Proposition 3.3, a solution of (20) will approximately satisfy the NCP condition

$$x \geq 0, \qquad F(x) \geq 0, \qquad x^T F(x) = 0.$$

Consequently, Proposition 3.3 can be used to establish the following useful theorem.

Theorem 3.3 *Consider a solvable nonlinear complementarity problem (9). Let $\delta_1 \geq \log 2$ and $\alpha > 0$. There exists x satisfy (20), and consequently the NCP conditions are approximately satisfied as follows:*

$$(-x)_+ \leq \frac{\delta_1}{\alpha}\mathbf{1}, \qquad (-\mathbf{F}(\mathbf{x}))_+ \leq \frac{\delta_1}{\alpha}\mathbf{1}, \qquad (\mathbf{x}^\mathbf{T}\mathbf{F}(\mathbf{x}))_+ \leq \frac{\mathbf{n}\mathbf{C}(\delta_1)}{\alpha^2},$$

where $C(\delta_1)$ is defined in Proposition 3.3.

We now specify our computational algorithm for solving the NCP by smoothing. The algorithm consists of a Newton method with an Armijo line search with parameters δ and σ such that $0 < \delta < 1$ and $0 < \sigma < \frac{1}{2}$.

Algorithm 3.1 Newton NCP Algorithm
Given $x_0 \in R^n$ and let $k = 0$.

(1) If $\|\nabla f(x_k)\| < \epsilon$, stop.

(2) **Direction** d_k

$$d_k = -\nabla R(x_k)^{-1} R(x_k)$$

(3) **Stepsize** λ_k *(Armijo)*

$$x_{k+1} = x_k + \lambda_k d_k, \lambda_k = \max\{1, \delta, \delta^2, \cdots\}, \ s.t.$$

$$f(x_k) - f(x_{k+1}) \geq \sigma \lambda_k |d_k^T \nabla f(x_k)|$$

$k = k + 1$ *go to step (1).*

The above algorithm is well defined for a monotone NCP with a continuously differentiable $F(x)$. We will state the following convergence theorem [7]. We omit the proof that is similar to the proof of Theorem 4.3.

Theorem 3.4 *Consider a solvable monotone nonlinear complementarity problem (9) with $F(x) \in LC_K^1(R^n)$. Then*

(1) The sequence $\{x_k\}$ defined in Algorithm 3.1 is will defined.

(2) Any accumulation point of the above sequence solves the nonlinear equation (10).

(3) If an accumulation point exists, the whole sequence $\{x_k\}$ converges to $\bar{x}$ quadratically.

(4) If, in addition, F is strongly monotone and Lipschitz continuous, then the sequence $\{x_k\}$ converges to $\bar{x}$, the solution of (10), at a quadratic rate.

4 The Mixed Complementarity Problem

The mixed complementarity problem (MCP) is defined as follows [8]:

Given a differentiable $F : R^n \longrightarrow R^n, \quad l, u \in \bar{R}^n, l < u$, where $\bar{R} = R \cup \{+\infty, -\infty\}$, find $x, w, v \in R^n$, such that

$$\begin{array}{l} F(x) - w + v = 0 \\ 0 \leq x - l \perp w \geq 0 \\ 0 \leq v \perp u - x \geq 0 \end{array} \tag{21}$$

This MCP model includes many classes of mathematical programming problems, such as nonlinear equations, nonlinear programming, nonlinear complementarity problems and variational inequalities.

By using the smooth function $\hat{p}(x, \beta)$ instead of the plus function, we reformulate the MCP approximately as follows.
For $i = 1, \cdots, n$:

Case 1. $l_i = -\infty$ and $u_i = \infty$:

$$F_i(x) = 0$$

Case 2. $l_i > -\infty$ and $u_i = \infty$:

$$x_i - l_i - \hat{p}(x_i - l_i - F_i(x), \beta) = 0$$

Case 3. $l_i = -\infty$ and $u_i < \infty$:

$$x_i - u_i + \hat{p}(u_i - x_i + F_i(x), \beta) = 0 \tag{22}$$

Case 4. $l_i > -\infty$ and $u_i < \infty$:

$$F_i(x) - w_i + v_i = 0$$
$$x_i - l_i - \hat{p}(x_i - l_i - w_i, \beta) = 0$$
$$u_i - x_i - \hat{p}(u_i - x_i - v_i, \beta) = 0.$$

We will denote the above 4 cases collectively by the nonlinear equation

$$R(x, w, v) = 0 \tag{23}$$

Note that the natural residual for the MCP is given by the left hand side of above relation with the $\hat{p}$ function replaced by the plus function. We denote collectively this natural residual by

$$r(x, w, v) \tag{24}$$

Now we give a lemma that bound the natural residual for MCP by residual of the equation (23) is vice versa. The proof is a simple application of the properties of the p function.

Lemma 4.1 *Let N be number of equations in (23) and $p(x, \alpha)$ defined in Example 2.1. Then*

$$\|R(x, w, v)\|_2 \leq \|r(x, w, v)\|_2 + \frac{\sqrt{N}\log 2}{\alpha_0}$$

and

$$\|r(x, w, v)\|_2 \leq \|R(x, w, v)\|_2 + \frac{\sqrt{N}\log 2}{\alpha_0}$$

Let $f(x, w, v)$ be the residual function of the nonlinear equation (23) defined as follows

$$f(x, w, v) = \frac{1}{2}R(x, w, v)^T R(x, w, v) \tag{25}$$

Now we state an existence result for the monotone MCP with $l, u \in R^n$.

Theorem 4.1 *Suppose that $d(x)$ satisfies (A1) - (A3) and $\hat{p}(x, \beta)$ is defined by (4). Consider a solvable mixed complementarity problem (21) with monotone $F(x)$ and $l, u \in R^n$. The nonlinear equation (23) has a solution for sufficiently small β.*

Proof We shall prove that a level set of $f(x, w, v)$ is nonempty and compact. First we will prove that the set $X = \{x | f(x, w, v) \leq C\}$ is compact for all $C \in R$. Since f is continuous, the level set X is closed. Hence we only need the show the set X is bounded. Suppose not, there exists $\{x^k\} \in X$ and there exists $1 \leq i \leq n$ such that x_i^k goes to $+\infty$ or $-\infty$. Without loss of generality, we assume that x_i^k goes to $+\infty$. Then the residual corresponding to the following equation approaches ∞:

$$u_i - x_i - \hat{p}(u_i - x_i - v_i, \beta) = 0.$$

This contradicts the fact that $x^k \in X$. Let $C = \frac{1}{4}n\max\{D_1, D_2\}^2\beta^2$, where D_1 and D_2 are nonnegative constants defined in (7) and (8), it is easy to show that the level set $Lev_C(f) = \{(x, w, v)|f(x, w, v) \leq C\}$ is not empty. Now we will prove that $Lev_C(f)$ is compact for

$$\beta < \frac{\min_{1\leq i\leq n}(u_i - l_i)}{\sqrt{n}\max\{D_1, D_2\}}.$$

We have proven that the x part must be bounded. Therefore, if the level set $Lev_C(f)$ is unbounded, there exists $(x^k, w^k, v^k) \in Lev_C(f)$ such that (w^k, v^k) are unbounded. Without loss of generality, we assume $x^k \longrightarrow \bar{x}$ and there exist $1 \leq i \leq n$ such $w_i^k \longrightarrow +\infty$ or $-\infty$ as $k \longrightarrow \infty$. If $w_i^k \longrightarrow -\infty$, the residual corresponding to the equation

$$x_i - l_i - \hat{p}(x_i - l_i - w_i, \beta) = 0.$$

goes to ∞ as $k \longrightarrow \infty$. But $(x^k, w^k, v^k) \in Lev_C(f)$, which is a contradiction. Otherwise, $w_i^k \longrightarrow +\infty$. By the equation

$$F_i(x) - w_i + v_i = 0$$

and the fact that x^k is bounded, we get that as $k \longrightarrow \infty$, $v_i^k \longrightarrow +\infty$. Hence, as $k \longrightarrow \infty$,

$$\begin{aligned} f(x^k, w^k, v^k) &\geq \frac{1}{2}((u_i - x_i^k - \hat{p}(u_i - x_i^k - v_i^k, \beta))^2 + (x_i^k - l_i - \hat{p}(x_i^k - l_i - w_i^k, \beta))^2) \\ &\longrightarrow \frac{1}{2}((u_i - \bar{x}_i)^2 + (\bar{x}_i - l_i)^2) \geq \frac{1}{4}(u_i - l_i)^2 > \frac{1}{4}n\max\{D_1, D_2\}^2\beta^2 = C \end{aligned}$$

for all

$$\beta < \frac{\min_{1\leq i\leq n}(u_i - l_i)}{\sqrt{n}\max\{D_1, D_2\}}.$$

This contradicts that $(x^k, w^k, v^k) \in Lev_C(f)$. Hence there exists a level set of $f(x, w, v)$ which is nonempty and compact. Therefore the problem

$$\min_{x,w,v} f(x, w, v)$$

must have a minimum, which satisfies

$$\nabla f(x, w, v) = \nabla R(x, w, v)^T R(x, w, v) = 0.$$

Let $\Lambda_1 = diag(\hat{p}'(x - l - w, \beta))$ and $\Lambda_2 = diag(\hat{p}'(u - x - v, \beta))$, then

$$\nabla R(x, w, v) = \begin{bmatrix} \nabla F(x) & -I & I \\ I - \Lambda_1 & \Lambda_1 & 0 \\ \Lambda_2 - I & 0 & \Lambda_2 \end{bmatrix}$$

Notice that

$$\begin{bmatrix} I & \Lambda_1^{-1} & -\Lambda_2^{-1} \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \nabla R(x, w, v) = \begin{bmatrix} \nabla F(x) + \Lambda_1^{-1} + \Lambda_2^{-1} - 2I & 0 & 0 \\ I - \Lambda_1 & \Lambda_1 & 0 \\ \Lambda_2 - I & 0 & \Lambda_2 \end{bmatrix}$$

and $\mathbf{0} < \hat{\mathbf{p}}'(\mathbf{x} - \mathbf{l} - \mathbf{w}, \beta)), \hat{\mathbf{p}}'(\mathbf{u} - \mathbf{x} - \mathbf{v}, \beta)) < \mathbf{1}$. When F is monotone, the Jacobian $\nabla R(x, w, v)$ is nonsingular. Therefore

$$\nabla f(x, w, v) = 0 \Longrightarrow R(x, w, v) = 0$$

□

From now on, we only consider the function $p(x, \alpha)$ defined in Example 2.1. The following theorem is a direct application of Proposition 3.3. that can be proved by applying the Proportion 3.3 to each equation.

