SMOOTHING METHODS IN MATHEMATICAL PROGRAMMING

By

Chunhui Chen

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

(Computer Sciences)

at the

UNIVERSITY OF WISCONSIN – MADISON

1995

Acknowledgements

My greatest thanks go to my advisor Professor Olvi L. Mangasarian for his guidance and advice. He has always been most generous with his time and encouragement. Without him, this thesis could not have been written.

I wish to express my thanks to Professor Robert R. Meyer, Professor Stephen M. Robinson, Professor Michael C. Ferris, Professor James G. Morris and Professor John C. Strikwerda for the time I spent in their classes, for reading a draft of this thesis and for being members of my examination committee.

I wish to thank my colleagues Steve Dirkse for providing the GAMS interface for testing our algorithm and Nick Street for many useful discussions. I also want to express my deepest gratitude to my family: my parents, Zhongkai Chen and Xioumei Jin, my husband, Yonghong Yang and my son, Michael for their love and encouragement all these years.

This research was supported by Air Force Office for Scientific Research Grant F49620–94–1–0036 and National Science Foundation Grants CCR–93222479 and CDA–9024618.

Abstract

A class of parametric smooth functions that approximate the fundamental plus function, $(x)_+ = \max\{0, x\}$, is obtained by twice integrating a probability density function. By means of this approximation, linear and convex inequalities are converted into smooth, convex unconstrained minimization problems, the solution of which approximates the solution of the original problem to a high degree of accuracy for sufficiently small positive value of the smoothing parameter β. In the special case when a Slater constraint qualification is satisfied, an exact solution can be obtained for finite β. Speedup over the linear/nonlinear programming package MINOS 5.4 was as high as 1142 times for linear inequalities of size 2000×1000, and 580 times for convex inequalities with 400 variables. Linear complementarity problems(LCPs) were treated by converting them into a system of smooth nonlinear equations and are solved by a quadratically convergent Newton method. For monotone LCPs with as many as 10,000 variables, the proposed approach was as much as 63 times faster than Lemke's method. Our smooth approach can also be used to solve nonlinear and mixed complementarity problems (NCPs and MCPs) by converting them to classes of smooth parametric nonlinear equations. 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 $\alpha = \beta^{-1}$. An efficient smooth algorithm, based on the Newton-Armijo approach with an adjusted smoothing parameter, is also given and its global and local quadratic convergence is established. For NCPs, 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 the interior point method. Computational results for 52 test problems compare favorably with those for another Newton-based method. The smooth technique is capable of efficiently solving all the test problems solved by Dirkse & Ferris (1993), Harker & Xiao (1990) and Pang & Gabriel (1993).

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The fundamental plus function

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

where x is a real number, plays an important role in mathematical programming. Many problems can be reformulated using this function. For example, a system of inequalities $g(x) \leq 0$, where g is a function from the n-dimensional real space R^n into R^m, can be reformulated as a unconstrained minimization problem: $\min_{x \in R^n} \|(g(x))_+\|$, where $(z)_+$ is taken to mean an m-vector of plus functions applied componentwise and $\|\cdot\|$ denotes any monotonic norm on R^m. Similarly 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, is satisfied if and only if

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

A great number of optimization problems involve such a complementarity condition. For example, most optimality conditions of mathematical programming [34] as well as variational inequalities [6] and extended complementarity problems [29, 12, 53] can be so formulated. In this sense, the plus function plays a key 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 thesis is to use a smooth function approximation to the plus function. With this approximation, many efficient algorithms, such as Newton and quasi-Newton methods, can be easily employed.

Smoothing techniques have already been applied to different problems, such as, l_1-minimization problems [24], multi-commodity flow problems [42], non-smooth programming [54, 22], linear and convex inequalities [4], and linear complementarity problems [2, 4, 19]. These successful techniques motivate a systematic study of the smoothing approach. Questions we wish to address include the following. How should new smoothing functions be constructed? What is a common property of useful smoothing functions?

In Chapter 2, we relate the plus function through a parametric smoothing procedure, to a probability density function with a positive 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 chapter 3 we treat linear and convex inequalities by converting them to unconstrained differentiable minimization problems. First we give a necessary and sufficient condition for existence of a solution, and then give a uniqueness condition for the unconstrained minimization problem. We prove that when β is small enough, the solution of the unconstrained problem can approximate the solution of the original inequalities to any desired accuracy. For the case when the solution set of the inequalities has an interior point, an exact solution to the inequalities is obtained for sufficiently small but positive β. Even for the case when the original linear inequalities are unsolvable, our method still gives an approximate solution in a least error sense.

In Chapter 4 we consider mixed complementarity problems. There are many Newton-based algorithms for solving nonlinear complementarity problems, variational inequalities and mixed complementarity problems. In [13] a good summary and references up to 1988 are given. Generalizations of the Newton method to nonsmooth equations can be found in [46, 47, 48]. Since then, several approaches based on B-differentiable equations were investigated in [14, 38, 39]. In addition, an algorithm based on nonsmooth equations and successive quadratic programming was given [40], as well as a Newton method with a path following technique [43, 9], and a trust region Newton method for solving a nonlinear least squares reformulation of the NCP [31]. With the exception of [31], 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. In Section 4.1, 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 $\alpha = \beta^{-1}$, we give a Newton-Armijo type algorithm and establish its convergence. In Section 4.2 we apply the above method to linear complementarity problem and prove that when the smoothing parameter α is large enough, we can get an exact solution to LCP by a purification of the approximate solution. In Section 4.3 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 [21]. 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 4.4, we treat the MCP, the mixed complementarity problem (4.14). 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. For a solvable MCP, there exists an arbitrarily accurate solution to the smooth nonlinear equation as well as the MCP for sufficiently large value of the smoothing parameter α. An efficient smooth algorithm based on the Newton-Armijo approach with an adjusted smoothing parameter is also given and convergence is established. In Section 4.5, encouraging numerical testing results are given for positive semidefinite linear complementarity problems up to 10,000 variables and 52 problems from the MCPLIB [10] which includes all the problems attempted in [14, 40, 9]. These problems range in size of up to 8192 variables. These examples include the difficult von Thünen NCP model [40, 52] 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 0 and l 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

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

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

Chapter 2

A Class of Smooth Functions

2.1 Twice the Integral of the Delta Distribution

We consider a class of smooth approximations to the fundamental plus function $(x)_+ = \max\{x, 0\}$ which underlies many of the optimality conditions of mathematical programming. Notice first that

$$(x)_+ = \int_{-\infty}^{x} \sigma(y) dy,$$

where $\sigma(x)$ is the step function:

$$\sigma(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases} \tag{2.1}$$

The step function $\sigma(x)$ can in turn be written as,

$$\sigma(x) = \int_{-\infty}^{x} \delta(y) dy,$$

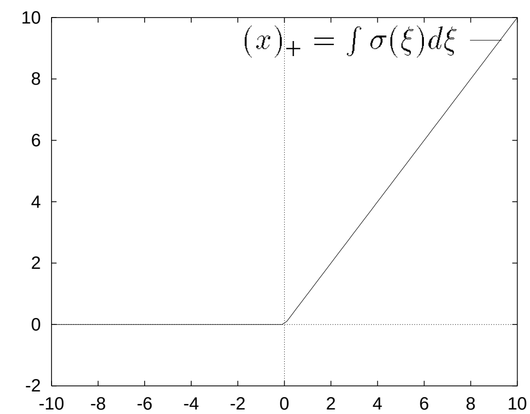

Figure 2.1: **The plus function** $(\mathbf{x})_+ = \max\{\mathbf{x}, \mathbf{0}\}$

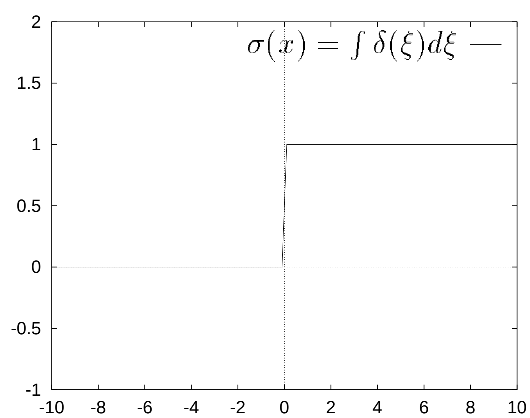

Figure 2.2: **The step function** $\sigma(\mathbf{x}) = \mathbf{1}$ **if** $\mathbf{x} > \mathbf{0}$**,** $\mathbf{0}$ **if** $\mathbf{x} \leq \mathbf{0}$

where $\delta(x)$ is the Dirac delta function that, among others, satisfies the following two properties

$$\delta(x) \geq 0, \quad \int_{-\infty}^{+\infty} \delta(y)dy = 1 \tag{2.2}$$

and can be thought as the limit of density functions. Figures 2.1 to 2.2 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 for smoothing the Dirac delta function and its integrals. Hence we consider the piecewise continuous function $d(x)$ with a finite number of pieces which is a density function, that is it satisfies

$$d(x) \geq 0 \text{ and } \int_{-\infty}^{\infty} d(x)\, dx = 1. \tag{2.3}$$

To parametrize the density function we define

$$\hat{t}(x, \beta) = \frac{1}{\beta} d(\frac{x}{\beta}) \tag{2.4}$$

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:

$$\hat{s}(x, \beta) = \int_{-\infty}^{x} \hat{t}(t, \beta)dt \approx \sigma(x) \tag{2.5}$$

The $\hat{s}(x, \beta)$ is an approximate of the step function $\sigma(x)$. And its integral is an approximate of the plus function $(x)_+$.

Definition 2.1.1 *Given a probability density function $d(x)$ that is parameterized by $\hat{t}(x,\beta) = \frac{1}{\beta}d(\frac{x}{\beta})$, where β is a positive parameter, define the function $\hat{p}(x,\beta)$ as follows*

$$\begin{aligned} \hat{p}(x,\beta) &= \int_{-\infty}^{x}\int_{-\infty}^{t}\hat{t}(\xi,\beta)d\xi dt \\ &= \int_{-\infty}^{x}\hat{s}(t,\beta)dt \end{aligned} \tag{2.6}$$

We will show under suitable assumptions that as β tends to zero, $\hat{p}(x,\beta)$ will be a better approximation of $(x)_+$ and hence we write $\hat{p}(x,\beta) \approx (x)_+$. Therefore, we can get an approximate plus function by twice integrating a density function.

2.2 Equivalent Convolution Representation

In fact, the definition of the previous section is equivalent to the convolution representation defined as follows:

Definition 2.2.1 *Given a probability density function $d(x)$ that is parameterized by $\hat{t}(x,\beta) = \frac{1}{\beta}d(\frac{x}{\beta})$, where β is a positive parameter, define the function $\hat{p}(x,\beta)$ as follows*

$$\begin{aligned} \hat{p}(x,\beta) &= \int_{-\infty}^{+\infty}(x-t)_+\hat{t}(t,\beta)dt \\ &= \int_{-\infty}^{x}(x-t)\hat{t}(t,\beta)dt. \end{aligned} \tag{2.7}$$

This formulation was given in [16, p.12] for a kernel function and in [22] for a density 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.

First we will prove that the function $\hat{p}(x,\beta)$ defined in Definitions 2.1.1 and 2.2.1 are equivalent.

Proposition 2.2.2 *Let $d(x)$ be a probability density function that is parameterized by $\hat{t}(x,\beta) = \frac{1}{\beta}d(\frac{x}{\beta})$, 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 (2.3).*

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

Then the definitions of $\hat{p}(x,\beta)$ given by 2.1.1 and 2.2.1 are consistent.

Proof By the definition of $\hat{t}$ and assumption (A2), we have that $\hat{p}(x,\beta)$ defined by 2.2.1 satisfies

$$\hat{p}(x,\beta) = x\int_{-\infty}^{\frac{x}{\beta}} d(t)dt - \beta\int_{-\infty}^{\frac{x}{\beta}} td(t)dt \tag{2.8}$$

By direct computation,

$$\begin{aligned} \hat{p}'(x,\beta) &= \textstyle\int_{-\infty}^{\frac{x}{\beta}} d(t)dt \\ &= \textstyle\int_{-\infty}^{x} \hat{t}(t,\beta)dt = \hat{s}(x,\beta) \end{aligned} \tag{2.9}$$

Hence the derivatives of $\hat{p}(x,\beta)$ defined by 2.1.1 and 2.2.1 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 (2.6) and (2.7), then $\hat{p}(x,\beta)$ approaches 0 in both, and hence $c = 0$. Therefore the $\hat{p}(x,\beta)$ given by 2.1.1 and 2.2.1 are consistent. $\square$

2.3 Properties of Smooth Plus Functions

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

Proposition 2.3.1 Properties of $\hat{\mathbf{p}}(\mathbf{x}, \beta)$, $\beta > \mathbf{0}$

Let $d(x)$ and $\hat{t}(x, \beta)$ be as in Proposition 2.2.2, 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

$$D_1 = \int_{-\infty}^{0} |x| d(x) dx \tag{2.10}$$

and

$$D_2 = \max\{\int_{-\infty}^{+\infty} x d(x) dx, 0\} \tag{2.11}$$

(3) $\hat{p}(0, \beta) = D_1\beta$.

(4) $\hat{p}(x, \beta)$ is nondecreasing and convex.

(5) $0 \leq \hat{p}'(x, \beta) \leq 1$.

Proof (1) By equation (2.9) in the proof of Proposition 2.2.2, the conclusion follows.

(2) If $x \geq 0$, by using (2.8), 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

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

and

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

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

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

Otherwise, $x < 0$, then

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

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

and

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

(3) By (2.8), we have

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

(4) By equation (2.9) 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.

(5) By formula (2.9), it is easy to see that $0 \leq \hat{p}'(x,\beta) \leq 1$. □

If we make the additional assumption that the density function $d(x)$ has infinite support, the $\hat{p}(x,\beta)$ will satisfy the additional properties stated in the following proposition.

Proposition 2.3.2 *Let $d(x)$ and $\hat{t}(x,\beta)$ be as in Proposition 2.3.1, and let $d(x)$ satisfy (A1), (A2) and (A3) defined as follows:*

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

Then $\hat{p}(x,\beta)$ has the following properties:

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

(2) $0 < \hat{p}'(x,\beta) < 1$.

(3) If the constant D_2 defined by (2.11) is zero, then

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

Proof (1) Since $d(x) > 0$ and $\hat{p}'(x,\beta) = \int_{-\infty}^{\frac{x}{\beta}} d(t)dt$,

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

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

(2) By (A3), $d(x) > 0$ and $\hat{p}'(x,\beta)$ is strictly increasing. Therefore $0 < \hat{p}'(x,\beta) < 1$.

(3) 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

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

□

If the density function $d(x)$ satisfies (A3), the function $\hat{p}(\cdot,\beta)$ is one to one. Therefore the inverse function exists and is denoted by $\hat{p}^{-1}(\cdot,\beta)$.

Now we have a systematic way for generating a class of smooth plus functions. Given any probability density function $d(x)$ satisfying conditions (A1) and (A2) of Proposition 2.2.2, we define $\hat{p}(x,\beta)$ as in 2.1.1 or 2.2.1. 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 the propositions 2.3.1 and 2.3.2 above.

