A BOX-CONSTRAINED DIFFERENTIABLE PENALTY METHOD FOR NONLINEAR COMPLEMENTARITY PROBLEMS

BOSHI TIAN*, YAOHUA HU†, AND XIAOQI YANG‡

Abstract. In this paper, we propose a box-constrained differentiable penalty method for nonlinear complementarity problems, which not only inherits the same convergence rate as the existing $\ell_{\frac{1}{p}}$-penalty method but also overcomes its disadvantage of non-Lipschitzianness. We introduce the concept of a uniform ξ-P-function with $\xi \in (1,2]$, and apply it to prove that the solution of box-constrained penalized equations converges to that of the original problem at an exponential order. Instead of solving the box-constrained penalized equations directly, we solve a corresponding differentiable least squares problem by using a trust-region Gauss-Newton method. Furthermore, we establish the connection between the local solution of the least squares problem and that of the original problem under mild conditions. We carry out the numerical experiments on the test problems from MCPLIB, and show that the proposed method is efficient and robust.

Key words. nonlinear complementarity problem, $\ell_{\frac{1}{p}}$-penalty method, differentiable penalty method, convergence rate, least squares method

AMS subject classifications. 90C33, 65K15, 49M30

1. Introduction. Consider the following nonlinear complementarity problem (NCP) of finding a vector $x \in \mathbb{R}^n$ satisfying the following conditions

$$x \leq 0,\ F(x) \leq 0,\ x^T F(x) = 0, \tag{1.1}$$

where the function $F : \mathbb{R}^n \to \mathbb{R}^n$ is assumed to be continuously differentiable. We use X^* to denote the solution set of problem (1.1) and $\mathcal{J}$ to denote the index set $\{1, 2, \ldots, n\}$. When F is an affine function, problem (1.1) is reduced to a linear complementarity problem (LCP). Complementarity problems play an important role in operations research, option pricing, economic equilibrium models and the engineering sciences; see, e.g., [12, 13, 16].

Comprehensive studies for the NCP have been done, see monographs [7, 10, 11] and the references therein. Particularly, we refer to two kinds of methods, which are efficient to solve the NCP and will be compared with the proposed method in the numerical experiments. Chen and Mangasarian [5] proposed a class of parametric smooth functions to smooth out the nonsmooth equations transformed by the NCP. These smoothing functions were refined by Chen and Harker in [4]. The semismooth Newton method for solving the NCP by virtue of the Fischer-Burmeister function [14] was widely studied in [8, 20, 21]. Penalty function methods play an important role in nonlinear programming, see [19, 29, 32, 33]. Recently, the $\ell_{\frac{1}{p}}(p > 1)$-penalty method has received a great deal of attention in solving problem (1.1) and some desirable results on the convergence rate were proved in [30]. Specifically, the $\ell_{\frac{1}{p}}$-penalty method for problem (1.1) is to find a vector $x \in \mathbb{R}^n$ satisfying the following system of nonlinear equations

$$\phi(x, \rho) := \rho F(x) + [x]_+^{\frac{1}{p}} = 0, \tag{1.2}$$

*Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, P.R. China (Bs.Tian@connect.polyu.hk).

†Department of Mathematics, Zhejiang University, Hangzhou 310027, P. R. China (hyh19840428@163.com).

‡Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, P.R. China (mayangxq@polyu.edu.hk). Research of this author was supported by grants from the Research Grants Council of Hong Kong(PolyU 5292/13E)

where $\rho > 0$ is the penalty parameter, $p \geq 1$ is the power, $[x]_+^{\frac{1}{p}}$ is a vector with components $([x]_+^{\frac{1}{p}})_i = \max\{x_i, 0\}^{\frac{1}{p}}$ for all $i \in \mathcal{J}$. Throughout this paper, we use $\|\cdot\|$ to indicate the Euclidean norm.

As $p = 1$, the $\ell_{\frac{1}{p}}$-penalty method is reduced to the ℓ_1(or linear)-penalty method, which was first proposed by Bensoussan and Lions [1] for solving a continuous variational inequality problem. They proved that the solution to the penalized equation converges to the exact one of the variational inequality problem at the rate of $\mathcal{O}(\rho^{\frac{1}{2}})$. This square root rate of convergence requires ρ to be sufficiently small so as to achieve a given accuracy of the approximate solution. However, researchers in [15, 36] pointed out that the small values of the penalty parameter ρ result in poorly conditional algebraic problems in solving nonlinear equations (1.2).

Wang et al. [31] proposed an $\ell_{\frac{1}{p}}$-penalty method to solve the LCP arising from American options. They proved that the solution x^ρ converges to x^* at a rate of $\mathcal{O}(\rho^{\frac{p}{2}})$, which significantly improves the theoretical result of the square root rate of convergence mentioned above. Furthermore, Huang and Wang [17] extended the $\ell_{\frac{1}{p}}$-penalty method to solve problem (1.1) and established that the convergence rate between the solution of penalized equations and that of problem (1.1) is of an order at least $\mathcal{O}(\rho^{\frac{p}{\xi}})$, provided that the function F is continuous and ξ-monotone for a positive constant $\xi \in (1, 2]$ (see Definition 2.1 for its definition). The same convergence rate has been established in [18] for the $\ell_{\frac{1}{p}}$-penalty method in solving a mixed complementarity problem.

However, all efficient methods for nonlinear equations cannot be used to solve the $\ell_{\frac{1}{p}}$-penalized equations directly as the $\ell_{\frac{1}{p}}$-penalized term is not locally Lipschitz. Some smoothing methods have been introduced to approximately solve the $\ell_{\frac{1}{p}}$-penalized equations in [31, 35]. A vital drawback of the smoothing methods is that their solutions become unstable as the smoothing parameter is sufficiently small.

In this paper, we introduce a new type of function F, called a uniform ξ-P-function, which is weaker than the ξ-monotonicity and coincides with a uniform P-function (or a P-function introduction in [10]) when the function F is linear. Under the assumption of a uniform ξ-P-function, we show that problem (1.1) has a unique solution, and moreover the penalized equations (1.2) have a unique solution for any $\rho > 0$. Then we introduce a box-constrained differentiable penalty method for solving problem (1.1), which not only inherits the convergence rate of the existing $\ell_{\frac{1}{p}}$-penalty method [30] but also removes the drawback of the non-Lipschitzianness of the $\ell_{\frac{1}{p}}$-penalized term. Specifically, we consider a differentiable system of nonlinear equations with box-constraints, whose solution converges to x^* at a rate of $\mathcal{O}(\rho^{\frac{k}{\xi}})$ provided the function F is a uniform ξ-P-function. In general, it needs some good starting point for solving the box-constrained nonlinear equations directly, which is unavailable in practice. In order to design globally convergent methods that allow arbitrary starting points to solve the NCP, we consider a corresponding least squares problem, instead of solving the box-constrained equations directly, and apply the trust-region Gauss-Newton method introduced by Macconi et al. [26] to solve it. Furthermore, we establish the connection between solutions of the least squares problem and that of problem (1.1).

We carry out numerical experiments on test problems from MCPLIB [3]. We first set $p = 2$ and compare the performance of the proposed method with the smoothed $\ell_{\frac{1}{2}}$-penalty method [17] and the ℓ_1-penalty method [1] in terms of the number of function evaluations and the values of the penalty parameter. Numerical results show that our proposed method is more efficient and robust than other two methods. Then different values of power $p = 1, 2, 100, 1000, 5000, 10000$ are chosen to test the efficiency of our method. Furthermore, we compare the proposed method with the smooth approximation method [4] and the nonsmooth equations method [20] in terms of the number of function evaluations.

