AN UNCONSTRAINED DIFFERENTIABLE PENALTY METHOD FOR IMPLICIT COMPLEMENTARITY PROBLEMS

BOSHI TIAN*, DONGHUI LI†, AND XIAOQI YANG‡

Abstract. In this paper, we introduce an unconstrained differentiable penalty method for solving implicit complementarity problems, which has an exponential convergence rate under the assumption of a uniform ξ-P-function. Instead of solving the unconstrained penalized equations directly, we consider a corresponding unconstrained optimization problem and apply the trust-region Gauss-Newton method to solve it. We prove that the local solution of the unconstrained optimization problem identifies that of the complementarity problems under monotone assumptions. We carry out numerical experiments on the test problems from MCPLIB, and show that the proposed method is efficient and robust.

Key words. implicit complementarity problems, lower-order penalty method, exponential convergence rate, trust-region Gauss-Newton method

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

1. Introduction.

Consider the following implicit complementarity problem (ICP, for short) [24, 25], which is to find a vector $x \in \mathbb{R}^n$ satisfying the following conditions,

$$H(x) \leq 0,\ F(x) \leq 0,\ \langle H(x), F(x)\rangle = 0, \tag{1.1}$$

where functions $H, F : \mathbb{R}^n \to \mathbb{R}^n$ are assumed to be continuously differentiable and $\langle y, z\rangle$ denotes the inner product for any vectors $y, z \in \mathbb{R}^n$. Specially, as $H(x) := x$, problem (1.1) reduces to the nonlinear complementarity problem (NCP, for short). Moreover, problem (1.1) becomes a linear complementarity problem (LCP, for short) as $H(x) := x$ and F is an affine function, i.e., $F(x) := Ax - b$ for a given matrix $A \in \mathbb{R}^{n\times n}$ and a vector $b \in \mathbb{R}^n$. Complementarity problems play an important role in operations research, option pricing, economic equilibrium models and the engineering sciences; see, e.g., [10, 11, 14].

Comprehensive studies for complementarity problems have been done, see monographs [6, 8, 9] and the references therein. Two differentiable minimization formulations for problem (1.1) were studied by Tseng et al. [29] by virtue of the Fukushima's merit function for variational inequality problems [13] and the Mangasarian and Solodov's implicit Lagrangian function [21]. Peng [26] not only extended Fukushima's merit function for variational inequality but also presented a new way to construct the merit functions for problem (1.1). Kanzow and Fukushima [19] employed the Fischer's function [12] to transform problem (1.1) into an unconstrained minimization formulation. They presented mild conditions to guarantee that the global minimizer of the unconstrained problem coincides with the solution of problem (1.1). The unconstrained formulation above was further investigated by Jiang et al. [17], where a trust-region method was proposed for solving

*Business School of Hunan University, Changsha, Hunan, China (tianboshi@126.com). Research of this author was partially supported by the Fundamental Research Funds for the Central Universities(531107050790), the NSF(71501067) of China and the State Key Program of National Natural Science of China(Grant No.71431008).

†School of Mathematical Sciences, South China Normal University, Guangzhou, 510631, P. R. China (dhli@scnu.edu.cn). Research of this author was supported by the NSF(11371154) of China.

‡Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, P.R. China (mayangxq@polyu.edu.hk). Research of this author was partially supported by the Research Committee of Hong Kong Polytechnic University (G-YBCN) and NSF(11431004) of China.

problem (1.1) and the global convergence and local Q-superlinear convergence were established under a nonsingularity assumption.

During the last decade, the $\ell_{\frac{1}{p}}(p > 1)$-penalty method has attracted wide attention for the LCP [31, 32, 34], the NCP [4, 15, 30] and the mixed nonlinear complementarity problem (MiCP, for short) [16, 35]. Furthermore, some desirable results on the convergence rate were established under mild conditions. However, the non-Lipschitzian of the $\ell_{\frac{1}{p}}$-penalized equations makes the classical numerical methods lose their efficiency. In order to overcome this drawback, a box-constrained differentiable penalty method was proposed in [28]. It is not proper to use the box-constrained differentiable penalty method to solve problem (1.1) directly as the corresponding constraint set is nonconvex. It is well known that optimization problems with the nonlinear and nonconvex constraint set are much harder to solve than optimization problems with the box constraint. Alternately, problem (1.1) can be reformulated as a MiCP by virtue of artificial variables, which can be solved by the $\ell_{\frac{1}{p}}$-penalty method [16]. The box-constrained differentiable penalty method can also be used to solve problem (1.1) by introducing artificial variables, which however doubles the number of nonlinear equations.

In this paper, we propose an unconstrained differentiable penalty method for solving problem (1.1) without introducing any artificial variables. Specifically, we consider the system of penalized equations as follows:

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

where $\rho > 0$ is the penalty parameter, $p \geq 1$ is the power, $[z]_+^\sigma$ denotes a vector with components $([z]_+^\sigma)_i = \max\{z_i, 0\}^\sigma$, for all $i \in \mathcal{I}$, for any given vector $z \in \mathbb{R}^n$ and constant $\sigma > 0$, and $H(x) \circ F(x)$ is the Hadamard (or Schur) product of vectors $H(x)$ and $F(x)$ with components $\big(H(x) \circ F(x)\big)_i = H_i(x)F_i(x)$, for all $i \in \mathcal{I}$. We establish the exponential convergence rate of $\mathcal{O}(\rho^{\frac{p}{\xi}})$ between a solution x^ρ of system (1.2) and the solution x^* of problem (1.1) under the assumption of a uniform ξ-P-function, where $\xi \in (1, 2]$ is a constant.

In order to design globally convergent methods allowing arbitrary starting points to solve problem (1.1), we do not solve system (1.2) directly and consider a corresponding unconstrained optimization problem and apply the trust-region Gauss-Newton method to solve it. Furthermore, we prove that the local solution of the unconstrained optimization problem identifies the solution of problem (1.1) under the assumption of monotonicity.

We carry out numerical experiments on test problems from MCPLIB [2]. We first compare the performance of the proposed method with $p = 2$ with the box-constrained differentiable penalty method and the ℓ_1-penalty method [1] in terms of the number of function evaluations and the values of the penalty parameter. Furthermore, different values of the power $p = 1, 2, 100, 1000, 5000, 10000$ are chosen to compare the efficiency of the proposed method. Finally, we compare the performance of the proposed method with the smooth approximation method [3], and the nonsmooth equations method [18] in terms of the number of function evaluations.