2.4 Examples of Smooth Plus Functions

We now give examples of smooth plus functions. The first example, which will be used throughout this paper, is based on the following classical sigmoid function of neural networks [15, 27, 4]:

$$s(x,\alpha) = \frac{1}{1+e^{-\alpha x}}, \quad \alpha > 0 \tag{2.12}$$

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 β^{-1} and we shall therefore take

$$\alpha = \frac{1}{\beta}. \tag{2.13}$$

Example 2.4.1 Neural Networks Smooth Plus Function [4, 5]

Let

$$d(x) = \frac{e^{-x}}{(1+e^{-x})^2} \tag{2.14}$$

Here $D_1 = \log 2, \quad D_2 = 0$ *and* $supp\{d(x)\} = R$*, where* D_1 *and* D_2 *are defined by (2.10) and (2.11). Integrating* $\frac{1}{\beta}d(\frac{x}{\beta})$ *twice gives*

$$\hat{p}(x,\beta) = x + \beta\log(1+e^{-\frac{x}{\beta}}) \tag{2.15}$$

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

$$p(x,\alpha) = \hat{p}(x, \frac{1}{\alpha}) = \int s(\xi,\alpha)d\xi = x + \frac{1}{\alpha}\log(1+e^{-\alpha x}) \tag{2.16}$$

$$s(x,\alpha) = \hat{s}(x, \frac{1}{\alpha}) = \frac{1}{1+e^{-\alpha x}} = \int t(\xi,\alpha)d\xi \tag{2.17}$$

$$t(x,\alpha) = \hat{t}(x, \frac{1}{\alpha}) = \frac{\alpha e^{-\alpha x}}{(1+e^{-\alpha x})^2} = \alpha s(x,\alpha)(1 - s(x,\alpha)) \tag{2.18}$$

Figures 2.3 to 2.5 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 2.4.2 Chen-Harker-Kanzow-Smale Smooth Plus Function [51], [19] and [2]

Let

$$d(x) = \frac{2}{(x^2+4)^{\frac{3}{2}}} \tag{2.19}$$

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

$$\hat{p}(x,\beta) = \frac{x + \sqrt{x^2 + 4\beta^2}}{2} \tag{2.20}$$

Example 2.4.3 Pinar-Zenios Smooth Plus Function [42]

Let

$$d(x) = \begin{cases} 1 & \textit{if } 0 \le x \le 1 \\ 0 & \textit{otherwise} \end{cases} \tag{2.21}$$

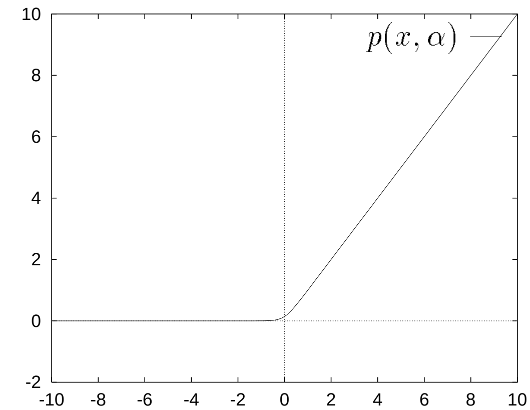

Figure 2.3: **The p function** $\mathbf{p(x,\alpha) = x + \frac{1}{\alpha}\log(1 + e^{-\alpha x})}$ **with** $\alpha = \mathbf{5}$

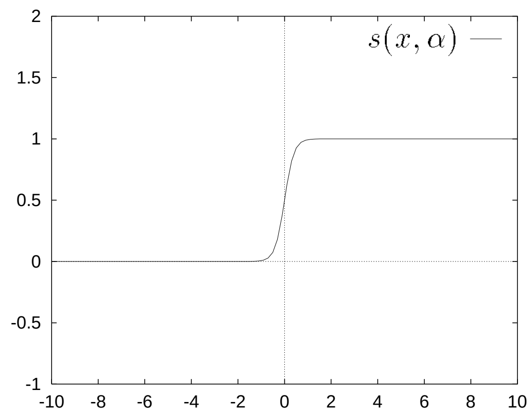

Figure 2.4: **The sigmoid function** $\mathbf{s(x,\alpha) = \frac{1}{1+e^{-\alpha x}}}$ **with** $\alpha = \mathbf{5}$

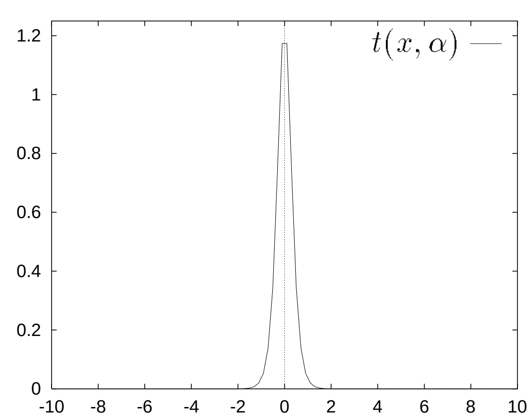

Figure 2.5: **The t function** $\mathbf{t(x,\alpha) = \frac{\alpha e^{-\alpha x}}{(1+e^{-\alpha x})^2}}$ **with** $\alpha = \mathbf{5}$

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

$$\hat{p}(x,\beta) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x^2}{2\beta} & \text{if } 0 \le x \le \beta \\ x - \frac{\beta}{2} & \text{if } x > \beta \end{cases} \tag{2.22}$$

This function can also be obtained by applying the Moreau-Yosida regularization [16, p.13] to the plus function.

Example 2.4.4 Zang Smooth Plus Function [54]

Let

$$d(x) = \begin{cases} 1 & \text{if } -\frac{1}{2} \le x \le \frac{1}{2} \\ 0 & \text{otherwise} \end{cases} \tag{2.23}$$

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

$$\hat{p}(x,\beta) = \begin{cases} 0 & \text{if } x < -\frac{\beta}{2} \\ \frac{1}{2\beta}(x + \frac{\beta}{2})^2 & \text{if } |x| \le \frac{\beta}{2} \\ x & \text{if } x > \frac{\beta}{2} \end{cases} \tag{2.24}$$

Note that in Examples 2.4.3 and 2.4.4 above, the density function $d(x)$ has compact support while the smooth function $\hat{p}(x,\beta)$ is only once continuously differentiable. In Examples 2.4.1 and 2.4.2, $d(x)$ 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:

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

In this thesis, we will assume that the density functions $d(x)$ which generate the smooth plus function $\hat{p}(x,\beta)$ by Definitions 2.1.1 and 2.2.1, satisfy the condition (A1) and (A2) defined in Proposition 2.2.2. Because of our favorable experience with the sigmoid function $s(x,\alpha)$ in neural network, we chose its integral $p(x,\alpha)$ for our numerical experiments. Further comparisons using different approximations to the plus function are left for future work.

Chapter 3

Inequalities

3.1 Linear Inequalities

We consider the following system of linear inequalities

$$Ax \leq b \tag{3.1}$$

where $A \in R^{m \times n}$ and $b \in R^m$ are given. Let X denote the solution set of (3.1). We shall employ two error functions for the linear inequalities (3.1) defined by:

$$f_1(x) = \|\hat{p}(Ax - b, \beta)\|_1 = l^T \hat{p}(Ax - b, \beta) \tag{3.2}$$

and

$$f_2(x) = \|\hat{p}(Ax - b, \beta)\|_2^2 \tag{3.3}$$

As an approximate solution to (3.1), we propose to solve

$$\min_{x \in R^n} f(x) \tag{3.4}$$

where $f(x)$ is either $f_1(x)$ or $f_2(x)$. Thus we replace the combinatorial problem of solving a system of linear inequalities by a deterministic unconstrained minimization of a differentiable function. The function $f(x)$ defined by (3.2) or (3.3) is convex on R^n. It is strictly convex on R^n if the matrix A has full column rank and the density function $d(x)$ satisfies (A3) of Proposition 2.3.2. We define the condition (B3) as follows:

(B3) supp { d(x) } is bounded.

Now we are ready to state the following theorem that characterizes the solvability of (3.4).

Theorem 3.1.1 Existence of Solution *Let $A \in R^{m \times n}$, $b \in R^m$, and let $f(x)$ be defined as in (3.2) or (3.3).*

1. *If the density function $d(x)$ satisfies (A3), the problem (3.4) has a solution if and only if $0 \neq Ax \leq 0$ has no solution.*

2. *If the density function $d(x)$ satisfies (B3), the problem (3.4) always has a solution.*

Proof We first prove 1.

($\Longrightarrow$) Suppose that $0 \neq Ax \leq 0$ has a solution x_0. Let $\overline{x}$ be the solution of problem (3.4). Then for any $\lambda > 0$, $A\overline{x} - b \neq A(\overline{x} + \lambda x_0) - b \leq A\overline{x} - b$. So $f(\overline{x} + \lambda x_0) < f(\overline{x})$, contradicts the fact that $\overline{x} \in \arg\min_{x \in R^n} f(x)$.

($\Longleftarrow$) Since $0 \neq Ax \leq 0$ has no solution, the set $Y = \{Ax - b | Ax - b \leq y_0 l\}$ is closed and bounded for any fixed $x_0 \in R^n$ and $y_0 \in R$ is large enough. By the continuity of the $\|\hat{p}(\cdot,\beta)\|$, there exists a point $\overline{y} \in Y$ such that $\|\hat{p}(\overline{y},\beta)\| \leq \|\hat{p}(y,\beta)\|, y \in Y$. Consequently, there exists a $\overline{x}$ such that $A\overline{x} - b = \overline{y}$. Hence problem (3.4) attains its minimum at $\overline{x}$.

Now we prove 2. Since $d(x)$ satisfies (B3), there exists d_x such that $d(x) = 0$ when $x \leq d_x$. By the definition of $\hat{p}(x,\beta)$, we know $\hat{p}(x,\beta) = 0$ when $x \leq d_x\beta$. The minimization problem (3.4) is equivalent to

$$\min_{y \in Y} f(y) \tag{3.5}$$

where $Y = \{Ax - b \mid d_x\beta l \leq Ax - b \leq Ax_0 - b\}$ for any fixed x_0. Since the set Y is closed and bounded, the minimum must be attained and therefore there exists $\bar{x}$ that is a minimizer of the problem (3.4). □

Now we will give conditions for the uniqueness for the solution of problem (3.4). Let $L_\mu(f) = \{x \in R^n | f(x) \leq \mu\}$ denote the level set of $f(x)$.

Theorem 3.1.2 *Let $A \in R^{m\times n}$, $b \in R^m$, and let $f(x)$ be defined as in (3.2) or (3.3). Assuming that the density function $d(x)$ satisfies (A3), the following are equivalent:*

1. *For some $\mu \in R$, $L_\mu(f)$ is compact and nonempty.*

2. *For all $\mu \in R$, $L_\mu(f)$ is compact.*

3. *$Ax \leq 0, x \neq 0$ has no solution.*

4. *Problem (3.4) has a unique solution.*

If, in addition, the solution set X of (3.1) is nonempty, then each of above is equivalent to

5. *The solution set X is bounded.*

Proof (**1** $\Longrightarrow$ **2**) Follows from the convexity and continuity of $f(x)$.

(**2** $\Longrightarrow$ **3**) If $Ax \leq 0$, $x \neq 0$ has a solution y, then for any $x \in R^n$, $x \in L_{f(x)}(f)$, and $\lambda \geq 0$, we have

$$A(x + \lambda y) - b \leq Ax - b. \tag{3.6}$$

Hence $f(x + \lambda y) \leq f(x)$. Therefore $x + \lambda y \in L_{f(x)}(f)$. This contradicts the compactness of $L_{f(x)}(f)$.

(**3** $\Longrightarrow$ **4**) Suppose problem (3.4) has no solution. Since $f(x)$ is continuous and bounded below by zero, there exists a sequence $\{x_k\}$ such that $\|x_k\| \to \infty$, as $k \to \infty$ and $f(x_k) \to \inf_{x \in R} f(x) \geq 0$. Hence there exists a $\mu > 0$ such that $f(x_k) \leq \mu$. Notice that the sequence $\{\frac{x_k}{\|x_k\|}\}$ has an accumulation point $\overline{x}$. Let $\{\frac{x_{k_i}}{\|x_{k_i}\|}\}$ denote the subsequence converging to $\overline{x}$. Since $f(x_{k_i}) \leq \mu$, for f defined by (3.2) or (3.3), we have that $Ax_{k_i} - b \leq p^{-1}(\mu, \beta)l$ or $Ax_{k_i} - b \leq p^{-1}(\sqrt{\mu}, \beta)l$ respectively. Dividing both sides by $\|x_{k_i}\|$, and letting $i \to \infty$, we get $A\overline{x} \leq 0$ and $\overline{x} \neq 0$, which contradicts 3. So problem (3.4) must have a solution. Since $Ax \leq 0$, $x \neq 0$ has no solution, the matrix A has full column rank. Therefore $f(x)$ is strictly convex, and the solution of (3.4) is unique.

($\mathbf{4} \Longrightarrow \mathbf{1}$) Let $x^* \in \arg\min_{x \in R^n} f(x)$. Then $L_{f(x^*)}(f) = \{x^*\}$ which is nonempty and compact.

($\mathbf{5} \Longleftrightarrow \mathbf{3}$) If, in addition, the solution set X of (3.1) is nonempty, the boundedness of X is equivalent to $Ax \leq 0, x \neq 0$ having no solution, which is 3. Hence that is equivalent to each of conditions 1 to 4. ∎

From the above two theorems, it is easy to see that if the matrix A is of full column rank and $d(x)$ satisfies (A3), the minimization problem (3.4) has a solution if and only if its solution is unique.

3.2 Accuracy of the Smooth Solution

We shall call a solution of the minimization problem (3.4) a smooth solution to the system of linear inequalities (3.1). The justification for this appellation is that such a solution gives an approximate solution of (3.1) and is obtained by minimizing a smooth function. First we will state an error bound lemma for the linear inequalities (3.1).

Lemma 3.2.1 Error bound [17] [25] *Suppose that the linear inequalities $Ax \leq b$ have a nonempty solution set X. For any x, there exists an $\overline{x} \in X$ such that*

$$\|x - \overline{x}\|_p \leq \mu_p(A)\|(Ax - b)_+\|_p, \tag{3.7}$$

for some positive constant $\mu_p(A)$ and any vector norm $\|\cdot\|_p$.

Since $\hat{p}(x,\beta)+D_2\beta$ majorizes x_+ (Proposition 2.3.1), $\hat{p}(Ax-b,\beta)+D_2\beta$ serves as an error bound also for any monotonic vector norm $\|\cdot\|_p$. We thus have that

$$\|x-\overline{x}\|_p \le \mu_p(A)\|\hat{p}(Ax-b,\beta)\|_p + c_p D_2\beta, \tag{3.8}$$

where $\overline{x}$ and $\mu_p(A)$ are the same as in Lemma 3.2.1 and c_p is a constant dependent on the norm $\|\cdot\|_p$ and dimension m . We now give an estimate of the error in satisfying the inequalities (3.1) by any exact solution of (3.4).

Theorem 3.2.2 *Let the solution set X of (3.1) be nonempty. Let $f(x)$ be the function defined in (3.2) or (3.3) and let $x^1(\beta)$ and $x^2(\beta)$ be solutions of (3.4) with $f=f_1$ and $f=f_2$ respectively. There exist $\bar{x}^1(\beta)$ and $\bar{x}^2(\beta)$, both in X, such that*

$$\|x^1(\beta)-\bar{x}^1(\beta)\|_1 \le \mu_1(A)m(D_1+D_2)\beta \tag{3.9}$$

and

$$\|x^2(\beta)-\bar{x}^2(\beta)\|_2 \le \mu_2(A)\sqrt{m}(D_1+D_2)\beta. \tag{3.10}$$

where $\mu_1(A)$ and $\mu_2(A)$ are the same as in Lemma 3.2.1.