Theorem 4.2 *Consider a solvable mixed complementarity problem (21). Let* $\delta_1 \geq \log 2$ *and* $\alpha > 0$*. Then the following system of inequalities For* $i = 1, \cdots, n$*:*

Case 1. $l_i = -\infty$ *and* $u_i = \infty$*:*

$$-\frac{\delta_1}{\alpha} \leq F_i(x) = 0 \leq \frac{\delta_1}{\alpha}$$

Case 2. $l_i > -\infty$ *and* $u_i = \infty$*:*

$$-\frac{\delta_1}{\alpha} \leq x_i - l_i - \hat{p}(x_i - l_i - F_i(x), \beta) \leq 0$$

Case 3. $l_i = -\infty$ *and* $u_i < \infty$*:*

$$0 \leq x_i - u_i + \hat{p}(u_i - x_i + F_i(x), \beta) \leq \frac{\delta_1}{\alpha} \tag{26}$$

Case 4. $l_i > -\infty$ *and* $u_i < \infty$*:*

$$-\frac{\delta_1}{\alpha} \leq F_i(x) - w_i + v_i \leq \frac{\delta_1}{\alpha}$$

$$-\frac{\delta_1}{\alpha} \leq x_i - l_i - \hat{p}(x_i - l_i - w_i, \beta) \leq 0$$

$$-\frac{\delta_1}{\alpha} \leq u_i - x_i - \hat{p}(u_i - x_i - v_i, \beta) \leq 0.$$

where $p(x, \alpha)$ *is defined in Example 2.1, always has a solution* (x, w, v) *and a solution* (x, w, v) *satisfies the MCP conditions approximately in the following sense*
For $i = 1, \cdots, n$*:*

Case 1. $l_i = -\infty$ *and* $u_i = \infty$*:*

$$|F_i(x)| \leq \frac{\delta_1}{\alpha}$$

Case 2. $l_i > -\infty$ *and* $u_i = \infty$*:*

$$(l_i - x_i)_+ \leq \frac{\delta_1}{\alpha}, \qquad (-F_i(x))_+ \leq \frac{\delta_1}{\alpha}, \qquad ((x_i - l_i)F_i(x))_+ \leq \frac{C(\delta_1)}{\alpha^2},$$

Case 3. $l_i = -\infty$ *and* $u_i < \infty$*:*

$$(x_i - u_i)_+ \leq \frac{\delta_1}{\alpha}, \qquad (F_i(x))_+ \leq \frac{\delta_1}{\alpha}, \qquad (-(u_i - x_i)F_i(x))_+ \leq \frac{C(\delta_1)}{\alpha^2},$$

Case 4. $l_i > -\infty$ *and* $u_i < \infty$*:*

$$|F_i(x) - w_i + v_i| \leq \frac{\delta_1}{\alpha},$$

$$(l_i - x_i)_+ \leq \frac{\delta_1}{\alpha}, \qquad (-w_i)_+ \leq \frac{\delta_1}{\alpha}, \qquad ((x_i - l_i)w_i)_+ \leq \frac{C(\delta_1)}{\alpha^2},$$

$$(x_i - u_i)_+ \leq \frac{\delta_1}{\alpha}, \qquad (-v_i)_+ \leq \frac{\delta_1}{\alpha}, \qquad ((u_i - x_i)v_i)_+ \leq \frac{C(\delta_1)}{\alpha^2},$$

where $C(\delta_1)$ *is defined in Proposition 3.3.*

Now we state the smooth method for the mixed complementarity problem based on the Newton Algorithm 3.1 in which the smoothing parameter will be adjusted. In the algorithm, we adjust the smoothing parameter α in reverse proportion to the natural residual $r(x, w, v)$ of (24) for the MCP in the following way. Let N be the total number of nonlinear equations in (23) and (x, w, v) be current point. Let

$$\alpha(x, w, v) = \begin{cases} \frac{\sqrt{N}}{\|r(x,w,v)\|_2} & \text{if } \|r(x, w, v)\|_2 < \sqrt{N} \\ \sqrt{\frac{\sqrt{N}}{\|r(x,w,v)\|_2}} & \text{otherwise} \end{cases} \tag{27}$$

The following smooth algorithm generates an $\epsilon-$accurate solution for the MCP, in the sense that the natural residual $r(x, w, v)$ of (24) satisfies $\|r(x, w, v)\|_\infty \leq \epsilon$.

In order to get an $\epsilon-$accurate solution for the MCP. We need α sufficient large. We will establish a simple lemma before we get the α.

Lemma 4.2 *Let real numbers a and b satisfy*

$$(-a)_+ \leq \frac{\delta_1}{\alpha}, \quad (-b)_+ \leq \frac{\delta_1}{\alpha} \text{ and } (ab)_+ \leq \frac{C(\delta_1)}{\alpha^2},$$

then

$$|\min(a, b)| \leq \frac{\max\{\delta_1, \sqrt{C(\delta_1)}\}}{\alpha},$$

where $C(\delta_1)$ is defined in Proposition 3.3.

Proof Without loss of generality, we assume that $a \leq b$. If $a \geq 0$, then $(ab)_+ \geq a^2$. Therefore

$$|\min(a, b)| = a \leq \sqrt{(ab)_+} \leq \frac{\sqrt{C(\delta_1)}}{\alpha}.$$

If $a < 0$,

$$|\min(a, b)| = -a = (-a)_+ \leq \frac{\delta_1}{\alpha}.$$

Combining the above two cases, the conclusion follows. $\square$

Therefore to satisfy $|\min(a, b)| \leq \epsilon$, we choose $\alpha \geq \dfrac{\max\{\delta_1, \sqrt{C(\delta_1)}\}}{\epsilon}$. By using $\delta_1 = \log 2$, we obtain from the definition of $C(\delta_1)$, when $\alpha \geq \alpha_{\max} \geq \dfrac{\sqrt{2}}{\epsilon}$, that $|\min(a, b)| \leq \epsilon$.

Algorithm 4.1 Smooth Algorithm for MCP *Input tolerance ϵ, parameter $\nu_1 \geq \nu > 1$ and initial guess $x_0 \in R^n$*

(1) **Initialization** *For $1 \leq i \leq n$ of Case 4 of (22), let $w_0^i = (F_i(x_0))_+$, $v_0^i = (-F_i(x_0))_+$, $k = 0$ and $\alpha_0 = \alpha(x_0, w_0, v_0)$. Choose $\alpha_{\max} \geq \dfrac{\sqrt{2}}{\epsilon}$*

(2) If $\|r(x_k, w_k, v_k)\|_\infty \leq \epsilon$, stop.

(3) **Newton Armijo Step** *Find $(x_{k+1}, w_{k+1}, v_{k+1})$ by a Newton-Armijo step applied to*

$$R(x, w, v) = 0.$$

(4) **Parameter Update** *If $\alpha(x_{k+1}, w_{k+1}, v_{k+1}) \geq \nu\alpha_k$, set*

$$\alpha_{k+1} = \alpha(x_{k+1}, w_{k+1}, v_{k+1}),$$

otherwise if $\|\nabla f(x_{k+1}, w_{k+1}, v_{k+1})\|_2 \leq \epsilon$, set

$$\alpha_{k+1} = \nu_1\alpha_k.$$

If $\alpha_{k+1} > \alpha_{\max}$, set $\alpha_{k+1} = \alpha_{\max}$. Let $k = k+1$, go to step (2).

Let I denote the index set of the F_i of Case 1, J of Case 2, K of Case 3 and L of Case 4 of (22). In order to characterize the nonsingularity of ∇R, we now give a definition of a regular MCP. Note that the monotone NCP is regular. More generally, an NCP with a P_0 Jacobian matrix is regular.

Definition 4.1 *An MCP is called regular if*

$$\begin{bmatrix} \nabla F_I(x) \\ \nabla F_J(x) \\ \nabla F_K(x) \\ \nabla F_L(x) \end{bmatrix} + \begin{bmatrix} 0 & & & \\ & D_J & & \\ & & D_K & \\ & & & D_L \end{bmatrix}$$

is nonsingular, for all positive diagonal matrices D_J,D_K and D_L, that has the dimension of $|J|$, $|K|$ and $|L|$ respectively.

Theorem 4.3 *Consider a solvable regular mixed complementarity problem (21) with with $F(x) \in LC^1_K(R^n)$. Then*

(1) The sequence $\{x_k, w_k, v_k\}$ defined in Algorithm 4.1 exists.

(2) Any accumulation point of the above sequence is an $\epsilon-$accurate solution of the MCP (21).

(3) If an accumulation point exists, the whole sequence $\{x_k, w_k, v_k\}$ converges to an $\epsilon-$accurate solution quadratically.

(4) If, in addition, the level set $\{(x, w, v) \mid \|r(x, w, v)\|_2 \leq \|r(x_0, w_0, v_0)\|_2 + \frac{\nu}{\nu-1}\frac{2\sqrt{N}\log 2}{\alpha_0}\}$ is compact, the sequence $\{x_k, w_k, v_k\}$ converges to an $\epsilon-$accurate solution at a quadratic rate.

