Smoothing Methods for Convex Inequalities and Linear Complementarity Problems

Chunhui Chen* & O. L. Mangasarian*

November 1993/ Revised November 1994

Abstract

A smooth approximation $p(x, \alpha)$ to the plus function: $\max\{x, 0\}$, is obtained by integrating the sigmoid function $1/(1 + e^{-\alpha x})$, commonly used in neural networks. 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 α sufficiently large. In the special case when a Slater constraint qualification is satisfied, an exact solution can be obtained for finite α. 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 are converted into a system of smooth nonlinear equations and are 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.

Key Words: Smoothing, convex inequalities, linear complementarity.
Abbreviated Title: Smoothing Methods in Mathematical Programming

1 Introduction

The plus function

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

where x is a real number, plays a fundamental role in mathematical programming in the sense that 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 the unconstrained minimization problem: $\min_{x \in R^n} \|(g(x))_+\|$, where $(z)_+$ is taken to mean an m-vector of plus functions applied componentwise. Similarly the nonlinear complementarity problem

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

where $F : R^n \rightarrow R^n$ and $\perp$ denotes orthogonality, is equivalent to $x - (x - F(x))_+ = 0$. Our basic idea in this work is to approximate the plus function by a smooth parametric approximation $p(x, \alpha)$

*Computer Sciences Department, University of Wisconsin, 1210 West Dayton Street, Madison, WI 53706, email: *chunhui@cs.wisc.edu, olvi@cs.wisc.edu.* This material is based on research supported by Air Force Office of Scientific Research Grant F49620-94-1-0036 and National Science Foundation Grant CCR-9101801.

where α is a positive number. Note that $(x)_+ = \int_{-\infty}^{x} \sigma(y)dy$, where $\sigma(x)$ is the step function:

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

In the extensive neural network literature [10], the step function is very effectively approximated by the sigmoid function

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

See Figures 1 and 3. In this work we utilize the integral of the sigmoid function as an approximation to the plus function $(x)_+$ as follows:

$$(x)_+ \approx p(x, \alpha) = \int_{-\infty}^{x} s(y, \alpha)dy = x + \frac{1}{\alpha}\log(1 + e^{-\alpha x}) \tag{1}$$

For even moderate value of α, the function $p(x, \alpha)$ is a good approximation to the plus function, see Figures 2 and 4. As α approaches infinity, $p(x, \alpha)$ approaches x_+ from above and remains continuously differentiable as many times as we wish. Hence first order and second order gradient methods can be used to solve the reformulated problem involving the p function. We treat α as a parameter in the function $p(\cdot, \alpha)$. Hence when we say p' or p^{-1}, we mean the derivative or inverse of p with respect to the first variable with the parameter α fixed. We immediately note the following basic properties of $p(x, \alpha)$ that are easy to verify.

Lemma 1.1 ***Properties of*** $\mathbf{p(x, \alpha)}$, $\alpha > \mathbf{0}$

1. *$p(x, \alpha)$ is k-times continuously differentiable for any positive integer k, with $p'(x, \alpha) = \frac{1}{1+e^{-\alpha x}}$ and $p''(x, \alpha) = \frac{\alpha e^{-\alpha x}}{(1+e^{-\alpha x})^2}$.*
2. *$p(x, \alpha)$ is strictly convex and strictly increasing on R.*
3. *$p(x, \alpha) > x_+$, for all $x \in R$.*
4. *$\max_{x \in R}\{p(x, \alpha) - x_+\} = p(0, \alpha) = \frac{\log 2}{\alpha}$.*
5. *$\lim_{|x| \to \infty} p(x, \alpha) - x_+ = 0$, for all $\alpha > 0$.*
6. *$\lim_{\alpha \to \infty} p(x, \alpha) = x_+$, for all $x \in R$.*
7. *$p(x, \alpha) \in (0, \infty)$ for all $x \in R, \alpha > 0$. The inverse function p^{-1} is well defined for $x \in (0, \infty)$.*
8. *$p(x, \alpha) > p(x, \beta)$, for $\alpha < \beta$, $x \in R$.*

Smoothing techniques have been used for l_1-minimization problems [7] and in multi-commodity flows problem [17] using a linear quadratic smoothing function with encouraging numerical results.

We now summarize our results. In Section 2 we treat linear 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 α is large enough, the solution of the unconstrained problem can approximate the solution of original linear inequalities to any desired accuracy. For the case when the solution set of the linear inequalities has an interior point, an exact solution to the linear inequalities is obtained for sufficiently large but finite α. Even for the case when

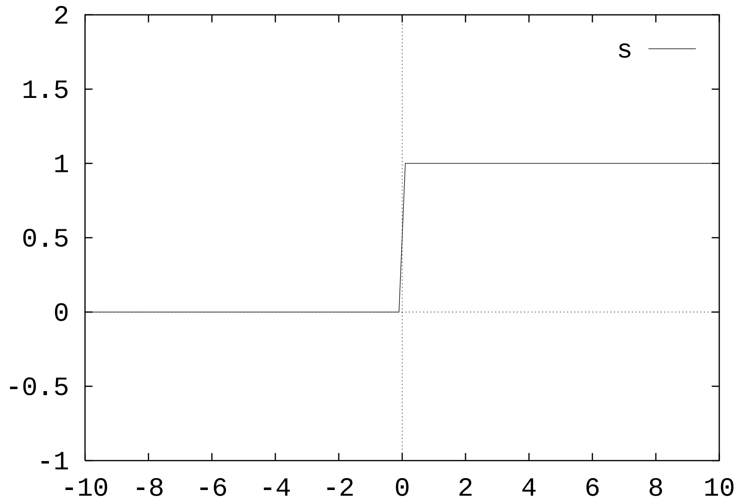

Figure 1: **The step function $\sigma(\mathbf{x}) = \mathbf{1}$ if $\mathbf{x} > \mathbf{0}$, $\mathbf{0}$ if $\mathbf{x} \leq \mathbf{0}$**