Proof By Lemma 3.2.1, there exists an $\overline{x}\in X$, such that

$$\|x(\beta)-\overline{x}\|_p \le \mu_p(A)\|(Ax(\beta)-b)_+\|_p \le \mu_p(A)\|\hat{p}(Ax(\beta)-b,\beta)\|_p + c_p D_2\beta \tag{3.11}$$

Since $x(\beta) \in \arg\min_{x \in R^n} f(x)$ and (3) of Proposition 2.3.1, it follows that

$$f(x(\beta)) \leq f(\overline{x}) \leq \begin{cases} \|\hat{p}(0,\beta)\|_1 = mD_1\beta & \text{if } f = f_1 \\ \|\hat{p}(0,\beta)\|_2^2 = m(D_1\beta)^2 & \text{if } f = f_2 \end{cases} \tag{3.12}$$

Combining the above two inequalities, the conclusion follows. □

Therefore, by choosing β sufficiently large, $x(\beta)$ can approximate a solution of (3.1) to any desired accuracy. In the case when X has an interior point, for β small enough, the solution $x(\beta)$ of (3.4) solves the linear inequalities (3.1) exactly. We give this result below in Theorem 3.2.4 after establishing a preliminary lemma.

Lemma 3.2.3 *For positive numbers δ and γ, there exists a positive $\overline{\beta}$ such that for all $\beta \leq \overline{\beta}$, $\hat{p}(-\delta,\beta) \leq \gamma\beta$.*

Proof By Definition 2.2.1,

$$\hat{p}(-\delta,\beta) = -\delta\int_{-\infty}^{-\frac{\delta}{\beta}} d(t)dt - \beta\int_{-\infty}^{-\frac{\delta}{\beta}} td(t)dt \leq \beta\int_{-\infty}^{-\frac{\delta}{\beta}} |t|d(t)dt \tag{3.13}$$

Hence

$$\frac{\hat{p}(-\delta,\beta)}{\beta} \leq \int_{-\infty}^{-\frac{\delta}{\beta}} |t|d(t)dt \longrightarrow 0, \text{ as } \beta \longrightarrow 0^+ \tag{3.14}$$

Therefore there exists a positive $\bar{\beta}$ such that for all $\beta \leq \bar{\beta}$, $\hat{p}(-\delta,\beta) \leq \beta\gamma$. □

Theorem 3.2.4 *Suppose that the solution set X of (3.1) has a nonempty interior and the density function $d(x)$ satisfies (A3). Let $x(\beta)$ denote a solution of (3.4). Then there exists an $\overline{\beta} > 0$, such that for any $0 < \beta \leq \overline{\beta}$, $x(\beta) \in X$.*

Proof Since $d(x)$ satisfies (A3), $D_1 > 0$. By assumption, there exists an $\hat{x} \in R^n$ and $\delta > 0$ such that $A\hat{x} - b \leq -\delta l$. Let $f(x)$ be defined by (3.2), and $x(\beta)$ denote a solution of (3.4). Let $0 < \gamma < \frac{D_1}{m}$, by Lemma 3.2.3, there is an $\overline{\beta} > 0$ such that for all $0 < \beta \leq \overline{\beta}$, $\hat{p}(-\delta, \beta) \leq \gamma\beta$. Hence

$$\|\hat{p}(Ax(\beta) - b, \beta)\|_1 = f(x(\beta)) \leq f(\hat{x}) = \|\hat{p}(A\hat{x} - b, \beta)\|_1$$

$$\leq m\hat{p}(-\delta, \beta) \leq D_1\beta.$$

Hence $Ax(\beta) - b \leq p^{-1}(D_1\beta)l = 0$. Therefore $x(\beta) \in X$. For f defined as in (3.3), let $0 < \gamma < \frac{D_1}{\sqrt{m}}$, a similar argument follows. □

3.3 Inconsistent Linear Inequalities

For an inconsistent system of linear inequalities, our proposed method will still give a useful result in the form of a point $x(\beta) \in \arg\min_{x \in R^n} f(x)$ that approximately minimizes the infeasibility. In fact a multiple of the value of $f(x)$ bounds the distance of x to the set of minimizers of $\|(Ax - b)_+\|_1$ for the case when $f = f_1$, see [28]. If we let x^1 and x^2 denote solutions of the inconsistent system $Ax \leq b$ in the sense of least l_1-norm and l_2-norm respectively, and if we let $x^1(\beta)$ and $x^2(\beta)$ be minimizers of f as defined in (3.2) and (3.3) respectively, then we have that

$$\|(Ax^1(\beta) - b)_+\|_1 \leq \|(Ax^1 - b)_+\|_1 + m\beta(D_1 + D_2) \qquad (3.15)$$

and

$$\|(Ax^2(\beta) - b)_+\|_2 \le \|(Ax^2 - b)_+\|_2 + \sqrt{m}\beta(D_1 + D_2) \tag{3.16}$$

In addition, we can bound the distance between $x^1(\beta)$ or $x^2(\beta)$ to the solution set as follows.

Theorem 3.3.1 *Let $x^1(\beta)$ denote the solutions of (3.4) with f defined as (3.2). Let X_1 denote the solution sets of $\min_{x\in R^n} \|(Ax-b)_+\|_1$. There exists $\sigma_1(A,b) > 0$, such that for some $\bar{x}^1(\beta) \in X_1$,*

$$\|x^1(\beta) - \bar{x}^1(\beta)\|_1 \le \sigma_1(A,b)m(D_1 + D_2)\beta \tag{3.17}$$

Proof It is easy to see that $x^1 \in X_1$ is equivalent to x^1 being a solution of the following linear programming problem:

$$\begin{array}{ll} \underset{x,z}{\text{minimize}} & l^T z \\ \text{subject to} & \begin{array}{r} Ax - b \le z \\ z \ge 0 \end{array} \end{array} \tag{3.18}$$

Let u be a dual solution of the above LP, then $(x^1(\beta), (Ax^1(\beta) - b)_+, u)$ is an approximate dual pair. By Lemma 5.2.1 of [44], there exists a $\bar{x}^1(\beta) \in X_1$ such that

$$\begin{aligned} \|x^1(\beta) - \bar{x}^1(\beta)\|_1 &\le \sigma_1(A,b)(\|(Ax^1(\beta) - b)_+\|_1 - \|(A\bar{x}^1(\beta) - b)_+\|_1)_+ \\ &\le \sigma_1(A,b)m(D_1 + D_2)\beta. \end{aligned} \tag{3.19}$$

□

In the remainder of this section, we consider the function $p(x,\alpha)$ defined in (2.16) of Example 2.4.1. Let $x^2(\alpha)$ be the minimizer of $f_2(x)$ defined in (3.3) with the function $p(x,\alpha)$, We will bound the distance $x^2(\alpha)$ to the least square solution set after we state a lemma. The proof of the lemma will be given in Appendix A.

Lemma 3.3.2 *Let $x \in R$, $\alpha > 0$. Then*

(i) $p(x,\alpha)^2(1-p'(x,\alpha)) \le \frac{4}{\alpha^2 e^2}$;

(ii) $p'(x,\alpha)p(x,\alpha)(p(x,\alpha)-x) \le \frac{1}{\alpha(e-1)}$;

(iii) $p(x,\alpha)(1-p'(x,\alpha)) \le \frac{1}{\alpha e}$;

(iv) $0 \le min\{p(x,\alpha)-x, p'(x,\alpha)p(x,\alpha)\} \le \frac{\log 2}{\alpha}$.

Theorem 3.3.3 *Let $x^2(\alpha)$ be the minimizer of $f_2(x)$ defined in (3.3) with the function $p(x,\alpha)$ and X_2 denote the solution sets of* $\min_{x\in R^n} \|(Ax-b)_+\|_2$. *There exists $\sigma_2(A,b) > 0$, such that for some $\bar{x}^2(\alpha) \in X_2$,*

$$\|x^2(\alpha) - \bar{x}^2(\alpha)\|_2 \le 2\sigma_2(A,b)\left(\frac{\sqrt{m}}{\alpha} + \frac{\sqrt{m}}{\alpha^2}\right). \tag{3.20}$$

Proof It is easy to see that $x^2 \in X_2$ is equivalent to x^2 being a solution of the following quadratic programming:

$$\begin{array}{ll} \underset{x,z}{\text{minimize}} & \frac{1}{2}z^T z \\ \text{subject to} & Ax - b \le z \\ & z \ge 0 \end{array} \tag{3.21}$$

Let $\epsilon(\alpha) = p(A(x^2(\alpha) - b, \alpha)$, $u(\alpha) = diag(p'(Ax^2(\alpha) - b, \alpha))p(Ax^2(\alpha) - b, \alpha)$ and $v(\alpha) = 0$. Then $w = (x^2(\alpha), \epsilon(\alpha), u(\alpha), v(\alpha))$ is an approximate dual pair. By Lemma 5.3.2 of [44], we have that

$$\|x^2(\alpha) - \bar{x}^2(\alpha)\|_2 \leq 2\sigma_2(A, b)(r(w) + s(w)) \tag{3.22}$$

where

$$r(w) = \| \min\{ \begin{pmatrix} A^T u(\alpha) \\ \epsilon(\alpha) - u(\alpha) - v(\alpha) \\ -A^T u(\alpha) \\ -\epsilon(\alpha) + u(\alpha) + v(\alpha) \\ -Ax^2(\alpha) + b + \epsilon(\alpha) \\ \epsilon(\alpha) \end{pmatrix}, \begin{pmatrix} x(\alpha)_+ \\ \epsilon(\alpha)_+ \\ x(\alpha)_- \\ \epsilon(\alpha)_- \\ u(\alpha) \\ v(\alpha) \end{pmatrix} \}\|_2 \tag{3.23}$$

Notice $A^T u(\alpha) = A^T diag(p'(Ax^2(\alpha) - b, \alpha))p(Ax^2(\alpha) - b, \alpha) = 0$ and substitute the definition of $u(\alpha)$, $v(\alpha)$ and $\epsilon(\alpha)$, and let $y = Ax^2(\alpha) - b$ we have

$$r(w) = \| \min\{ \begin{pmatrix} 0 \\ diag(l - p'(y, \alpha))p(y, \alpha) \\ 0 \\ -diag(l - p'(y, \alpha))p(y, \alpha) \\ -y + p(y, \alpha) \\ p(y, \alpha) \end{pmatrix}, \begin{pmatrix} x(\alpha)_+ \\ p(y, \alpha) \\ x(\alpha)_- \\ 0 \\ diag(p'(y, \alpha))p(y, \alpha) \\ 0 \end{pmatrix} \}\|_2$$

$$= \| \begin{pmatrix} 0 \\ diag(l - p'(y,\alpha))p(y,\alpha) \\ 0 \\ -diag(l - p'(y,\alpha))p(y,\alpha) \\ \min\{p(y,\alpha) - y, diag(p'(y,\alpha))p(y,\alpha)\} \\ 0 \end{pmatrix} \|_2 \qquad (3.24)$$

By (iii) and (iv) of Lemma 3.3.2, we have

$$r(w) \leq \sqrt{m}\frac{2}{\alpha e} + \sqrt{m}\frac{\log 2}{\alpha} \qquad (3.25)$$

and

$$\begin{aligned} s(w) &= \|(Ax^2(\alpha) - b - \epsilon(\alpha), -\epsilon(\alpha), -u(\alpha), -v(\alpha))_+, \\ &\quad (x^2(\alpha)^T A^T u(\alpha) + \epsilon(\alpha)^T(\epsilon(\alpha) - u(\alpha) - v(\alpha)) \\ &\quad -u(\alpha)^T(Ax^2(\alpha) - b - \epsilon(\alpha)) + v(\alpha)^T\epsilon(\alpha))_+, \\ &\quad A^T u(\alpha), \epsilon(\alpha) - u(\alpha) - v(\alpha)\|_2 \\ &= \|0, 0, 0, 0, ((l - p'(y,\alpha))^T diag(p(y,\alpha))p(y,\alpha) \qquad (3.26) \\ &\quad -p(y,\alpha)^T diag(p'(y,\alpha))(p(y,\alpha) - y))_+, \\ &\quad 0, diag(l - p'(y,\alpha))p(y,\alpha)\|_2 \end{aligned}$$

By (i),(ii) and (iii) of Lemma 3.3.2, we have

$$s(w) \leq \sqrt{m}(\frac{4}{\alpha^2 e^2} + \frac{1}{\alpha(e-1)} + \frac{1}{\alpha e}) \qquad (3.27)$$

Combining (3.22), (3.25) and (3.27), we get the desired conclusion. □

Remark 3.3.4 *Suppose that the solution set of (3.1) is nonempty and bounded and the density function $d(x)$ satisfies (A3), then the level sets of $f(x)$ are compact and $f(x)$ is strongly convex on its level sets. Also note that $f(x)$ is differentiable as many times as we wish, hence we can apply any first or second order algorithm of unconstrained minimization to get linear, super-linear, or a local quadratic rate of convergence.*

3.4 Convex Inequalities

In this section, we consider system of convex inequalities

$$g(x) \leq 0 \tag{3.28}$$

where $g : R^n \to R^m$. We shall assume that $g(x)$ is convex and continuous on R^n. Let X be the solution set of (3.28).

In a similar manner to the case of linear inequalities, we consider the following functions:

$$f(x) = f_1(x) = \|\hat{p}(g(x), \beta)\|_1 = l^T \hat{p}(g(x), \beta) \tag{3.29}$$

and

$$f(x) = f_2(x) = \|\hat{p}(g(x), \beta)\|_2^2, \tag{3.30}$$

where $\hat{p}(\cdot, \beta)$ is defined in (2.16). Again we solve

$$\min_{x \in R^n} f(x) \tag{3.31}$$

to get an approximate solution to the convex inequalities (3.28). Let $rc(g)$ denote the recession cone of a proper convex function g, that is

$$rc(g) = \{y | \sup_{x \in dom\, g} (g(x+y) - g(x)) \leq 0\},$$

where $dom\, g$ is the domain of g [49]. Now we will state a condition under which (3.31) has a solution.

Theorem 3.4.1 *Let $g : R^n \to R^m$ be continuous and convex and let $f(x)$ be defined as in (3.29) or (3.30). The following are equivalent:*

1. *For some $\mu \in R$, $L_\mu(f)$ is compact and nonempty.*

2. *For all $\mu \in R$, $L_\mu(f)$ is compact.*

3. $\cap_{i=1}^{i=m} rc(g_i) = \{0\}$

4. *Problem (3.31) has nonempty compact solution set.*

   If, in addition, the solution set X of (3.28) is nonempty, then all above are equivalent to

5. *The solution set X is bounded.*

Proof (**1** $\Longrightarrow$ **2**) Follows from that $f(x)$ is closed proper convex.

(**2** $\Longrightarrow$ **3**) Suppose there exists a nonzero vector $y \in \cap_{i=1}^{i=m} rc(g_i)$. For arbitrary fixed $x \in R^n$, $y \in rcL_{g_i(x)}(g_i)$, $i = 1, \cdots, m$. Hence for any $\lambda > 0$,

$x + \lambda y \in L_{g_i(x)}(g_i)$, $g_i(x + \lambda y) \leq g_i(x)$ and $f(x + \lambda y) \leq f(x)$. Therefore $x + \lambda y \in L_{f(x)}(f)$. This contradicts the compactness of level sets.

(**3** $\Longrightarrow$ **4**) Suppose not, then there exists $\{x_k\}$ such that $\|x_k\| \to \infty$, as $k \to \infty$, and $f(x_k) \to \inf_{x \in R^n} f(x) \geq 0$. Therefore there exists a μ such that $L_\mu(f)$ is nonempty and unbounded, hence $rc(f) \neq \{0\}$. Let $0 \neq y \in rc(f)$, for f defined in (3.29). We have that $g_i(x + \lambda y) \leq f(x + \lambda y) \leq f(x)$. Hence $x + \lambda y \in L_{f(x)}(g_i), i = 1, \cdots, m$. Hence $0 \neq y \in \cap_{i=1}^{i=m} rc(g_i)$. This contradicts 3. Similarly, the case of f defined by (3.30) can be proved.

