LOWER-ORDER PENALTY METHODS FOR NONLINEAR OPTIMIZATION AND COMPLEMENTARITY PROBLEMS

BOSHI TIAN

Ph.D

The Hong Kong Polytechnic University

2014

The Hong Kong Polytechnic University
Department of Applied Mathematics

Lower-Order Penalty Methods for Nonlinear Optimization and Complementarity Problems

Boshi Tian

A thesis submitted in partial fulfilment of
the requirements for the degree of Doctor of Philosophy

January 2014

CERTIFICATE OF ORIGINALITY

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of a university or other institute of higher learning, except where due acknowledgment is made in the text.

__________________________(Signed)

Boshi TIAN (Name of student)

Abstract

The main purpose of this thesis is to propose efficient numerical methods to solve inequality constrained nonlinear programming problems and complementarity problems by virtue of the $\ell_{\frac{1}{p}}(p > 1)$-penalty function.

We propose an interior-point $\ell_{\frac{1}{p}}$-penalty method for inequality constrained optimization problems by introducing a technique of the p-order relaxation to the nonconvex and non-Lipschitzian $\ell_{\frac{1}{p}}$-penalty function and combining with an interior-point method. We introduce different kinds of constraint qualifications to establish first-order necessary conditions for the relaxed problem. We employ the modified Newton method to solve a sequence of logarithmic barrier subproblems and detail three reliable algorithms by using the Armijo line search. We prove that the iteration sequence generated by the proposed method converges to some KKT (or FJ) point of the original problem under mild conditions. Preliminary numerical experiments on small, medium and large test problems in the literature show that, comparing with some existing interior-point ℓ_1-penalty methods, the proposed method is competitive in terms of the iteration numbers, better when comparing the number of updating the penalty parameters and more reliable when comparing the relative error.

We introduce 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 the non-Lipschitzianness. We introduce a concept of a uniform ξ-P-function with $\xi \in [1, 2)$, under which we prove that the solution of box-constrained penalized equations converges to a solution 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 to design the globally convergent

method that allows arbitrary starting points for solving the complementarity problems. Furthermore, we establish the connection between the local solution of the least squares problem and the solution of the original problem under mild conditions. We carry out the numerical experiments on the test problems from MCPLIB, which show that the proposed method is efficient and robust.

We investigate an unconstrained differentiable penalty method for general complementarity problems without introducing artificial variables, which shares the exponential convergence rate under the assumption of a uniform ξ-P-function. Instead of solving the unconstrained penalized equations directly, we solve a corresponding differentiable least squares problem by using a trust-region Gauss-Newton method. Preliminary numerical experiments show that the proposed method is more robust than the box-constrained differentiable penalty method.

Acknowledgments

I would like to thank my supervisor Prof. Xiaoqi Yang for providing the vision and support for this work throughout my years in Hong Kong. His support not only includes mathematical guides to my research, but also providing adequate funds for conducting my research. I am grateful for his kindness, encouragement and patience. I wish to thank Prof. Cedric, Ka-Fai, Yiu (Hong Kong Polytechnic University) as the Chair of my oral defense and thank Prof. Boris Mordukhovich (Wayne State University) and Prof. Xiaoming, Yuan (Hong Kong Baptist University) as the examiners of my oral defense.

I would like to thank Prof. Donghui Li (South China Normal University) and Prof. Jinping Zeng (Dongguan University of Technology) for their sustained help and encouragement. I wish to thank Prof. Weijun Zhou (Changsha University of Science & Technology) and Prof. Zhe Sun (Jiangxi Normal University) for their suggestions during the course of this research. I had many fruitful and stimulating discussions with them. To my academic brothers, Prof. Kai Zhang (Shenzhen University), Prof. Kaiwen Meng (Southwest Jiaotong University), Dr. Yaohua Hu and Dr. Zhangyou Chen, I am grateful to them for their help and friendship. In addition, I wish to thank all fellow graduate students for their friendship. They made my time at The Hong Kong Polytechnic University more enjoyable.

Last but not least, I would like to thank my parents. They gave me the whole life and have always encouraged me to keep on studying. To my sister Jingjing Tian and brother Shiye Tian, I am grateful to their love and support. Many thanks to my wife Qiaoling Li. I thank her for her love, encouragement and tolerance.