This paper is organized as follows. In Section 2, we introduce an unconstrained differentiable penalty method and establish its exponential convergence rate. In Section 3, we present our unconstrained optimization problem formulation to solve problem (1.1). In the last section, we show our numerical results.

Notation Throughout this paper, we write $\mathcal{I} := \{1, 2, \ldots, n\}$ to indicate the index set and use $\|\cdot\|$ to indicate the Euclidean norm. Given vectors $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$. We write $x \circ y$ to indicate the Hadamard product of vectors x and y, that is, $x \circ y := (x_1 y_1, \ldots, x_n y_n)^T$. we write $\binom{x}{y}$ to indicate a vector in $\mathbb{R}^{n \times n}$, that is, $\binom{x}{y} := (x^T\ y^T)^T$. We write $\max x$ to indicate that the maximum element of vector x, that is, $\max x := \max\limits_{1 \le i \le n} x_i$.

2. Unconstrained Differentiable Penalty Method. In this section, an unconstrained differentiable penalty method is proposed for problem (1.1). Then, we establish the exponential convergence rate under mild assumptions. Some important definitions are recalled as follows.

DEFINITION 2.1 ([8, Definition 2.3.1]). *A function $S : \mathbb{R}^n \to \mathbb{R}^n$ is said to be*

- *monotone if*

$$(x-y)^T(S(x)-S(y)) \ge 0,\ \forall\, x, y \in \mathbb{R}^n;$$

- *strictly monotone if*

$$(x-y)^T(S(x)-S(y)) > 0,\ \forall\, x, y \in \mathbb{R}^n \text{ and } x \neq y;$$

- *ξ-monotone for some $\xi \in (1,2]$, if there exists a constant $\alpha > 0$ such that*

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

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

DEFINITION 2.2 ([28, Definition 2.4]). *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 $\mathbb{R}^n$,*

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

Definitions 2.1 and 2.2 indicate that the uniform ξ-P-function is weaker than the ξ-monotonicity.

PROPOSITION 2.3 ([8, Proposition 2.3.2]). *Let $S : D \subset \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable on the open convex set D,*

- *S is monotone on D if and only if $\nabla S(x)$ is positive semidefinite for all $x \in D$;*
- *S is strictly monotone on D if and only if $\nabla S(x)$ is positive definite for all $x \in D$.*

Let $y = H(x)$. Problem (1.1) can be recast as a MiCP as follows:

$$\begin{aligned} H(x) - y &= 0, \\ F(x) &\le 0, \\ \langle F(x), y\rangle &= 0, \\ y &\le 0. \end{aligned} \tag{2.1}$$

We consider the penalized equations of problem (2.1) as follows

$$\begin{pmatrix} H(x) - y \\ \rho F(x) \circ y + [y]_+^{1+\frac{1}{p}} + [F(x)]_+^{1+\frac{1}{p}} \end{pmatrix} = 0. \tag{2.2}$$

where $\rho > 0$ is the penalty parameter.

A close relationship between the solution of system (1.2) and that of system (2.2) is stated in the next proposition. Here, we omit its proof.

PROPOSITION 2.4. *For each $\rho > 0$, we say that $x^\rho \in \mathbb{R}^n$ is a solution of penalized equations (1.2) if and only if there exists some $y^\rho \in \mathbb{R}^n$ satisfying $y^\rho = H(x^\rho)$ such that $\binom{x^\rho}{y^\rho}$ is a solution of system (2.2).*

We present an example to show that the solution of system (1.2) is not unique even if problem (1.1) has a unique solution.

EXAMPLE 2.1. *Let $F(x) = x + 1$ and $H(x) = x$ with $x \in \mathbb{R}$ in problem (1.1). We have that $x^* = -1$ is the unique solution. Take $p = 1$ in (1.2). The penalized equation is $\rho x(x+1) + [x]_+^2 + [x+1]_+^2 = 0$. After computing, we see that $\bar{x}^\rho = -1$ and $\hat{x}^\rho = -\frac{1}{\rho+1}$ are two solutions of this penalized equation.*

Next, we prove that the solution of system (2.2) converges to a solution of problem (2.1) at an exponential convergence rate under the assumption of a uniform ξ-P-function. Before doing this, we first show some useful lemmas. We define a vector-valued function $Z : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ as

$$Z(x, y) = \begin{pmatrix} H(x) - y \\ F(x) \end{pmatrix}.$$

LEMMA 2.5. *Assume that Z is a uniform ξ-P-function. For each $\rho > 0$, let $\binom{x^\rho}{y^\rho}$ be a solution of system (2.2). Then there exists a constant $M > 0$, independent of $\binom{x^\rho}{y^\rho}$, ρ and p, such that*

$$\left\| \begin{pmatrix} x^\rho \\ y^\rho \end{pmatrix} \right\| \le M.$$

Proof. It follows from $\binom{x^\rho}{y^\rho}$ solving system (2.2) that we have $H(x^\rho) - y^\rho = 0$ and

$$\rho F_i(x^\rho) y_i^\rho + [y_i^\rho]_+^{1+\frac{1}{p}} + [F_i(x^\rho)]_+^{1+\frac{1}{p}} = 0, \ \forall\, i \in \mathcal{I}. \tag{2.3}$$

Then we see that $\big(H(x^\rho) - y^\rho\big) \circ x^\rho = 0$ and $F(x^\rho) \circ y^\rho \le 0$, which means that $\binom{x^\rho}{y^\rho} \circ Z(x^\rho, y^\rho) \le 0$. By Definition 2.2 of a uniform ξ-P-function, there exist constants $\alpha > 0$ and $\xi > 1$ such that

$$\begin{aligned} \alpha \left\| \begin{pmatrix} x^\rho \\ y^\rho \end{pmatrix} \right\|^\xi &\le \max \begin{pmatrix} x^\rho \\ y^\rho \end{pmatrix} \circ \Big(Z(x^\rho, y^\rho) - Z(0,0) \Big) \\ &\le \max - \begin{pmatrix} x^\rho \\ y^\rho \end{pmatrix} \circ Z(0,0) \\ &\le \left\| \begin{pmatrix} x^\rho \\ y^\rho \end{pmatrix} \right\| \|Z(0,0)\|_\infty. \end{aligned}$$