(**4** $\Longrightarrow$ **1**) Let $x^* \in \arg\min_{x \in R^n} f(x)$. Then $L_{f(x^*)}(f)$ which is nonempty and compact.

(**5** $\Longleftrightarrow$ **3**) If, in addition, solution set X of (3.28) is nonempty, X is bounded if and only if $\cap_{i=1}^{i=m} rc(g_i) = \{0\}$, which is 3. Hence condition 5 is equivalent to each of conditions 1 to 4. $\square$

Following are some results similar to those of the sections 3.1 and 3.2. We omit the proofs.

Theorem 3.4.2 *Suppose that the solution set X of (3.28) is nonempty. Let $f(x)$ be the function defined in (3.29) or (3.30) and let $x^1(\beta)$ and $x^2(\beta)$ be solutions of (3.31) with $f = f_1$ and $f = f_2$ respectively.*

(i) Let X be bounded and let g satisfy the Slater constraint qualification: $g(\hat{x}) < 0$ or let $g(x)$ be differentiable and satisfy the Slater and asymptotic constraint qualification [26]. Then there exist $\bar{x}^1(\beta)$ and $\bar{x}^2(\beta)$, both

in X, such that

$$\|x^1(\beta) - \bar{x}^1(\beta)\|_1 \leq mC_1(D_1 + D_2)\beta \tag{3.32}$$

and

$$\|x^2(\beta) - \bar{x}^2(\beta)\|_2 \leq \sqrt{m}C_2(D_1 + D_2)\beta, \tag{3.33}$$

where C_1 and C_2 are constants dependent on $g(x)$ [45, 26].

(ii) If the Slater constraint qualification is satisfied by $g(x) \leq 0$ and the density function $d(x)$ satisfies (A3), then there exists an $\overline{\beta} > 0$ such that for any $\beta \leq \overline{\beta}$, $x^1(\beta)$ and $x^2(\beta)$ solve the convex inequalities (3.28) exactly.

Note that $f(x)$ is convex, and is continuously differentiable as many times as $g(x)$ is. However, in the case when $d(x)$ satisfies (A3), $f(x)$ is not strictly convex in general as was the case for linear inequalities. In the following we will give a condition which ensures the strict convexity of $f(x)$.

Theorem 3.4.3 *Suppose that $g(x)$ is convex and twice continuously differentiable on R^n and the density function $d(x)$ satisfies (A3). Let*

$$\nabla g(x)y = 0,\ y \neq 0 \implies y^T \left(\sum_{i=1}^{i=m} \nabla^2 g_i(x)\right) y > 0 \tag{3.34}$$

for each $x \in R^n$. Then $f(x)$ is continuously twice differentiable, strictly convex on R^n and strongly convex on any bounded set.

We note the following simple conditions that ensure the satisfaction of condition (3.34)

1. For some i, $g_i(x)$ is strongly convex on R^n.

2. $\sum_{i=1}^{i=m} g_i(x)$ is strongly convex on R^n.

3. Let $I \subset \{1, \cdots, m\}$ denote the index set of linear inequalities, and $g_i(x) = a_i^T x - b_i, i \in I$. Then $\{a_i\}_{i \in I}$ has rank n.

3.5 Numerical Results

We now give a summary of our computational experience with the smooth algorithms described in this chapter. The smooth algorithms were implemented in C. The CPU times for all the algorithms do not include the time to input data. The time of MINOS 5.4 [33] is the execution time for subroutine M5SOLV and also does not include the input time. MINOS is a pivot-based solver, which employs sparsity by updating the basis using sparse linear algebra. The smooth algorithms employ sparsity by evaluating the function and gradients using sparse matrix computation.

For linear inequalities, we compared the smooth algorithm with MINOS as well as with the relaxation method of Motzkin and Schoenberg [32]. The relaxation method was implemented in C. All the algorithms for linear inequalities were run on a Sun SPARCstation 10. We used the truncated Newton algorithm [35] to solve the smooth unconstrained minimization problem. We started with $\alpha = 1000.0$ and increased it by a factor of 2 at each major iteration. The algorithms terminate when the infeasibilities are less than 1.0e-7. All problems were generated randomly. Problems of size up to 500 variables are dense while larger problems are sparse with density in the range of 2 percent to 5 percent. The linear inequalities generated were of the following form:

$$\begin{aligned} A_1 x &\le b_1 \\ A_2 x &\le b_2 \\ -r \le l^T(A_1 x - b_1) &\le r \end{aligned} \tag{3.35}$$

where the matrices A_1, A_2 and a solution vector x_0 were randomly generated and b_1 and b_2 were determined by $b_1 = A_1 x_0, b_2 = A_2 x_0 + c$, where the vector c was randomly generated such that about half of the entries were zeros and the others were random nonnegative real numbers. The size of the block A_1 is 1 percent of the total number of the inequalities for sparse problems, and 5 percent for dense problems. The number r was chosen to be 1.0e-5. Figures 3.1 to 3.3 depict the CPU times taken by the smooth, MINOS and relaxation algorithms. For the cases $m = 2n$ and $m = 4n$, the smooth algorithm is faster than the other two algorithms. For the case $m = n$, the smooth algorithm is faster than MINOS and comparable with the relaxation method.

For convex inequalities, we use the BFGS algorithm [8] to solve the unconstrained minimization problem for variables up to 150, and the limited memory BFGS algorithm [36] for larger problems. Starting with $\alpha = 5$, we increased α by a factor of 1.05 to 1.2 at each minor iteration. The algorithm terminates when the infeasibilities are less than 1.0e-7. We compared the smooth algorithm with MINOS. Both algorithms were run on a DECstation 3100. The problems generated were in the following form:

$$\begin{aligned} Ax &\leq b_1 \\ g(x) &\leq b_2 \end{aligned} \tag{3.36}$$

Here 90 percent of inequalities were linear and 10 percent were nonlinear. The nonlinear inequalities were of the form:

$$g_i(x) = e^{xMx+qx-c} + ax - b, \tag{3.37}$$

where M is a positive semidefinite matrix. All the matrices and the vector x_0 were randomly generated and b_1 and b_2 were determined by $b_1 = Ax_0 + c_1$ and $b_2 = g(x_0) + c_2$. The vectors c_1 and c_2 were randomly generated such that about half of the entries were zeros and the others were random nonnegative real numbers. Figure 3.4 depicts the ratio of CPU time taken by MINOS to the time taken by the smooth algorithm as a function of problem size n.

We conclude this section with Table 3.1, that shows the potential of our smoothing methods as indicated by the maximum speedup that was achieved over standard algorithms. This table shows that smoothing techniques can be very effective for solving linear and convex inequalities.

Problem	*Speedup*	*Over*	*Size*
Linear Inequalities	1142	MINOS	2000×1000
Convex Inequalities	580	MINOS	210×400

Table 3.1: **Maximum speedup over Minos for Inequalities**

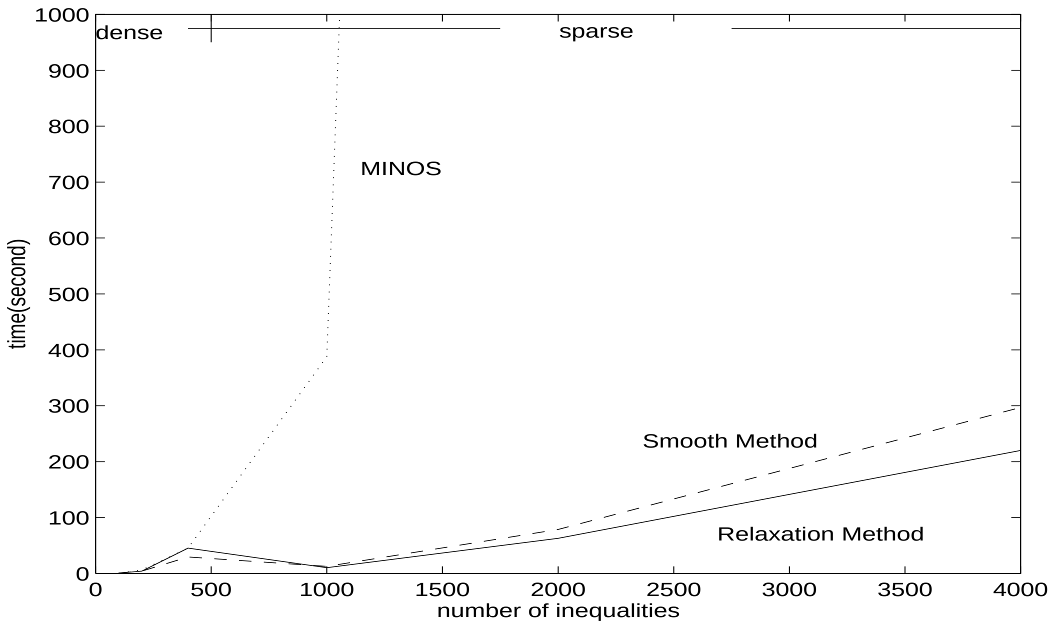

Figure 3.1: **Dense and Sparse Linear Inequalities: Comparison of solution times for MINOS, relaxation and smooth methods with m=n**

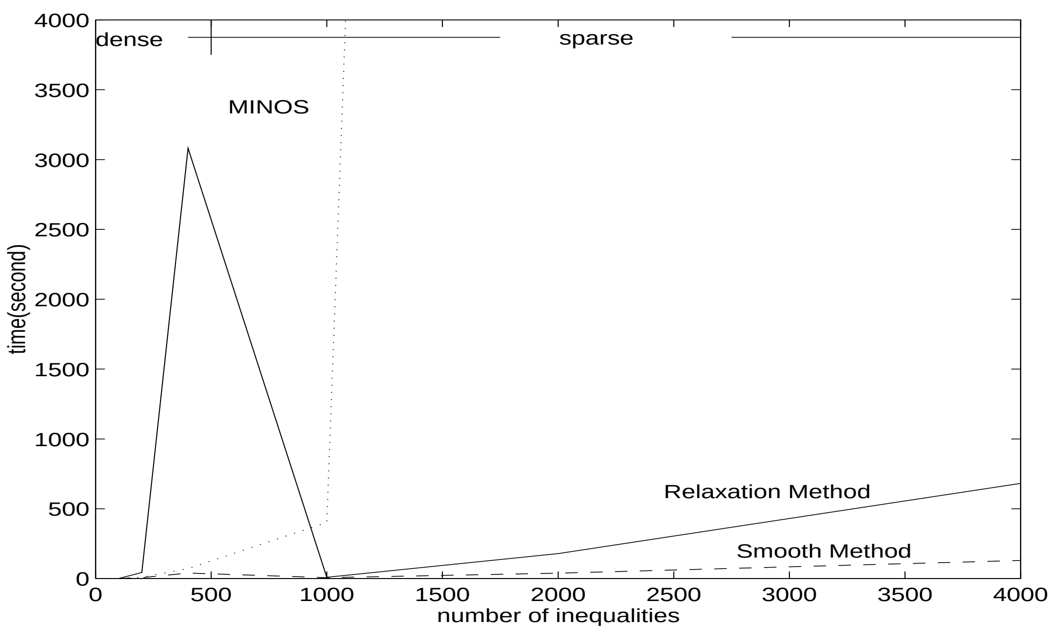

Figure 3.2: **Dense and Sparse Linear Inequalities: Comparison of solution times for MINOS, relaxation and smooth methods with m=2n**

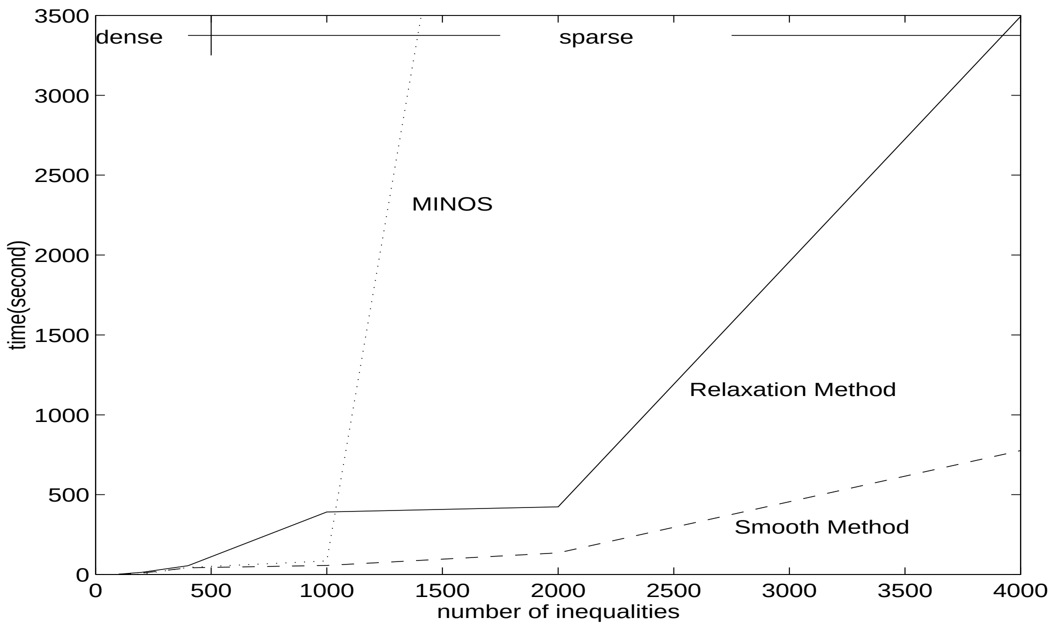

Figure 3.3: **Dense and Sparse Linear Inequalities: Comparison of solution times for MINOS, relaxation and smooth methods with m=4n**

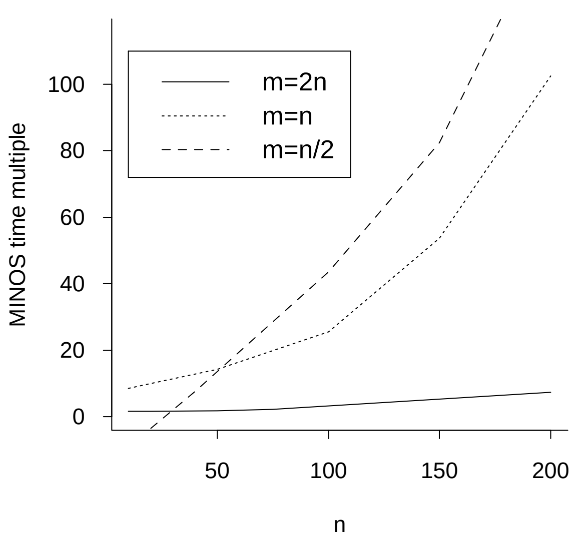

Figure 3.4: **Comparison of solution times for MINOS and the smooth method for convex inequalities**

Chapter 4

Complementarity Problems

4.1 The Nonlinear Complementarity Problem

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

$$0 \leq x \perp F(x) \geq 0 \tag{4.1}$$

Here $F(x)$ is a continuous differentiable function from R^n to R^n. By using the smooth function $\hat{p}(x, \beta)$ introduced in the Chapter 2, we consider the smooth nonlinear equation

$$R(x) = x - \hat{p}(x - F(x), \beta) = 0 \tag{4.2}$$

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

$$x = (x - F(x))_+ \tag{4.3}$$

We first show that a natural residual for the NCP is easily bounded by a corresponding residual for the nonlinear equation (4.2).

Lemma 4.1.1

$$\|x-(x-F(x))_+\|_p \leq \|x-\hat{p}(x-F(x),\beta)\|_p + \gamma_p \max\{D_1, D_2\}\beta, \quad p=1,2,\infty, \tag{4.4}$$

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 (2.3) and (2.4).*

Proof

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

□

The above result is also true for any monotone norm [37].