Proof We denote (x_k, w_k, v_k) by y_k and $(\bar{x}, \bar{w}, \bar{v})$ by $\bar{y}$.
(1) Let $\Lambda_J = diag(p'(x_J - l_J - F_J(x), \alpha))$, $\Lambda_K = diag(p'(u_K - x_K - F_K(x), \alpha))$, $\Lambda_{L_1} = diag(p'(x_L - l_L - w, \alpha))$ and $\Lambda_{L_2} = diag(p'(u_L - x_L - v, \alpha))$, then

$$\nabla R(x, w, v) = \begin{bmatrix} I & & & & & \\ & \Lambda_J & & & & \\ & & \Lambda_K & & & \\ & & & I & & \\ & & & & I & \\ & & & & & I \end{bmatrix} \begin{bmatrix} \nabla_I F_I(x) & \nabla_J F_I(x) & \nabla_K F_I(x) & \nabla_L F_I(x) & & \\ \nabla_I F_J(x) & \nabla_J F_J(x) & \nabla_K F_J(x) & \nabla_L F_J(x) & & \\ \nabla_I F_K(x) & \nabla_J F_K(x) & \nabla_K F_K(x) & \nabla_L F_K(x) & & \\ \nabla_I F_L(x) & \nabla_J F_L(x) & \nabla_K F_L(x) & \nabla_L F_L(x) & -I & I \\ & & & I - \Lambda_{L_1} & \Lambda_{L_1} & \\ & & & \Lambda_{L_2} - I & & \Lambda_{L_2} \end{bmatrix}$$

$$+\begin{bmatrix} 0 & & & & & \\ & I-\Lambda_J & & & & \\ & & I-\Lambda_K & & & \\ & & & 0 & & \\ & & & & 0 & \\ & & & & & 0 \end{bmatrix}$$

Therefore

$$\begin{bmatrix} I & & & & & \\ & I & & & & \\ & & I & & & \\ & & & I & \Lambda_{L_1}^{-1} & -\Lambda_{L_2}^{-1} \\ & & & & I & \\ & & & & & I \end{bmatrix} \nabla R(x,w,v) =$$

$$\begin{bmatrix} I & & & & & \\ & \Lambda_J & & & & \\ & & \Lambda_K & & & \\ & & & I & & \\ & & & & I & \\ & & & & & I \end{bmatrix} \left(\begin{bmatrix} \nabla F_I(x) & 0 \\ \nabla F_J(x) & 0 \\ \nabla F_K(x) & 0 \\ \nabla F_L(x) & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & & & & & \\ & \Lambda_J^{-1}-I & & & & \\ & & \Lambda_K^{-1}-I & & & \\ & & & \Lambda_{L_1}^{-1}+\Lambda_{L_2}^{-1}-2I & & \\ & & & I-\lambda_{L_1} & \Lambda_{L_1} & \\ & & & \lambda_{L_2}-I & & \Lambda_{L_2} \end{bmatrix} \right)$$

Hence $\nabla R(x,w,v)$ is nonsingular if the MCP is regular. If $F \in LC_K^1(R^n)$, then $\{(x_k, w_k, v_k)\}$ are well defined.

(2) For an accumulation point $\bar{y}$, we have $y_{k_i} \longrightarrow \bar{y}$. Since the parameter α can be changed only finite many times, then $\alpha_k = \bar{\alpha}$, for all $k \geq \bar{k}$. Therefore, without loss generality, we consider the sequence $\{y_{k_i}\}$ for fixed $\bar{\alpha}$. In case that $f(\bar{y}) = 0$, for $\bar{\alpha}$, we have $\bar{\alpha} = \alpha_{\max}$. Otherwise, since $\|\nabla f(y_{k_i})\|_2 \longrightarrow 0$, there exists an $\bar{i}$ such that $\|\nabla f(y_{k_{\bar{i}}})\|_2 \leq \epsilon$. By (4) of Algorithm 4.1, $\bar{\alpha}$ will change to $\nu_1 \bar{\alpha}$. That contradicts the definition of $\bar{\alpha}$. Hence $R(\bar{y}) = 0$ for $\bar{\alpha} = \alpha_{\max}$, $\bar{y}$ is an $\epsilon-$accurate solution of the MCP (21). The other case is that $f(\bar{y}) > 0$ for $\bar{\alpha}$. Since $F \in LC_K^1(R^n)$, for a compact set S whose interior contains $\{y_{k_i}\}$ and $\bar{y}$, we have that $R(y) \in LC_{K_1}^1(S)$ for some K_1. By Quadratic Bound Lemma [27, p.144], we have

$$\|f(y_{k_i} + \lambda_{k_i} d_{k_i}) - f(y_{k_i}) - \nabla f(y_{k_i})^T \lambda_{k_i} d_{k_i}\|_2 \leq \frac{K_1}{2} \|\lambda_{k_i} d_{k_i}\|_2^2.$$

Since $\nabla R(y)$ is nonsingular, on the compact S, there exists $K(S)$ and $k(S)$ such that

$$x^T \nabla R^{-T}(y) \nabla R^{-1}(y) x \leq K(S) x^T x, \quad \forall y \in S, x \in R^n$$

and

$$x^T \nabla R^{-1}(y) \nabla R^{-T}(y) x \geq k(S) x^T x, \quad \forall y \in S, x \in R^n.$$

Consequently

$$f(y_{k_i}) - f(y_{k_i} + \lambda_{k_i} d_{k_i}) \geq -\lambda_{k_i} \nabla f(y_{k_i})^T d_{k_i} - \frac{K_1}{2} \lambda_{k_i}^2 R(y_{k_i})^T \nabla R(y_{k_i})^{-T} \nabla R(y_{k_i})^{-1} R(y_{k_i})$$

$$\geq -\lambda_{k_i} \nabla f(y_{k_i})^T d_{k_i} - \frac{K_1 K(S)}{2} \lambda_{k_i}^2 R(y_{k_i})^T R(y_{k_i}) = \lambda_{k_i} (1 - \frac{K_1 K(S)}{2} \lambda_{k_i}) |\nabla f(y_{k_i})^T d_{k_i}|$$

$$\geq \lambda_{k_i}\sigma|\nabla f(y_{k_i})^T d(y_{k_i})|, \qquad \text{if } \lambda_{k_i} \leq \frac{2(1-\sigma)}{K_1 K(S)}$$

By the rule of choosing λ_{k_i}, we have $\lambda_{k_i} \geq \delta\frac{2(1-\sigma)}{K_1K(S)}$, where δ is the constant used in Armijo stepsize. Therefore

$$f(y_{k_i}) - f(y_{k_i} + \lambda_{k_i} d_{k_i}) \geq 2\sigma\delta\frac{1-\sigma}{K_1K(S)}|\nabla f(y_{k_i})^T d(y_{k_i})|$$

$$= 2\sigma\delta\frac{1-\sigma}{K_1K(S)}|\nabla f(y_{k_i})^T \nabla R(y_{k_i})^{-1}\nabla R(y_{k_i})^{-T}\nabla f(y_{k_i})|$$

$$\geq 2\sigma\delta k(S)\frac{1-\sigma}{K_1K(S)}\|\nabla f(y_{k_i})\|_2^2$$

Since $y_{k_i} \longrightarrow \bar{y}$, we have $\nabla f(\bar{y}) = 0$. Thus $R(\bar{y}) = 0$ and $f(\bar{y}) = 0$. This contradicts the assumption $f(\bar{y}) > 0$. This case cannot occur.

(3) By the analysis in (2), we have $y_{k_i} \longrightarrow \bar{y}$, $R(\bar{y}) = 0$ and $R(y) \in LC_{K_1}(S)$. Therefore

$$\|R(y+d) - R(y) - \nabla R(y)^T d\|_2 \leq \frac{K_1}{2}\|d\|_2^2$$

for $y, y+d \in S$. For $d = \nabla R(y)^{-1}R(y)$, we have

$$\|R(y)\|_2^2 - \|R(y+d)\|_2^2 \geq \|R(y)\|_2^2 - (\frac{K_1}{2}\|d\|_2^2)^2 \geq (1 - \frac{K_1^2K(S)^2}{4}\|R(y)\|_2^2)\|R(y)\|_2^2$$

$$= (1 - \frac{K_1^2K(S)^2}{4}\|R(y)\|_2^2)|\nabla f(y)^T d| \geq \sigma|\nabla f(y)^T d|, \qquad \text{if } \|R(y)\|_2 \leq \frac{2\sqrt{1-\sigma}}{K_1K(S)}$$

Hence, if y is close enough to $\bar{y}$, the Newton step accepted. According to the standard result of local quadratic convergence for Newton Method, Theorem 5.2.1 in [7], the conclusion follows.

(4) Let $\alpha^i, i = 0, 1, \cdots$, be the sequence of different parameters α used in Algorithm 4.1. Let $\{k_i\}$, $i = 0, 1, \cdots$, with $k_0 = 0$, be the indices such that the parameter α changes, that is for $k_i \leq k \leq k_{i+1} - 1$, $\alpha_k = \alpha^i$. For α_0 and y_0, by Lemma 4.1, we have

$$\|R(y_0)\|_2 \leq \|r(y_0)\|_2 + \frac{\sqrt{N}\log 2}{\alpha_0}$$

For $k_0 \leq k < k_1$, since $f(y_k) \leq f(y_0)$ with α_0,

$$\|R(y_k)\|_2 \leq \|R(y_0)\|_2 \leq \|r(y_0)\|_2 + \frac{\sqrt{N}\log 2}{\alpha_0}$$

By Lemma 4.1,

$$\|r(y_k)\|_2 \leq \|R(y_k)\|_2 + \frac{\sqrt{N}\log 2}{\alpha_0} \leq \|r(y_0)\|_2 + \frac{2\sqrt{N}\log 2}{\alpha_0}$$

For α^1 and y_{k_1-1}, by Lemma 4.1,

$$\|R(y_{k_1-1})\|_2 \leq \|r(y_{k_1-1})\|_2 + \frac{\sqrt{N}\log 2}{\alpha_1} \leq \|r(y_0)\|_2 + \frac{2\sqrt{N}\log 2}{\alpha_0} + \frac{\sqrt{N}\log 2}{\nu\alpha_0}$$

For $k_1 \le k < k_2$, since $f(y_k) \le f(y_{k_1-1})$ with α_1,

$$\|R(y_k)\|_2 \le \|R(y_{k_1-1})\|_2 \le \|r(y_0)\|_2 + \frac{2\sqrt{N}\log 2}{\alpha_0} + \frac{\sqrt{N}\log 2}{\nu\alpha_0}$$

By Lemma 4.1,

$$\|r(y_k)\|_2 \le \|R(y_k)\|_2 + \frac{\sqrt{N}\log 2}{\alpha_1} \le \|r(y_0)\|_2 + \frac{2\sqrt{N}\log 2}{\alpha_0} + \frac{2\sqrt{N}\log 2}{\nu\alpha_0}$$

Inductively, for α^i and y_{k_i-1},

$$\|R(y_{k_i-1})\|_2 \le \|r(y_0)\|_2 + \frac{2\sqrt{N}\log 2}{\alpha_0} + \frac{2\sqrt{N}\log 2}{\nu\alpha_0} + \cdots + \frac{2\sqrt{N}\log 2}{\nu^{i-1}\alpha_0} + \frac{\sqrt{N}\log 2}{\nu^i\alpha_0}$$

for $k_i \le k < k_{i+1}$,

$$\|R(y_k)\|_2 \le \|r(y_0)\|_2 + \frac{2\sqrt{N}\log 2}{\alpha_0} + \frac{2\sqrt{N}\log 2}{\nu\alpha_0} + \cdots + \frac{2\sqrt{N}\log 2}{\nu^{i-1}\alpha_0} + \frac{\sqrt{N}\log 2}{\nu^i\alpha_0}$$

$$\|r(y_k)\|_2 \le \|r(y_0)\|_2 + \frac{2\sqrt{N}\log 2}{\alpha_0} + \frac{2\sqrt{N}\log 2}{\nu\alpha_0} + \cdots + \frac{2\sqrt{N}\log 2}{\nu^{i-1}\alpha_0} + \frac{2\sqrt{N}\log 2}{\nu^i\alpha_0}$$

Therefore, for all k we have

$$\|r(y_k)\|_2 \le \|r(y_0)\|_2 + \frac{2\sqrt{N}\log 2}{\alpha_0}(1 + \frac{1}{\nu} + \frac{1}{\nu^2} + \cdots) \le \|r(y_0)\|_2 + \frac{\nu}{\nu-1}\frac{2\sqrt{N}\log 2}{\alpha_0}$$

If the level set $\{y | \|r(y)\|_2 \le \|r(y_0)\|_2 + \frac{\nu}{\nu-1}\frac{2\sqrt{N}\log 2}{\alpha_0}\}$ is compact, there exists an accumulation point. By (2) and (3), the whole sequence converges to an $\epsilon-$accurate solution of MCP (21). $\square$

We shall give our numerical test results for Algorithm 4.1 in Section 6, after relating our smooth approach to the central path of the interior point method [19] in Section 5.