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

Lemma 2.5 shows that the solution of problem (2.2) always lies in a bounded closed set for each $\rho > 0$. By the continuity of Z, there exists a positive constant L, independent of $\binom{x^\rho}{y^\rho}$, ρ and p, such that

$$\|Z(x^\rho, y^\rho)\| \le L. \tag{2.4}$$

Lemma 2.6. *Assume that Z is a uniform ξ-P-function. For each $\rho > 0$, let $\binom{x^\rho}{y^\rho}$ be a solution of system (2.2). Then there exist constants $C_1 > 0$ and $C_2 > 0$, independent of $\binom{x^\rho}{y^\rho}$ and ρ, such that*

$$\|[y^\rho]_+\| \leq C_1\rho^p \text{ and } \|[F(x^\rho)]_+\| \leq C_2\rho^p.$$

Proof. It follows from (2.3) that we have

$$\begin{aligned}[y_i^\rho]_+^{1+\frac{1}{p}} &= -\rho F_i(x^\rho)y_i^\rho - \rho[F_i(x^\rho)]_+^{1+\frac{1}{p}} \\ &\leq -\rho F_i(x^\rho)y_i^\rho \leq \rho\|F(x^\rho)\|_\infty\|y^\rho\|_\infty,\end{aligned}$$

for all $i \in \mathcal{I}$. Thus, $\|[y^\rho]_+\|_\infty \leq \rho^p\|F(x^\rho)\|_\infty^p$. By the fact that all norms in $\mathbb{R}^n$ are equivalent, there exists a constant $\widetilde{C} > 0$ such that $\|[y^\rho]_+\| \leq \widetilde{C}\|[y^\rho]_+\|_\infty$. It follows from inequality (2.4) that we have $\|F(x^\rho)\|_\infty^p \leq L^p$. Thus, $\|[y^\rho]_+\| \leq C_1\rho^p$ with $C_1 = \widetilde{C}L^p$. Similarly, we can prove that $\|[F(x^\rho)]_+\| \leq C_2\rho^p$ with $C_2 = \widetilde{C}M^p$. ◻

Theorem 2.7. *Assume that Z is a uniform ξ-P-function. For each $\rho > 0$, let $\binom{x^\rho}{y^\rho}$ be a solution of system (2.2) and $\binom{x^*}{y^*}$ be a solution of problem (2.1). Then there exists a constant $\widehat{C} > 0$, independent of $\binom{x^\rho}{y^\rho}$ and ρ, such that*

$$\left\| \begin{pmatrix} x^* \\ y^* \end{pmatrix} - \begin{pmatrix} x^\rho \\ y^\rho \end{pmatrix} \right\| \leq \widehat{C}\rho^{\frac{p}{\xi}}.$$

Proof. We define the index sets at $\binom{x^\rho}{y^\rho}$ as follows

$$\begin{aligned}
\mathcal{Y}_a^\rho &= \{i \in \mathcal{I} \mid y_i^\rho = 0,\ F_i(x^\rho) > 0\};\\
\mathcal{Y}_b^\rho &= \{i \in \mathcal{I} \mid y_i^\rho = 0,\ F_i(x^\rho) = 0\};\\
\mathcal{Y}_c^\rho &= \{i \in \mathcal{I} \mid y_i^\rho = 0,\ F_i(x^\rho) < 0\};\\
\mathcal{Y}_d^\rho &= \{i \in \mathcal{I} \mid y_i^\rho > 0,\ F_i(x^\rho) > 0\};\\
\mathcal{Y}_e^\rho &= \{i \in \mathcal{I} \mid y_i^\rho > 0,\ F_i(x^\rho) = 0\};\\
\mathcal{Y}_f^\rho &= \{i \in \mathcal{I} \mid y_i^\rho > 0,\ F_i(x^\rho) < 0\};\\
\mathcal{Y}_g^\rho &= \{i \in \mathcal{I} \mid y_i^\rho < 0,\ F_i(x^\rho) > 0\};\\
\mathcal{Y}_h^\rho &= \{i \in \mathcal{I} \mid y_i^\rho < 0,\ F_i(x^\rho) = 0\};\\
\mathcal{Y}_s^\rho &= \{i \in \mathcal{I} \mid y_i^\rho < 0,\ F_i(x^\rho) < 0\}.
\end{aligned}$$

Since that $\binom{x^\rho}{y^\rho}$ solves system (2.2), the sets $\mathcal{Y}_a^\rho$, $\mathcal{Y}_d^\rho$, $\mathcal{Y}_e^\rho$ and $\mathcal{Y}_s^\rho$ are empty. Let $\Lambda := \mathcal{Y}_b^\rho \cup \mathcal{Y}_c^\rho \cup \mathcal{Y}_f^\rho$ and $\Gamma := \mathcal{Y}_g^\rho \cup \mathcal{Y}_s^\rho$. Then $\mathcal{I} = \Lambda \cup \Gamma$.

We first prove that the next inequality holds for all $i \in \Lambda$,

$$\Big(y_i^* + [y_i^\rho]_-\Big)\Big(F_i(x^*) - F_i(x^\rho)\Big) \leq 0, \tag{2.5}$$

where $[a]_- := \max\{-a, 0\}$ for all $a \in \mathbb{R}$.