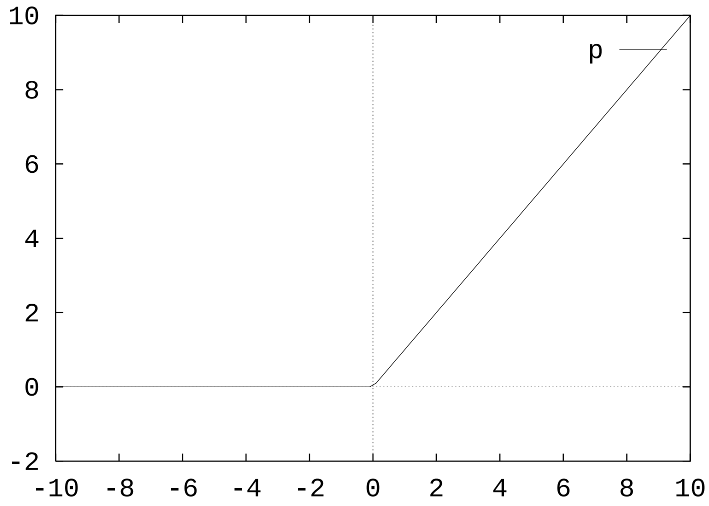

Figure 2: **The plus function $(\mathbf{x})_+ = \max\{\mathbf{x}, \mathbf{0}\}$**

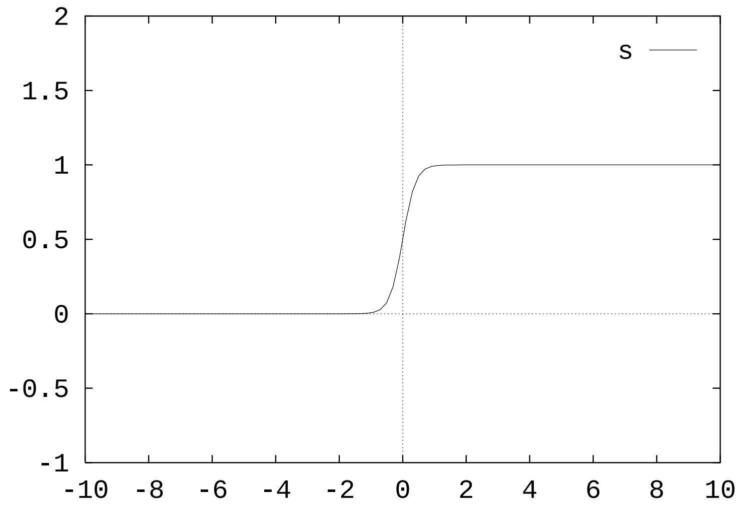

Figure 3: **The sigmoid function** $\mathbf{s(x,\alpha) = \frac{1}{1+e^{-\alpha x}}}$ **with** $\alpha = \mathbf{5}$

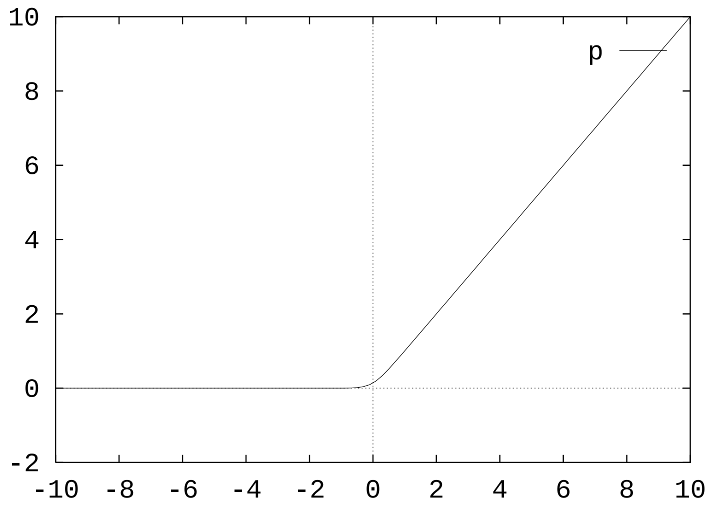

Figure 4: **The p function** $\mathbf{p(x,\alpha) = x + \frac{1}{\alpha}\log(1+e^{-\alpha x})}$ **with** $\alpha = \mathbf{5}$

the original linear inequalities are unsolvable, our method still gives an approximate solution in a least error sense. In Section 3 we treat convex inequalities in a similar manner to that of Section 2. In Section 4 we consider the linear complementarity problem (LCP) by solving a system of differentiable nonlinear equations. We give a sufficient condition for the existence of solution for the nonlinear equations and bound the distance between this solution and the solution set of LCP. Also, we prove that if α is large enough, we can get an exact solution to the LCP by a one-step purification of the approximate solution. In Section 5, some numerical results are given for linear and convex inequalities as well as monotone LCP's.

We now briefly describe our notation. For $f : R \to R$ and $x \in R^n$, the notation $f(x)$ will denote a vector in R^n defined by its components $(f(x))_i = f(x_i), i = 1, \cdots, n$. The notation $R^{m\times n}$ will denote the set of m-by-n real matrices. The notations $\mathbf{0}$ and $\mathbf{1}$ will represent vectors with components 0 and 1 of appropriate dimensions. The notation $\|\cdot\|_1$ and $\|\cdot\|_2$ will denote the l_1 and l_2 norms respectively. A monotonic norm is any norm such that $\|y\| \leq \|x\|$ whenever $|y| \leq |x|$. For a differentiable function $f\colon R^n \to R^m$, ∇f will denote the $m \times n$ Jacobian matrix of partial derivatives. For a twice differentiable function $f\colon R^n \to R$, $\nabla^2 f$ is the $n \times n$ Hessian. The set of minimizers of $\min_{x\in X} f(x)$ will be denoted by $\arg\min_{x\in X} f(x)$.

2 Linear Inequalities

We consider the following system of linear inequalities

$$Ax \leq b \tag{2}$$

where $A \in R^{m\times n}$ and $b \in R^m$ are given. Let X denote the solution set of (2). We shall employ two error functions for the linear inequalities (2) defined by:

$$f_1(x) = \|p(Ax - b, \alpha)\|_1 = \mathbf{1}^T p(Ax - b, \alpha) \tag{3}$$

and

$$f_2(x) = \|p(Ax - b, \alpha)\|_2^2 \tag{4}$$

As an approximate solution to (2), we propose to solve

$$\min_{x\in R^n} f(x) \tag{5}$$

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) or (4) is convex on R^n. It is strictly convex on R^n if the matrix A has full column rank. The following theorem characterizes the solvability of (5).