This paper is organized as follows. In Section 2, we propose a differentiable penalty method for solving problem (1.1). Moreover, we establish the main convergence rate theorem for the proposed method under the assumption of a uniform ξ-P-function. We present a numerical algorithm to solve problem (1.1) in Section 3. In the last section, preliminary numerical experiments are shown.

2. Box-Constrained Differentiable Penalty Method. In this section, we first introduce the concept of a uniform ξ-P-function with $\xi \in (1, 2]$ that is used as the basic assumption in this paper. Then we propose a box-constrained differentiable penalty method and establish its convergence rate theorem.

2.1. Uniform ξ-P-function. To begin, we first recall some useful definitions.

DEFINITION 2.1 ([10, Definition 2.3.1]). *A function $S : \mathbb{R}^n \to \mathbb{R}^n$ is said to be ξ-monotone for some $\xi \in (1, 2]$, if there exists a constant $\alpha > 0$ such that*

$$(x - y)^T (S(x) - S(y)) \geq \alpha \|x - y\|^\xi, \ \forall\, x, y \in \mathbb{R}^n.$$

When $\xi = 2$, the ξ-monotonicity is called the 2-monotonicity.

DEFINITION 2.2 ([10, Definition 3.5.8]). *A function $S : \mathbb{R}^n \to \mathbb{R}^n$ is said to be*

- *a P_0-function, if for all pairs of distinct vectors x and y in $\mathbb{R}^n$, there exists an index $\kappa = \kappa(x, y) \in \mathcal{J}$ such that*

$$x_\kappa \neq y_\kappa \text{ and } (x_\kappa - y_\kappa)(S_\kappa(x) - S_\kappa(y)) \geq 0;$$

- *a P-function, if for all pairs of distinct vectors x and y in $\mathbb{R}^n$,*

$$\max_{1 \leq \kappa \leq n} (x_\kappa - y_\kappa)(S_\kappa(x) - S_\kappa(y)) > 0;$$

- *a uniform P-function, if there exists constant $\alpha > 0$ such that for all pairs of vectors x and y in $\mathbb{R}^n$,*

$$\max_{1 \leq \kappa \leq n} (x_\kappa - y_\kappa)(S_\kappa(x) - S_\kappa(y)) \geq \alpha \|x - y\|^2.$$

DEFINITION 2.3 ([23, Definition 2(b)]). *A function $S : \mathbb{R}^n \to \mathbb{R}^n$ is said to be a strict P-function, if there exists $\gamma > 0$ such that $S - \gamma I$ is a P-function, where $I : \mathbb{R}^n \to \mathbb{R}^n$ is an identical mapping.*

DEFINITION 2.4 ([10]). *A matrix $A \in \mathbb{R}^{n \times n}$ is said to be*

- *a P_0-matrix if for any vector $x \neq 0$ in $\mathbb{R}^n$, and $y = Ax$, there is at least one index $\kappa \in \mathcal{J}$ such that $x_\kappa \neq 0$ and $x_\kappa y_\kappa \geq 0$;*
- *a P-matrix if for any $x \neq 0$ in $\mathbb{R}^n$, and $y = Ax$, there is at least one index $\kappa \in \mathcal{J}$ such that $x_\kappa \neq 0$ and $x_\kappa y_\kappa > 0$;*
- *an M-matrix if $a_{i,j} \leq 0$ whenever $i \neq j$ and all principal minors of A are positive.*

Extending the definition of the ξ-monotonicity, we introduce a new notion of function F, called a uniform ξ-P-function, which is stronger than the P-function.

DEFINITION 2.5. *A function $S : \mathbb{R}^n \to \mathbb{R}^n$ is said to be a uniform ξ-P-function for some $\xi \in (1, 2]$, if there exists a constant $\alpha > 0$ such that for all pairs of vectors x and y in $\mathbb{R}^n$,*

$$\max_{1 \le \kappa \le n} (x_\kappa - y_\kappa)(S_\kappa(x) - S_\kappa(y)) \ge \alpha \|x - y\|^\xi.$$

We see that a ξ-P-function is a P_0-function and is weaker than the ξ-monotonicity. The ξ-monotonicity has been utilized in [17] to establish the convergence rate of $\mathcal{O}(\rho^{\frac{p}{\xi}})$ by which the solution of problem (1.2) converges to that of problem (1.1). The following propositions are useful to investigate properties of a uniform ξ-P-function.

PROPOSITION 2.6 ([10, Proposition 3.5.9]). *Let $S : D \subset \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable P_0-function on the open set D. Then $\nabla S(x)$ is a P_0-matrix for each $x \in D$.*

COROLLARY 2.7. *Let $S : D \subset \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable ξ-P-function on the open set D. Then $\nabla S(x)$ is a P_0-matrix for each $x \in D$.*

PROPOSITION 2.8 ([7, Theorem 3.4.2]). *A matrix $A \in \mathbb{R}^{n \times n}$ is a P_0-matrix if and only if for every nonzero vector x, there exists an index $i \in \mathcal{J}$ such that $x_i \neq 0$ and $x_i \big(Ax\big)_i \ge 0$.*

PROPOSITION 2.9 ([10]). *Let the linear function $S : \mathbb{R}^n \to \mathbb{R}^n$ be $S(x) = Ax + b$ with a given matrix $A \in \mathbb{R}^{n \times n}$ and a vector $b \in R^n$. Then*
(a) S is ξ-monotone if and only if matrix A is positive definite;
(b) S is a (uniform) P-function if and only if A is a P-matrix.

It follows from Definition 2.5 and the last proposition, we conclude the next corollary.

COROLLARY 2.10. *Let the linear function $S : \mathbb{R}^n \to \mathbb{R}^n$ be $S(x) = Ax + b$ with a given matrix $A \in \mathbb{R}^{n \times n}$ and a vector $b \in \mathbb{R}^n$. Then S is a uniform ξ-P-function if and only if A is a P-matrix.*

PROPOSITION 2.11 ([10, Proposition 2.3.2]). *Let $S : D \subset \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable on the open convex set D. Then S is 2-monotone on D if and only if its Jacobian matrix $\nabla S(x)$ is uniformly positive definite for all x in D, i.e., there exists a constant $c' > 0$ such that*

$$y^T \nabla S(x) y \ge c' \|y\|^2, \ \forall\, y \in \mathbb{R}^n,$$

for all $x \in D$.

We present an example from [7, Example 3.3.2] below to show that the uniform ξ-P-function is strictly weaker than the ξ-monotonicity.

EXAMPLE 2.1. *Let $S(x) = Ax + b$ with*

$$A = \begin{pmatrix} 1 & -3 \\ 0 & 1 \end{pmatrix}$$

and a vector $b \in \mathbb{R}^n$. Clearly, A is a P-matrix. Letting $x = (1, 1)^T$, we note that $x^T Ax = -1 < 0$, which shows that A is not positive definite. Therefore, it follows from Proposition 2.9 that we know function $S(x)$ is a uniform ξ-P-function, but not ξ-monotone.

We further describe a nonlinear example to show that the uniform P-function is strictly weaker than the 2-monotonicity.