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

$$(F(x)-F(y))^T(x-y) \geq k\|x-y\|^2 \tag{4.5}$$

Since the NCP is strongly monotone, it has a unique solution [13]. The following error bound for the strongly monotone NCP is given as Theorem 3.2.1 in [44].

Lemma 4.1.2 *Let the NCP be strongly monotone and let $F(x)$ be Lipschitz continuous. Then for any $x \in R^n$*

$$\|x - \bar{x}\|_p \leq C_p \|x - (x - F(x))_+\|_p, \quad p = 1, 2, \infty, \tag{4.6}$$

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 4.1.1 and Lemma 4.1.2, it is easy to get the following lemma.

Lemma 4.1.3 *Let the NCP be strongly monotone and let $F(x)$ be Lipschitz continuous. Then for any $x \in R^n$*

$$\|x - \bar{x}\|_p \leq C_p(\|x - \hat{p}(x - F(x), \beta)\|_p + \gamma_p \max\{D_1, D_2\}\beta), \quad p = 1, 2, \infty, \tag{4.7}$$

where $\bar{x}$ and C_p are defined in Lemma 4.1.2, and γ_p and the constants D_1, D_2 are defined in Lemma 4.1.1.

Let the residual $f(x)$ of the nonlinear equation (4.2) be defined as follows

$$f(x) = \frac{1}{2} R(x)^T R(x) = \frac{1}{2} \|x - \hat{p}(x - F(x), \beta)\|_2^2 \tag{4.8}$$

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 (4.2), and hence by (4.7), x is an approximate solution of the NCP.

Proposition 4.1.4 *Suppose that $d(x)$ satisfies (A1) - (A3) and $\hat{p}(x,\beta)$ is defined by Definitions 2.1.1 or 2.2.1. 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 (4.2).*

Proof By definition,

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

By (2) of Proposition 2.3.2, 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, $R(x) = 0$, which means x satisfies (4.2). □

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 4.1.5 *Consider the strongly monotone NCP with Lipschitz continuous $F(x)$. Then $f(x)$ defined by (4.8) 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 $\|x_k - \hat{p}(x_k - F(x_k),\beta)\|_2 \le M$. Then by Lemma 4.1.3 ,

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

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. $\square$

We now show that, for the strongly monotone NCP with Lipschitz continuous function $F(x)$, the nonlinear equation (4.2) always has a unique solution.

Theorem 4.1.6 *Suppose that $d(x)$ satisfies (A1) - (A3), $\hat{p}(x, \beta)$ is defined by Definitions 2.1.1 or 2.2.1, $F(x)$ is strongly monotone and Lipschitz continuous. Then the nonlinear equation (4.2) has a unique solution.*

Proof By Proposition 4.1.5, 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 4.1.4 we get that x satisfies (4.2). If y is another solution of (4.2), 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 4.1.4, it follows that $x = y$. Therefore equation (4.2) has a unique solution. $\square$

Let $x(\beta)$ be a solution of (4.2). Then $x(\beta) = \hat{p}(x(\beta) - F(x(\beta)), \beta)$. By Lemma 4.1.3, we have the following theorem which bounds the distance between the solution $x(\beta)$ of (4.2) and the solution point of the original NCP (4.1).

Theorem 4.1.7 *Consider a strongly monotone NCP with Lipschitz continuous $F(x)$. Let $x(\beta)$ be a solution of (4.2). Then, for the solution $\bar{x}$ of the NCP (4.1), 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 4.1.2, γ_p and D_1, D_2 are constants defined in Lemma 4.1.1.

By the above result, we know that if β is sufficiently small, a solution of (4.2) can approximate the solution of NCP to any desired accuracy. Hence we can solve (4.2) to get an approximate solution of the NCP.

For the most part of this remaining section, we consider only the function $p(x, \alpha)$ defined in Example 2.4.1. 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 (4.2) that generates an approximate solution of the NCP. We claim that such x and y will approximately satisfy

$$0 \le x \perp y \ge 0.$$

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

Lemma 4.1.8 *Let $h(x)$ be defined as follows:*

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

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

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

(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 4.1.9. Note that as $\alpha \longrightarrow \infty$, the complementarity condition $0 \leq x \perp y \geq 0$ is exactly satisfied.

Proposition 4.1.9 *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

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

we have

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

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

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

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

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

Similarly,

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

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

$$y \geq 0 \Longleftrightarrow 1 + \delta - e^{-\alpha x} \leq 1 \Longleftrightarrow e^{-\alpha x} \geq \delta \tag{4.9}$$

and

$$y \leq 0 \Longleftrightarrow 1 + \delta - e^{-\alpha x} \geq 1 \Longleftrightarrow e^{-\alpha x} \leq \delta \tag{4.10}$$

Case 1 $0 < \delta < 1$.

If $y \geq 0$, then by (4.9),

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

Hence

$$\begin{aligned}(xy)_+ &\leq \max_{x\in[0,-\frac{\log\delta}{\alpha}]} -\frac{x}{\alpha}\log(1+\delta-e^{-\alpha x}) \\ &= \frac{1}{\alpha^2}\max_{y\in[0,-\log\delta]} h(y) \qquad (\text{let } y=\alpha x) \\ &\leq \frac{2}{\alpha^2} \qquad\qquad (\text{by (i) of Lemma 4.1.8})\end{aligned}$$

Otherwise $y \leq 0$, then

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

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

Case 2 $\delta \geq 1$.

If $y \geq 0$, then by (4.9),

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

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 4.1.8,

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

Case 3 $\delta = 0$.

In this case, we have $x \geq 0$ and $y \geq 0$. By using (iv) of Lemma 4.1.8,

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

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. $0 \leq x \perp F(x) \geq 0, \quad F(x) := 0$), the nonlinear equation (4.2) 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

$$-\delta_1 \beta l \leq x - \hat{p}(x - F(x), \beta) \leq \delta_2 \beta l, \tag{4.11}$$

always has a solution for $\beta > 0$. In particular, for the $p(x, \alpha)$ defined in Example 2.4.1 we have that for all $\delta_1 \geq \log 2$, the following system of inequalities

$$-\frac{\delta_1}{\alpha} l \leq x - p(x - F(x), \alpha) \leq 0, \tag{4.12}$$

always has a solution. Hence by Proposition 4.1.9, a solution of (4.12) will approximately satisfy the NCP condition

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

Consequently, Proposition 4.1.9 can be used to establish the following useful theorem.

Theorem 4.1.10 *Consider a solvable nonlinear complementarity problem (4.1). Let* $\delta_1 \geq \log 2$ *and* $\alpha > 0$. *There exists* x *satisfying (4.12), for which the NCP conditions are approximately satisfied as follows:*

$$(-x)_+ \leq \frac{\delta_1}{\alpha} l, \qquad (-F(x))_+ \leq \frac{\delta_1}{\alpha} l, \qquad (x^T F(x))_+ \leq \frac{nC(\delta_1)}{\alpha^2},$$

where $C(\delta_1)$ *is defined in Proposition 4.1.9.*

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 4.1.11 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 [8] and omit its proof.

Theorem 4.1.12 *Consider a solvable monotone nonlinear complementarity problem (4.1) with $F(x) \in LC^1_K(R^n)$. Then*

(1) The sequence $\{x_k\}$ defined in Algorithm 4.1.11 is well defined.

(2) Any accumulation point of the above sequence solves the nonlinear equation (4.2).

(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 (4.2), at a quadratic rate.

4.2 Linear Complementarity Problem

As a special case of the previous section, the linear complementarity problem is defined by $F(x) := Mx + q$. That is find an x in R^n such that

$$Mx + q \geq 0, x \geq 0, x^T(Mx + q) = 0 \tag{4.13}$$

where $M \in R^{n \times n}$ and $q \in R^n$. We shall denote this problem by $LCP(M, q)$.

We will show that under the assumption that M is a P_0 matrix, that is a matrix with nonnegative principal minors [7], all stationary points of (4.8) are solutions of (4.2). First we will state a simple lemma for P_0 matrices.

Lemma 4.2.1 *Suppose $M \in R^{n\times n}$ is a P_0 matrix. For any positive diagonal matrix $D = diag(d_1, \cdots, d_n)$, the matrix D+M is nonsingular.*

Proof Suppose $D + M$ is singular, then there exist a nonzero $x \in R^n$ such that $(D + M)x = 0$. Therefore $x_i(Mx)_i = x_i(-d_i x_i) = -d_i x_i^2$, which is negative whenever $x_i \neq 0$, $i = 1, \cdots, n$. This contradicts Theorem 3.4.2 of [7]. □

Theorem 4.2.2 *Suppose that $d(x)$ satisfies (A1) - (A3) and $\hat{p}(x, \beta)$ is defined by Definitions 2.1.1 or 2.2.1. Consider $LCP(M, q)$ with $M \in P_0$. Let $x(\alpha)$ be a stationary point of $\min_{x\in R^n} f(x)$, where $f(x)$ is defined by (4.8). Then $x(\alpha)$ is a solution of (4.2).*

This theorem can be proved easily by Lemma 4.2.1 and the formulation of $\nabla f(x)$ that is given in the proof of Proposition 4.1.4. Note that the class of P_0 matrices contains the classes of P matrices, positive semi-definite matrices and row-sufficient matrices [7]. For this class of matrices, if $f(x)$ defined by (4.8) has a stationary point, that point is also a solution of (4.2).

Now we establish the existence of a solution to (4.2) for $P_0 \cap R_0$ matrices. A matrix M is called an R_0 matrix if the only solution to $LCP(M, 0)$ is the zero vector [7].

Theorem 4.2.3 *Suppose that $d(x)$ satisfies (A1) - (A3) and $\hat{p}(x, \beta)$ is defined by Definitions 2.1.1 or 2.2.1. Consider $LCP(M, q)$ with $M \in P_0 \cap R_0$. The system of nonlinear equations (4.2) always has a solution.*

Proof First we will prove that the level set of $f(x)$ defined in (4.8) is compact if $M \in R_0$. Suppose not, then there exists a sequence $\{x_k\} \subset R^n$ and a positive number C such that $\|x_k\|_2 \to \infty$ as $k \to \infty$, and $\|x_k - \hat{p}(x_k - Mx_k - q, \beta)\|_2 \leq C$. Then

$$\begin{aligned} \|x_k - (x_k - Mx_k - q)_+\|_2 &\leq \|x_k - \hat{p}(x_k - Mx_k - q, \beta)\|_2 \\ &\quad + \|(x_k - Mx_k - q)_+ - \hat{p}(x_k - Mx_k - q, \beta)\|_2 \\ &\leq C + \sqrt{n}\max\{D_1, D_2\}\beta \end{aligned}$$

Note that there exists a subsequence $\{k_i\}$ such that $\{\frac{x_{k_i}}{\|x_{k_i}\|_2}\}$ converges to some $\overline{x} \in R^n$. Dividing both sides of the above inequality by $\|x_{k_i}\|_2$ and letting $i \to \infty$, we get $\|\overline{x} - (\overline{x} - M\overline{x})_+\|_2 = 0$. So $\overline{x}$ solves $LCP(M, 0)$ and $\overline{x} \neq 0$. This contradicts the fact that M is an R_0 matrix. Since $f(x)$ is continuously differentiable and the level sets of $f(x)$ are compact, $\min_{x \in R^n} f(x)$ must have a solution x, by the Theorem 4.2.2 x satisfies (4.2). $\square$

Now we give an error bound for the solution of the original $LCP(M, q)$ in terms of a solution to (4.2), but skip the proof.

Theorem 4.2.4 *Consider a solvable $LCP(M, q)$ with $M \in R_0$. Let $x(\beta)$ be a solution of (4.2). Then there exists an $\bar{x}(\beta)$ which is a solution of $LCP(M, q)$ such that*

$$\|x(\beta) - \bar{x}(\beta))\|_2 \leq \tau(M, q)\sqrt{n}\max\{D_1, D_2\}\beta,$$

where $\tau(M,q)$ is a constant, see Theorem 2.2.1 [44].

For this remaining section, we consider only the function $p(x,\alpha)$ defined in Example 2.4.1. The following theorem proves that if α is sufficiently large, then a solution of (4.2) can be purified to an exact solution of $LCP(M,q)$. In the following theorem, we assume that all the elements of matrix M and vector q are integers and $n \geq 2$. Let L be the size of $LCP(M,q)$ defined by [21]

$$L = \lfloor \sum_{i=1}^{i=n} \sum_{j=1}^{j=n} \log(|a_{ij}|) + \sum_{i=1}^{i=n} \log(|q_i|) + \log(n^2) \rfloor + 1.$$

Theorem 4.2.5 *Suppose that $LCP(M,q)$ is solvable. Let $x(\alpha)$ be a solution of (4.2) with $\alpha \geq \overline{\alpha} = \sqrt{n}2^L$. Then $x(\alpha)$ can be purified to a solution of $LCP(M,q)$.*

Proof Since $x(\alpha)$ is a solution of (4.2). we have $x(\alpha) > 0$, $Mx(\alpha)+q > 0$ and $x(\alpha) = -\frac{\log(l - e^{-\alpha(Mx(\alpha)+q)})}{\alpha}$. Hence

$$x(\alpha)(Mx(\alpha)+q) \overset{y=\alpha(Mx(\alpha)+q)}{=} -\frac{y\log(l-e^{-y})}{\alpha^2} < \frac{n}{\alpha^2} \leq 2^{-2L}$$

for all $\alpha \geq \overline{\alpha} = \sqrt{n}2^L$. By the purification procedure described in Appendix B [21], $x(\alpha)$ can be purified to a solution of $LCP(M,q)$. □

4.3 Relation to the Interior Point Method

In this section, we consider the NCP (4.1). Let the density function $d(x)$ satisfy (A1)-(A3) and $D_2 = 0$, and let let $\hat{p}(x,\beta)$ be defined by Definition 2.1.1 or 2.2.1.

If x solves the nonlinear equation (4.2) exactly, then

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

where the last inequality follows from the fact that $\hat{p}(\xi, \beta) > \xi$, (3) of Proposition 2.3.2. Hence

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

and x belongs to the interior of the feasible region $\{x \mid F(x) \geq 0, x \geq 0\}$ of the NCP. Hence an exact solution of (4.2) is interior to the feasible region. However the iterates of the smooth method, which are only approximate solutions of (4.2), are not necessarily feasible. For the function $\hat{p}$ defined in Example 2.4.2 [51, 19, 2] , the exact solution x of the equation (4.2) 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 [51], [2] and [19]. In [20], the relation between Smale's method [51] and the central path was pointed out. For our function $\hat{p}$ defined in Example 2.4.1, the solution x of the nonlinear equation (4.2), for different values of β, constitutes another path in the interior of the feasible region that satisfies:

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

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

Example 4.3.1 *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 4.1 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 (4.2). Figure 4.2 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.