Theorem 2.1 ***Existence of Solution*** *Let $A \in R^{m\times n}$, $b \in R^m$, and let $f(x)$ be defined as in (3) or (4). Problem (5) has a solution if and only if $0 \neq Ax \leq 0$ has no solution.*

Proof ($\Longrightarrow$) Suppose that $0 \neq Ax \leq 0$ has a solution x_0. Let $\overline{x}$ be the solution of problem (5). 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 $Y = \{Ax - b | Ax - b \leq Ax_0 - b\}$ is closed and bounded for any fixed $x_0 \in R^n$. By the continuity of the $\|p(\cdot, \alpha)\|$, there exists a point $\overline{y} \in Y$ such that $\|p(\overline{y},\alpha)\| \leq \|p(y,\alpha)\|, y \in Y$. Consequently, there exists a $\overline{x}$ such that $A\overline{x} - b = \overline{y}$. Hence problem (5) attains its minimum at $\overline{x}$. □

Now we will give conditions for the uniqueness for the solution of problem (5). Let $L_\mu(f) = \{x \in R^n | f(x) \leq \mu\}$ denote the level set of $f(x)$.

Theorem 2.2 *Let $A \in R^{m\times n}$, $b \in R^m$, and let $f(x)$ be defined as in (3) or (4). 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 (5) has a unique solution.*

   If, in addition, the solution set X of (2) is nonempty, then each of above is equivalent to

5. *The solution set X is bounded.*

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

($\mathbf{2} \Longrightarrow \mathbf{3}$) 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

$$A(x + \lambda y) - b \leq Ax - b. \tag{6}$$

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

($\mathbf{3} \Longrightarrow \mathbf{4}$) Suppose problem (5) 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) or (4), we have that $Ax_{k_i} - b \leq p^{-1}(\mu,\alpha)\mathbf{1}$ or $Ax_{k_i} - b \leq p^{-1}(\sqrt{\mu},\alpha)\mathbf{1}$ 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 (5) 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 (5) 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 (2) 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, the minimization problem (5) has a solution if and only if its solution is unique.

In the following, we will prove that a solution of (5) gives an approximate solution of (2). First we will state an error bound lemma for linear inequalities.

Lemma 2.1 ***Error bound [4] [8]*** *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*

$$\|x - \overline{x}\|_\beta \leq \mu_\beta(A)\|(Ax - b)_+\|_\beta, \tag{7}$$

for some positive constant $\mu_\beta(A)$ and any vector norm $\|\cdot\|_\beta$.

Since $p(x,\alpha)$ majorizes x_+, $p(Ax-b,\alpha)$ serves as an error bound also for any monotonic vector norm $\|\cdot\|_\beta$. We thus have that

$$\|x-\overline{x}\|_\beta \le \mu_\beta(A)\|p(Ax-b,\alpha)\|_\beta, \tag{8}$$

where $\overline{x}$ and $\mu_\beta(A)$ are the same as in Lemma 2.1. We now give an estimate of the error in satisfying the inequalities (2) by any exact solution of (5).

Theorem 2.3 *Let the solution set X of (2) be nonempty. Let $f(x)$ be the function defined in (3) or (4) and let $x^1(\alpha)$ and $x^2(\alpha)$ be solutions of (5) with $f=f_1$ and $f=f_2$ respectively. There exist $\bar{x}^1(\alpha)$ and $\bar{x}^2(\alpha)$, both in X, such that*

$$\|x^1(\alpha)-\bar{x}^1(\alpha)\|_1 \le \frac{\log 2}{\alpha} m\mu_1(A)$$

and

$$\|x^2(\alpha)-\bar{x}^2(\alpha)\|_2 \le \frac{\log 2}{\alpha} \sqrt{m}\mu_2(A).$$

where $\mu_1(A)$ and $\mu_2(A)$ are the same as in Lemma 2.1.

Proof By Lemma 2.1, there exists an $\overline{x}\in X$, such that

$$\|x(\alpha)-\overline{x}\|_\beta \le \mu_\beta(A)\|(Ax(\alpha)-b)_+\|_\beta \le \mu_\beta(A)\|p(Ax(\alpha)-b,\alpha)\|_\beta \tag{9}$$

Since $x(\alpha)\in \arg\min_{x\in R^n} f(x)$, it follows that

$$f(x(\alpha)) \le f(\overline{x}) \le \begin{cases} \|p(\mathbf{0},\alpha)\|_1 = \frac{m\log 2}{\alpha} & \text{if } f=f_1 \\ \|p(\mathbf{0},\alpha)\|_2^2 = m(\frac{\log 2}{\alpha})^2 & \text{if } f=f_2 \end{cases}$$

Combining the above two inequalities, the conclusion follows. □

Therefore, by choosing α sufficiently large, $x(\alpha)$ can approximate a solution of (2) to any desired accuracy. In the case when X has an interior point, for α large enough, the solution $x(\alpha)$ of (5) solves the linear inequalities (2) exactly. We give this result below in Theorem 2.4 after establishing a preliminary lemma.

Lemma 2.2 *For positive numbers δ and β, there exists a positive $\overline{\alpha}$ such that for all $\alpha\ge\overline{\alpha}$, $p(-\delta,\alpha)\le\frac{\beta}{\alpha}$.*

Proof By definition, $\alpha p(-\delta,\alpha)=\alpha(-\delta+\log(1+e^{\alpha\delta})/\alpha)=\log(1+e^{-\alpha\delta})$. Since $\lim_{\alpha\to\infty}\log(1+e^{-\alpha\delta})=0$, there exists a positive number $\overline{\alpha}$ such that for all $\alpha\ge\overline{\alpha}$, $p(-\delta,\alpha)\le\frac{\beta}{\alpha}$. □

Theorem 2.4 *Suppose that the solution set X of (2) has a nonempty interior. Let $x(\alpha)$ denote a solution of (5). Then there exists an $\overline{\alpha}>0$, such that for any $\alpha\ge\overline{\alpha}$, $x(\alpha)\in X$.*

Proof By assumption, there exists an $\hat{x}\in R^n$ and $\delta>0$ such that $A\hat{x}-b\le-\delta e$. Let $f(x)$ be defined by (3), and $x(\alpha)$ denote a solution of (5). Let $\beta=\frac{\log 2}{m}$, by Lemma 2.2, there is an $\overline{\alpha}>0$ such that for all $\alpha\ge\overline{\alpha}$, $mp(-\delta,\alpha)\le\frac{\log 2}{\alpha}$. Hence