5 Relation to the Interior Point Method

In this section, we consider the NCP (9). Let the density function $d(x)$ satisfy (A1)-(A3) and $D_2 = 0$, and let let $\hat{p}(x, \beta)$ be defined by (4). If x solves the nonlinear equation (10) exactly, then

$$x = \hat{p}(x - F(x), \beta) > x - F(x)$$

where the last inequality follows from the fact that $\hat{p}(\xi, \beta) > \xi$, (5) of Proposition 2.2. Hence

$$x > 0 \qquad F(x) > 0,$$

and x belongs to the interior of the feasible region $\{x \mid F(x) \ge 0, x \ge 0\}$ of the NCP. Hence an exact solution of (10) is interior to the feasible region. However the iterates of the smooth method, which are only approximate solutions of (10), are not necessarily feasible. For the function $\hat{p}$ defined in Example 2.2 [38, 17, 4] , the exact solution x of the equation (10) satisfies

$$x > 0, \quad F(x) > 0, \quad x_i F_i(x) = \beta^2, \;\; i = 1, \cdots, n$$

which is precisely the central path of the interior point method for solving NCP. Methods that trace this path but allow iterates to be exterior to the feasible region have been proposed in [38], [4] and [17]. In [18], the relation between Smale's method [38] and the central path was pointed out. For our function $\hat{p}$ defined in Example 2.1, the solution x of the nonlinear equation (10), for different values of β, constitutes another path in the interior of the feasible region that satisfies:

$$x > 0, \quad F(x) > 0, \quad x_i F_i(x) \le 2\beta^2, \ \ i = 1, \cdots, n$$

We now compare our path and the central path of the interior point method by using a very simple example.

Example 5.1 *Let $F(x) = Mx + q$, where*

$$M = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}, \quad q = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

The unique solution is $(1, 0)$. Figure 7 depicts the central path of the interior point method as well as the smooth path generated by an exact solution of the smooth nonlinear equation (10). Figure 8 depicts the error along the central path and along our smooth path as a function of the smoothing parameter β. The error is measured by the distance to the solution point. For this example, the error along our smooth path is smaller than that along the central path for the same value of the parameter β.

6 Numerical Results

In this section, we give our computational experience with the smooth Algorithm 4.1 for the MCP. We implemented the smooth Algorithm 4.1 with an SOR preprocessor if all diagonal elements of the Jacobian matrix are positive. An initial scaling of the function $F_i(x)$, inversely proportional to the absolute value of the diagonal elements of the Jacobian matrix, is performed if $|\nabla_i F_i(x_0)| \ge 100$. The details of implementing the smooth algorithm are given in Appendix 2. For comparison, we also give the results for the PATH solver [8]. Both algorithms were run on a DECstation 5000/125. Among the 52 test problems, which includes all the problems attempted in [13], [30] and [8], 51 problems are from the MCPLIB [9], and one is the generalized von Thünen model from [30] and [39]. Our smooth algorithm was run using one set of default parameters and so was the PATH solver. The smooth algorithm is written in the C language and implemented by using the GAMS/CPLIB [10]. A MINOS routine [25] was used to perform a sparse LU decomposition for solving sparse linear equations. Both algorithms use the same convergence tolerance of $\epsilon = 1.0e-6$.

Table 1 gives a simple description of the test problems [9].

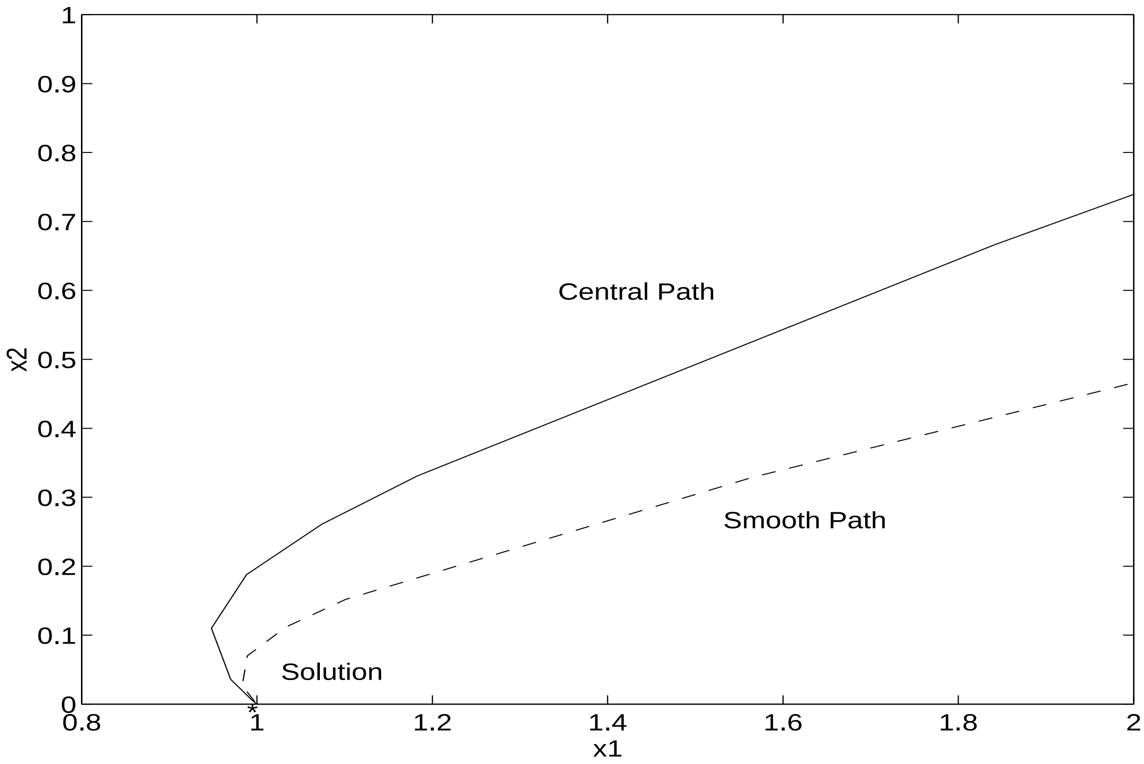

Figure 7: **Comparison of the interior smooth path generated by an exact solution of the smooth nonlinear equation (10) versus the central path for Example 5.1**

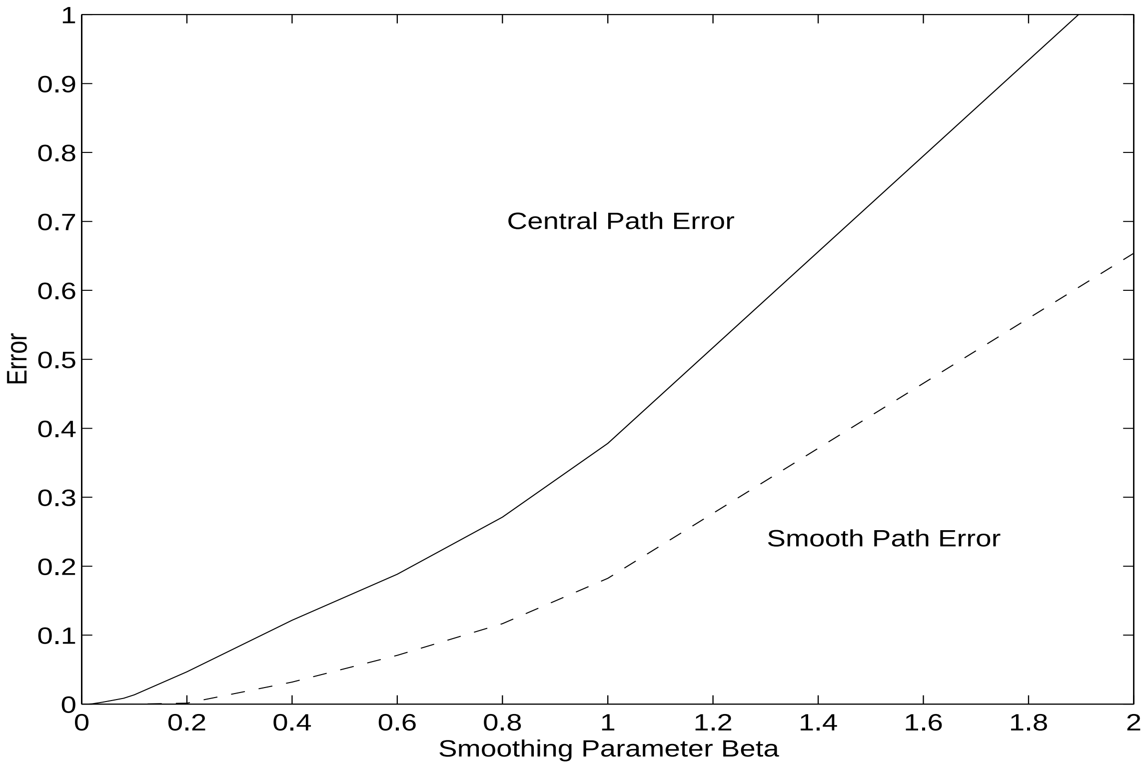

Figure 8: **Error comparison for the central path versus the smooth path for Example 5.1**