EXAMPLE 2.2. *Consider function $G : \mathbb{R}^2 \to \mathbb{R}^2$ as*

$$G(x) = \begin{pmatrix} x_1^3 \\ x_2^3 \end{pmatrix} + S(x),$$

where $S(x)$ is the linear function defined in Example 2.1. The Jacobian matrix of function $G(x)$ is

$$\nabla G(x) = \begin{pmatrix} 3x_1^2 & 0 \\ 0 & 3x_2^2 \end{pmatrix} + \begin{pmatrix} 1 & -3 \\ 0 & 1 \end{pmatrix}.$$

Take $x = (0,0)^T$. Then $\nabla G(x) = \begin{pmatrix} 1 & -3 \\ 0 & 1 \end{pmatrix}$. Example 2.1 shows that the matrix $\nabla G(0)$ is not positive definite. Therefore, by Proposition 2.11, we conclude that the function $G(x)$ is not 2-monotone. By Example 2.1, we have the function $S(x)$ is a uniform P-function. Then there exists constant $\alpha > 0$ such that for all pairs of vectors x and y in $\mathbb{R}^2$ the inequality $\max_{1\le\kappa\le 2}(x_\kappa - y_\kappa)(S_\kappa(x) - S_\kappa(y)) \ge \alpha\|x-y\|^2$ holds. We also notice that the inequality $(x_\kappa - y_\kappa)(x_\kappa^3 - y_\kappa^3) \ge 0$ holds for all pairs of vectors x and y in $\mathbb{R}^2$ and any $1 \le \kappa \le 2$. Therefore, we have

$$\begin{aligned}
&\max_{1\le\kappa\le 2}(x_\kappa - y_\kappa)(G_\kappa(x) - G_\kappa(y)) \\
&= \max_{1\le\kappa\le 2}\Big((x_\kappa - y_\kappa)(x_\kappa^3 - y_\kappa^3) + (x_\kappa - y_\kappa)(S_\kappa(x) - S_\kappa(y))\Big) \\
&\ge \max_{1\le\kappa\le 2}(x_\kappa - y_\kappa)(S_\kappa(x) - S_\kappa(y)) \ge \alpha\|x-y\|^2.
\end{aligned}$$

Consequently, the function $G(x)$ is a uniform P-function.

In the following, assuming the function F is a uniform ξ-P-function, we show that the solution of penalized equations (1.2) is unique. Before doing so, we first prove an auxiliary proposition.

Proposition 2.12. *Assume that the function $F : \mathbb{R}^n \to \mathbb{R}^n$ is a uniform ξ-P-function. Then problem (1.1) has a unique solution.*

Proof. It follows from [10, Proposition 1.1.3] that problem (1.1) is equivalent to the following variational inequality problem: find a vector $x \in \mathcal{K}$ such that for all vectors $y \in \mathcal{K}$

$$(y-x)^T F(x) \ge 0, \tag{2.1}$$

where $\mathcal{K} = \{y \in \mathbb{R}^n \mid y \le 0\}$.

Since a uniform ξ-P-function is a P-function, it follows from [10, Proposition 3.5.10] that the variational inequality problem (2.1) has at most one solution. Thus, to prove that problem (1.1) has a unique solution, it suffices to show that the variational inequality problem (2.1) has a solution. Using [10, Proposition 3.5.1], we only need to prove that there exists a vector $x^{\rm ref} \in \mathcal{K}$ such that the set

$$L'_{\le} := \{x \in \mathcal{K} \mid F_\nu(x)(x_\nu - x_\nu^{\rm ref}) \le 0,\ \forall\, \nu \in \mathcal{J} \text{ such that } x_\nu \ne x_\nu^{\rm ref}\}$$

is nonempty and bounded. Let $x^{\rm ref} \in \mathcal{K}$ and $\|x^{\rm ref}\| \ne 0$. By the continuity of function F on the closed convex set $\mathcal{K}$, we obtain that the set $L'_{\le}$ is nonempty via the intermediate value theorem. Now, assume on the contrary that the set $L'_{\le}$ is unbounded. There exists a sequence $\{x^k\} \subset \mathcal{K}$ such that for all k,

$$F_\nu(x^k)(x_\nu^k - x_\nu^{\rm ref}) \le 0,\ \forall\, \nu \in \mathcal{J} \text{ such that } x_\nu^k \ne x_\nu^{\rm ref}, \tag{2.2}$$

and $\lim_{k\to\infty}\|x^k\| = +\infty$.

Since the function F is a uniform ξ-P-function, it follows that there exist constants $\alpha > 0$, $\xi > 1$ and an index $\nu = \nu(x^k, x^{\text{ref}}) \in \mathcal{J}$ with $x_\nu^k \neq x_\nu^{\text{ref}}$ such that

$$(F_\nu(x^k) - F_\nu(x^{\text{ref}}))(x_\nu^k - x_\nu^{\text{ref}}) \geq \alpha \|x^k - x^{\text{ref}}\|^\xi.$$

Dividing on both sides of the last inequality by the term $\|x^k\|^{\frac{\xi+1}{2}}$, we have

$$\lim_{\|x^k\| \to +\infty} \frac{F_\nu(x^k)(x_\nu^k - x_\nu^{\text{ref}})}{\|x^k\|^{\frac{\xi+1}{2}}} = +\infty,$$

which contradicts with inequality (2.2). Therefore, the set $L'_{\leq}$ is bounded for the given $x^{\text{ref}} \in \mathcal{K}$. By [10, Proposition 3.5.1], we conclude that the variational inequality problem (2.1) has a solution. Hence, we conclude that problem (1.1) has a unique solution. ▯

Proposition 2.13. *Assume the function $F : \mathbb{R}^n \to \mathbb{R}^n$ is a uniform ξ-P-function. Then the penalized nonlinear equations (1.2) have a unique solution for any $\rho > 0$.*

Proof. For any vectors $x, y \in \mathbb{R}^n$ and index $i \in \mathcal{J}$, we have

$$\begin{aligned}(x_i - y_i)(\phi_i(x,\rho) - \phi_i(y,\rho)) &= \rho(x_i - y_i)(F_i(x) - F_i(y)) + (x_i - y_i)([x_i]_+^{\frac{1}{p}} - [y_i]_+^{\frac{1}{p}}) \\ &\geq \rho(x_i - y_i)(F_i(x) - F_i(y)),\end{aligned}$$

since the function $[x]_+^{\frac{1}{p}}$ is monotone. There exist constants $\alpha > 0$ and $\xi > 1$ such that

$$\begin{aligned}\max_{1 \leq \kappa \leq n} (x_\kappa - y_\kappa)(\phi_\kappa(x,\rho) - \phi_\kappa(y,\rho)) &\geq \rho \max_{1 \leq \kappa \leq n} (x_\kappa - y_\kappa)(F_\kappa(x) - F_\kappa(y)) \\ &\geq \rho\alpha \|x - y\|^\xi,\end{aligned}$$

where the last inequality follows from Definition 2.5. Therefore, the function $\phi(x,\rho)$ is a uniform ξ-P-function for any $\rho > 0$, and thus the following variational inequality problem: find a vector $x \in \mathbb{R}^n$ such that

$$(y - x)^T \phi(x,\rho) \geq 0, \ \forall\, y \in \mathbb{R}^n$$

has a unique solution by Proposition 2.12. We proved that the penalized equations (1.2) have a unique solution. ▯

Remark 2.1. *Using [23, Propositions 2 and 3], we can achieve the same conclusion as Proposition 2.13 under the assumption of a strict P-function for F.*

2.2. Box-Constrained Differentiable Penalty Method. In the following, we introduce a box-constrained differentiable penalty method for solving problem (1.1), which not only shares the same convergence rate as the existing $\ell_{\frac{1}{p}}$-penalty method but also can be implemented easily. We consider the system of box-constrained equations as follows:

$$\mathcal{F}(x,\rho) := \begin{pmatrix} \rho x_1 F_1(x) & + & [F_1(x)]_+^q \\ \rho x_2 F_2(x) & + & [F_2(x)]_+^q \\ \vdots & \vdots & \vdots \\ \rho x_n F_n(x) & + & [F_n(x)]_+^q \end{pmatrix} = 0, \quad x \in \Omega, \tag{2.3}$$

where $q = 1 + \frac{1}{p}$ and $\Omega := \{x \in \mathbb{R}^n \mid x \leq 0\}$. Since the composite function $[g(x)]_+^q$ is first order continuously differentiable as long as the function $g : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and $q > 1$. Under our assumption, we see that the function $\mathcal{F}(\cdot, \rho) : \mathbb{R}^n \to \mathbb{R}^n$ is first order continuously differentiable for each ρ. The system (2.3) can be solved efficiently by algorithms [22, 26].