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

$$\leq mp(-\delta,\alpha) \leq \frac{\log 2}{\alpha}.$$

Hence $Ax(\alpha) - b \leq p^{-1}(\frac{\log 2}{\alpha})\mathbf{1} = \mathbf{0}$. Therefore $x(\alpha) \in X$. For f defined as in (4), let $\beta = \frac{\log 2}{\sqrt{m}}$, a similar argument follows. $\square$

For an inconsistent system of linear inequalities, our proposed method will still give a useful result in the form of a point $x(\alpha) \in \arg\min_{x\in R^n} f(x)$ that minimizes the infeasibility approximately. In fact a multiple of 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 [11]. 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 $x^1(\alpha)$ and $x^2(\alpha)$ be minimizers of f as defined in (3) and (4) respectively, then we have that

$$\|(Ax^1(\alpha) - b)_+\|_1 \leq \|(Ax^1 - b)_+\|_1 + \frac{m \log 2}{\alpha}$$

and

$$\|(Ax^2(\alpha) - b)_+\|_2 \leq \|(Ax^2 - b)_+\|_2 + \frac{\sqrt{m}\log 2}{\alpha}$$

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

Theorem 2.5 *Let $x^1(\alpha)$ and $x^2(\alpha)$ denote the solutions of (5) with f defined as (3) and (4). Let X_1 and X_2 denote the solution sets of $\min_{x\in R^n}\|(Ax-b)_+\|_1$ and $\min_{x\in R^n}\|(Ax-b)_+\|_2$ respectively. There exist $\sigma_1(A,b) > 0$ and $\sigma_2(A,b) > 0$, such that for some $\bar{x}^1(\alpha) \in X_1$ and $\bar{x}^2(\alpha) \in X_2$,*

$$\|x^1(\alpha) - \bar{x}^1(\alpha)\|_1 \leq \sigma_1(A,b)\frac{m\log 2}{\alpha}$$

and

$$\|x^2(\alpha) - \bar{x}^2(\alpha)\|_2 \leq 2\sigma_2(A,b)(\frac{\sqrt{m}}{\alpha} + \frac{\sqrt{m}}{\alpha^2}).$$

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:

$$\begin{array}{ll} \underset{x,z}{minimize} & \mathbf{1}^T z \\ \text{subject to} & \begin{array}{r} Ax - b \leq z \\ z \geq \mathbf{0} \end{array} \end{array} \tag{10}$$

Let u be a dual solution of the above LP, then $(x^1(\alpha), (Ax^1(\alpha)-b)_+, u)$ is an approximate dual pair. By Lemma 5.2.1 of [18], there exists a $\bar{x}^1(\alpha) \in X_1$ such that

$$\|x^1(\alpha) - \bar{x}^1(\alpha)\|_1 \leq \sigma_1(A,b)(\|(Ax^1(\alpha)-b)_+\|_1 - \|(A\bar{x}^1(\alpha)-b)_+\|_1)_+ \leq \sigma_1(A,b)\frac{m\log 2}{\alpha}.$$

Similarly, $x^2 \in X_2$ is equivalent to x^2 being a solution of the following quadratic programming problem:

$$\begin{array}{ll} \underset{x,z}{minimize} & \frac{1}{2}z^T z \\ \text{subject to} & \begin{array}{r} Ax - b \leq z \\ z \geq \mathbf{0} \end{array} \end{array} \tag{11}$$

Then $(x^2(\alpha), p(A(x^2(\alpha)-b),\alpha), diag(p'(Ax^2(\alpha)-b,\alpha))p(Ax^2(\alpha)-b,\alpha), \mathbf{0})$ is an approximate dual pair. By Lemma 5.3.2 of [18] and some tedious computation, we get the desired conclusion. $\square$

Remark 2.1 *Suppose that the solution set of (2) is nonempty and bounded, 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 Convex Inequalities

In this section, we consider system of convex inequalities

$$g(x) \leq 0 \tag{12}$$

where $g : R^n \rightarrow R^m$. We shall assume that $g(x)$ is convex and continuous on R^n. Let X be the solution set of (12).

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

$$f(x) = f_1(x) = \|p(g(x), \alpha)\|_1 = \mathbf{1}^T p(g(x), \alpha) \tag{13}$$

and

$$f(x) = f_2(x) = \|p(g(x), \alpha)\|_2^2, \tag{14}$$

where $p(., \alpha)$ is defined in (1). Again we solve

$$\min_{x \in R^n} f(x) \tag{15}$$

to get an approximate solution to the convex inequalities (12). 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 [20]. Now we will state a condition under which (15) has a solution.

Theorem 3.1 *Let $g : R^n \rightarrow R^m$ be continuous and convex and let $f(x)$ be defined as in (13) or (14). 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 (15) has nonempty compact solution set.*

   If, in addition, the solution set X of (12) 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 $\|x_k\| \rightarrow \infty$, as $k \rightarrow \infty$, and $f(x_k) \rightarrow \inf_{x \in R^n} f(x) \geq 0$. 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 (13). 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 (14) 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 (12) 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. □

Following are some results similar to those of Section 1. We omit the proofs.

Theorem 3.2 *Suppose that solution set X of (12) be nonempty. Let $f(x)$ be the function defined in (13) or (14) and let $x^1(\alpha)$ and $x^2(\alpha)$ be solutions of (15) with $f = f_1$ and $f = f_2$ respectively.*

(i) Let X be bounded and let g satisfy Slater constraint qualification: $g(\hat{x}) < 0$ or let $g(x)$ be differentiable and satisfy the Slater and asymptotic constraint qualification [9]. Then there exist $\bar{x}^1(\alpha)$ and $\bar{x}^2(\alpha)$, both in X, such that

$$\|x^1(\alpha) - \bar{x}^1(\alpha)\|_1 \leq \frac{\log 2}{\alpha} m C_1$$

and

$$\|x^2(\alpha) - \bar{x}^2(\alpha)\|_2 \leq \frac{\log 2}{\alpha} \sqrt{m} C_2,$$

where C_1 and C_2 are constants dependent on $g(x)$ [19, 9].

(ii) If the Slater constraint qualification is satisfied by $g(x) \leq 0$, then there exists an $\overline{\alpha} > 0$ such that for any $\alpha \geq \overline{\alpha}$, $x^1(\alpha)$ and $x^2(\alpha)$ solve the convex inequalities (12) exactly.