Table 1: MCPLIB Problems

Model origin	*GAMS file*	*Size*
Distillation column modeling (NLE)	hydroc20.gms	99
Distillation column modeling (NLE)	hydroc06.gms	29
Distillation column modeling (NLE)	methan08.gms	31
NLP problem form Powell (NLP)	powell_mcp.gms	8
NLP problem form Powell (NLP)	powell.gms	16
NLP test problem form Colville (NLP)	colvncp.gms	15
Dual of Colville problem (NLP)	colvdual.gms	20
Obstacle problem (NLP)(6 cases)	obstacle.gms	$\leq$ 5625
Obstacle Bratu problem (NLP)(6 cases)	bratu.gms	$\leq$ 5625
(NCP)	cycle.gms	1
(NCP)	josephy.gms	4
(NCP)	kojshin.gms	4
(LCP)	explcp.gms	16
Elastohydrodynamic lubrication (NCP)	ehl_kost.gms	101
Nash equilibrium (VI)	nash.gms	10
Nash equilibrium (VI)	choi.gms	13
Spatial price equilibrium (VI)	sppe.gms	27
Spatial price equilibrium (VI)	tobin.gms	42
Walrasian equilibrium (VI)(2 cases)	mathi*.gms	4
Walrasian equilibrium (VI)(2 cases)	scarfa*.gms	14
Walrasian equilibrium (VI)(2 cases)	scarfb*.gms	40
Traffic assignment (VI)	gafni.gms	5
Traffic assignment (VI)	bertsekas.gms	15
Traffic assignment (VI)	freebert.gms	15
Invariant capital stock (VI)	hanskoop.gms	14
Project Independence energy system (VI)	pies.gms	42
Optimal control (Extended LQP)(6 cases)	opt_cont.gms	$\leq$ 8192
Optimal control from Bertsekas (MCP)(6 cases)	bert_oc.gms	$\leq$ 5000

The average CPU times taken by PATH solver and smooth algorithm for all small problems are depicted in Figure 9. Figures 10, 11 and 12 depict the CPU times for all remaining problems except the von Thünen model. We note that the PATH solver [8] is faster than Josephy's Newton method [16] and Rutherford's GAMS [10] mixed inequality and linear equation solver (MILES) [37] which is also Newton-based. Figures 9 to 12 indicate that our smooth algorithm is faster than PATH solver for the larger problems, whereas PATH solver is faster on smaller problems.

The newest version of PATH (PATH 2.7) that uses a Newton method on the active set [2] as a preprocessor, improves solution times on the larger problems. Our smooth method can be similarly improved by adding the projected Newton preprocessor. We have compared PATH and SMOOTH with a Newton preprocessor on a Sun SPARCstation 20. The results are given in Figures 13 to 16. It can be seen that with a Newton preprocessor, the solution times are very similar for PATH and SMOOTH for larger problems, whereas PATH is still better for the smaller problems.

As mentioned in [30], the generalized von Thünen model is an NCP with 106 variables. This is a very difficult problem that has challenged many of the recently proposed algorithms [30, 39]. In order to guarantee that the function $F(x)$ is well defined, we added a lower bound of 1.0e-7 to variables x_1 to x_{26} as suggested by Jong-Shi Pang. We used three starting points. In the first, we