(I) Let $i \in \mathcal{Y}_b^\rho$. Then

$$\Big(y_i^* + [y_i^\rho]_-\Big)\Big(F_i(x^*) - F_i(x^\rho)\Big) = y_i^*F_i(x^*) \leq 0.$$

(II) Let $i \in \mathcal{Y}_c^\rho$. Then

$$\begin{aligned}&\Big(y_i^* + [y_i^\rho]_-\Big)\Big(F_i(x^*) - F_i(x^\rho)\Big)\\ &= y_i^* F_i(x^*) - y_i^* F_i(x^\rho) + [y_i^\rho]_- F_i(x^*) - [y_i^\rho]_- F_i(x^\rho)\\ &= -y_i^* F_i(x^\rho) \leq 0.\end{aligned}$$

(III) Let $i \in \mathcal{Y}_f^\rho$. Then

$$\begin{aligned}&\Big(y_i^* + [y_i^\rho]_-\Big)\Big(F_i(x^*) - F_i(x^\rho)\Big)\\ &= y_i^* F_i(x^*) - y_i^* F_i(x^\rho) + [y_i^\rho]_- F_i(x^*) - [y_i^\rho]_- F_i(x^\rho)\\ &= -y_i^* F_i(x^\rho) \leq 0.\end{aligned}$$

Next, we prove that the following inequality holds for all $i \in \Gamma$,

$$\Big(y_i^* - y_i^\rho\Big)\Big(F_i(x^*) + [F_i(x^\rho)]_-\Big) \leq 0. \tag{2.6}$$

(I) Let $i \in \mathcal{Y}_g^\rho$. Then

$$\begin{aligned}&\Big(y_i^* - y_i^\rho\Big)\Big(F_i(x^*) + [F_i(x^\rho)]_-\Big)\\ &= y_i^* F_i(x^*) + y_i^* [F_i(x^\rho)]_- - y_i^\rho F_i(x^*) - y_i^\rho [F_i(x^\rho)]_-\\ &= -y_i^\rho F_i(x^*) \leq 0.\end{aligned}$$

(II) Let $i \in \mathcal{Y}_h^\rho$. Then

$$\begin{aligned}&\Big(y_i^* - y_i^\rho\Big)\Big(F_i(x^*) + [F_i(x^\rho)]_-\Big)\\ &= y_i^* F_i(x^*) + y_i^* [F_i(x^\rho)]_- - y_i^\rho F_i(x^*) - y_i^\rho [F_i(x^\rho)]_-\\ &= -y_i^\rho F_i(x^*) \leq 0.\end{aligned}$$

Therefore, we have

$$\begin{aligned}&\max_{i\in\Lambda}\, (y_i^* - y_i^\rho)(F_i(x^*) - F_i(x^\rho))\\ &= \max_{i\in\Lambda} (y_i^* + [y_i^\rho]_- - [y_i^\rho]_+)(F_i(x^*) - F_i(x^\rho))\\ &\leq \max_{i\in\Lambda} -[y_i^\rho]_+ (F_i(x^*) - F_i(x^\rho))\\ &\leq \|[y^\rho]_+\| \|\big(F(x^*) - F(x^\rho)\big)\|_\infty\\ &\leq C_1 \rho^p \|\big(F(x^*) - F(x^\rho)\big)\|_\infty\\ &\leq 2C_1 L \rho^p,\end{aligned}$$

where the first inequality is from (2.5) and the third inequality comes from Lemma 2.6.

Moreover, we have

$$\begin{aligned}&\max_{i\in\Gamma} (y_i^* - y_i^\rho)(F_i(x^*) - F_i(x^\rho))\\ &= \max_{i\in\Gamma} (y_i^* - y_i^\rho)(F_i(x^*) - [F_i(x^\rho)]_+ + [F_i(x^\rho)]_-)\\ &\leq \max_{i\in\Gamma} -[F_i(x^\rho)]_+ (y_i^* - y_i^\rho)\\ &\leq \|[F(x^\rho)]_+\| \|y^* - y^\rho\|_\infty\\ &\leq C_2 \rho^p \|y^* - y^\rho\|_\infty\\ &\leq 2C_2 M_1 \rho^p,\end{aligned}$$

where the first inequality and the third inequality are from (2.6) and Lemma 2.6, respectively.

By Definition 2.2 of a uniform ξ-P-function, there exist constants $\alpha > 0$ and $\xi > 1$ such that

$$\begin{aligned}
&\alpha \left\| \begin{pmatrix} x^* \\ y^* \end{pmatrix} - \begin{pmatrix} x^\rho \\ y^\rho \end{pmatrix} \right\|^\xi \\
&\leq \max \begin{pmatrix} x^* - x^\rho \\ y^* - y^\rho \end{pmatrix} \circ \Big(Z(x^*, y^*) - Z(x^\rho, y^\rho) \Big) \\
&= \max \Big\{ 0, \max_{i \in \Lambda \cup \Gamma} (y_i^* - y_i^\rho) \big(F_i(x^*) - F_i(x^\rho) \big) \Big\} \\
&\leq \widehat{C} \rho^p,
\end{aligned}$$

where $\widehat{C} = \max \left\{ \sqrt[\xi]{\frac{2C_1 L}{\alpha}}, \sqrt[\xi]{\frac{2C_2 M_1}{\alpha}} \right\}$. ◻

THEOREM 2.8. *Assume that H and F are uniform ξ-P-functions. For each $\rho > 0$, let x^ρ be a solution of system (1.2) and x^* be the solution of problem (1.1). Then there exists a constant $\widetilde{C}_1 > 0$, independent of x^ρ and ρ, such that*

$$\|x^* - x^\rho\| \leq \widetilde{C}_1 \rho^{\frac{p}{\xi}}.$$

Proof. It follows from x^* solving problem (1.1) that there exists $y^* \in \mathbb{R}^n$ satisfying $y^* = H(x^*)$ such that $\begin{pmatrix} x^* \\ y^* \end{pmatrix}$ is a solution of problem (2.1). Since x^ρ is a solution of system (1.2), by Proposition 2.4, there exists some $y^\rho \in \mathbb{R}^n$ such that $\begin{pmatrix} x^\rho \\ y^\rho \end{pmatrix}$ is a solution of system (2.2). Therefore, by use of Theorem 2.7, we conclude that $\|x^* - x^\rho\| \leq \left\| \begin{pmatrix} x^* \\ y^* \end{pmatrix} - \begin{pmatrix} x^\rho \\ y^\rho \end{pmatrix} \right\| \leq \widetilde{C}_1 \rho^{\frac{p}{\xi}}$ with $\widetilde{C}_1 = \max \left\{ \sqrt[\xi]{\frac{2C_1 L}{\alpha}}, \sqrt[\xi]{\frac{2C_2 M_1}{\alpha}} \right\}$. ◻

Problem (2.1) also can be solved by the box-constrained differentiable penalty method proposed in [28]. Specifically, we consider the next system of penalized equations with box-constraint

$$\begin{pmatrix} H(x) - y \\ \rho F(x) \circ y + [F(x)]_+^{1+\frac{1}{p}} \end{pmatrix} = 0, \ y \in \Omega, \tag{2.7}$$

where $\Omega := \{y \in \mathbb{R}^n \mid y \leq 0\}$.