Note that $f(x)$ is convex, and is continuously differentiable as many times as $g(x)$ is. However, $f(x)$ 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 3.3 *Suppose that $g(x)$ is convex and twice continuously differentiable on R^n. Let*

$$\nabla g(x) y = 0,\ y \neq 0 \implies y^T \left(\sum_{i=1}^{i=m} \nabla^2 g_i(x)\right) y > 0 \tag{16}$$

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 (16)

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.

4 The Linear Complementarity Problem

Consider the linear complementarity problem of finding an x in R^n such that

$$Mx + q \geq 0, x \geq 0, x^T(Mx + q) = 0 \tag{17}$$

where $M \in R^{n \times n}$ and $q \in R^n$. We shall denote this problem by $LCP(M, q)$. It is very easy to see that x is a solution of $LCP(M, q)$ if and only if x is a solution of the following system

$$x = (x - Mx - q)_+ \tag{18}$$

If we introduce the $p(\cdot, \alpha)$ function as an approximation to the plus function, then we have the smooth system

$$x = p(x - Mx - q, \alpha) \tag{19}$$

as an approximation to (18). We begin with a simple preliminary result.

Lemma 4.1 *The following are equivalent:*

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

(ii) $e^{-\alpha x} + e^{-\alpha y} = 1$

(iii) $x = -\frac{\log(1-e^{-\alpha y})}{\alpha}$.

Proof

$$\begin{aligned}
& x = p(x-y,\alpha) = x - y + \tfrac{\log(1+e^{-\alpha(x-y)})}{\alpha} \\
\iff & \alpha y = \log(1+e^{\alpha y - \alpha x}) \\
\iff & e^{\alpha y} = 1 + e^{\alpha y - \alpha x} \\
\iff & e^{-\alpha x} + e^{-\alpha y} = 1 \\
\iff & e^{-\alpha x} = 1 - e^{-\alpha y} \\
\iff & x = -\tfrac{\log(1-e^{-\alpha y})}{\alpha}.
\end{aligned}$$

□

By using (19) and the above lemma, we know that x solves (19) if and only if x solves the following system of nonlinear equations.

$$e^{-\alpha x} + e^{-\alpha(Mx+q)} = \mathbf{1} \tag{20}$$

This is an interesting symmetric reformulation of an approximate solution to $LCP(M,q)$. Let

$$f(x) = \frac{1}{2}\|e^{-\alpha x} + e^{-\alpha(Mx+q)} - \mathbf{1}\|_2^2 \tag{21}$$

We will show that under the assumption that M is a P_0 matrix, that is a matrix with nonnegative minors [1], then all the stationary points of (21) are solutions of (20). First we will state a simple lemma for P_0 matrices.

Lemma 4.2 *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 $x_i(Mx)_i = x_i(-d_ix_i) = -d_ix_i^2$, which is negative whenever $x_i \neq 0$, $i = 1,\cdots,n$. This contradicts Theorem 3.4.2 of [1]. □

Theorem 4.1 *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 (21). Then $x(\alpha)$ is a solution of (20).*

Proof For $f(x)$ defined in (21),

$$\begin{aligned}
\nabla f(x) &= -\alpha(diag(e^{-\alpha x}) + M^T diag(e^{-\alpha(Mx+q)}))(e^{-\alpha x} + e^{-\alpha(Mx+q)} - \mathbf{1}) \\
&= -\alpha(diag(e^{\alpha(Mx+q)-\alpha x}) + M^T)diag(e^{-\alpha(Mx+q)})(e^{-\alpha x} + e^{-\alpha(Mx+q)} - \mathbf{1}).
\end{aligned}$$

Since $diag(e^{\alpha(Mx+q)-\alpha x})$ and $diag(e^{-\alpha(Mx+q)})$ are positive diagonal matrices, it follows by by Lemma 4.2 that $x(\alpha)$ is a solution of (20). □

Note that the class of P_0 matrices contains the classes of P matrices, positive semi-definite matrices and row-sufficient matrices [1]. For this class of matrices, if $f(x)$ defined by (21) has a stationary point, that point is also a solution of (20). Now we establish the existence of a solution to (20) 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 [1].

Theorem 4.2 *Consider $LCP(M,q)$ with $M \in P_0 \cap R_0$. The system of nonlinear equations (20) always has a solution.*

Proof Let $f(x) = \frac{1}{2}\|x - p(x - Mx - q, \alpha)\|_2^2$. First we will prove that the level set of $f(x)$ 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 $\|x_k - p(x_k - Mx_k - q, \alpha)\|_2 \leq C$. Then

$$\begin{aligned} \|x_k - (x_k - Mx_k - q)_+\|_2 &\leq \|x_k - p(x_k - Mx_k - q, \alpha)\|_2 + \|(x_k - Mx_k - q)_+ - p(x_k - Mx_k - q, \alpha)\|_2 \\ &\leq C + \tfrac{\sqrt{n}\log 2}{\alpha}. \end{aligned}$$

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 $x(\alpha)$, which satisfies $\nabla f(x) = 0$ with

$$\nabla f(x) = (diag(p'^{-1}(x - Mx - q, \alpha) - \mathbf{1}) + M^T) diag(p'(x - Mx - q, \alpha))(x - p(x - Mx - q, \alpha)).$$

Since $0 < p'(x, \alpha) < \mathbf{1}$ and $M \in P_0$, then by using Lemma 4.1 and Lemma 4.2 we get that $x(\alpha)$ satisfies (20). ∎

Now we give an error bound for the solution of the original $LCP(M,q)$ in terms of a solution to (20) but skip the proof.

Theorem 4.3 *Consider a solvable $LCP(M,q)$ with $M \in R_0$. Let $x(\alpha)$ be a solution of (20). Then there exists an $x(x(\alpha))$ which is a solution of $LCP(M,q)$ such that*

$$\|x(\alpha) - x(x(\alpha))\|_2 \leq \tau(M,q)\frac{\sqrt{n}\log 2}{\alpha},$$

where $\tau(M,q)$ is a constant, see Theorem 2.2.1 [18].