Contents

List of Abbreviations

KKT	Karush-Kuhn-Tucker
FJ	Fritz-John
CQ	Constraint qualification
GCQ	Guignard constraint qualification
ACQ	Abadie constraint qualification
LICQ	Linear independence constraint qualification
MFCQ	Mangasarian-Fromovitz constraint qualification
EMFCQ	Extended Mangasarian-Fromovitz constraint qualification
SQP	Sequential quadratic programming
Sℓ_1QP	Sequential ℓ_1 quadratic programming
GCP	General complementarity problem
NCP	Nonlinear complementarity problem
LCP	Linear complementarity problem
MiCP	Mixed complementarity problem
OPDCs	Optimization problem with degenerate constraints

List of Algorithms

IPLOP	Interior-Point Lower-Order Penalty Method
PIPAL-c	Penalty-Interior-Point Algorithm with conservative updates
PIPAL-a	Penalty-Interior-Point Algorithm with aggressive updates
CDLOP	Constrained Differentiable Lower-Order Penalty Method
$\mathbf{SLOP}_{1/2}$	Smoothing Lower-Order Penalty Method with $p = 2$
$\mathbf{SSOOP}_{1}$	Semismooth One-Order Penalty Method
SAM	Smoothing Approximation Method
NSEM	Nonsmooth Equations Method
UDLOP	Unconstrained Differentiable Lower-Order Penalty Method
$\mathbf{EGA}_{12}$	Extra-Gradient Method with Modifications 1 and 2

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Nonlinear Programming Problems

Consider the inequality constrained nonlinear programming problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & c_i(x) \leq 0,\ i \in \mathcal{I}, \end{aligned} \tag{1.1.1}$$

where the functions f and $c_i : \mathbb{R}^n \to \mathbb{R}$ are assumed to be twice continuously differentiable and $\mathcal{I} = \{1, 2, \ldots, m\}$. We define the feasible set $\mathcal{F}$ to be the set of the points x satisfying the constraints, that is, $\mathcal{F} := \{x \mid c_i(x) \leq 0,\ i \in \mathcal{I}\}$. A vector $x^* \in \mathbb{R}^n$ is called a local solution of problem (1.1.1) if $x^* \in \mathcal{F}$ and there is a neighborhood $\mathcal{N}$ of x^* such that $f(x^*) \leq f(x)$ for all $x \in \mathcal{N} \cap \mathcal{F}$. Similarly, a point x^* is called a strict local solution of problem (1.1.1) if $x^* \in \mathcal{F}$ and there is a neighborhood $\mathcal{N}$ of x^* such that $f(x^*) < f(x)$ for all $x \in \mathcal{N} \cap \mathcal{F}$ with $x \neq x^*$.

1.1.1 Constraint Qualifications

To state the first-order necessary conditions for x^* to be a local solution of problem (1.1.1), the Lagrange function of problem (1.1.1) is defined as

$$\mathcal{L}(x, \lambda) := f(x) + \sum_{i \in \mathcal{I}} \lambda_i c_i(x), \tag{1.1.2}$$

where $\lambda = (\lambda_1, \ldots, \lambda_m)^T \in \mathbb{R}^m$ is the Lagrange multiplier vector. The first order necessary conditions hold at x^* if there exists a vector $\lambda^* \in \mathbb{R}^m$ such that (x^*, λ^*) is a solution to the nonlinear system

$$\nabla f(x) + A(x)\lambda = 0, \tag{1.1.3a}$$

$$c_i(x)\lambda_i = 0, \ \forall\, i \in \mathcal{I}, \tag{1.1.3b}$$

$$\lambda_i \geq 0, \ \ -c_i(x) \geq 0, \ \forall\, i \in \mathcal{I}, \tag{1.1.3c}$$

where $\nabla f(x)$ denotes the gradient of $f(x)$ and $A(x)$ is the transpose of the Jacobian matrix of $c(x) := (c_1(x), \ldots, c_m(x))^T$ at x, i.e., $A(x) := [\nabla c_1(x), \ldots, \nabla c_m(x)]$. The first-order necessary conditions are known as the Karush-Kuhn-Tucker (KKT, for short) conditions, which were derived independently by Karush [99] and by Kuhn and Tucker [100]. Such a point x^* is called a KKT point of problem (1.1.1). The Fritz John (FJ, for short) conditions are said to be satisfied if there exist a constant $\lambda_0^* \in \mathbb{R}$ and a vector $\lambda^* \in \mathbb{R}^m$ such that $(x^*, \lambda_0^*, \lambda^*)$ is a solution to the nonlinear system

$$\begin{aligned} \lambda_0 \nabla f(x) + A(x)\lambda = 0,& \\ c_i(x)\lambda_i = 0, \ \forall\, i \in \mathcal{I},& \\ \lambda_0 \geq 0, \ \lambda_i \geq 0, \ -c_i(x) \geq 0, \ \forall\, i \in \mathcal{I}.& \end{aligned}$$

Such x^* is called a FJ point, which was introduced by Fritz John [67]. Algorithms for solving problem (1.1.1) often focus on producing points that satisfy the KKT conditions. It is worth noting that the KKT conditions may not hold at local solutions of problem (1.1.1) unless some constraint qualification (CQ, for short) is satisfied.