REMARK 2.2. *Alternately, we can consider another system of constrained equations for problem (1.1) as follows*

$$\begin{pmatrix} \rho x_1 F_1(x) & + & [x_1]_+^q \\ \rho x_2 F_2(x) & + & [x_2]_+^q \\ \vdots & \vdots & \vdots \\ \rho x_n F_n(x) & + & [x_n]_+^q \end{pmatrix} = 0, \quad x \in \widehat{\Omega}, \tag{2.4}$$

where $\widehat{\Omega} := \{x \in \mathbb{R}^n \mid F(x) \leq 0\}$. However, the feasible set $\widehat{\Omega}$ is not convex. It is not easy to solve the system (2.4) when the function F is nonlinear.

PROPOSITION 2.14. *Let $x^* \in \mathbb{R}^n$ be a solution of problem (1.1). Then x^* solves $\mathcal{F}(x, \rho) = 0$ for any given $\rho > 0$.*

We present an example that shows the converse of Proposition 2.14 is not true.

EXAMPLE 2.3. *Let $F(x) = 0$ for all $x \in \mathbb{R}$. It is obvious that x^* solves the equation $\mathcal{F}(x, \rho) = 0$ for any $x^* \in \mathbb{R}$. But x^* is not the solution of problem (1.1) when $x^* > 0$.*

REMARK 2.3. *Example 2.3 indicates that the constraint set Ω in the system (2.3) is vital to the box-constrained differentiable penalty method for problem (1.1).*

Given the penalty parameter ρ and power p. The solution of the system (2.3) in general is not unique even if problem (1.1) has a unique solution, which is verified by the next example.

EXAMPLE 2.4. *Let $F(x) = x + 1$ with $x \in \mathbb{R}$ and $q = 2$. It is clear that $x^* = -1$ is the unique solution of this linear complementarity problem. Its box-constrained equation is $\rho x(x + 1) + [x + 1]_+^2 = 0$ with $x \leq 0$. The constrained equation has two solutions, one is $\bar{x}^\rho = -1$ and the other one is $\hat{x}^\rho = -\frac{1}{\rho+1}$.*

2.3. Convergence Rate Analysis. In this subsection, we establish that the solution x^ρ of system (2.3) converges to a solution x^* of problem (1.1) at a rate of $\mathcal{O}(\rho^{\frac{p}{\xi}})$, provided that the function F is a uniform ξ-P-function. We first show some useful lemmas as follows.

LEMMA 2.15. *For each $\rho > 0$, assume that the function $F : \mathbb{R}^n \to \mathbb{R}^n$ is a uniform ξ-P-function and let the vector $x^\rho \in \mathbb{R}^n$ be a solution of system (2.3). Then there exists a positive constant $M_1 > 0$, independent of x^ρ, ρ and p, such that*

$$\|x^\rho\| \leq M_1.$$

Proof. Given $\rho > 0$. Since x^ρ is a solution of system (2.3), it follows that $\rho x_i^\rho F_i(x^\rho) + F_i(x^\rho)[F_i(x^\rho)]_+^{\frac{1}{p}} = 0$, which means $x_i^\rho F_i(x^\rho) \leq 0$, for all $i \in \mathcal{J}$.

By the uniform ξ-P-function of function F, we see that there exist constants $\alpha > 0$ and $\xi > 1$ such that

$$\alpha \|x^\rho\|^\xi \leq \max_{1 \leq i \leq n} x_i^\rho \Big(F_i(x^\rho) - F_i(0)\Big) \leq \max_{1 \leq i \leq n} \big(-x_i^\rho F_i(0)\big) \leq \|x^\rho\| \|F(0)\|_\infty.$$

Consequently, we proved this lemma with $M_1 = \sqrt[\xi-1]{\frac{1}{\alpha}\|F(0)\|_\infty}$. $\square$

Lemma 2.15 implies that, for any $\rho > 0$, the solution of system (2.3) always lies in a bounded closed set. Since F is continuous, we have that there exists a positive constant L, independent of x^ρ, ρ and p, such that $\|F(x^\rho)\| \le L$, for all $\rho > 0$.

LEMMA 2.16. *For each $\rho > 0$, assume that the function $F : \mathbb{R}^n \to \mathbb{R}^n$ is a uniform ξ-P-function and let the vector $x^\rho \in \mathbb{R}^n$ be a solution of system (2.3). Then there exists a positive C_1, independent of x^ρ and ρ, such that*

$$\|[F(x^\rho)]_+\| \le C_1 \rho^p.$$

Proof. Since x^ρ is a solution of system (2.3), it follows that $[F_i(x^\rho)]_+^q = -\rho F_i(x^\rho) x_i^\rho \le \rho \|F(x^\rho)\|_\infty \|x^\rho\|_\infty$ for all index $i \in \mathcal{J}$. Therefore, we have $\|[F(x^\rho)]_+\|_\infty \le \rho^p \|x^\rho\|_\infty^p$. By Lemma 2.15 and the fact that there exists some constant $\widetilde{C} > 0$ such that $\|[F(x^\rho)]_+\| \le \widetilde{C} \|[F(x^\rho)]_+\|_\infty$, we have

$$\|[F(x^\rho)]_+\| \le C_1 \rho^p$$

where $C_1 = \widetilde{C} M_1^p$. □

Now, we establish our main convergence rate theorem.

THEOREM 2.17. *For each $\rho > 0$, assume that the function $F : \mathbb{R}^n \to \mathbb{R}^n$ is a uniform ξ-P-function. Let vectors x^* and x^ρ in $\mathbb{R}^n$ be the solutions of problem (1.1) and system (2.3), respectively. Then there exist constants $\widehat{C} > 0$ and $\xi > 1$, independent of x^ρ and ρ, such that*

$$\|x^* - x^\rho\| \le \widehat{C} \rho^{\frac{p}{\xi}}.$$

Proof. Since x^ρ is a solution of system (2.3), the index set at x^ρ can be split into the following two sets:

$$\alpha^\rho = \{i \in \mathcal{J} \mid x_i^\rho = 0,\ F_i(x^\rho) \le 0\};$$
$$\gamma^\rho = \{i \in \mathcal{J} \mid x_i^\rho < 0,\ F_i(x^\rho) \ge 0\}.$$

We first show that the inequality holds for any index $i \in \mathcal{J}$

$$\left(x_i^* - x_i^\rho\right)\left(F_i(x^*) - F_i(x^\rho) + [F_i(x^\rho)]_+\right) = \left(x_i^* - x_i^\rho\right)\left(F_i(x^*) + [F_i(x^\rho)]_-\right) \le 0 \quad (2.5)$$

where $[a]_- := \max\{-a, 0\}$ for all $a \in \mathbb{R}$. Note that x^* is a solution of problem (1.1), the following two cases are considered.

(I) $i \in \alpha^\rho$. We have

$$\left(x_i^* - x_i^\rho\right)\left(F_i(x^*) - F_i(x^\rho) + [F_i(x^\rho)]_+\right) = x_i^* [F_i(x^\rho)]_- \le 0;$$

(II) $i \in \gamma^\rho$. We have

$$\left(x_i^* - x_i^\rho\right)\left(F_i(x^*) - F_i(x^\rho) + [F_i(x^\rho)]_+\right) = -x_i^\rho F_i(x^*) \le 0.$$

Therefore, we proved that the inequality (2.5) holds for all index $i \in \mathcal{J}$.