The following theorem proves that if α is sufficiently large, then a solution of (20) can be purified to a 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 [5]

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

Theorem 4.4 *Suppose that $LCP(M,q)$ is solvable. Let $x(\alpha)$ be a solution of (20) 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 (20). we have $x(\alpha) > 0, Mx(\alpha) + q > 0$ and by Lemma 4.1 $x(\alpha) = -\frac{\log(\mathbf{1} - e^{-\alpha(Mx(\alpha)+q)})}{\alpha}$. Hence

$$x(\alpha)(Mx(\alpha) + q) \overset{y=\alpha(Mx(\alpha)+q)}{=} -\frac{y\log(\mathbf{1} - e^{-y})}{\alpha^2} < \frac{n}{\alpha^2} \leq 2^{-2L}$$

for all $\alpha \geq \overline{\alpha} = \sqrt{n}2^L$. By the purification procedure described in Appendix B [5], $x(\alpha)$ can be purified to a solution of $LCP(M,q)$. ∎

5 Numerical Results

We now give a summary of our computational experience with the smooth algorithms described in this paper. 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 [14] 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 and solving linear equations by a sparse LU decomposition from MINOS.

For linear inequalities, we compared the smooth algorithm with MINOS as well as with the relaxation method of Motzkin and Schoenberg [13]. 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 [15] 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:

$$\begin{gathered} A_1 x \le b_1 \\ A_2 x \le b_2 \\ -r \le \mathbf{1}^T (A_1 x - b_1) \le r \end{gathered}$$

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 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 5 to 7 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 [2] to solve the unconstrained minimization problem for variables up to 150, and the limited memory BFGS algorithm [16] for lager 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:

$$\begin{gathered} Ax \le b_1 \\ g(x) \le b_2 \end{gathered}$$

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

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

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 8 depicts the ratio of CPU time taken by MINOS to the time taken by the smooth algorithm as a function of problem size n.

For monotone linear complementarity problems, we used a Newton method with a safeguarded linear search to solve the nonlinear equation (19). This approach is computationally more efficient than solving the nonlinear equation (20) . We compared the smooth method with Lemke's method [1]. 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: $\| \min\{x, Mx+q\}\|_2$ [12] . 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 [3], which employs sparse basis updating techniques. The SOR method of Delone and Tork Roth [6] does not apply to this class of nonsymmetric LCP nor do other splitting methods described in [1]. In fact, the SOR method of [6] failed on all test problems. Figures 9 and 10 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.

We conclude with Table 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 as well as linear complementarity problems and hence warrant further study. We plan to generalize our smoothing techniques to various problems such as linear programs, convex programs, nonlinear complementarity problems and neural networks.

Problem	*Speedup*	*Over*	*Size*
Linear Inequalities	1142	MINOS	2000×1000
Convex Inequalities	580	MINOS	210×400
LCP	63	Lemke	10000

Table 1 Maximum speedup over Minos or Lemke

Acknowledgement

We are grateful to Michael C. Ferris and Steven P. Dirkse for letting us use their implementation of the sparse Lemke method for comparison purposes.

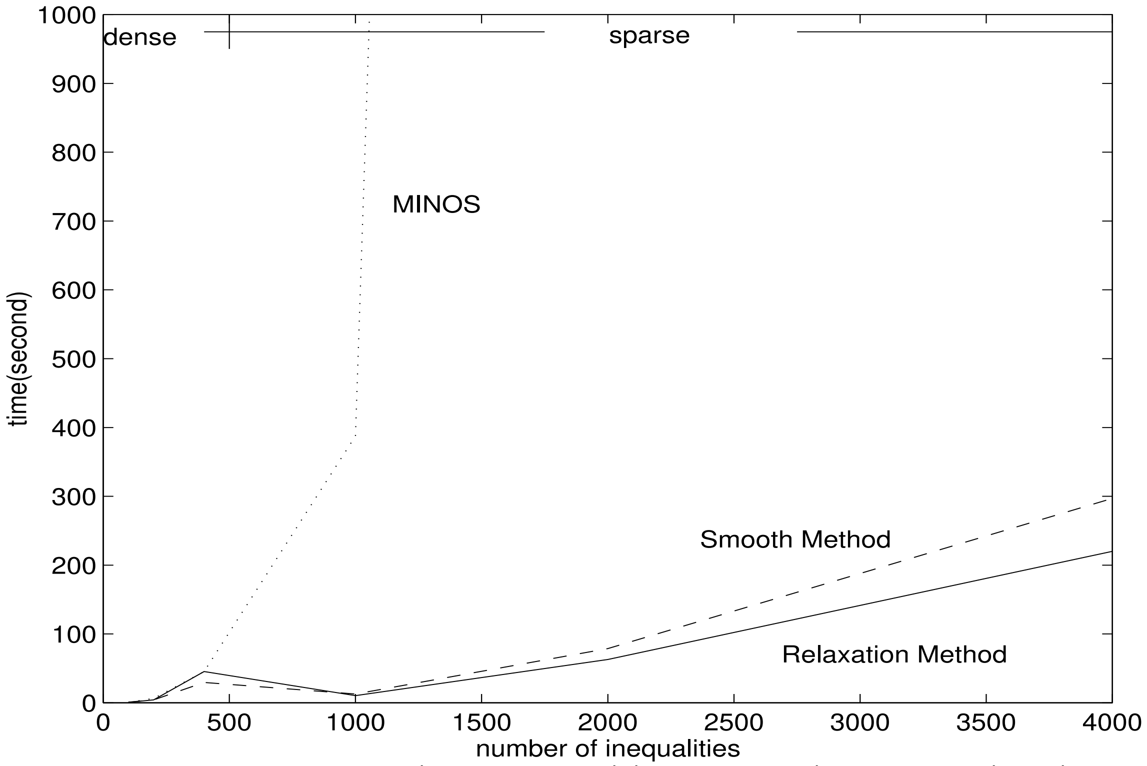

Figure 5: **Dense and Sparse Linear Inequalities: Comparison of solution times for MINOS, relaxation and smooth methods with m=n**