4.4 The Mixed Complementarity Problem

The mixed complementarity problem (MCP) is defined as follows [9]:

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

$$\begin{aligned} F(x) - w + v &= 0 \\ 0 \leq x - l \perp w &\geq 0 \\ 0 \leq v \perp u - x &\geq 0 \end{aligned} \tag{4.14}$$

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

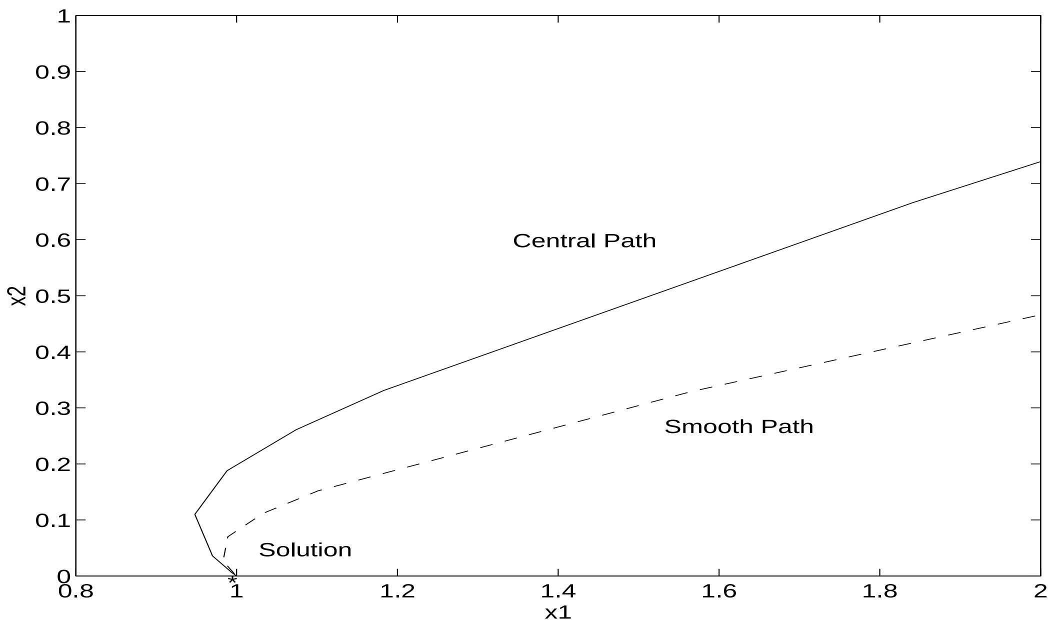

Figure 4.1: **Comparison of the interior smooth path generated by an exact solution of the smooth nonlinear equation (4.2) versus the central path for Example 4.3.1**

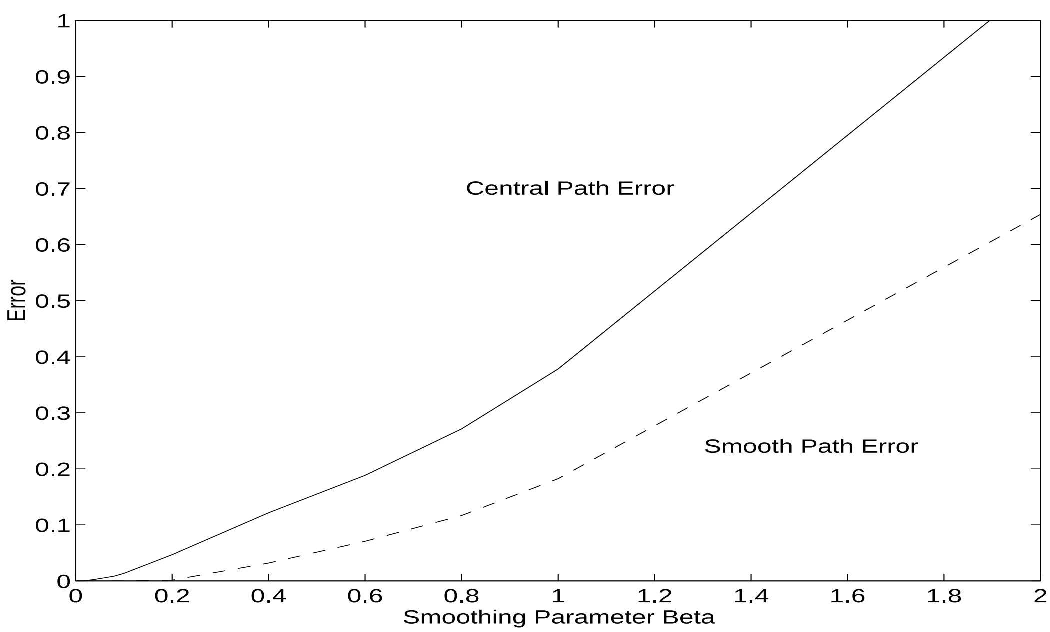

Figure 4.2: **Error comparison for the central path versus the smooth path for Example 4.3.1**

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 $u_i < \infty$:

$$x_i - u_i + \hat{p}(u_i - x_i + F_i(x), \beta) = 0 \tag{4.15}$$

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

$$R(x, w, v) = 0 \tag{4.16}$$

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

$$r(x, w, v) \tag{4.17}$$

Now we give a lemma that bounds the natural residual for MCP by the residual of equation (4.16), and vice versa. The proof is a simple application of the properties of the p function.

Lemma 4.4.1 *Let N be number of equations in (4.16) and $p(x,\alpha)$ as defined in Example 2.1. Then*

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

and

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

Let $f(x,w,v)$ be the residual function of the nonlinear equation (4.16) defined as follows

$$f(x,w,v) = \frac{1}{2}R(x,w,v)^T R(x,w,v) \tag{4.18}$$

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

Theorem 4.4.2 *Suppose that $d(x)$ satisfies (A1) - (A3) and $\hat{p}(x,\beta)$ is defined by Definition 2.1.1 or 2.2.1. Consider a solvable mixed complementarity problem (4.14) with monotone $F(x)$ and $l, u \in R^n$. The nonlinear equation (4.16) 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 $X = \{x | f(x,w,v) \le C\}$ 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 (2.3) and (2.4), 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 $k \longrightarrow \infty$,

$$f(x^k, w^k, v^k) \geq \frac{1}{2}((u_i - x_i^k - \hat{p}(u_i - x_i^k - v_i^k, \beta))^2 + (x_i^k - l_i - \hat{p}(x_i^k - l_i - w_i^k, \beta))^2)$$

$$\longrightarrow \frac{1}{2}((u_i - \bar{x}_i)^2 + (\bar{x}_i - l_i)^2) \geq \frac{1}{4}(u_i - l_i)^2 > \frac{1}{4} n \max\{D_1, D_2\}^2 \beta^2 = C$$

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 $0 < \hat{p}'(x - l - w, \beta)), \hat{p}'(u - x - v, \beta)) < l$. 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 2.4.1. The following theorem is a direct application of Proposition 4.1.9. that can be proved by applying the Proposition 4.1.9 to each equation.

Theorem 4.4.3 *Consider a solvable mixed complementarity problem (4.14). 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* $u_i < \infty$:

$$0 \leq x_i - u_i + \hat{p}(u_i - x_i + F_i(x), \beta) \leq \frac{\delta_1}{\alpha} \tag{4.19}$$

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 2.4.1, 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 4.1.9.

Now we state the smooth method for the mixed complementarity problem based on the Newton Algorithm 4.1.11 in which the smoothing parameter will be adjusted. In the algorithm, we adjust the smoothing parameter α in inverse proportion to the natural residual $r(x,w,v)$ of (4.17) for the MCP in the following way. Let N be the total number of nonlinear equations in (4.16) and (x,w,v) be current point. Let

$$\alpha(x,w,v) = \begin{cases} \frac{\sqrt{N}}{\|r(x,w,v)\|_2} & \text{if } \|r(x,w,v)\|_2 < \sqrt{N} \\ \sqrt{\frac{\sqrt{N}}{\|r(x,w,v)\|_2}} & \text{otherwise} \end{cases} \tag{4.20}$$

The following smooth algorithm generates an $\epsilon-$accurate solution for the MCP, in the sense that the natural residual $r(x,w,v)$ of (4.17) 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 4.4.4 *Let real numbers a and b satisfy*

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

then

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

where $C(\delta_1)$ is defined in Proposition 4.1.9.

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 4.4.5 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 (4.15), 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 (4.15). 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 4.4.6 *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 4.4.7 *Consider a solvable regular mixed complementarity problem (4.14) with with $F(x) \in LC^1_K(R^n)$. Then*

(1) The sequence $\{x_k, w_k, v_k\}$ defined in Algorithm 4.4.5 exists.

(2) Any accumulation point of the above sequence is an $\epsilon-$accurate solution of the MCP (4.14).

(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

$$\{(x,w,v) \mid \|r(x,w,v)\|_2 \le \|r(x_0,w_0,v_0)\|_2 + \frac{\nu}{\nu-1}\frac{2\sqrt{N}\log 2}{\alpha_0}\}$$

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 4.4.5, $\bar{\alpha}$ 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 (4.14). The other case is that $f(\bar{y}) > 0$ for $\bar{\alpha}$. Since $F \in LC^1_K(R^n)$, for a compact set S whose interior contains $\{y_{k_i}\}$ and $\bar{y}$, we have that $R(y) \in LC^1_{K_1}(S)$ for some K_1. By the Quadratic Bound Lemma [37, 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 $\lambda_{k_i} \geq \delta \frac{2(1-\sigma)}{K_1 K(S)}$, where δ is the constant used in the Armijo stepsize. Therefore

$$f(y_{k_i}) - f(y_{k_i} + \lambda_{k_i} d_{k_i}) \geq 2\sigma\delta \frac{1-\sigma}{K_1 K(S)} |\nabla f(y_{k_i})^T d(y_{k_i})|$$

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

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

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

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

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

Hence, if y is close enough to $\bar{y}$, the Newton step is accepted. According to the standard result of local quadratic convergence for the Newton Method, Theorem 5.2.1 in [8], the conclusion follows.

(4) Let $\alpha^i, i = 0, 1, \cdots$, be the sequence of different parameters α used in Algorithm 4.4.5. 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 4.4.1, 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 4.4.1,

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

For α^1 and y_{k_1-1}, by Lemma 4.4.1,

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

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

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

By Lemma 4.4.1,

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

Inductively, for α^i and y_{k_i-1},

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

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

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

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

Therefore, for all k we have

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

If the level set $\{y \mid \|r(y)\|_2 \leq \|r(y_0)\|_2 + \frac{\nu}{\nu-1}\frac{2\sqrt{N}\log 2}{\alpha_0}\}$ is compact, there exists an accumulation point. By (2) and (3), the whole sequence converges to an $\epsilon-$accurate solution of MCP (4.14). $\square$

4.5 Numerical Results

In the section, we first give numerical result for monotone linear complementarity problems. We used a Newton method with a safeguarded linear search to solve the smooth nonlinear equation (4.2). We compared the smooth method with Lemke's method [7]. The smooth algorithms were implemented in C. All problems were generated randomly. The matrix M was determined by: $M = AA^T + C$, where A is an $n \times r$ random matrix, r is a random number between 1 to n and C is a random skew-symmetric matrix. For the sparse problems, if the i-th row and column of M were all zeros, we replaced m_{ii} by a random nonnegative real number. The vectors x, w were randomly generated nonnegative vectors that were made complementary to each other by setting to zero a random 50 percent entries of x and the remaining 50 percent of w. The vector q was then defined by: $q = w - Mx$. We chose the parameter α inversely proportional to the 2-norm of the natural residual: $\| \min\{x, Mx+q\}\|_2$ [30]. The algorithm terminates when the infinity-norm of the natural residual is less than 1.0e-6. For dense problems, we compared the smooth algorithm with Lemke's method which was implemented in FORTRAN. For sparse problems with density between 0.012 and 0.15 percent, we compared the smooth algorithm with a sparse version of Lemke's method [9], which employs sparse basis updating techniques. We note that the SOR method of De Leone and Tork Roth [23] does not apply to this class of nonsymmetric LCP nor do other splitting methods described in [7]. In fact, the SOR method of [23] failed on all test problems.

Figures 4.3 and 4.4 show the CPU times for the smooth algorithm and Lemke's method. The smooth algorithm is always better than Lemke's method for both dense as well as the sparse problems.

Next 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 $|\nabla_i F_i(x_0)| \geq 100$. The details of implementing the smooth algorithm are given in Appendix C. For comparison, we also give the results for the PATH solver [9]. Both algorithms were run on a DECstation 5000/125. Among the 52 test problems, which includes all the problems attempted in [14], [40] and [9], 51 problems are from the MCPLIB [10], and one is the generalized von Thünen model from [40] and [52]. 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 [11]. A MINOS routine [33] 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 4.5.1 gives a simple description of the test problems [10].

The average CPU times taken by PATH solver and smooth algorithm for all small problems are depicted in Figure 4.5. Figures 4.6, 4.7 and 4.8 depict the CPU times for all remaining problems except the von Thünen model. We

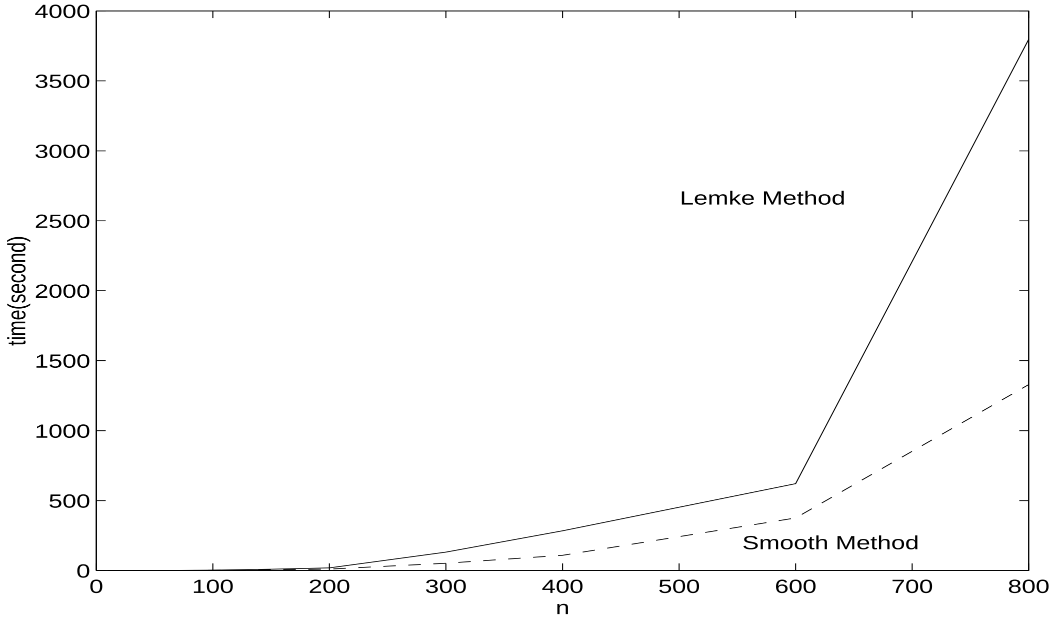

Figure 4.3: **Dense Monotone Linear Complementarity Problems: Comparison of solution times for Lemke and smooth methods. SOR method fails on all these problems.**

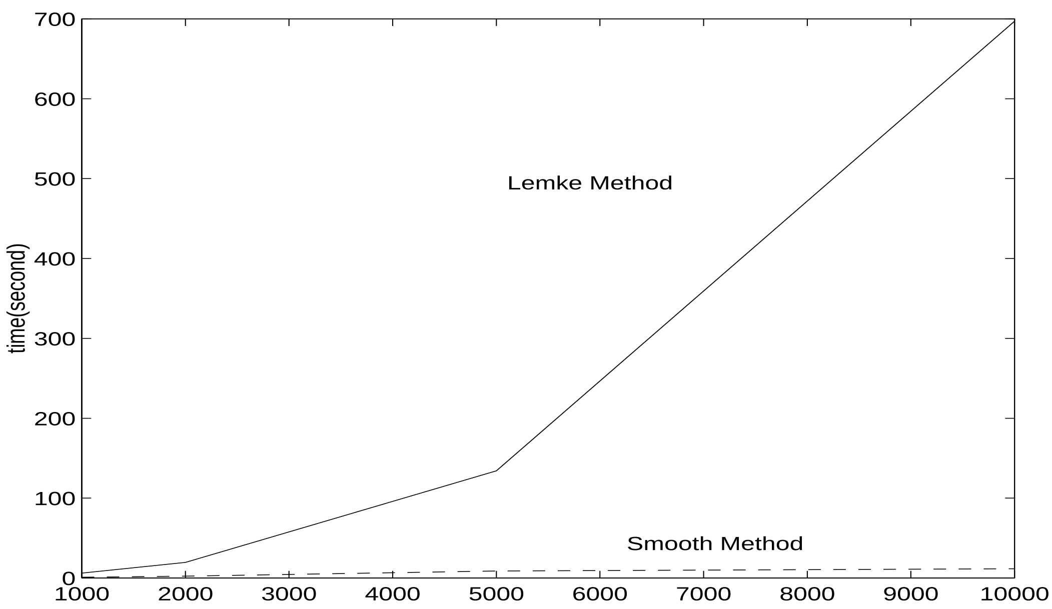

Figure 4.4: **Sparse Monotone Linear Complementarity Problems: Comparison of solution times for sparse Lemke and smooth methods. SOR method fails on all these problems.**

note that the PATH solver [9] is faster than Josephy's Newton method [18] and Rutherford's GAMS [11] mixed inequality and linear equation solver (MILES) [50] which is also Newton-based. Figures 4.6 to 4.8 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 [1] 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 4.9 to 4.12. 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 [40], 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 [40, 52]. Starting with the initial point provided by Jong-shi Pang [41], PATH failed to give a solution while SMOOTH obtained a solution with some small negative components in 27 iterations and 2.92 seconds. The values x and $F(x)$ returned by GAMS satisfied

$$\|\min\{x, F(x)\}\|_\infty \leq 8.50746e-07. \tag{4.21}$$

However, because GAMS returns 0.0e+00 for negative numbers raised to a noninteger power and 0.1e+05 for division by zero, the error residual given in