Since the function F is a uniform ξ-P-function, it follows that there exist constants $\alpha > 0$ and $\xi > 1$ such that

$$\begin{aligned}
\alpha\|x^* - x^\rho\|^\xi &\le \max_{1\le i\le n} \big(x_i^* - x_i^\rho\big)\big(F_i(x^*) - F_i(x^\rho)\big) \\
&\le \max_{1\le i\le n} \big(- [F_i(x^\rho)]_+ (x_i^* - x_i^\rho)\big) \\
&\le C_1 \rho^p \|x^* - x^\rho\|_\infty \\
&\le 2C_1 M_1 \rho^p,
\end{aligned}$$

where the second inequality is from inequality (2.5), the third one is from Lemma 2.16 and the last one is from Lemma 2.15. Therefore, we proved this theorem with $\widehat{C} = \sqrt[\xi]{\frac{2C_1 M_1}{\alpha}}$. □

Similar to the proof of Theorem 2.17, we can establish the convergence rate of $\mathcal{O}(\rho^{\frac{p}{\xi}})$ for the existing $\ell_{\frac{1}{p}}$-penalty method under the assumption of a uniform ξ-P-function (or a P-matrix for the LCP), which is weaker than that of the ξ-monotonicity of the function F (or a positive definite matrix for the LCP) used in [17]. Here, the details are omitted.

THEOREM 2.18. *For each $\rho > 0$, assume that the function $F : \mathbb{R}^n \to \mathbb{R}^n$ is a uniform ξ-P-function. Let vectors x^* and x^ρ in $\mathbb{R}^n$ be the solutions of the problem (1.1) and system (1.2), respectively. Then there exist constants $\widehat{C} > 0$ and $\xi > 1$, independent of x^ρ and ρ, such that*

$$\|x^* - x^\rho\| \le \widehat{C}\rho^{\frac{p}{\xi}}.$$

COROLLARY 2.19. *For each $\rho > 0$, assume that the linear function $F : \mathbb{R}^n \to \mathbb{R}^n$ is $F(x) = Ax + b$ with the matrix $A \in \mathbb{R}^{n\times n}$ being a P-matrix and any vector $b \in \mathbb{R}^n$. Let vectors x^* and x^ρ in $\mathbb{R}^n$ be the solutions of problem (1.1) and system (1.2), respectively. Then, there exists a constant $\widehat{C} > 0$, independent of x^ρ and ρ, such that*

$$\|x^* - x^\rho\|_\infty \le \widehat{C}\rho^p.$$

REMARK 2.4. *We note that the assumption of a P-matrix is weaker than the assumption of a M-matrix used in [30] and the assumption of positive definiteness used in [17].*

REMARK 2.5. *It has been proved in [25] that the class of P-matrices contains not only the positive definiteness matrix but also the M-matrix; furthermore, any strictly or irreducibly diagonally dominant matrix with non-negative elements is likewise a P-matrix.*

We present an example from [7, Example 3.3.10] to verify the conclusions in Remarks 2.4 and 2.5.

EXAMPLE 2.5. *Let*

$$A = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & -17 \\ 4 & 0 & 1 \end{pmatrix}.$$

Three eigenvalues of matrix A are 5 and $-1 \pm i\sqrt{13}$. Thus, the matrix A is neither positive definite nor an M-matrix. However, it is a P-matrix.

3. Numerical Method. In this section, we describe specific algorithm to solve the NCP. In order to use the global convergent methods that allow arbitrary starting points to solve the NCP, we consider the corresponding least squares problem

$$\min_{x\in\Omega} \ \Psi(x,\rho) := \frac{1}{2}\|\mathcal{F}(x,\rho)\|^2. \tag{3.1}$$

The first order necessary condition for a vector $x^\rho \in \Omega$ to be a local solution of problem (3.1) for given $\rho > 0$ is stated in the following proposition.

PROPOSITION 3.1 ([2, Proposition 2.1.2]). *For each $\rho > 0$, assume that x^ρ is a local solution of problem (3.1). Then, we have*

$$\nabla\Psi(x^\rho,\rho)^T(x - x^\rho) \geq 0, \ \ \forall\, x \in \Omega, \tag{3.2}$$

where $\nabla\Psi$ is the gradient of function Ψ.

For each $\rho > 0$, the Jacobian matrix $\nabla\mathcal{F}$ of function $\mathcal{F}(x,\rho)$ can be expressed as

$$\nabla\mathcal{F}(x,\rho) := \rho\Theta(x) + \Pi(x,\rho)\nabla F(x), \tag{3.3}$$

where $\nabla F(x)$ is the Jacobian matrix of function F, $\Theta(x)$ and $\Pi(x,\rho)$ are diagonal matrices, i.e., $\Theta(x) := \mathrm{diag}(F_1(x),\ldots,F_n(x))$, $\Pi(x,\rho) := \mathrm{diag}(G_1(x,\rho),\ldots,G_n(x,\rho))$, and $G_i(x,\rho) := \rho x_i + (1+\frac{1}{p})[F_i(x)]_+^{\frac{1}{p}}$ for all $i \in \mathcal{J}$.

3.1. Trust-Region Gauss-Newton Method. In the following, we apply a trust-region Gauss-Newton method to solve the least squares problem (3.1) for given $\rho > 0$; more details can be found in [6, 24, 27]. At the k-th iteration, we consider a quadratic approximation $m^k(d,\rho)$ for $\Psi(x,\rho)$ at $x^k \in \Omega$ and replace the problem (3.1) by a trust region problem

$$\min \quad m^k(d,\rho) \quad \text{s.t. } \|d\| \leq \Delta^k, \tag{3.4}$$

with the objective function

$$m^k(d,\rho) := \frac{1}{2}\|\mathcal{F}(x^k,\rho) + \nabla\mathcal{F}(x^k,\rho)d\|^2, \tag{3.5}$$

where Δ^k is the trust-region radius.

A formal description of the trust-region Gauss-Newton method for problem (3.1) for given $\rho > 0$ is can be found in [26, Algorithm 3.1]. Here we omit the details.

We present a box-constrained differentiable penalty algorithm for problem (1.1). Before doing this, we define the termination criterion for this algorithm as follows

$$\text{Termination}(x) := \min\{\|[x]_+\|, \ \|[F(x)]_+\|, \|F(x)\circ x\|\} \leq \epsilon, \tag{3.6}$$

where $\epsilon > 0$ is the tolerance parameter, which should be small enough, "$\circ$" denotes the component-wise multiplication. Now, a formal description of the box-constrained differentiable penalty algorithm for problem (1.1) is given as follows.

Algorithm 1: Box-constrained differentiable penalty method for the NCP.

Initializing $\rho^0 > 0$, ρ^{min}; $\sigma \in (0,1)$, $\epsilon > 0$ and an initial point x^0 and let $i := 0$;
while $\rho^i > \rho^{min}$ **do**
    if Termination$(x^i) \le \epsilon$ **then**
        Stop;
    else
        Using [26, Algorithm 3.1] to solve problem (3.1) with starting point x^i and penalty parameter ρ^i, we obtain x^{i+1};
    end
    Letting $\rho^{i+1} := \sigma\rho^i$ and $i := i+1$;
end

3.2. Convergence Analysis. In this subsection, we establish the connection between solutions of the least squares problem (3.1) and solutions of problem (1.1).