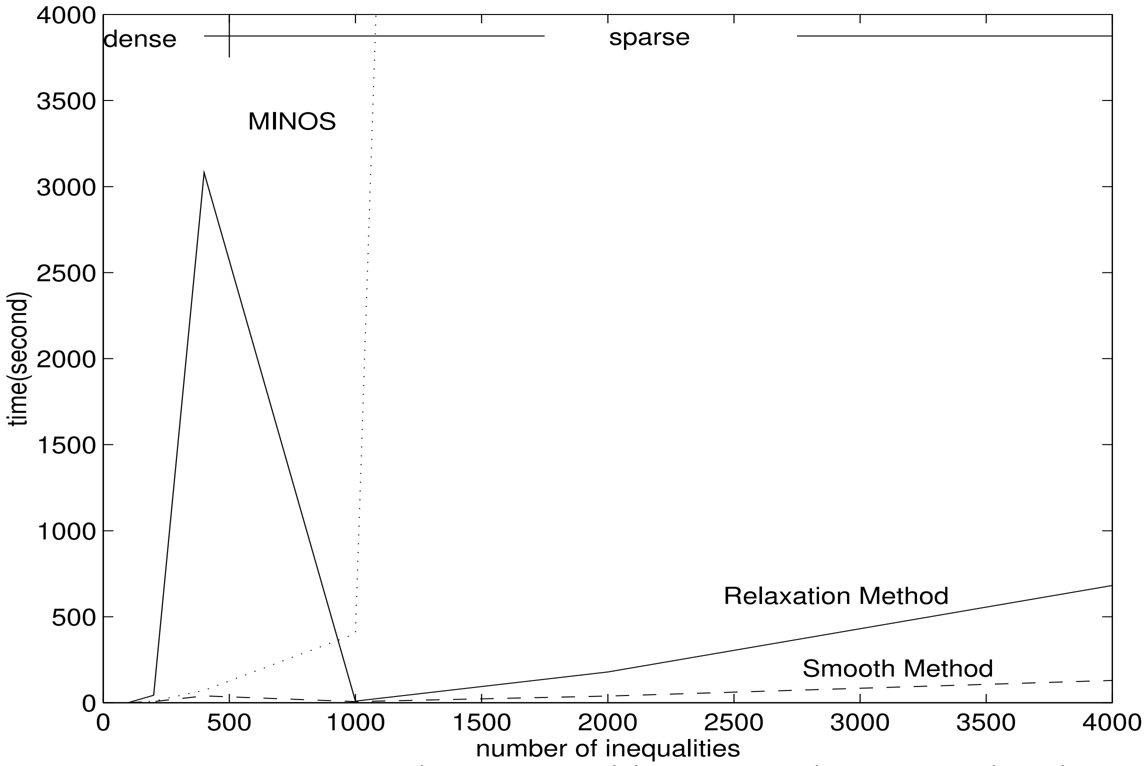

Figure 6: **Dense and Sparse Linear Inequalities: Comparison of solution times for MINOS, relaxation and smooth methods with m=2n**

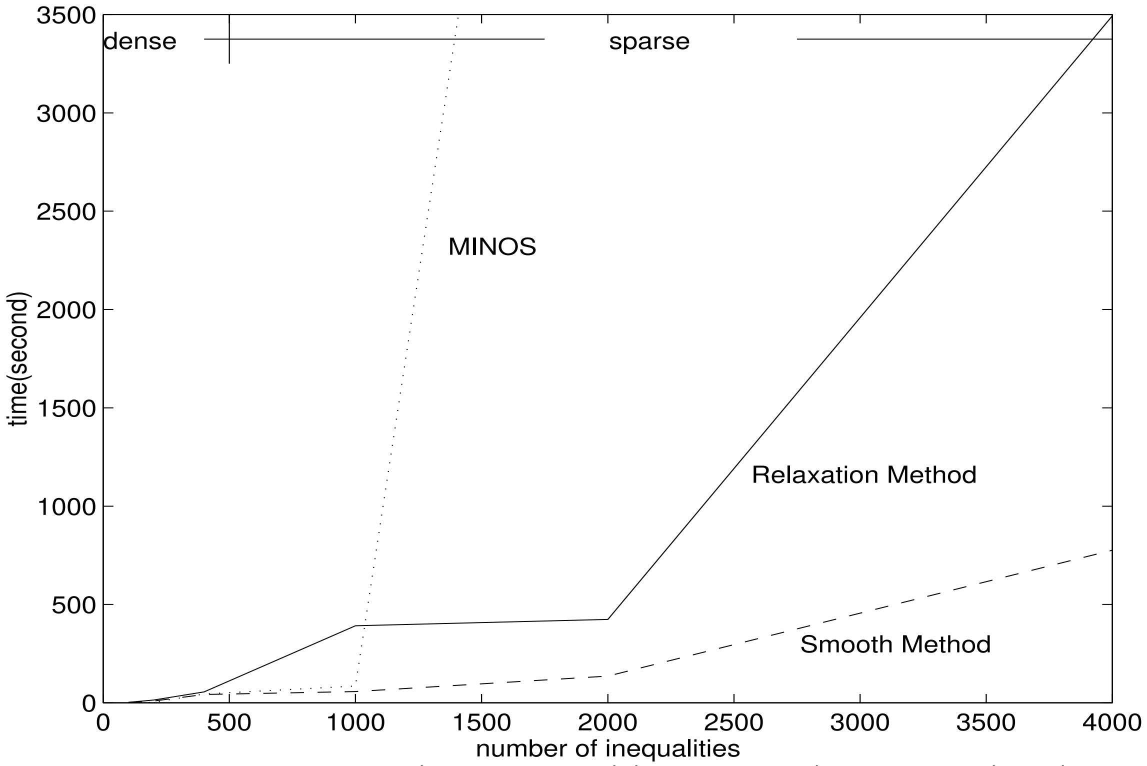

Figure 7: **Dense and Sparse Linear Inequalities: Comparison of solution times for MINOS, relaxation and smooth methods with m=4n**