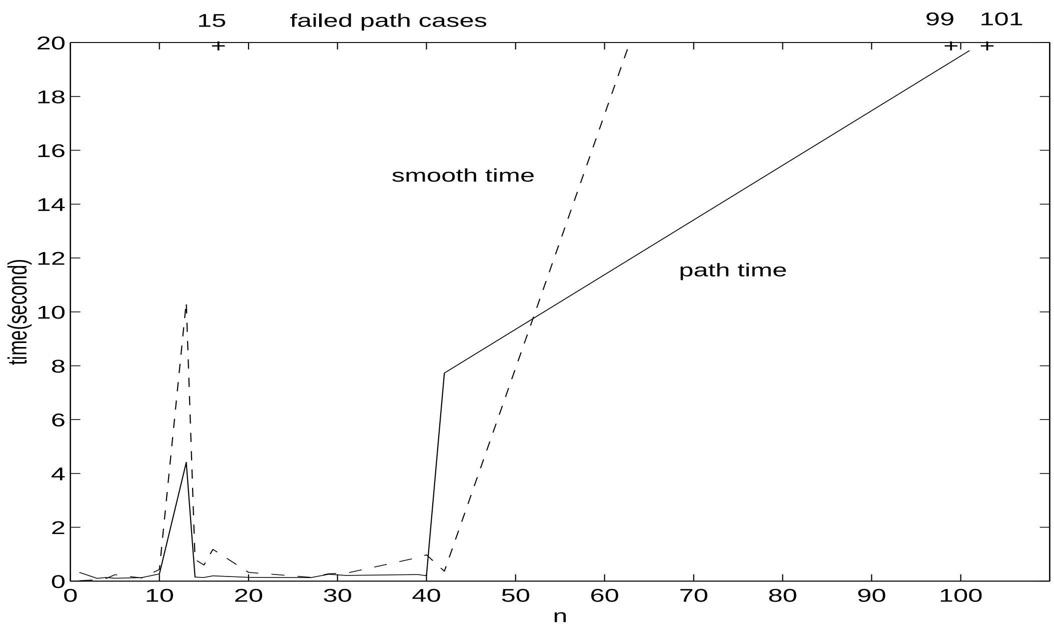

Figure 4.5: **Smooth versus PATH for Small MCP**

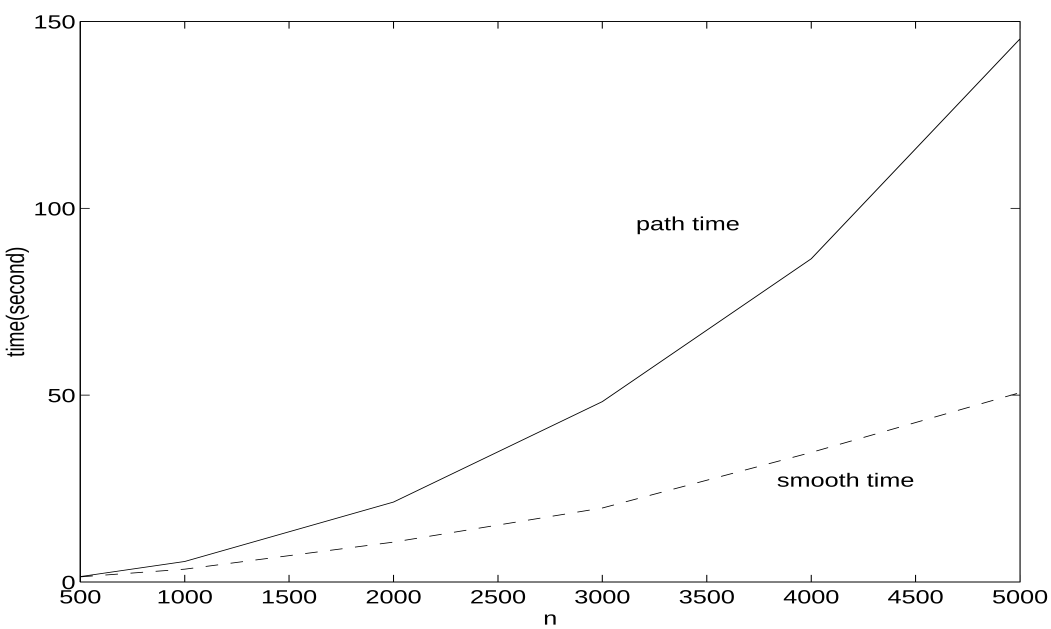

Figure 4.6: **Smooth versus PATH for Optimal Control Problem (bert_oc.gms)**

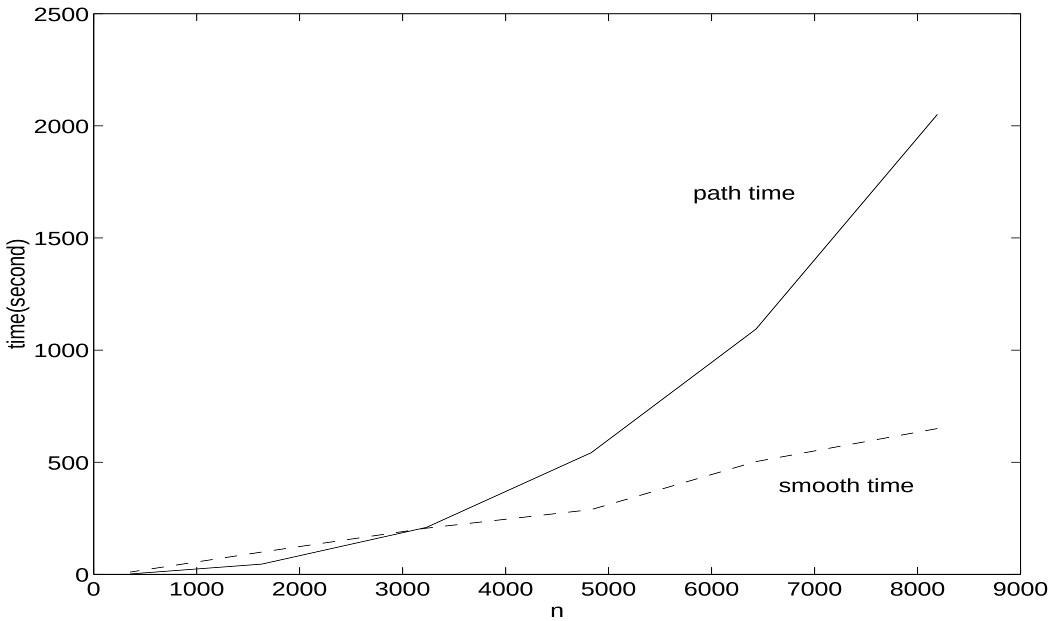

Figure 4.7: **Smooth versus PATH for Optimal Control Problem (opt_cont.gms)**

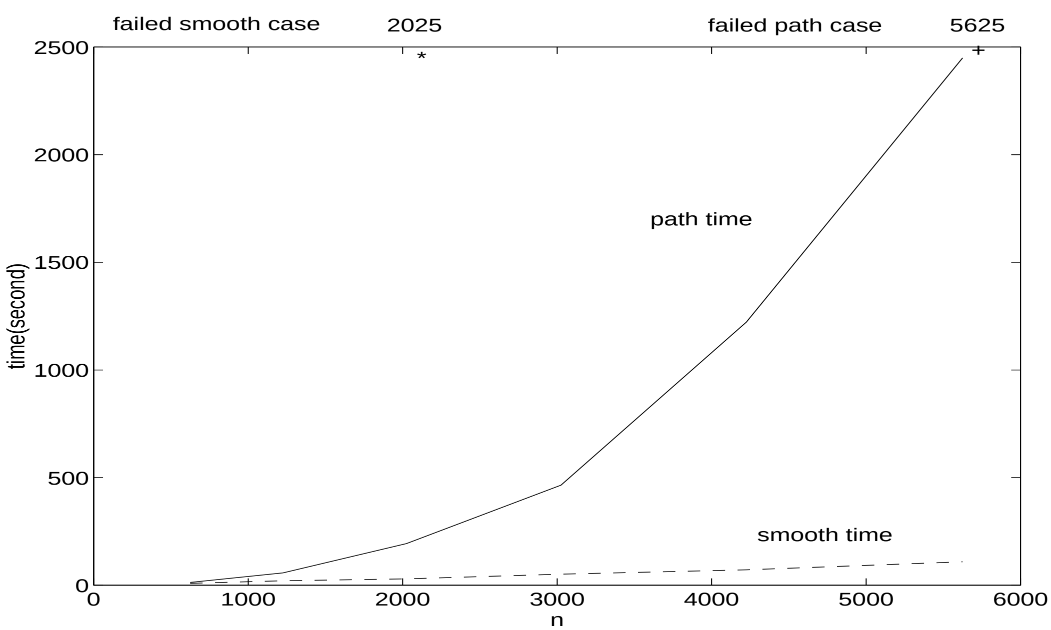

Figure 4.8: **Smooth versus PATH for Obstacle Problems**

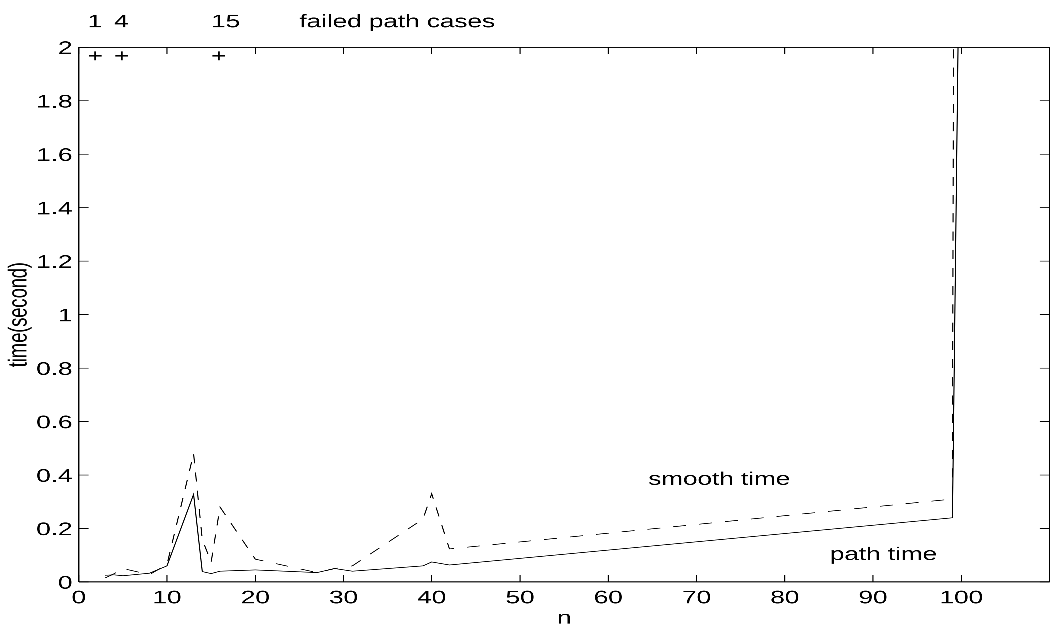

Figure 4.9: **Smooth with Newton preprocessor versus PATH 2.7 for Small MCP**

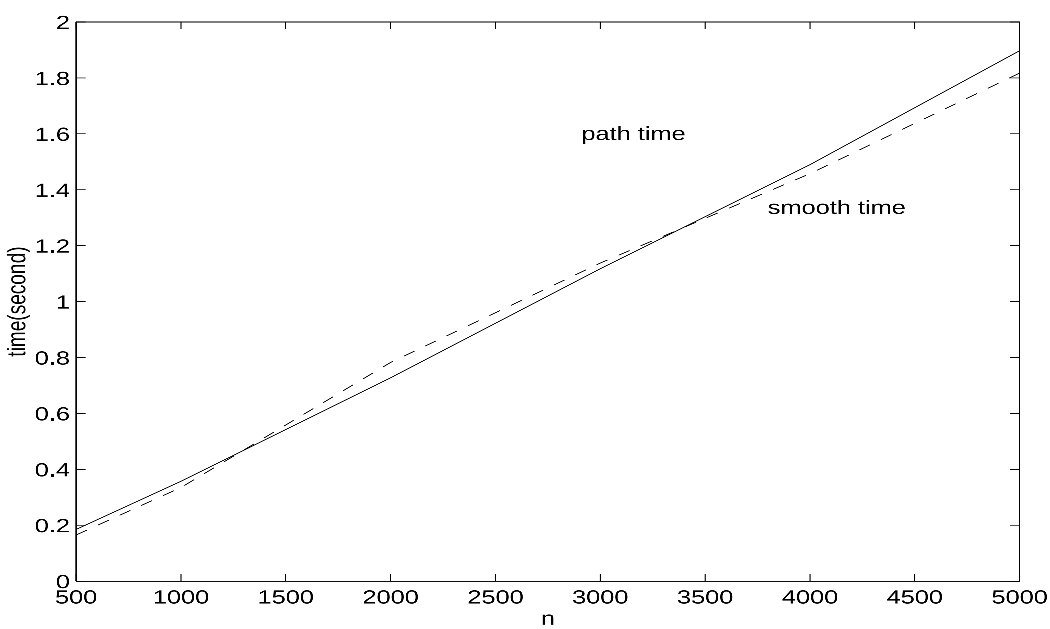

Figure 4.10: **Smooth with Newton preprocessor versus PATH 2.7 for Optimal Control Problem(bert_oc.gms)**