We establish its exponential convergence rate as follows.

THEOREM 2.9. *Assume that H and F are uniform ξ-P-functions. For each $\rho > 0$, let $\begin{pmatrix} x^\rho \\ y^\rho \end{pmatrix}$ be a solution of system (2.7) and x^* be the solution of problem (1.1). Then there exists a constant $\widehat{C} > 0$, independent of $\begin{pmatrix} x^\rho \\ y^\rho \end{pmatrix}$ and ρ, such that*

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

Proof. Since $\begin{pmatrix} x^\rho \\ y^\rho \end{pmatrix}$ solves system (2.7), we see that it is a solution of system (2.2). By use of Theorems 2.7 and 2.8, we conclude that this theorem holds. ◻

3. Numerical Method. In this section, instead of solving the penalized equations (2.2) directly, we consider the corresponding optimization problem

$$\min_x \Psi(x,\rho) := \frac{1}{2}\|\mathcal{G}(x,\rho)\|^2. \tag{3.1}$$

For each $\rho > 0$, a local solution x^ρ of problem (3.1) satisfies that $\nabla\mathcal{G}(x^\rho,\rho)^T\mathcal{G}(x^\rho,\rho) = 0$, where $\nabla\mathcal{G}(x,\rho)$ is the Jacobian matrix of $\mathcal{G}(x,\rho)$, which can be expressed as

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

where $\nabla F(x)$ and $\nabla H(x)$ are the Jacobian matrices of $F(x)$ and $H(x)$, respectively, $\Theta(x,\rho) := \mathbf{diag}(R_1(x,\rho),\ldots,R_n(x,\rho))$ and $\Pi(x,\rho) := \mathbf{diag}(Q_1(x,\rho),\ldots,Q_n(x,\rho))$ are diagonal matrices with

$$R_i(x,\rho) := \rho F_i(x)+(1+\frac{1}{p})[H_i(x)]_+^{\frac{1}{p}} \text{ and } Q_i(x,\rho) := \rho H_i(x)+(1+\frac{1}{p})[F_i(x)]_+^{\frac{1}{p}},\ \forall\, i\in\mathcal{I}.$$

3.1. Convergence Analysis. In this subsection, we establish the connection between the solution of the problem (3.1) and that of problem (1.1).

THEOREM 3.1. *Assume that H and F are uniform ξ-P-functions. Moreover, suppose that $x^i \in \mathbb{R}^n$ is a global solution of problem (3.1) for each $\rho^i > 0$ 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 a solution of problem (1.1). Then we have $\Psi(x^*,\rho) = 0$ for each $\rho > 0$. Therefore, we have $\Psi(x^i,\rho^i) \le \Psi(x^*,\rho^i) = 0$, which implies that $\Psi(x^i,\rho^i) = 0$. Specifically, we have

$$\begin{aligned}
&\frac{1}{2}\sum_{l=1}^n \left((\rho^i)^2 H_l(x^i)^2F_l(x^i)^2 + [H_l(x^i)]_+^{2+\frac{2}{p}} + [F_l(x^i)]_+^{2+\frac{2}{p}}\right)\\
&+\sum_{l=1}^n \left(\rho^i F_l(x^i)[H_l(x^i)]_+^{2+\frac{1}{p}} + \rho^i H_l(x^i)[F_l(x^i)]_+^{2+\frac{1}{p}}\right)\\
&+\sum_{l=1}^n [H_l(x^i)]_+^{1+\frac{1}{p}}[F_l(x^i)]_+^{1+\frac{1}{p}} = 0.
\end{aligned} \tag{3.2}$$

Suppose that $\bar{x}$ is a limit point of the sequence $\{x^i\}$, so there exists an infinite subsequence $\mathcal{K}$ such that $\bar{x} = \lim_{i\xrightarrow{\mathcal{K}}\infty} x^i$. By taking the limit as $i \xrightarrow{\mathcal{K}} \infty$ on both sides of the last equation, we obtain

$$\frac{1}{2}\sum_{l=1}^n \left([H_l(\bar{x})]_+^{2+\frac{2}{p}} + [F_l(\bar{x})]_+^{2+\frac{2}{p}}\right) + \sum_{l=1}^n [H_l(\bar{x})]_+^{1+\frac{1}{p}}[F_l(\bar{x})]_+^{1+\frac{1}{p}} = 0.$$

Therefore, we conclude that $F(\bar{x}) \le 0$ and $H(\bar{x}) \le 0$. Taking the limit as $i \xrightarrow{\mathcal{K}} \infty$, we see that

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

where the second equality follows from the fact that

$$F_l(x^i)[H_l(x^i)]_+^{2+\frac{1}{p}} + H_l(x^i)[F_l(x^i)]_+^{2+\frac{1}{p}} = -\rho^i\big([F_l(x^i)]_+[H_l(x^i)]_+\big)^2,\ \forall\, l \in \mathcal{I}.$$

Therefore, we conclude that $\langle F(\bar{x}), H(\bar{x})\rangle = 0$. The proof is complete. □

It is difficulty to find a global solution of problem (3.1) without the assumption of the convexity on $\Psi(x,\rho)$ for each $\rho > 0$. We mainly focus on the local solution of problem (3.1) in practice. Next, we prove that a local solution of problem (3.1) can identify a solution of problem (1.1) under the monotone assumptions.