THEOREM 3.2. *Suppose that vector $x^i \in \Omega$ is the exact global solution of problem (3.1), and that $\rho^i \to 0^+$. Then every limit point of the sequence $\{x^i\}$ is a solution of problem (1.1).*

Proof. Let x^* be the solution of problem (1.1). It follows from Proposition 2.14 that $\Psi(x^*, \rho) = 0$ for any $\rho > 0$. Since x^i is the exact global solution of problem (3.1) for given $\rho^i > 0$, we have $\Psi(x^i, \rho^i) \le \Psi(x^*, \rho^i)$, which means that $\Psi(x^i, \rho^i) = 0$. Specifically, we have

$$\frac{1}{2}\sum_{l=1}^{n}(x_l^i F_l(x^i))^2 + \frac{1}{\rho^i}\Big(\sum_{l=1}^{n} x_l^i [F_l(x^i)]_+^{q+1} + \frac{1}{2\rho^i}\sum_{l=1}^{n}[F_l(x^i)]_+^{2q}\Big) = 0. \tag{3.7}$$

By rearranging this expression, we obtain

$$\begin{aligned}\frac{1}{2}(\frac{1}{\rho^i})^2\sum_{l=1}^{n}[F_l(x^i)]_+^{2q} &= -\frac{1}{2}\sum_{l=1}^{n}(x_l^i F_l(x^i))^2 - \frac{1}{\rho^i}\sum_{l=1}^{n} x_l^i[F_l(x^i)]_+^{q+1} \\ &\le -\frac{1}{\rho^i}\sum_{l=1}^{n} x_l^i[F_l(x^i)]_+^{q+1},\end{aligned}$$

which means that

$$\sum_{l=1}^{n}[F_l(x^i)]_+^{2q} \le -2\rho^i \sum_{l=1}^{n} x_l^i [F_l(x^i)]_+^{q+1}. \tag{3.8}$$

Suppose that $\bar{x}$ is a limit point of the sequence $\{x^i\}$, so there is an infinite subsequence $\mathcal{K}$ such that $\bar{x} = \lim_{i \xrightarrow{\mathcal{K}} \infty} x^i \le 0$, which implies $\bar{x} \in \Omega$. By taking the limit as $i \xrightarrow{\mathcal{K}} \infty$, on both sides of (3.8), we have

$$\sum_{l=1}^{n}[F_l(\bar{x})]_+^{2q} = \lim_{i \xrightarrow{\mathcal{K}} \infty}\sum_{l=1}^{n}[F_l(x^i)]_+^{2q} \le -\lim_{i \xrightarrow{\mathcal{K}} \infty} 2\rho^i \sum_{l=1}^{n} x_l^i [F_l(x^i)]_+^{q+1} = 0,$$

where the last equality follows from $\rho^i \to 0^+$. Therefore, we have $F_l(\bar{x}) \le 0$ for all

$l \in \mathcal{J}$. Moreover, by taking the limit as $i \xrightarrow{\mathcal{K}} \infty$ in (3.7), we have

$$\begin{aligned}
&\sum_{l=1}^{n}(\bar{x}_l F_l(\bar{x}))^2 = \lim_{i \xrightarrow{\mathcal{K}} \infty} \sum_{l=1}^{n}(x_l^i F_l(x^i))^2 \\
&= -\lim_{i \xrightarrow{\mathcal{K}} \infty} \Big(\frac{2}{\rho^i}\sum_{l=1}^{n} x_l^i [F_l(x^i)]_+^{q+1} + (\frac{1}{\rho^i})^2 \sum_{l=1}^{n}[F_l(x^i)]_+^{2q}\Big), \\
&= -\lim_{i \xrightarrow{\mathcal{K}} \infty} \Big(-2\sum_{l=1}^{n}(x_l^i)^2 [F_l(x^i)]_+^{2} + (\frac{1}{\rho^i})^2 \sum_{l=1}^{n}[F_l(x^i)]_+^{2q}\Big) \le 0,
\end{aligned}$$

where the last equality follows from (2.3) that $x_l^i[F_l(x^i)]_+^{q+1} = -\rho^i (x_l^i)^2 [F_l(x^i)]_+^2$ for all $l \in \mathcal{J}$.

Therefore, we have proved that $\bar{x} \le 0$, $F(\bar{x}) \le 0$ and $\sum\limits_{l=1}^{n}(\bar{x}_l F_l(\bar{x}))^2 = 0$, that is, $\bar{x}$ is a solution of problem (1.1). □

THEOREM 3.3. *Suppose that the function $F : \mathbb{R}^n \to \mathbb{R}^n$ is a uniform ξ-P-function. Moreover, assume that x^ρ is a local solution of problem (3.1) for given $\rho > 0$ and satisfies $F(x^\rho) \le 0$. Then x^ρ is the solution of problem (1.1).*

Proof. Applying Proposition 3.1 at x^ρ for given $\rho > 0$, we have

$$\begin{cases} \frac{\partial \Psi(x^\rho,\rho)}{\partial x_i} = 0, & \text{if } x_i^\rho < 0, \\ \frac{\partial \Psi(x^\rho,\rho)}{\partial x_i} \le 0, & \text{if } x_i^\rho = 0, \end{cases}$$

which can be expressed by virtue of (3.3) as follows

$$\begin{cases} \big(\rho\Theta(x^\rho)\mathcal{F}(x^\rho,\rho) + \nabla F(x^\rho)^T \Pi(x^\rho,\rho)\mathcal{F}(x^\rho,\rho)\big)_i = 0, & \text{if } x_i^\rho < 0, \\ \big(\rho\Theta(x^\rho)\mathcal{F}(x^\rho,\rho) + \nabla F(x^\rho)^T \Pi(x^\rho,\rho)\mathcal{F}(x^\rho,\rho)\big)_i \le 0, & \text{if } x_i^\rho = 0. \end{cases} \tag{3.9}$$

Since x^ρ satisfies $F(x^\rho) \le 0$, it follows that $\Pi(x^\rho,\rho) = \rho \mathrm{diag}(x_1^\rho, \dots, x_n^\rho)$.

We first prove that $\mathcal{F}(x^\rho,\rho) = 0$. Assume on the contrary that $\mathcal{F}(x^\rho,\rho) \ne 0$. Then there exists at least one index $i \in \mathcal{J}$ such that $\mathcal{F}_i(x^\rho,\rho) \ne 0$. Without loss of generality, we assume $\mathcal{F}_1(x^\rho,\rho) \ne 0$ and $\mathcal{F}_i(x^\rho,\rho) = 0$ for all $i = 2, \dots, n$. Since $\mathcal{F}_1(x^\rho,\rho) = \rho x_1^\rho F_1(x^\rho)$, we see that $F_1(x^\rho) \ne 0$ and $x_1^\rho \ne 0$. It follows from (3.9) that

$$\Big(\rho\Theta(x^\rho)\mathcal{F}(x^\rho,\rho) + \nabla F(x^\rho)^T \Pi(x^\rho,\rho)\mathcal{F}(x^\rho,\rho)\Big)_1 = 0. \tag{3.10}$$

Thus,

$$\big(\Theta(x^\rho)\mathcal{F}(x^\rho,\rho)\big)_1 = \rho x_1^\rho F_1(x^\rho)^2 < 0 \text{ and } \big(\Pi(x^\rho,\rho)\mathcal{F}(x^\rho,\rho)\big)_1 = \rho^2 (x_1^\rho)^2 F_1(x^\rho) < 0.$$

It follows from equality (3.10) that

$$\big(\Pi(x^\rho,\rho)\mathcal{F}(x^\rho,\rho)\big)_1 \big(\nabla F(x^\rho)^T \Pi(x^\rho,\rho)\mathcal{F}(x^\rho,\rho)\big)_1 = -\rho^4 (x_1^\rho)^3 F_1(x^\rho)^3 < 0,$$

which contradicts the fact that $\nabla F(x^\rho)^T$ is a P_0-matrix (because the uniform ξ-P-function F is a P_0-function). Therefore, we proved that $\mathcal{F}(x^\rho,\rho) = 0$. Since further $x^\rho \le 0$ and $F(x^\rho) \le 0$, it follows from (2.3) that x^ρ is the solution of problem (1.1). □

In the next theorem, under the assumption of a uniform ξ-P-function on the function F, we prove that the merit function Ψ has bounded level sets for given $\rho > 0$.