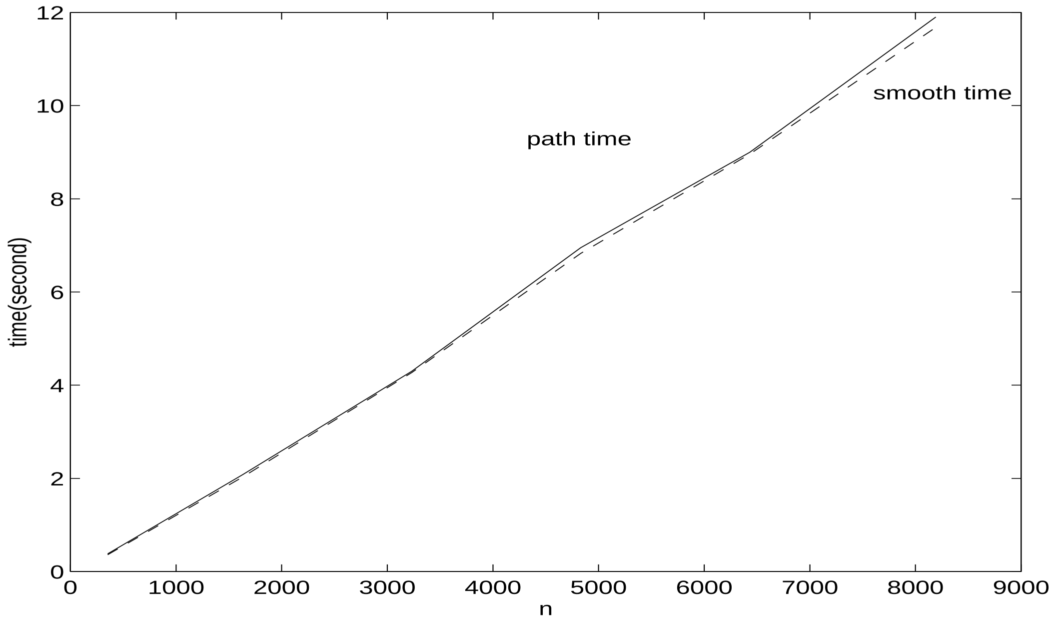

Figure 4.11: **Smooth with Newton preprocessor versus PATH 2.7 for Optimal Control Problem(opt_cont.gms)**

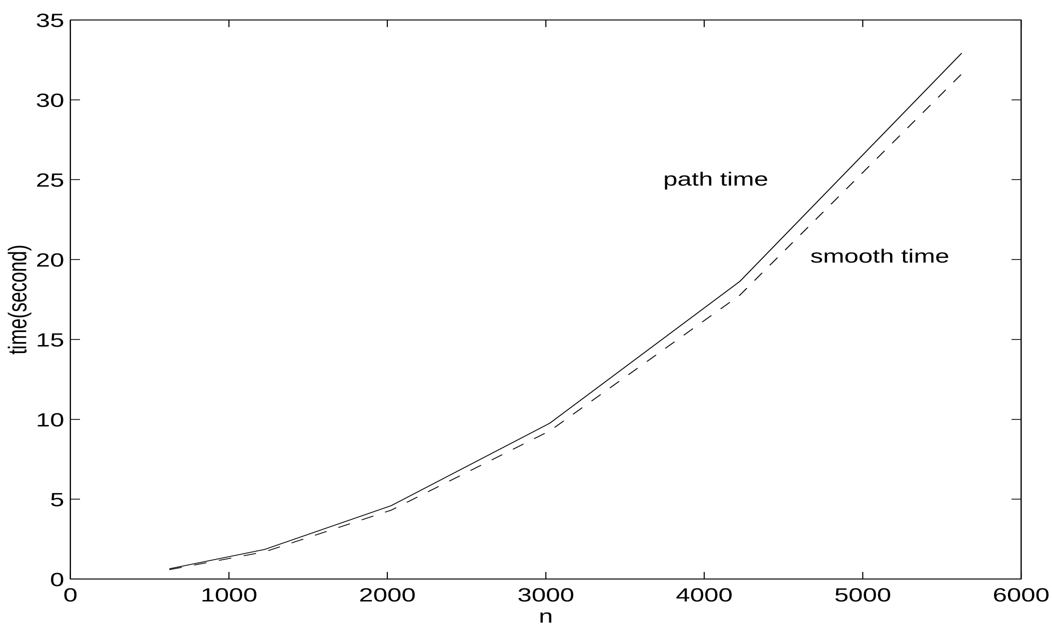

Figure 4.12: **Smooth with Newton preprocessor versus PATH 2.7 for Obstacle Problems**

(4.21) is incorrect. When the solution x obtained by our algorithm is passed to a FORTRAN routine, we get undefined numbers for some components of $F(x)$. When the negative components of x were set to 1.0e-9, it gave a solution accurate to 1.79539e-7. 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 set all variables to 1, as suggested by Michael C. Ferris; the second one is a starting point suggested in [52], while the third is the point suggested in [52] 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.

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.

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

Table 4.1: **MCPLIB Problems**

Chapter 5

Conclusion

A class of parametric smooth functions that approximate the fundamental plus function, $(x)_+ = \max\{0, x\}$, is obtained by twice integrating a probability density function. With this smooth function, most optimality conditions of mathematical programming as well as mixed and extended complementarity problems, can be approximated to any desired accuracy by a system of smooth nonlinear equations. This reformulation avoids the combinatorial difficulty inherent in all mathematical programming optimality conditions. Since the reformulated problem is smooth, many efficient algorithms, such as Newton and quasi-Newton methods, can be directly applied to the smooth system.

More specifically, by using the smooth approximation, linear and convex inequalities are converted into smooth, convex unconstrained minimization problems, the solution of which approximates the solution of the original problem to a high degree of accuracy for $\beta > 0$ sufficiently small. In the special case when a Slater constraint qualification is satisfied, an exact solution can be obtained for finite $\beta > 0$. Speedup over MINOS 5.4 was as high as 1142 times for linear inequalities of size 2000×1000, and 580 times for convex inequalities with 400 variables.

Linear complementarity problems can also be handled by converting them to a system of smooth nonlinear equations and solved by a quadratically convergent Newton method. For monotone LCP's with as many as 10,000 variables, the proposed approach was as much as 63 times faster than Lemke's method.

This approach can be extended to solve mixed complementarity problems. These problems include many classes of mathematical programs, such as nonlinear equations, nonlinear programming, nonlinear complementarity problems, variational inequalities and equilibrium problems. Existence of an arbitrarily accurate solution to the smooth nonlinear equation as well as the MCP, is established for sufficiently large value of a smoothing parameter α. An efficient smooth algorithm based on the Newton-Armijo approach with an adjusted smoothing parameter, is also given and its convergence is established. Very encouraging numerical testing results are given for 52 problems from the MCPLIB [10] which includes all the problems attempted in [14, 40]. These problems range in size of up to 8192 variables. These examples include the difficult von Thünen NCP model [40, 52] which is solved here to an accuracy of 1.0e-7.

For the nonlinear complementarity problem, an exact solution of our parametric smooth nonlinear equations traces a path in the interior of the feasible region when the smoothing parameter is varied. This path is different from the central path of the interior point method. However, unlike the interior point method, the iterates of the smooth method, which are only approximate solutions to the system of nonlinear equations, are not necessarily feasible for the original problem. The path generated by the smooth algorithm has been compared with the central path of the interior point method using a simple linear complementarity problem. This comparison shows that the smooth path gives a smaller error than the central path for corresponding values of the smoothing and penalty parameters.

We have shown that smoothing methods constitute a powerful computational tool for solving broad classes of optimization and related problems. Further study and application of these methods to problems in related areas such as machine learning [27, 3] appears to be promising.

Bibliography

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

[2] 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.

[3] Chunhui Chen and O. L. Mangasarian. Hybrid misclassification minimization. Technical Report 95-05, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin, February 1995. Advances in Computational Mathematics, submitted. Available from ftp://ftp.cs.wisc.edu/math-prog/tech-reports/95-05.ps.Z.

[4] 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.

[5] Chunhui Chen and O.L. Mangasarian. A class of smoothing functions for nonlinear and mixed complementarity problems. Technical Report 94-11, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin 53706, August 1994. Computational Optimization and Applications, to appear.

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

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

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

[9] 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.

[10] 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.

[11] 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.

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

[13] 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.

[14] 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.

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

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

[17] A. J. Hoffman. On approximate solutions of systems of linear inequalities. *Journal of Research of the National Bureau of Standards*, 49:263–265, 1952.

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

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

[20] 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.

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

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

[23] R. De Leone and M.A. Tork Roth. Massively parallel solution of quadratic programs via successive overrelaxation. *Concurrency: Practice and Experience*, 5:623–634, 1993.

[24] 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.

[25] O.L. Mangasarian. A condition number for linear inequalities and linear programs. In G. Bamberg and O. Opitz, editors, *Proceedings of 6. Symposium uber Operations Research, Augsburg, 7-9 September 1981*, pages 3–15, Konigstein, 1981. Verlagsgruppe Athenaum/Hain/Scriptor/Hanstein.

[26] O.L. Mangasarian. A condition number for differentiable convex inequalities. *Mathematics of Operations Research*, 10(2):175–179, 1985.

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

[28] O.L. Mangasarian. Error bounds for inconsistent linear inequalities and programs. *Operations Research Letters*, 15:187–192, 1994.

[29] 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.

[30] O.L. Mangasarian and J. Ren. New error bounds for the linear complementarity problem. Technical Report 1156, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin 53706, June 1993. Mathematical Programming, to appear.

[31] 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.

[32] T. S. Motzkin and I. J. Schoenberg. The relaxation method for linear inequalities. *Canadian Journal of Mathematics*, 6:393–404, 1954.

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

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

[35] S.G. Nash. Newton-type minimization via the Lanczos method. *SIAM Journal of Numerical Analysis*, 21:770–778, 1984.

[36] J. Nocedal. Theory of algorithms for unconstrained optimization. *Acta Numerica*, pages 199–242, 1992.

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

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

[39] 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.

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

[41] J.S. Pang. Private communication, October 3, 1994.

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

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

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

[45] S.M. Robinson. Application of error bounds for convex programming in a linear space. *SIAM Journal on Control and Optimization*, 13:271–273, 1975.

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

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

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

[49] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, New Jersey, 1970.

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

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

[52] 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.

[53] 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.

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

Appendix A

Proof of Lemma 3.3.2

(i) By definition,

$$p(x,\alpha)^2(1-p'(x,\alpha)) = (x+\frac{1}{\alpha}\log(1+e^{-\alpha x}))^2\frac{e^{-\alpha x}}{1+e^{-\alpha x}} = \frac{1}{\alpha^2}\log^2(1+e^{\alpha x})\frac{1}{1+e^{\alpha x}}$$

$$= \frac{1}{\alpha^2}\frac{\log^2 y}{y} \qquad (\text{Let } y = 1+e^{\alpha x} \geq 1)$$

Hence

$$p(x,\alpha)^2(1-p'(x,\alpha)) \leq \frac{1}{\alpha^2}\max_{x\geq 1} g(x), \tag{A.1}$$

where $g(x) = \frac{\log^2 x}{x}$. Since

$$\lim_{x\to+\infty} g(x) = 0, \;\; g(1) = 0$$

and $g'(x) = \frac{2\log x - \log^2 x}{x^2} = \frac{\log x}{x^2}(2-\log x)$, so $g'(x) = 0 \Longrightarrow x = 1$ or $x = e^2$. Hence

$$\max_{x\geq 1} g(x) = g(e^2) = \frac{4}{e^2}$$

Combining the above formula and (A.1), we get (i).

(ii) We know that

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

$$= \frac{1}{\alpha^2}\frac{1}{y}\log(\frac{y}{y-1})\log y \qquad (\text{Let } y = 1+e^{-\alpha x} \geq 1)$$

Hence

$$p'(x,\alpha)p(x,\alpha)(p(x,\alpha)-x) \leq \frac{1}{\alpha^2}\max_{x\geq 1} g(x), \tag{A.2}$$

where $g(x) = \frac{1}{x}\log(\frac{x}{x-1})\log x$. It is easy to see that

$$\lim_{x\to+\infty} g(x) = 0, \ \lim_{x\to 1} g(x) = 0$$

and $g'(x) = -\frac{1}{x^2}(\log(\frac{x}{x-1})\log x + \frac{\log x}{x-1} - \log(\frac{x}{x-1}))$. Every stationary point $x_0 > 1$ of g(x) must satisfies

$$\log(\frac{x}{x-1})(1-\log x) = \frac{\log x}{x-1}.$$

Since $x_0 > 1$, the first term and the right hand side of the above equality are positive. Hence $1-\log x$ must be positive also. Therefore, we get $x_0 \leq e$. So

$$\max_{x\geq 1} g(x) = \max_{1\leq x\leq e} g(x) \leq \max_{1\leq x\leq e} h(x) \tag{A.3}$$

where $h(x) = \log(\frac{x}{x-1})\log x$. We know

$$h(e) = 1-\log(e-1), \ \lim_{x\to 1} h(x) = 0$$

and $h'(x) = \frac{1}{x}(-\frac{\log x}{x-1} + \log(\frac{x}{x-1}))$, the stationary point x of $h(x)$ must satisfies $\log(\frac{x}{x-1}) = \frac{\log x}{x-1}$. Hence

$$\max_{1\leq x\leq e} h(x) \leq \max\{0, 1-\log(e-1), \max_{1\leq x\leq e} t(x)\} \tag{A.4}$$

where $t(x) = \frac{\log^2 x}{x-1}$. So $t'(x) = \frac{\log x}{(x-1)^2}(2(1-\frac{1}{x}) - \log x)$, it is easy to verify that $t'(x) > 0$ whenever $1 < x < e$. Hence

$$\max_{1\leq x\leq e} t(x) = \max\{\lim_{x\to 1} t(x), t(e)\} = \frac{1}{e-1}$$

Combining the above equality with (A.2), (A.3) and (A.4), and noticing that $1 - \log(e-1) \leq \frac{1}{e-1}$, we get the conclusion.

(iii) By definition,

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

$$= \frac{1}{\alpha}\frac{\log y}{y} \qquad (\text{Let } y = 1+e^{\alpha x} \geq 1)$$

Hence

$$p(x,\alpha)(1-p'(x,\alpha)) \leq \frac{1}{\alpha}\max_{x\geq 1} g(x), \tag{A.5}$$

where $g(x) = \frac{\log x}{x}$.

$$\lim_{x\to+\infty} g(x) = 0, \;\; g(1) = 0$$

and $g'(x) = \frac{1-\log x}{x^2}$. So the stationary point of $g(x)$ is $x = e$. Hence

$$\max_{x\geq 1} g(x) = g(e) = \frac{1}{e}$$

Combining the above equality and (A.5), conclusion follows.

(iv) By definition, we know

$$p(x, \alpha) - x \geq 0, \;\; p(x, \alpha)p'(x, \alpha) \geq 0.$$

In the case when $x \geq 0$, we have

$$\min\{p(x, \alpha) - x, p(x, \alpha)p'(x, \alpha)\} \leq \max_{x \geq 0} p(x, \alpha) - x = \frac{\log 2}{\alpha}$$

In the case when $x < 0$, we have

$$\min\{p(x, \alpha) - x, p(x, \alpha)p'(x, \alpha)\} \leq \max_{x < 0} p(x, \alpha)p'(x, \alpha) \leq \frac{\log 2}{\alpha}$$

Therefore, we have

$$0 \leq min\{p(x, \alpha) - x, p'(x, \alpha)p(x, \alpha)\} \leq \frac{\log 2}{\alpha}$$

□

Appendix B

Proof of Lemma 4.1.8

In order to prove Lemma 4.1.8, we need the following lemma.

Lemma B.0.1 *(i) Let $t(x) = xe^{-x}$, then*

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

(ii) Let

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

then

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

(iii) Let $h(x)$ be defined in Lemma 4.1.8, $-\log(1+\delta) \notin [a,b]$, then

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

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

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

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

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

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

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

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

(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),

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

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

□

Proof of Lemma 4.1.8

(i) If $0 < \delta < 1$, $-\log(1+\delta) \notin [0, -\log\delta]$. By (iii) of Lemma B.0.1 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

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

It is easy to get

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

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 B.0.1,

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

(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

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

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

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

Let ϵ approaches 0, we have

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

□

Appendix C

Implementation of the Smooth Algorithm 4.4.5

Here is the actual implementation of the smooth algorithm 4.4.5. In the following algorithm (x_k, w_k, v_k) is simply denoted by y_k.

Algorithm C.0.1 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 (4.15), 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 (4.15), 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.