THEOREM 3.2. *Suppose that F is monotone and H is strictly monotone. For each $\rho > 0$, let x^ρ be a local solution of problem (3.1). Moreover, assume that x^ρ satisfies $F(x^\rho) \le 0$ and $H(x^\rho) \le 0$. Then x^ρ is a solution of problem (1.1).*

Proof. Since x^ρ is a local solution of problem (3.1) for a given $\rho > 0$, we see that

$$\nabla H(x^\rho)^T\Theta(x^\rho,\rho)\mathcal{G}(x^\rho,\rho) + \nabla F(x^\rho)^T\Pi(x^\rho,\rho)\mathcal{G}(x^\rho,\rho) = 0. \tag{3.3}$$

By assumptions of $F(x^\rho) \le 0$ and $H(x^\rho) \le 0$, we have

$$\begin{aligned}
\mathcal{G}(x^\rho,\rho) &= \rho(H_1(x^\rho)F_1(x^\rho),\ldots,H_n(x^\rho)F_n(x^\rho))^T,\\
\Theta(x^\rho,\rho) &= \rho\mathrm{diag}(F_1(x^\rho),\ldots,F_n(x^\rho)),\\
\Pi(x^\rho,\rho) &= \rho\mathrm{diag}(H_1(x^\rho),\ldots,H_n(x^\rho)).
\end{aligned}$$

In order to prove this theorem, we need to prove that $\mathcal{G}(x^\rho,\rho) = 0$. Assume on the contrary that $\mathcal{G}(x^\rho,\rho) \neq 0$. We see that there exists at least some $i \in \mathcal{I}$ such that $\mathcal{G}_i(x^\rho,\rho) \neq 0$, which means that $H_i(x^\rho) \neq 0$ and $F_i(x^\rho) \neq 0$. Thus, we have $\big(\Theta(x^\rho,\rho)\mathcal{G}(x^\rho,\rho)\big)_i = \rho^2 H_i(x^\rho)F_i(x^\rho)^2 \neq 0$. By Proposition 2.3 and the assumption of strict monotonicity on H, we see that

$$\big(\Theta(x^\rho,\rho)\mathcal{G}(x^\rho,\rho)\big)^T\nabla H(x^\rho)^T\Theta(x^\rho,\rho)\mathcal{G}(x^\rho,\rho) > 0.$$

Multiply $(\Theta(x^\rho,\rho)\mathcal{G}(x^\rho,\rho))^T$ on both sides of (3.3) to get

$$\begin{aligned}
&\big(\Theta(x^\rho,\rho)\mathcal{G}(x^\rho,\rho)\big)^T\nabla H(x^\rho)^T\Theta(x^\rho,\rho)\mathcal{G}(x^\rho,\rho)\\
&= -\big(\Theta(x^\rho,\rho)\mathcal{G}(x^\rho,\rho)\big)^T\nabla F(x^\rho)^T\Pi(x^\rho,\rho)\mathcal{G}(x^\rho,\rho)\\
&= -\mathcal{G}(x^\rho,\rho)^T\Big(\Theta(x^\rho,\rho)\nabla F(x^\rho)^T\Pi(x^\rho,\rho)\Big)\mathcal{G}(x^\rho,\rho).
\end{aligned}$$

By the assumption of monotonicity on F and the fact of positive semidefiniteness of the matrices $-\Theta(x^\rho,\rho)$ and $-\Pi(x^\rho,\rho)$, we see that the matrix $\Theta(x^\rho,\rho)\nabla F(x^\rho)^T\Pi(x^\rho,\rho)$ is positive semidefinite, which implies that

$$-\mathcal{G}(x^\rho,\rho)^T\Big(\Theta(x^\rho,\rho)\nabla F(x^\rho)^T\Pi(x^\rho,\rho)\Big)\mathcal{G}(x^\rho,\rho) \le 0.$$

We have a contradiction, which means that $\mathcal{G}(x^\rho,\rho) = 0$. Since $F(x^\rho) \le 0$ and $H(x^\rho) \le 0$, we see that x^ρ is a solution of problem (1.1). The proof is complete. □

3.2. A Trust-Region Gauss-Newton Method. this section, we apply the trust-region Gauss-Newton method [5, 20, 23] to solve problem (3.1) with each $\rho > 0$. At the k-th iteration, we consider a quadratic approximation $m^k(d,\rho)$ for $\Psi(x,\rho)$ at x^k and replace problem (3.1) by a trust region problem

$$\min_d\ m^k(d,\rho) \quad \text{s.t. } \|d\| \le \Delta^k,$$

with the objective function

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

where $\Delta^k > 0$ is the trust-region radius. A formal description of the trust-region Gauss-Newton method can be found in [22, Algorithm 3.1]. Here we omit the details.

Before presenting the unconstrained differentiable penalty method for problem (1.1), we define the termination criterion as follows

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

where $\epsilon > 0$ is the tolerance parameter. A formal description of the proposed method for problem (1.1) is given as follows.

Algorithm 1: Unconstrained differentiable penalty method for the ICP.

```
Initializing ρ^0 > 0, ρ^min, σ ∈ (0,1), ϵ > 0 and an initial point x^0 and let
  i := 0;
while ρ^i > ρ^min do
    if Termination(x^i) ≤ ϵ then
        Stop;
    else
        Using [22, Algorithm 3.1] to solve the unconstrained problem (3.1)
        with starting point x^i and penalty parameter ρ^i, we obtain x^{i+1};
    end
    Letting ρ^{i+1} := σρ^i and i := i + 1;
end
```

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

We refer to the implementation of Algorithm 1 as the UDLOP method, which stands for the Unconstrained Differentiable Lower-Order Penalty method. For convenience, we write the UDLOP method with $p = 1$, 2 and 100 as UDLOP_1, $\text{UDLOP}_{1/2}$ and $\text{UDLOP}_{1/100}$ methods, respectively. The abbreviations for some existing penalty methods are presented in Table 4.1.

Table 4.1: Abbreviations for some existing methods.

CDLOP_1 [28]	constrained differentiable lower-order penalty method with $p = 1$
$\text{CDLOP}_{1/2}$	constrained differentiable lower-order penalty method with $p = 2$
$\text{CDLOP}_{1/100}$	constrained differentiable lower-order penalty method with $p = 100$
SSOOP_1 [1]	semismooth one-order penalty method
SAM [3]	smooth approximate method
NSEM [18]	nonsmooth equations method

The semismooth Newton method [27] is employed to solve the ℓ_1-penalized equations for the SSOOP_1 method and the semismooth equations for the NSEM. The Zang smooth plus function [33] is used in the SAM to smooth its normal equations. The solver TRESNEI [22] is used to solve the corresponding optimization problems for all methods mentioned in this paper. The fact that the box-constrained differentiable penalty method is more efficient than the smoothed $\ell_{\frac{1}{p}}$-penalty method was proved in [28]. Therefore, we do not consider the smoothed $\ell_{\frac{1}{p}}$-penalty method in our numerical experiments.