THEOREM 3.4. *Suppose that the function $F : \mathbb{R}^n \to \mathbb{R}^n$ is a uniform ξ-P-function. Then the merit function $\Psi(x,\rho)$ is level-bounded for each $\rho > 0$.*

Proof. Suppose on the contrary that the level sets of $\Psi(x,\rho)$ are unbounded for given $\rho > 0$. Then there exist a sequence $\{x^k\}$ and a constant $\hat{\alpha} \geq 0$ such that $\lim_{k\to\infty} \|x^k\| = \infty$ and

$$\Psi(x^k,\rho) \leq \hat{\alpha}. \tag{3.11}$$

We define the index set $\mathcal{T} := \{i \in \mathcal{J} \mid \{x_i^k\} \text{ is unbounded}\}$. Since $\{x^k\}$ is unbounded, it follows that $\mathcal{T} \neq \emptyset$. Let $\{z^k\}$ denote a bounded sequence defined by:

$$z_i^k = \begin{cases} 0 & \text{if } i \in \mathcal{T}, \\ x_i^k & \text{if } i \notin \mathcal{T}. \end{cases}$$

By the definition of sequence $\{z^k\}$ and the assumption of a uniform ξ-P-function on F, there exist constants $\alpha > 0$, $\xi > 1$ and an index $\nu = \nu(x^k, z^k) \in \mathcal{J}$ such that

$$\begin{aligned} \alpha \sum_{i\in\mathcal{T}} (x_i^k)^\xi &= \alpha \|x^k - z^k\|^\xi \\ &\leq (x_\nu^k - z_\nu^k)(F_\nu(x^k) - F_\nu(z^k)) \\ &\leq \max_{i\in\mathcal{T}} x_i^k (F_i(x^k) - F_i(z^k)) \\ &= x_j^k (F_j(x^k) - F_j(z^k)) \\ &\leq |x_j^k||F_j(x^k) - F_j(z^k)|, \end{aligned} \tag{3.12}$$

where j is one of the indices at which the max is attained. Since $j \in \mathcal{T}$, we can assume, without loss of generality, that

$$\{|x_j^k|\} \to \infty. \tag{3.13}$$

Dividing by $|x_j^k|$ on both sides of inequality (3.12), we have

$$\alpha |x_j^k|^{\xi-1} \leq |F_j(x^k) - F_j(z^k)|,$$

this, in turn, implies

$$\{|F_j(x^k)|\} \to \infty, \tag{3.14}$$

since $F_j(z^k)$ is bounded. However, (3.13) and (3.14) imply that $\{|\mathcal{F}_j(x^k,\rho)|\} \to \infty$, which contradicts with (3.11). ◻

4. Numerical Experiments. In this section, we present numerical results of our proposed method described in Section 3 by using MATLAB R2011b. We conduct numerical testing on Windows XP with 3.00GB of main memory and Intel(R) Core(TM) 2 Duo 3.0GHz processors. We carry out the numerical experiments on the test problems from MCPLIB [3].

We refer to the implementation of Algorithm 1 as the CDLOP method, which stands for the Constrained Differentiable Lower Order Penalty method. For convenience, we write the CDLOP method with $p = 2$ and 100 as the CDLOP$_{1/2}$ and CDLOP$_{1/100}$ methods, respectively. We first compare the performance of the

DLOPP$_{1/2}$ method with the $\ell_{\frac{1}{2}}$-penalty method [31] and ℓ_1-penalty method [1] in terms of the number of function evaluations and the values of the penalty parameter ρ. Using the same terms, we compare the performance of the CDLOP method with different values of power $p = 1, 2, 100, 1000, 5000, 10000$. Finally, based on the number of function evaluations, we compare the performance of the proposed method with some well known approaches, such as the smooth approximation method [4, 5] and the nonsmooth equations method [20].

Before presenting our numerical results, we illustrate the implementation details for our method and other existing methods used in this section as follows.

A smoothing strategy in [31] is used to smooth out the non-Lipschitzian term in the $\ell_{\frac{1}{2}}$-penalized term. The smoothing $\ell_{\frac{1}{2}}$-penalty method is abbreviated as SLOP$_{1/2}$ method. The ℓ_1-penalty method employs the semismooth Newton method [28] to solve the corresponding ℓ_1-penalized equations. We refer to the implementation of ℓ_1-penalty method as the SSOOP$_1$ method, which stands for the SemiSmooth One Order Penalty method. A Matlab solver TRESNEI[1] developed by Morini and Porcelli [26] for bound-constrained (or unconstrained) nonlinear least squares problems is used to solve the corresponding least squares problems for the SLOP$_{1/2}$ and SSOOP$_1$ methods.

Throughout the experiments, we set parameters $\rho^0 = 1$, $\rho^{min} = 10^{-16}$, $\sigma = 0.1$ and $\epsilon = 10^{-6}$ in Algorithm 1. We use $\varepsilon_s = 10^{-22}$ for the value of smoothing factor in the SLOP$_{1/2}$ method. We follow all default parameters in the solver TRESNEI. Details can be found in [26].

We select 22 test problems from MCPLIB shown in Table 4.1, in which there are 7 linear complementarity problems. For each problem, we perform 100 runs from randomly generated starting points by a uniform distribution in a given interval. Therefore, we run each method on a set of 2200 test problems.

Table 4.1: Problem characteristics and starting intervals.

Problem	**Dim**	**Char**	**Interval**	**Problem**	**Dim**	**Char**	**Interval**
colvnlp	15	N	[-10,0]	cycle	1	N	[-10,0]
josephy	4	N	[-10,0]	kojshin	4	N	[-10,0]
mathisum	4	N	[-10,0]	powell	16	N	[-10,0]
scarfanum	13	N	[-1,0]	scarfsum	14	N	[-1,0]
sppe	27	N	[-10,0]	tobin	42	N	[-10,0]
billups	1	N	[-10,0]	colvdual	20	L	[-10,0]
degen	2	L	[-10,0]	hanskoop	14	N	[-10,0]
nash	10	N	[-10,0]	tinloi	146	L	[-1,0]
colvtemp	20	N	[-1,0]	oligomcp	6	N	[-10,0]
fathi	100	L	[-10,0]	murty	100	L	[-10,0]
primaldual	6	L	[-10,0]	explcp	16	L	[-10,0]

In Table 4.1, the **Problem** denotes the name of test problem, the **Dim** denotes the dimension of problem (1.1), the **Char** denotes the characterization of problem (1.1) where **N** denotes that F is nonlinear and **L** denotes that F is linear, and the **Interval** denotes the interval in which a starting point is generated by a uniform distribution.

[1] http://tresnei.de.unifi.it/.

Using the performance profiles of Dolan and Moré in [9], we plot Figure **??**, where the plots $\pi_s(\tau)$ denote the scaled performance profile

$$\pi_s(\tau) := \frac{\text{number of problems } \hat{p} \text{ where } \log_2(r_{\hat{p},s}) \leq \tau}{\text{total number of problems}},\ \tau \geq 0,$$

where $\log_2(r_{\hat{p},s})$ is the scaled performance ratio between the number of function evaluations to solve problem $\hat{p}$ by solver s over the fewest number of function evaluations required by the three solvers. It is clear that $\pi_s(\tau)$ is the probability for solver s that a scaled performance ratio $\log_2(r_{\hat{p},s})$ is within a factor $\tau \geq 0$ of the best possible ratio. See [9] for more details regarding the performance profiles.

Figure **??** indicates that the $\text{CDLOP}_{1/2}$ method is the most efficient method among them as its performance profile lies above all others for all performance ratios. Moreover, the $\text{CDLOP}_{1/2}$ method can solve the most test problems (about 88%) successfully. The $\text{SLOP}_{1/2}$ method is the weakest solver among them and can only solve about 55% test problems.