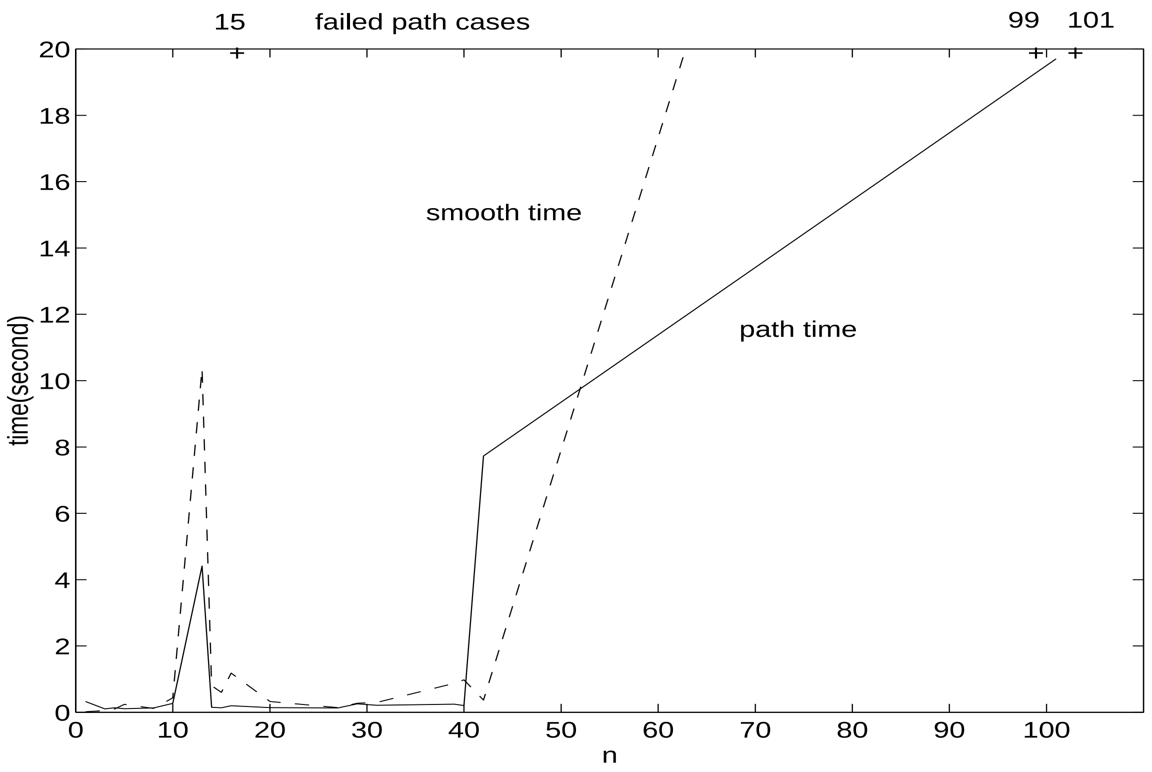

Figure 9: **Smooth versus PATH for Small MCP**

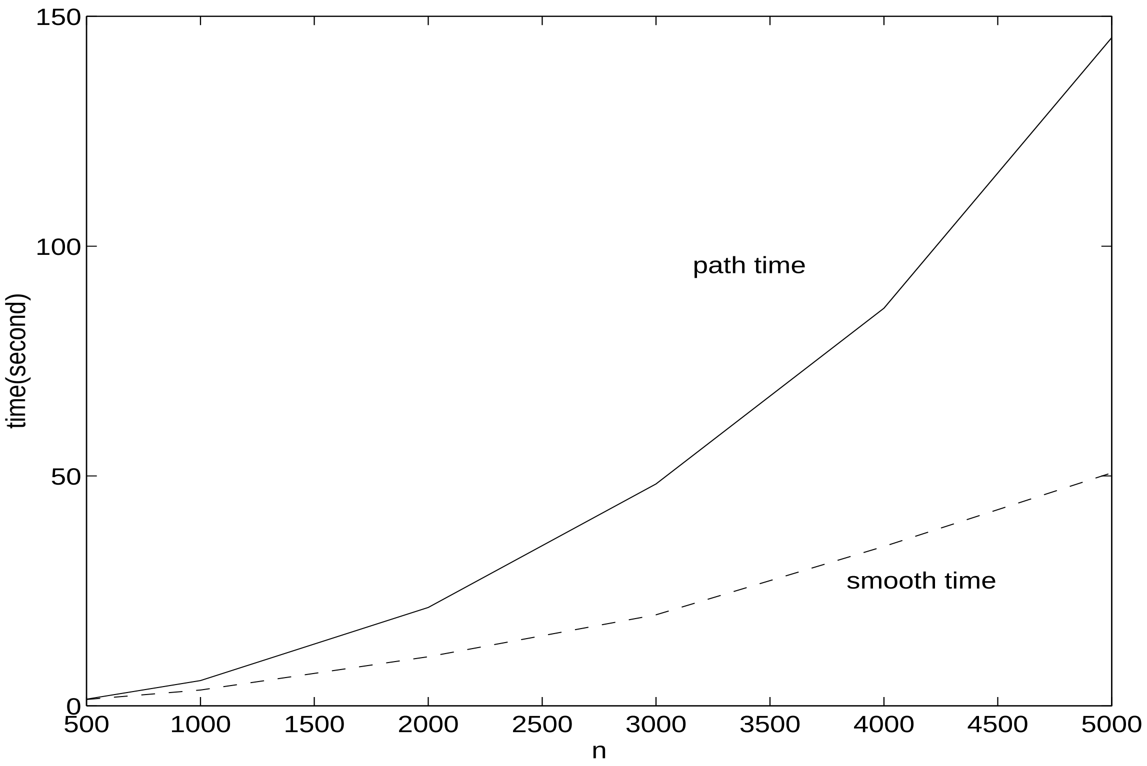

Figure 10: **Smooth versus PATH for Optimal Control Problem(bert_oc.gms)**

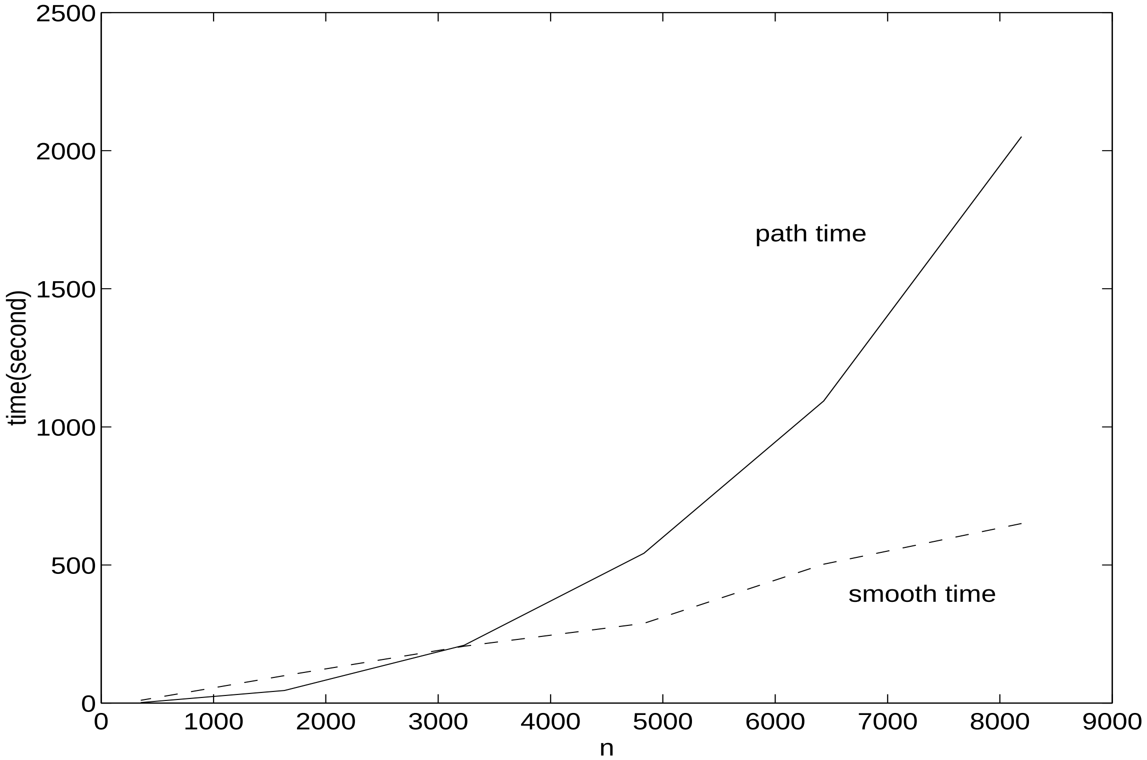

Figure 11: **Smooth versus PATH for Optimal Control Problem(opt_cont.gms)**

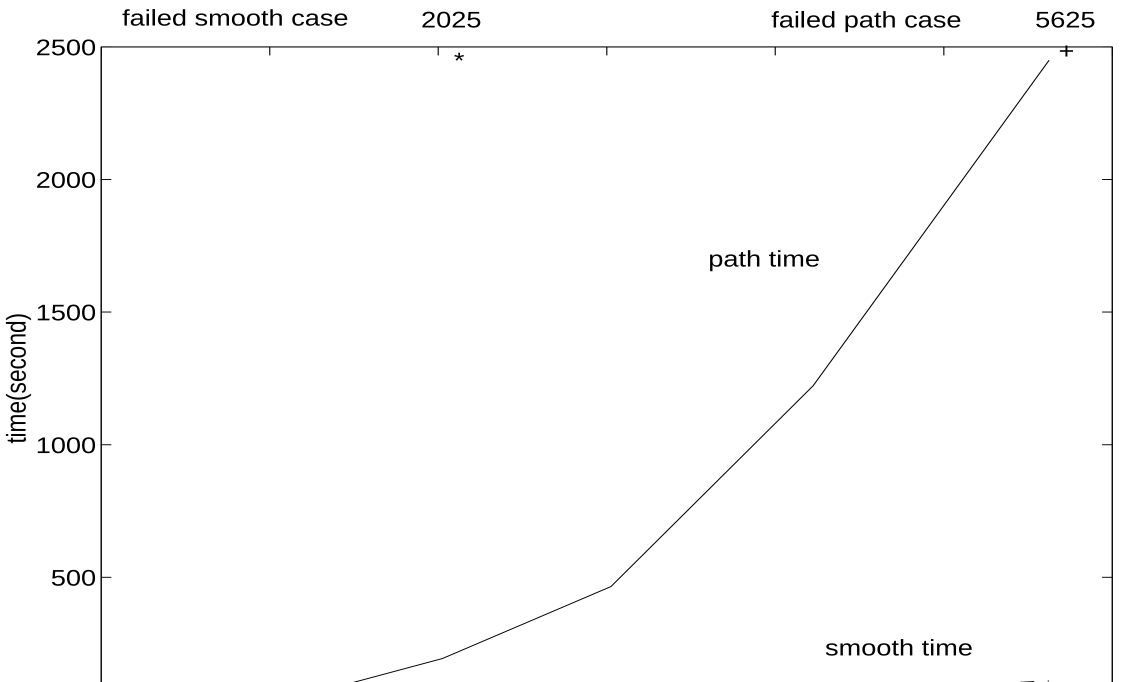

Figure 12: **Smooth versus PATH for Obstacle Problems**

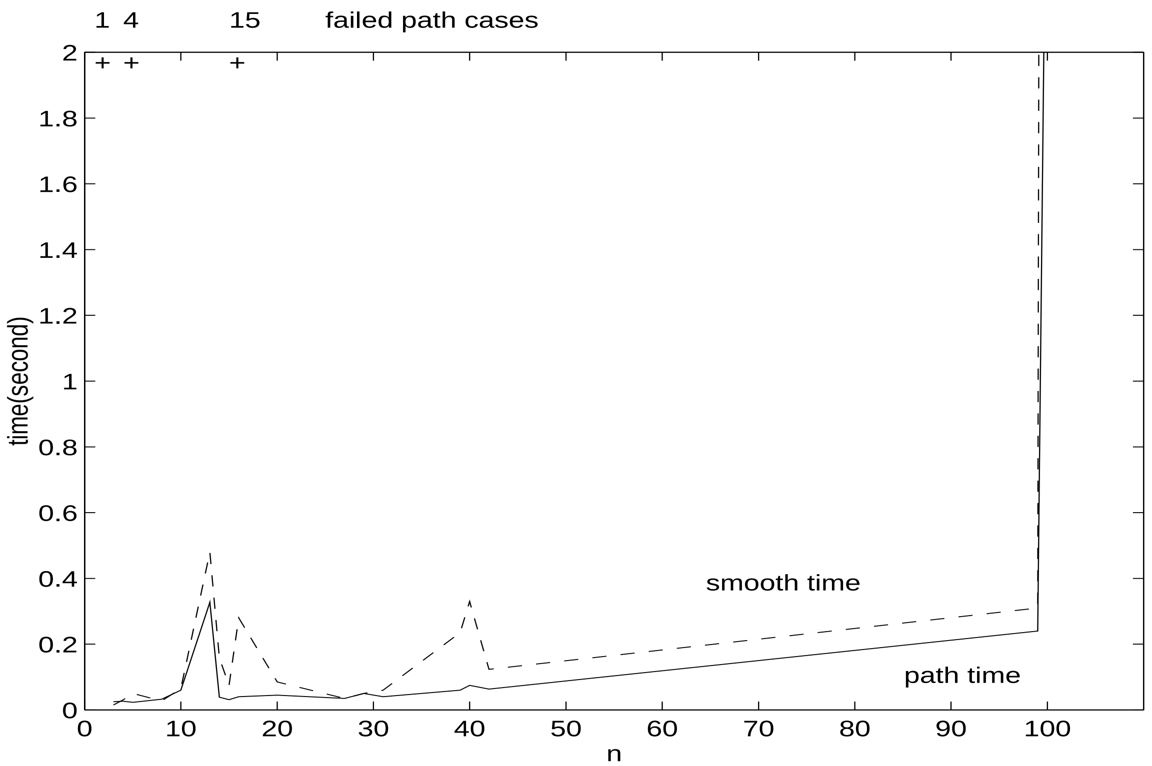

Figure 13: **Smooth with Newton preprocessor versus PATH 2.7 for Small MCP**

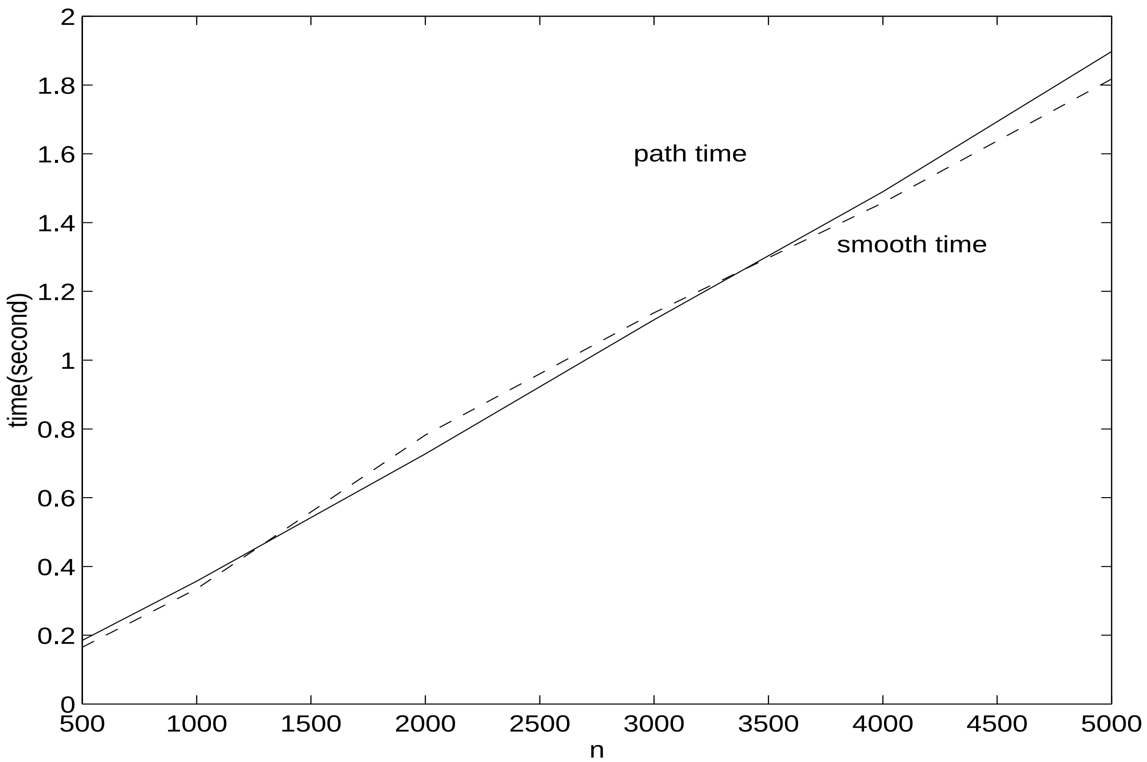

Figure 14: **Smooth with Newton preprocessor versus PATH 2.7 for Optimal Control Problem(bert_oc.gms)**