Throughout the experiments, we set parameters $\rho^0 = 1$, $\rho^{min} = 10^{-16}$, $\sigma = 0.1$ and $\epsilon = 10^{-6}$ in Algorithm 1. We follow all default parameters in the solver TRESNEI. The details can be found in [22]. We select 22 test problems from MCPLIB shown in Table 4.2. 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 4.2: Problem characteristics and starting intervals.

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

In Table 4.2, the **Problem** denotes the name of test problem, the **Dim** denotes the dimension of problem (1.1) and the **Interval** denotes the interval in which a starting point is generated by a uniform distribution.

Using the performance profiles of Dolan and Moré in [7], we plot Figure 4.1, 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 [7] for more details regarding the performance profiles.

We first compare the performance of the $\text{UDLOP}_{1/2}$ method with the $\text{CDLOP}_{1/2}$ and SSOOP_1 methods in terms of the number of function evaluations and the values of the penalty parameter.

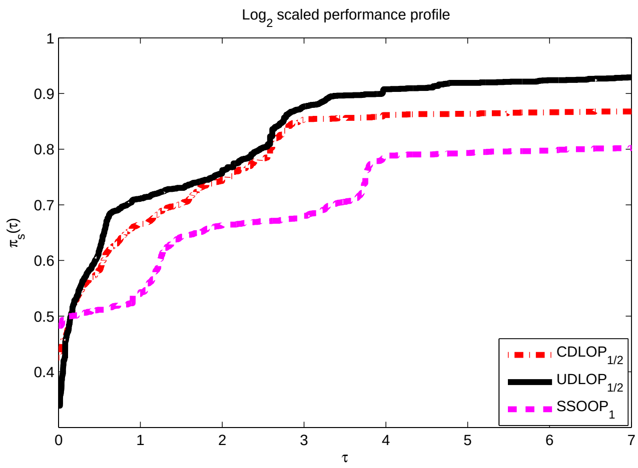

Fig. 4.1: Performance profiles based on the number of function evaluations for the $\mathrm{CDLOP}_{1/2}$, $\mathrm{UDLOP}_{1/2}$ and SSOOP_1 methods.

Figure 4.1 indicates that the SSOOP_1 method is the weakest solver as it can only solve 80% test problems. The $\mathrm{UDLOP}_{1/2}$ method is the most robust and can solve about 93% test problems.

We plot $\pi(\tau)$ for different values of $\frac{1}{\rho}$ in Figure 4.2, which shows that the SSOOP_1 method employs smaller values of the penalty parameter than that of the $\mathrm{CDLOP}_{1/2}$ method in order to achieve an approximate solution within the given accuracy.

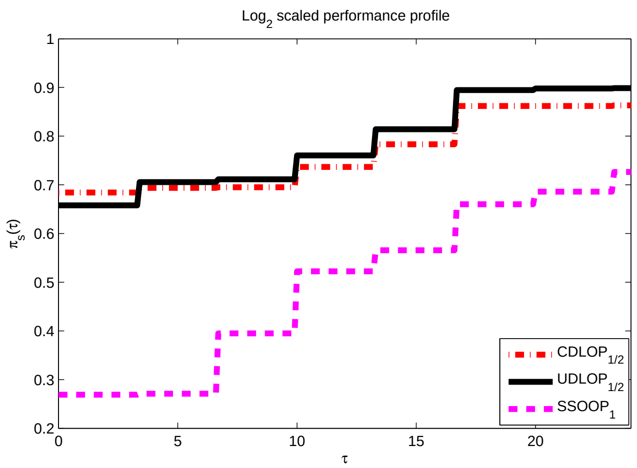

Fig. 4.2: Performance profiles based on the values of the penalty parameter for the $\mathrm{CDLOP}_{1/2}$, $\mathrm{UDLOP}_{1/2}$ and SSOOP_1 methods.

We plot Figures 4.3 and 4.4 to compare the performance of the UDLOP method with different values of the power p in terms of the number of function evaluations and the values of the penalty parameter.

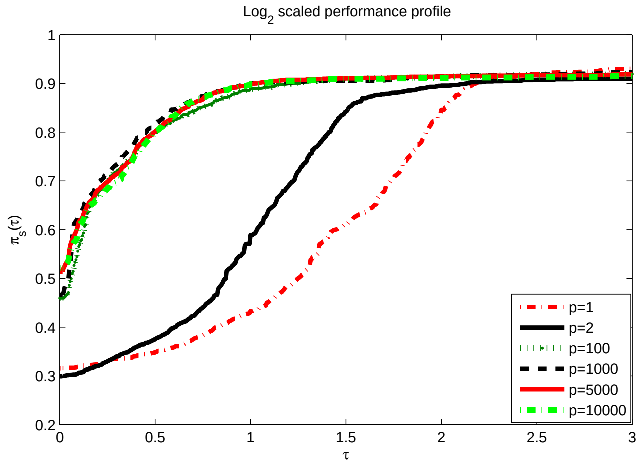

Fig. 4.3: Performance profiles based on the number of function evaluations for the UDLOP method with different p.

Figure 4.3 indicates that the number of function evaluations for the UDLOP method decreases dramatically as the power p increases from 1 to 100. However, slight difference happens on the performance profiles as we increase p from 100 to 10000.

We plot $\pi(\tau)$ for different values of $\frac{1}{\rho}$ in Figure 4.4, which indicates that the UDLOP$_1$ method uses the smallest values of the penalty parameter among them.

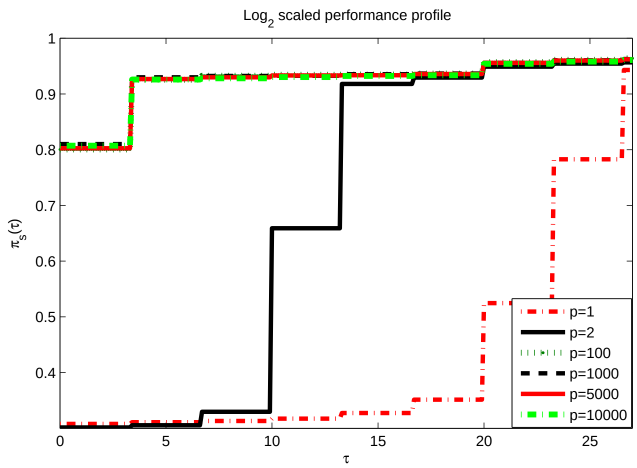

Fig. 4.4: Performance profiles based on the values of the penalty parameter for the UDLOP method with different p.