The active set $\mathcal{I}(x)$ at any feasible point x consists of indices of inequality constraints i for which $c_i(x) = 0$, that is, $\mathcal{I}(x) := \{i \in \mathcal{I} \mid c_i(x) = 0\}$. The Bouligand tangent cone and linearized tangent cone of $\mathcal{F}$ at x^* are defined, respectively, by

$$T_{\mathcal{F}}(x^*) := \{d \in \mathbb{R}^n \mid \exists t_k \to 0^+,\ \exists d_k \to d,\ \text{s.t. } x^* + t_k d_k \in \mathcal{F},\ \forall\, k\}$$

and

$$L_{\mathcal{F}}(x^*) := \{d \in \mathbb{R}^n \mid \nabla c_i(x^*)^T d \le 0,\ \forall\, i \in \mathcal{I}(x^*)\}.$$

It is important to note that $L_{\mathcal{F}}(x^*)$ only uses the information of gradients of constraints and $T_{\mathcal{F}}(x^*) \subset L_{\mathcal{F}}(x^*)$. The polar cone of $L_{\mathcal{F}}(x^*)$ is given by (see [151, Chapter 6])

$$L_{\mathcal{F}}(x^*)^* := \Big\{v \in \mathbb{R}^n \mid v = \sum_{i \in \mathcal{I}(x^*)} \lambda_i \nabla c_i(x^*),\ \lambda_i \ge 0\Big\}.$$

The tangent cone is composed by the limits of directions that move inward of the feasible set. By this fact, a necessary condition for a local solution x^* of problem (1.1.1) is presented by

$$-\nabla f(x^*) \in T_{\mathcal{F}}(x^*)^*, \tag{1.1.5}$$

which is also called a geometric necessary condition, as the tangent cone relays only on the geometric specification of the feasible set $\mathcal{F}$. The linearized tangent cone does, however, depend on the algebraic specification of the feasible set $\mathcal{F}$, and hence it can be directly used in algorithms. If $L_{\mathcal{F}}(x^*)^* = T_{\mathcal{F}}(x^*)^*$, then the necessary condition (1.1.5) can be rewritten as

$$-\nabla f(x^*) \in L_{\mathcal{F}}(x^*)^*,$$

which is exactly the KKT conditions. The Guignard constraint qualification (GCQ, for short) [79] holds at x^* if $L_{\mathcal{F}}(x^*)^* = T_{\mathcal{F}}(x^*)^*$. The Abadie constraint qualification (ACQ, for short) [1] holds at x^* if $L_{\mathcal{F}}(x^*) = T_{\mathcal{F}}(x^*)$.

The ACQ obviously implies the GCQ, whereas the converse in general is not true, see [139] for a counterexample. It was noted in [4, 75] that the GCQ is the weakest constraint qualification in the sense that the GCQ holds at x^* if and only if the KKT conditions hold at x^* whenever a continuously differentiable objective function f has a local solution at x^* relative to the feasible set $\mathcal{F}$. A common challenge of the ACQ and the GCQ in numerical algorithms is that they are extremely difficult to be verified.

Apart from the GCQ and the ACQ, another two well known constraint qualifications are presented as follows, which are important in establishing convergence results in numerical algorithms [8, 128].

The linear independence constraint qualification (LICQ, for short) [56] holds at x^* if the vectors $\{\nabla c_i(x^*),\ i \in \mathcal{I}(x^*)\}$ are linearly independent. The Mangasarian-Fromovitz constraint qualification (MFCQ, for short) [112] holds at x^* if there exists a vector $d \in \mathbb{R}^n$ such that

$$\nabla c_i(x^*)^T d < 0, \text{ for all } i \in \mathcal{I}(x^*).$$

The MFCQ is a weaker condition than the LICQ and it is easy to construct examples in which the MFCQ is satisfied but the LICQ is not; see [128, Exercise 12.13]. If x^* is a local solution of problem (1.1.1), then the KKT conditions hold at x^* provided that the LICQ or the MFCQ is satisfied. It was reported in [72] that the MFCQ is equivalent to boundedness of the set of Lagrange multiplier vectors λ^* for which the KKT conditions are satisfied. In the case of the LICQ, this set consists of a unique vector λ^*.