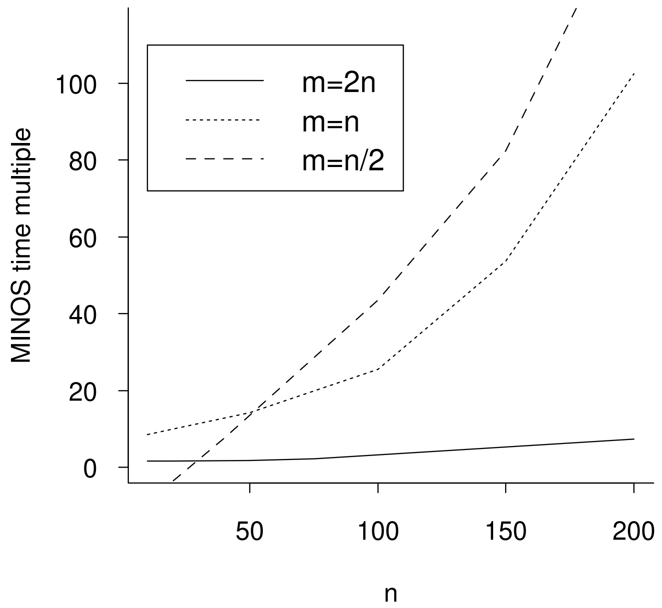

Figure 8: **Comparison of solution times for MINOS and the smooth method**

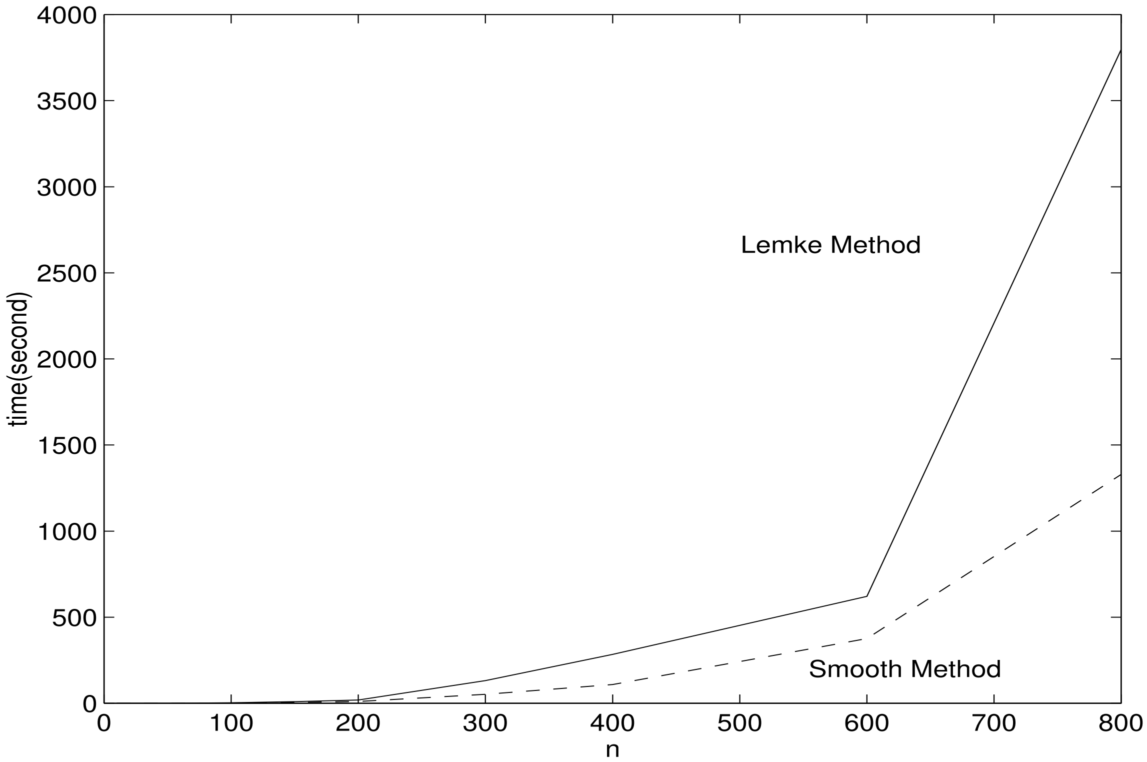

Figure 9: **Dense Monotone Linear Complementarity Problems: Comparison of solution times for Lemke and smooth methods. SOR method fails on all these problems.**

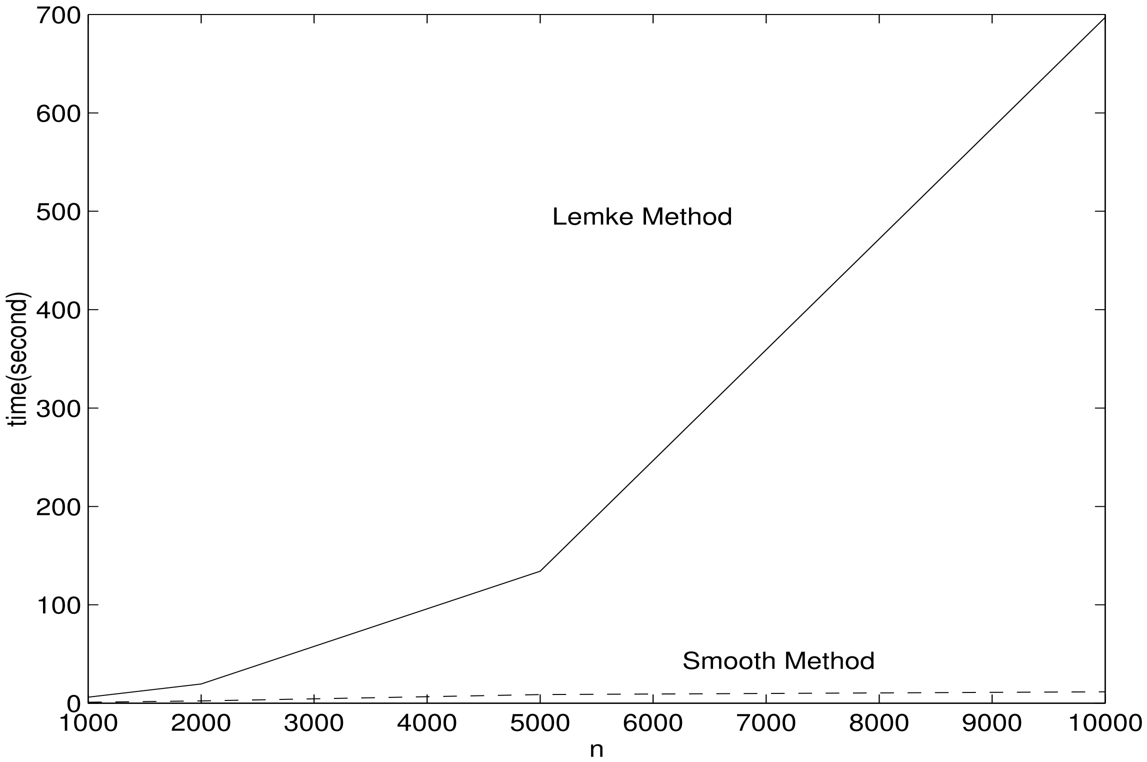

Figure 10: **Sparse Monotone Linear Complementarity Problems: Comparison of solution times for sparse Lemke and smooth methods. SOR method fails on all these problems.**

References

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

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

[3] 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.

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

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

[6] 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.

[7] 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.

[8] 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.

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

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

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

[12] 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.

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

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

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

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

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

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

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

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