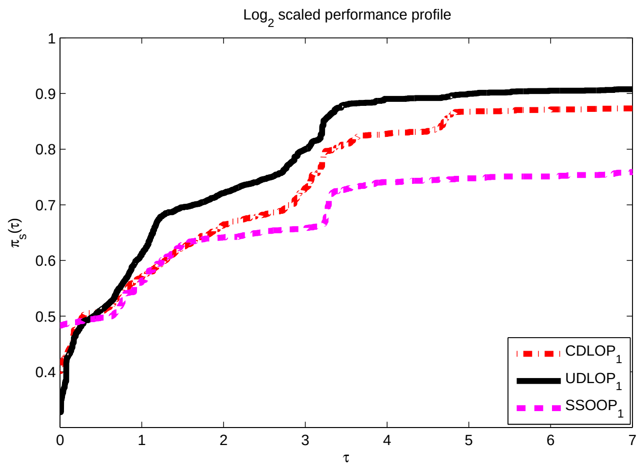

Fig. 4.5: Performance profiles based on the number of function evaluations for the CDLOP_1, UDLOP_1 and SSOOP_1 methods.

Figure 4.5 is plotted using the number of function evaluations, which indicates that the UDLOP_1 method uses the least number of function evaluations and the SSOOP_1 method uses the most number of function evaluations among all these three solvers.

Finally, using the number of function evaluations, we compare the performance of the $\mathrm{CDLOP}_{1/100}$ and $\mathrm{UDLOP}_{1/100}$ methods with the SAM and NSEM.

Figure 4.6 shows that the SAM can solve about 47% test problems with the least number of function evaluations, but this method only can solve about 75% test problems. The $\mathrm{UDLOP}_{1/100}$ method is the most robust among them and can solve about 89% test problems.

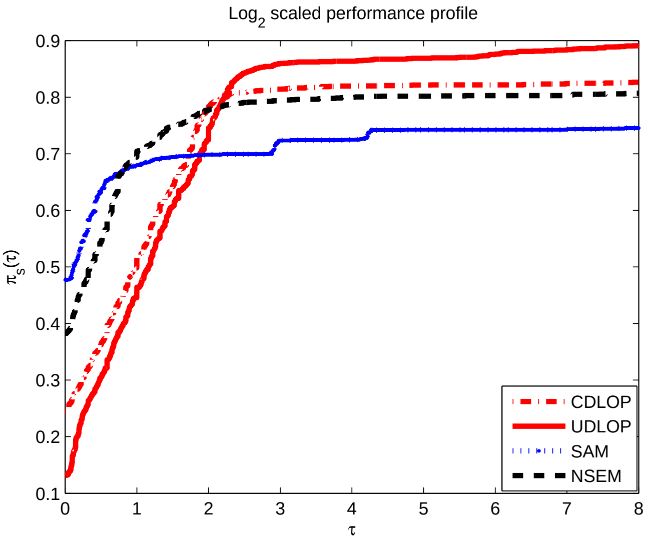

Fig. 4.6: Performance profiles based on the number of function evaluations for the $\text{CDLOP}_{1/100}$, $\text{UDLOP}_{1/100}$, SAM and NSEM.

REFERENCES

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

[2] 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.

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

[4] M. Chen and C. C. Huang. A power penalty method for the general traffic assignment problem with elastic demand. *Journal of Industral and Management Optimization*, 10(4):1019–1030, 2014.

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

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

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

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

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

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

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

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

[13] M. Fukushima. Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems. *Mathematical Programming*, 53(1-3):99–110, 1992.

[14] 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.

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

[16] 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.

[17] H. Y. Jiang, M. Fukushima, L. Q. Qi, and D. F. Sun. A trust region method for solving generalized complementarity problems. *SIAM Journal on Optimization*, 8(1):140–157, 1998.

[18] 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.

[19] C. Kanzow and M. Fukushima. Equivalence of the generalized complementarity problem to differentiable unconstrained minimization. *Journal of Optimization Theory and Applications*, 90(3):581–603, 1996.

[20] 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.

[21] O. L. Mangasarian and M. V. Solodov. Nonlinear complementarity as unconstrained and constrained minimization. *Mathematical Programming*, 62(1):277–297, 1993.

[22] 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.

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

[24] J.S. Pang. On the convergence of a basic iterative method for the implicit complementarity problem. *Journal of Optimization Theory and Applications*, 37(2):149–162, 1982.

[25] J.S. Pang. Complementarity problems. In *Handbook of global optimization*, pages 271–338. Springer, 1995.

[26] J.M. Peng. Equivalence of variational inequality problems to unconstrained minimization. *Mathematical Programming*, 78(3):347–355, 1997.

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

[28] B. S. Tian, Y. H. Hu, and X. Q. Yang. A box-constrained differentiable penalty method for nonlinear complementarity problems. *Journal of Global Optimization*, DOI: 10.1007/s10898-015-0275-6.

[29] P. Tseng, N. Yamashita, and M. Fukushima. Equivalence of complementarity problems to differentiable minimization: A unified approach. *SIAM Journal on Optimization*, 6(2):446–460, 1996.

[30] S. Wang. A penalty method for a finite-dimensional obstacle problem with derivative constraints. *Optimization Letters*, 8(6):1799–1811, 2014.

[31] S. Wang and X. Q. Yang. A power penalty method for linear complementarity problems. *Operations Research Letters*, 36(2):211–214, 2008.

[32] 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.

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

[34] K. Zhang, X. Q. Yang, S. Wang, and K. L. Teo. Numerical performance of penalty method for American option pricing. *Optimization Methods and Software*, 25(5):737–752, 2010.

[35] Y. Y. Zhou, S. Wang, and X. Q. Yang. A penalty approximation method for a semilinear parabolic double obstacle problem. *Journal of Global Optimization*, 60(3):531–550, 2014.