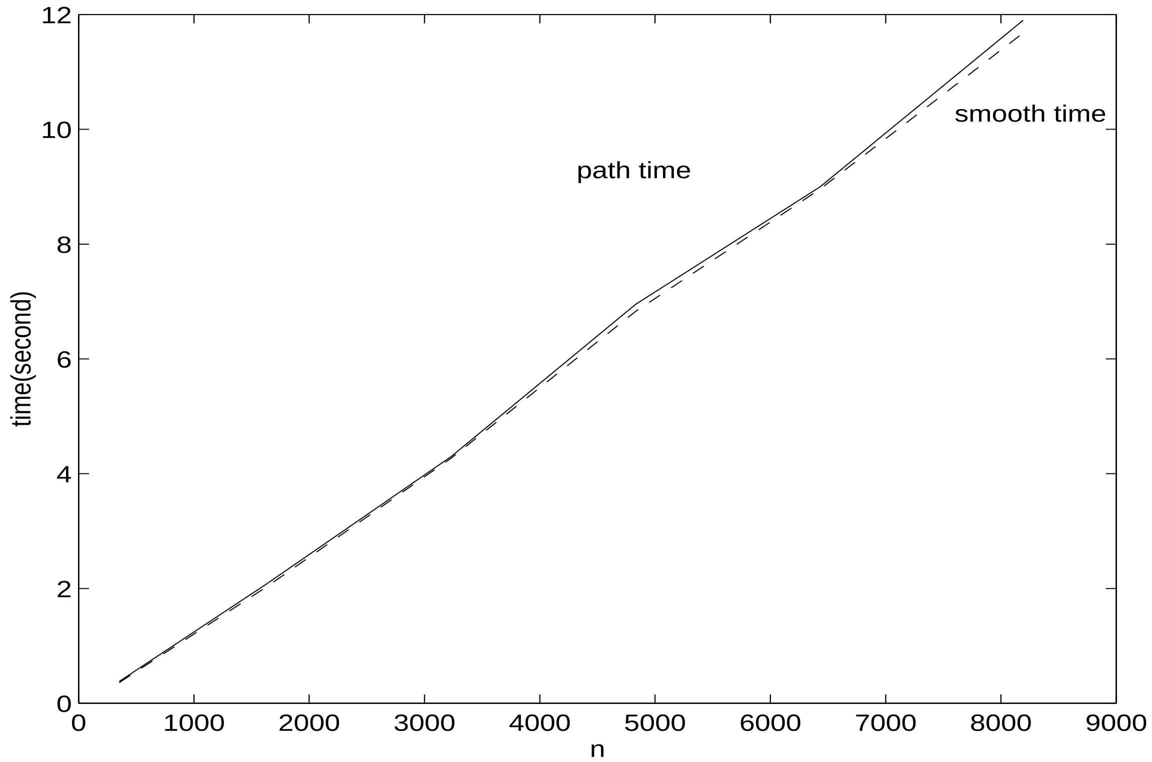

Figure 15: **Smooth with Newton preprocessor versus PATH 2.7 for Optimal Control Problem(opt_cont.gms)**

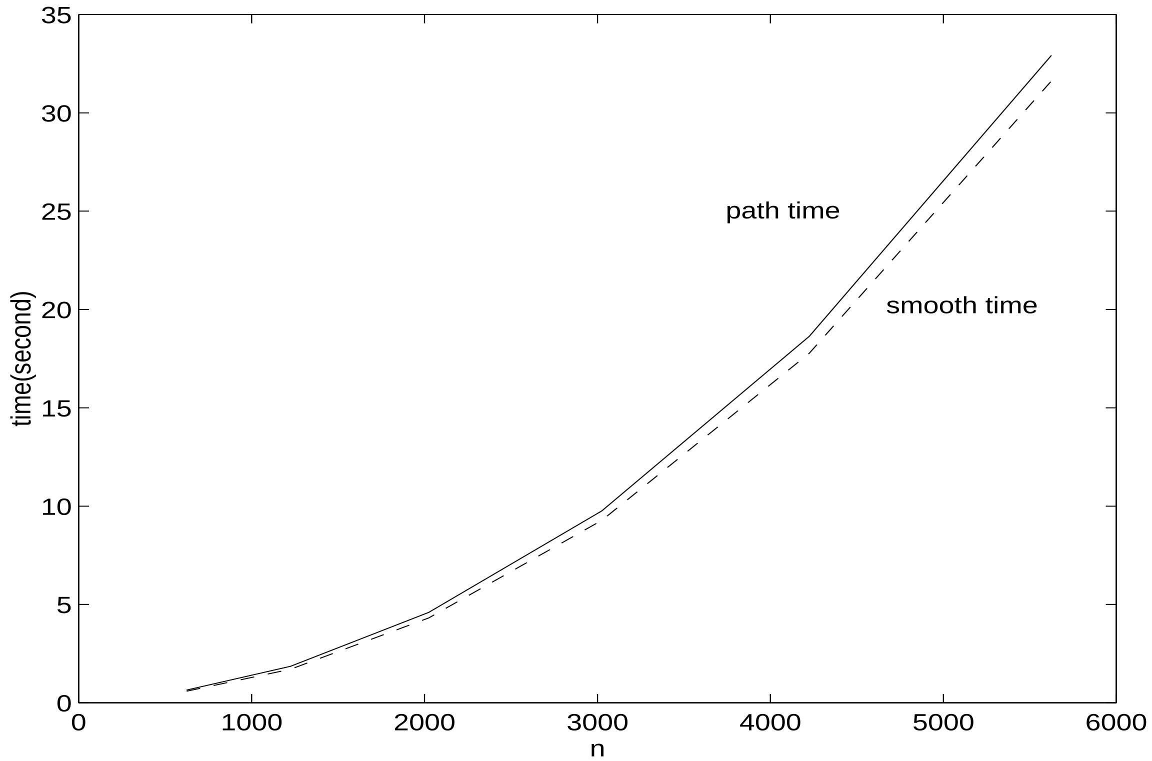

Figure 16: **Smooth with Newton preprocessor versus PATH 2.7 for Obstacle Problems**

set all variables to 1, as suggested by Michael C. Ferris; the second one is a starting point suggested in [39], while the third is the point suggested in [39] and modified by Jong-Shi Pang. SMOOTH, with or without the Newton preprocessor, solved the problem from all the three starting points. Solution times did not change by adding the Newton preprocessor. We report times for SMOOTH with the preprocessor. Starting with the first point, SMOOTH took a long time, 95.44 seconds to solve the problem. From the second point, we obtained a solution in 36 iterations and 3.70 seconds and from the third point, we obtained a solution in 49 iterations and 7.01 seconds. PATH 2.7 solved the problem 7 times out of 10 from the first starting point, 6 times out of 10 from the second starting point, and 5 times out of 10 from the third starting point. The average times of the successful PATH runs were 2.59, 3.94 and 3.21 seconds for the first, second and third starting points respectively. We note that when the artificially imposed lower bounds on x_1 to x_{26} were removed, PATH failed from all starting points and SMOOTH failed from the first and second starting points. However from the third starting point, SMOOTH generated a solution with some small negative components which, when set to 1.0e-9, gave a solution accurate to 1.79539e-7.

Summing up the numerical experiments with PATH and SMOOTH, we believe that comparisons between the two methods without a Newton preprocessor is more indicative of their relative effectiveness. With the Newton preprocessor, a lot of the work for the larger problems is performed by the Newton preprocessor and hence the nearly equal performance of the two methods on these problems.

7 Conclusion

Smoothing is an effective approximate reformulation for a wide range of complementarity problems to any desired degree of accuracy. Newton-type methods for solving smooth problems constitute a powerful computational approach for solving these problems. Parallelization of smooth methods for large-scale problems and their extension to other nonconvex and nonsmooth problems, are two promising research areas worth investigating.

Appendix 1

In order to prove Lemma 3.4, we need the following lemma.

Lemma 7.1 *(i) Let $t(x) = xe^{-x}$, then*

$$\max_{x\in[a,b]} t(x) \le \max\{t(a), t(b), \frac{1}{e}\}.$$

(ii) Let

$$g(x) = \frac{x^2}{(1+\delta)e^x - 1}, \quad \delta \ge 0, \quad -\log(1+\delta) \notin [a,b]$$

then

$$\max_{x\in[a,b]} g(x) \le \max\{g(a), g(b), \frac{2}{1+\delta}t(a), \frac{2}{1+\delta}t(b), \frac{2}{(1+\delta)e}\}.$$

(iii) Let $h(x)$ be defined in Lemma 3.4, $-\log(1+\delta) \notin [a,b]$, then

$$\max_{x\in[a,b]} h(x) \le \max\{h(a), h(b), g(a), g(b), \frac{2}{1+\delta}t(a), \frac{2}{1+\delta}t(b), \frac{2}{(1+\delta)e}\}.$$

Proof (i) By definition, $t'(x) = e^{-x} - xe^{-x} = (1-x)e^{-x}$, hence $t'(x) = 0$ implies $x = 1$. Therefore

$$\max_{x\in[a,b]} t(x) \le \max\{t(a), t(b), t(1)\} = \max\{t(a), t(b), \frac{1}{e}\}.$$

(ii) Notice $-\log(1+\delta) \notin [a,b]$ and $\delta \ge 0$, we know

$$g'(x) = \frac{2x((1+\delta)e^x - 1) - x^2(1+\delta)e^x}{((1+\delta)e^x - 1)^2}.$$

Hence $g'(x) = 0$ implies $x = 0$ or

$$(1+\delta)e^x - 1 = \frac{1+\delta}{2} x e^x.$$

By (i) and $g(0) = 0$

$$\max_{x\in[a,b]} g(x) \le \max\{g(a), g(b), g(0), \frac{2}{1+\delta} \max_{x\in[a,b]} t(x)\}$$

$$\le \max\{g(a), g(b), \frac{2}{1+\delta} t(a), \frac{2}{1+\delta} t(b), \frac{2}{(1+\delta)e}\}.$$

(iii) Since

$$h'(x) = -\log(1+\delta - e^{-x}) - \frac{xe^{-x}}{1+\delta - e^{-x}},$$

the point x is a stationary point of $h(x)$ if and only if

$$\log(1+\delta - e^{-x}) = -\frac{xe^{-x}}{1+\delta - e^{-x}} = -\frac{x}{(1+\delta)e^x - 1}.$$

Therefore, by (ii),

$$\max_{x\in[a,b]} h(x) \le \max\{h(a), h(b), \max_{x\in[a,b]} g(x)\}$$

$$\le \max\{h(a), h(b), g(a), g(b), \frac{2}{1+\delta} t(a), \frac{2}{1+\delta} t(b), \frac{2}{(1+\delta)e}\}.$$

$\square$

Proof of Lemma 3.4

(i) If $0 < \delta < 1$, $-\log(1+\delta) \notin [0, -\log\delta]$. By (iii) of Lemma 7.1 and notice

$$h(0) = 0,\ h(-\log\delta) = 0,\ g(0) = 0,\ g(-\log\delta) = \delta\log^2\delta,\ t(0) = 0,\ t(-\log\delta) = -\delta\log\delta,$$

we have

$$\max_{x\in[0,-\log\delta]} h(x) \le \max\{\delta\log^2\delta, -\frac{2\delta}{1+\delta}\log\delta, \frac{2}{(1+\delta)e}\}.$$

It is easy to get

$$\max_{\delta\in[0,1]} \delta\log^2\delta \le \frac{4}{e^2}, \quad \max_{\delta\in[0,1]} -\frac{2\delta}{1+\delta}\log\delta \le 2.$$

Combining the above inequalities, we get the conclusion.