The performance profiles in Figure **??** are plotted on the values of $\frac{1}{\rho}$. Figure **??** indicates the $\text{CDLOP}_{1/2}$ method can solved about 68% test problems with the biggest values of the penalty parameter ρ. The fewest test problems (about 27%) can be solved by the SSOOP_1 method with the biggest penalty parameter ρ. However, the SSOOP_1 method is more robust than the $\text{SLLOP}_{1/2}$ method.

We plot Figures **??** and **??** to compare performance of the CDLOP method with different values of p in terms of the number of function evaluations and the values of the penalty parameter.

Figure **??** indicates that the CDLOP method with $p = 100$ can solve about 60% test problems with the least number of function evaluations and is the most efficient solver among them. We also see that the number of function iterations used by the CDLOP method decreases dramatically as the power p increases from 2 to 100. Slight changes happen on the performance profiles as we increase p from 100 to 10000. Furthermore, there are nearly the same test problems (about 90%) that can be solved successfully by the CDLOP method with different values of p.

The performance profiles in Figure **??** are plotted on the values of $\frac{1}{\rho}$, which indicates that the CDLOP method with $p = 1$ uses the smallest values of penalty parameter. Bigger values of the penalty parameter ρ are used by the CDLOP method as we increase p from 1 to 100, which verifies the conclusion of Theorem 2.17.

Next, we use the CDLOP method with $p = 100$ to compare its performance with the smooth approximation method and the nonsmooth equations method in terms of the number of function evaluations. The Zang smooth plus function [34] is used in the smooth approximation method to smooth its normal equations. The nonsmooth equations method employs the semismooth Newton method [28] to solve its nonsmooth equations. We write SAM and NSEM to denote the Smooth Approximation and Nonsmooth Equations Methods, respectively. Moreover, the solver TRESNEI is used to solve the corresponding least squares problems for the last two methods.

Figure **??** indicates that the SAM can solve about 47% test problems with the least number of function evaluations. However, the fewest problems can be successfully solved by this method. The NSEM is more efficient than the SAM. The CDLOP with $p = 100$ can successfully solve the most test problems among them.

Acknowledgments The authors would like to thank the referees for the valuable

suggestions which have improved the early version of the manuscript.

REFERENCES

[1] A. Bensoussan and J. L. Lions. *Applications of Variational Inequalities in Stochastic Control*. North Holland, 1982.

[2] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1999.

[3] S. C. Billups, S. P. Dirkse, and M. C. Ferris. A comparison of large scale mixed complementarity problem solvers. *Computational Optimization and Applications*, 7(1):3–25, 1997.

[4] B. Chen and P. T. Harker. Smooth approximations to nonlinear complementarity problems. *SIAM Journal on Optimization*, 7(2):403–420, 1997.

[5] C. Chen and O. L. Mangasarian. A class of smoothing functions for nonlinear and mixed complementarity problems. *Computational Optimization and Applications*, 5(2):97–138, 1996.

[6] A. R. Conn, N. I. M. Gould, and P. L. Toint. *Trust Region Methods*. SIAM, 1987.

[7] R. W. Cottle, J. S. Pang, and R. E. Stone. *The Linear Complementarity Problem*. SIAM, 2009.

[8] T. De Luca, F. Facchinei, and C. Kanzow. A semismooth equation approach to the solution of nonlinear complementarity problems. *Mathematical Programming*, 75(3):407–439, 1996.

[9] E. D. Dolan and J. J. Moré. Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91(2):201–213, 2002.

[10] F. Facchinei and J. S. Pang. *Finite-dimensional Variational Inequalities and Complementarity Problems, Volume I* . Springer Verlag, 2003.

[11] F. Facchinei and J. S. Pang. *Finite-dimensional Variational Inequalities and Complementarity Problems, Volume II*. Springer Verlag, 2003.

[12] P. L. Fackler. *Applied Computational Economics and Finance*. The MIT Press, 2002.

[13] M. C. Ferris and J. S. Pang. Engineering and economic applications of complementarity problems. *SIAM Review*, 39(4):669–713, 1997.

[14] A. Fischer. A special Newton-type optimization method. *Optimization*, 24(3-4):269–284, 1992.

[15] R. Fletcher. *Practical Methods of Optimization*. Wiley, 2013.

[16] 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(1-3):161–220, 1990.

[17] C. C. Huang and S. Wang. A power penalty approach to a nonlinear complementarity problem. *Operations Research Letters*, 38(1):72–76, 2010.

[18] C. C. Huang and S. Wang. A penalty method for a mixed nonlinear complementarity problem. *Nonlinear Analysis: Theory, Methods and Applications*, 75(2):588–597, 2012.

[19] X. X. Huang and X. Q. Yang. A unified augmented Lagrangian approach to duality and exact penalization. *Mathematics of Operations Research*, 28(3):533–552, 2003.

[20] H. Y. Jiang and L. Q. Qi. A new nonsmooth equations approach to nonlinear complementarity problems. *SIAM Journal on Control and Optimization*, 35(1):178–193, 1997.

[21] C. Kanzow, N. Yamashita, and M. Fukushima. New NCP-functions and their properties. *Journal of Optimization Theory and Applications*, 94(1):115–135, 1997.

[22] C. Kanzow, N. Yamashita, and M. Fukushima. Levenberg–Marquardt methods with strong local convergence properties for solving nonlinear equations with convex constraints. *Journal of Computational and Applied Mathematics*, 173(2):321–343, 2005.

[23] I. V. Konnov. Properties of gap functions for mixed variational inequalities. *Siberian J. Numerical Mathematics*, 3(3):259–270, 2000.

[24] M. Macconi, B. Morini, and M. Porcelli. Trust-region quadratic methods for nonlinear systems of mixed equalities and inequalities. *Applied Numerical Mathematics*, 59(5):859–876, 2009.

[25] J. Moré and W. Rheinboldt. On P- and S-functions and related classes of n-dimensional nonlinear mappings. *Linear Algebra and its Applications*, 6:45–68, 1973.

[26] B. Morini and M. Porcelli. TRESNEI, a Matlab trust-region solver for systems of nonlinear equalities and inequalities. *Computational Optimization and Applications*, 51(1):27–49, 2012.

[27] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Verlag, 2006.

[28] L. Q. Qi and J. Sun. A nonsmooth version of Newton's method. *Mathematical Programming*, 58(1):353–367, 1993.

[29] A. M. Rubinov and X. Q. Yang. *Lagrange-type Functions in Constrained Non-convex Optimization*. Springer, 2003.

[30] S. Wang and X. Q. Yang. A power penalty method for linear complementarity problems.

Operations Research Letters, 36(2):211–214, 2008.

[31] S. Wang, X. Q. Yang, and K. L. Teo. Power penalty method for a linear complementarity problem arising from American option valuation. *Journal of Optimization Theory and Applications*, 129(2):227–254, 2006.

[32] X. Q. Yang and X. X. Huang. A nonlinear Lagrangian approach to constrained optimization problems. *SIAM Journal on Optimization*, 11(4):1119–1144, 2001.

[33] C. J. Yu, K. L. Teo, L. S. Zhang, and Y. Q. Bai. A new exact penalty function method for continuous inequality constrained optimization problems. *Journal of Industrial and Management Optimization*, 6(4):895, 2010.

[34] I. Zang. A smoothing-out technique for min-max optimization. *Mathematical Programming*, 19(1):61–77, 1980.

[35] K. Zhang. *American Option Pricing and Penalty Methods.* PhD thesis, The Hong Kong Polytechnic University, 2006.

[36] R. Zvan, P. A. Forsyth, and K. R. Vetzal. Penalty methods for American options with stochastic volatility. *Journal of Computational and Applied Mathematics*, 91(2):199–218, 1998.