(ii) If $\delta \geq 1$, $-\log(1+\delta) \notin [-\log\delta, 0]$. Similarly with (i), we know

$$h(0)=0,\ h(-\log\delta)=0,\ g(0)=0,\ g(-\log\delta)=\delta\log^2\delta,\ t(0)=0,\ t(-\log\delta)=-\delta\log\delta.$$

By (iii) of Lemma 7.1,

$$\max_{x\in[-\log\delta,0]} h(x) \leq \max\{\delta\log^2\delta, -\frac{2\delta}{1+\delta}\log\delta, \frac{2}{(1+\delta)e}\} \leq \max\{\delta\log^2\delta, \frac{1}{e}\}.$$

(iv) If $\delta = 0$,

$$\lim_{x\longrightarrow 0} h(x)=0,\ \lim_{x\longrightarrow\infty} h(x)=0,\ \lim_{x\longrightarrow 0} g(x)=0,\ \lim_{x\longrightarrow\infty} g(x)=0,\ \lim_{x\longrightarrow 0} t(x)=0,\ \lim_{x\longrightarrow\infty} t(x)=0,$$

For any $\epsilon > 0$, we have $0 = -\log(1+\delta) \notin [\epsilon, +\infty)$. And there exists $\epsilon_0 > 0$ such that

$$h(\epsilon) < \frac{2}{e},\ g(\epsilon) < \frac{2}{e},\ t(\epsilon) < \frac{1}{e},\quad \text{for } 0 < \epsilon < \epsilon_0$$

Therefore, for $0 < \epsilon < \epsilon_0$

$$\max_{x\in[\epsilon,\infty)} h(x) \leq \frac{2}{e}$$

Let ϵ approaches 0, we have

$$\max_{x\in[0,\infty)} h(x) \leq \frac{2}{e}$$

$\square$

Appendix 2

Here is the actual implementation of the smooth algorithm 4.1. In the following algorithm (x_k, w_k, v_k) is simply denoted by y_k.

Algorithm 7.1 Smooth Algorithm for MCP *Input tolerance* $\epsilon = 1.0e-6$, *and initial guess* $x_0 \in R^n$

(1) **Initialization** *For* $1 \leq i \leq n$ *of Case 4 of (22), let* $w_0^i = (F_i(x_0))_+$, $v_0^i = (-F_i(x_0))_+$, $k = 0$ *and* $\alpha_0 = \alpha(y_0)$.

(2) If $\|r(y_0)\|_\infty \leq \epsilon$, *stop.*

(3) **Newton Direction** d_k

$$d_k = -\nabla R(y_k)^{-1} R(y_k)$$

In order to avoid nonsingularity, for the Case 2-4 of (22), if $\nabla_i R_i(y_k) < 1.0e-9$, *let* $\nabla_i R_i(y_k) = 1.0e-9$.

(4) **Stepsize** λ_k *(Armijo)*

$$y_{k+1} = y_k + \lambda_k d_k, \lambda_k = \max\{1, \delta, \delta^2, \cdots\},\ s.t.$$

$$f(y_{k+1}) \leq f(y_k)$$

where $\delta = 0.75$.

(5) **Parameter Update** *If* $\alpha(y_{k+1}) \geq \alpha_k$, *set*

$$\alpha_{k+1} = \alpha(y_{k+1}),$$

otherwise if $\|\nabla f(y_{k+1})\|_2 \leq \epsilon$, *set*

$$\alpha_{k+1} = 2\alpha_k.$$

Let $k = k+1$, *go to step (2).*

For some of the test problems, the function is not well defined outside the feasible region. In such cases, the line search step (5) may fail. If this occurs, we will try to push the next point inside the feasible region by setting the smooth parameter α to a very small value, such as 1.0e-10.

Acknowledgement

We are grateful to Michael C. Ferris and Steven P. Dirkse for letting us use their GAMS interface routines, their PATH solver for comparison purposes and their Newton preprocessor that we incorporated into a later version of SMOOTH. We are also grateful to Jong-Shi Pang for providing us with the generalized von Thünen model and an initial starting point, and to Marc Teboulle for pointing out that our function $p(x, \alpha)$ of Example 2.1 can also be obtained by using an appropriate recession function given in Example 2.2 of [1].

References

[1] A. Ben-Tal and M. Teloulle. A smoothing technique for nondifferentiable optimization problems. In S. Deldeki, editor, *Optimization*, pages 1–11, Berlin, 1989. Springer–Verlag. Lecture Notes in Mathematics 1405.

[2] Dimitri P. Bertsekas. Projected Newton methods for optimization problems with simple constraints. *SIAM Journal on Control and Optimization*, 20:221–246, 1982.

[3] D.P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods.* Academic Press, New York, 1982.

[4] Bintong Chen and P. T. Harker. A non-interior-point continuation method for linear complementarity problems. *SIAM Journal on Matrix Analysis and Applicatons*, 14:1168–1190, 1993.

[5] Chunhui Chen and O.L. Mangasarian. Smoothing methods for convex inequalities and linear complementarity problems. Technical Report 1191, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin 53706, November 1993. Mathematical Programming, to appear.

[6] R.W. Cottle, F. Giannessi, and J.-L. Lions. *Variational Inequalities and Complementariy Problems.* John Wiley & Sons, New York, 1980.

[7] J.E. Dennis and R.B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall, Englewood Cliffs, N.J., 1983.

[8] S.P. Dirkse and M.C. Ferris. The path solver: A non-monotone stabilization scheme for mixed complementarity problems. Computer Sciences Department Technical Report 1179, University of Wisconsin, Madison, Wisconsin, 1993.

[9] S.P. Dirkse and M.C. Ferris. MCPLIB: A collection of nonlinear mixed complementarity problems. Computer Sciences Department Technical Report 1215, University of Wisconsin, Madison, Wisconsin, 1994.

[10] S.P. Dirkse, M.C. Ferris, P.V. Preckel, and T. Rutherford. The GAMS callable program library for variational and complementarity solvers. Manuscript, University of Wisconsin, 1993.

[11] M.S. Gowda. On the extended linear complementarity problem. Technical report, Department of Mathematics & Statistics, University of Maryland Baltimore County, Baltimore, Maryland, 1994.

[12] P. T. Harker and J.-S. Pang. Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications. *Mathematical Programming*, 48:161–220, 1990.

[13] P.T. Harker and B. Xiao. Newton's method for the nonlinear complementarity problem: A B–differentiable equation approach. *Mathematical Programming*, 48:339–357, 1990.

[14] J. Hertz, A. Krogh, and R. G. Palmer. *Introduction to the Theory of Neural Computation.* Addison-Wesley, Redwood City, California, 1991.

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

[16] N.H. Josephy. Newton's method for generalized equations. Technical Summary Report 1965, Mathematics Research Center, University of Wisconsin, Madison, Wisconsin, 1979.

[17] C. Kanzow. Some tools allowing interior-point methods to become noninterior. Technical report, Institue of Applied Mathematics, Unversity of Hamburg, Germany, 1994.

[18] M. Kojima and N. Megiddo. The relation between the path of centers and smale's regularization of the linear programming problem. *Linear Algrbra and Its Applications*, 152:135–139, 1991.

[19] M. Kojima, S. Mizuno, and A. Yoshise. A polynomial–time algorithm for a class of linear complementarity problems. *Mathematical Programming*, 44:1–27, 1989.

[20] J. Kreimer and R.Y. Rubinstein. Nondifferentiable optimization via smooth approximation: General analytical approach. *Annals of Operations Research*, 39:97–119, 1992.

[21] K. Madsen and H.B. Nielsen. A finite smoothing algorithm for linear l_1 estimation. *SIAM Journal on Optimization*, 3(2):223–235, May 1993.

[22] O.L. Mangasarian. Mathematical programming in neural networks. *ORSA Journal on Computing*, 5(4):349–360, 1993.

[23] O.L. Mangasarian and J.-S. Pang. The extended linear complementarity problem. Technical Report 1188, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin 53706, November 1993. SIAM Journal on Matrix Analysis and Applications, to appear.

[24] Jorge J. Moré. Global methods for nonlinear complementarity problems. Technical Report Preprint MCS-P429-0494, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois, 1994.

[25] B.A. Murtagh and M.A. Saunders. MINOS 5.0 user's guide. Technical Report SOL 83.20, Stanford University, December 1983.

[26] K.G. Murty. *Linear Complementarity, Linear and Nonlinear Programming.* Helderman–Verlag, Berlin, 1988.

[27] J.M. Ortega. *Numerical Analysis, a Second Course.* Academic Press, 1972.

[28] J.-S. Pang. Newton's method for B–differentiable equations. *Mathematics of Operations Research*, 15:331–341, 1990.

[29] J.-S. Pang. A B–differentiable equation based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems. *Mathematical Programming*, 51:101–131, 1991.

[30] J.-S. Pang and S.A. Gabriel. NE/SQP: A robust algorithm for the nonlinear complementarity problem. *Mathematical Programming*, 60:295–337, 1993.

[31] M. C. Pinar and S. A. Zenios. On smoothing exact penalty functions for convex constrained optimization. *SIAM Journal on Optimization*, 4:486–511, 1994.

[32] D. Ralph. Global convergence of damped newton's method for nonsmooth equations via the path search. *Mathematics of Operations Research*, 1993.

[33] J. Ren. Computable error bounds in mathematical programming. Technical Report 1173, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin 53706, August 1993.

[34] S.M. Robinson. Generalized equations and their solution: Part I: Basic thoery. *Mathematical Programming Study*, 10:128–140, 1979.

[35] S.M. Robinson. Strongly regular generalized equations. *Mathematics of Operations Research*, 5:43–62, 1980.

[36] S.M. Robinson. Newton's method for a class of nonsmooth functions. *Set-Valued Analysis*, 2:291–305, 1994.

[37] Thomas F. Rutherford. MILES: A mixed inequality and nonlinear equation solver. Working Paper, Department of Economics, University of Colorado, Boulder, 1993.

[38] S. Smale. Algorithms for solving equations. In *Proceedings of the International Congress of Mathematicians*, pages 172–195. Ameri. Math. Soc., Providence, 1987.

[39] B. Xiao. Global newton methods for nonlinear programs and variational inequalities. Technical report, Ph. D. thesis, Department of Decision Sciences, The Wharton School, University of Pennsylvania, Philadelphia, PA, 1990.

[40] Y. Ye. A fully polynomial-time approximation algorithm for computing a stationary point of the general linear complementarity problem. *Mathematics of Operations Research*, 18:334–345, 1993.

[41] Israel Zang. A smoothing-out technique for min-max optimization. *Mathematical Programming*, 19:61–77, 1980.