1.1.2 Penalty Methods

Penalty methods are an important class of numerical optimization methods for solving problem (1.1.1). Such methods essentially eliminate the constraints and replace them with cost terms in the objective function so as to penalize violations in the original constraints. The penalty functions associated with problem (1.1.1) can be written in general as

$$P(x, \rho) := f(x) + \rho Q(\|[c(x)]_+\|), \tag{1.1.6}$$

where $\rho > 0$ is the penalty parameter, $([c(x)]_+)_i = \max\{0, c_i(x)\}$, for all $i \in \mathcal{I}$, $\|\cdot\|$ is any fixed vector norm in $\mathbb{R}^m$, and Q is some function from the nonnegative real line $\mathbb{R}_+$ into itself with the property that $Q(t) = 0$ if and only if $t = 0$. By making this parameter ρ bigger and bigger, the penalization of constraint violations is more and more severely, thereby forcing the minimizer of the penalty function closer and closer to the feasible region of the original problem. The simplest penalty function of this type is the quadratic penalty function

$$P_2(x, \rho) := f(x) + \rho \sum_{i \in \mathcal{I}} ([c_i(x)]_+)^2,$$

which is a natural result by setting $Q(t) = t^2$ and using the ℓ_2-norm in (1.1.6) and was first used by Courant [41]. Extensive studies on the quadratic penalty function method can be found in Fiacco and McCormick's monograph [56]. Given penalty parameter ρ^k, an approximate solution x^k can be identified by minimizing the function $P_2(x, \rho^k)$ by nonsmooth Newton methods [132, 146]. As $\rho^k \to \infty$, the KKT conditions hold at the limit point x^* if the LICQ holds at x^*. However, the minimization of the function $P_2(x, \rho^k)$ becomes more and more difficult to perform when ρ^k becomes very large as the approximate Hessian matrix becomes ill-conditioned near x^*.

In order to overcome the drawback of the quadratic penalty function method, exact penalty functions were proposed to solve problem (1.1.1). The exact penalization means that there exists a threshold $\hat{\rho} > 0$ such that for any $\rho \geq \hat{\rho}$, the unconstrained minimizing points of penalty functions are also solutions of problem (1.1.1). This property is desirable because it makes the performance of penalty methods less dependent on the strategy for updating the penalty parameter. The quadratic penalty function is not exact because its minimizer is generally not the same as the solution of problem (1.1.1) for any finite penalty parameter ρ.

The classical ℓ_1-penalty function is included in this class of exact penalty functions

$$P_1(x, \rho) := f(x) + \rho \sum_{i \in \mathcal{I}} [c_i(x)]_+, \tag{1.1.7}$$

which is obtained from (1.1.6) by setting $Q(t) = t$ and using the ℓ_1-norm. Under the convexity of the objective function and constraints, and the assumption that the strict feasible set is nonempty, Zangwill [184] proved that the solutions of problem (1.1.1) are the unconstrained minimizing points of the ℓ_1-penalty function for all sufficiently big values of the penalty parameter ρ. Pietrzykowski [141] proved that, if x^* is a strict local solution of problem (1.1.1) and the LICQ holds at x^*, then x^* is a local solution of the ℓ_1-penalty function for all ρ sufficiently big. The same result was shown by Han and Mangasarian [80] by assuming the MFCQ, a weaker condition than the LICQ.

Furthermore, Han and Mangasarian established a well-known theorem that if the ℓ_1-penalty function is exact at x^*, then the KKT conditions hold. Using this theorem, we may interpret the existence of a local solution to the exact penalty function as constraint qualification which ensures the satisfaction of the KKT conditions at a local solution of problem (1.1.1).

The existence of exact penalty functions can be viewed as a consequence of regularity conditions such as error bounds [134] and metric regularity [10, 34, 120]. One of the weakest conditions was known as the calmness which was originally formulated by Rockafellar and first appeared in the literature of Clarke [33]. The calmness can be utilized to provide some full characterizations of the exact penalization. Consider the perturbed nonlinear programming problem

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in M(u), \end{array} \tag{1.1.8}$$

where $M : \mathbb{R}^m \to \mathbb{R}^n$ is a set-valued mapping defined by

$$M(u) := \{x \in \mathbb{R}^n \mid c_i(x) \leq u_i, \ \forall\, i \in \mathcal{I}\}.$$

It is clear that problem (1.1.8) with $u = 0$ is exactly the same as problem (1.1.1). Let x^* be a solution of problem (1.1.1). According to Clarke [34, Definition 6.4.1], problem (1.1.1) is calm at x^* provided that there exist positive constants ϵ and $\bar{\alpha}$ such that for all $u \in \epsilon\mathbb{B}$ and $x \in x^* + \epsilon\mathbb{B}$ which are feasible for problem (1.1.8), one has

$$f(x) + \bar{\alpha}\|u\| \geq f(x^*),$$

where $\mathbb{B}$ is a unit ball centered at origin and $\|z\|$ stands for the norm of z in $\mathbb{R}^m$ and here we can specify it to be the ℓ_1-norm without loss of generality. Burke [11, Definition 1.1] introduced another definition of the calmness, which varies from the definition given above in that the variable u is not restricted to an ϵ neighborhood of the origin in order for the above inequality to hold. It was proved that the restriction on the choice of perturbation u is redundant when $c_i(x)$ are continuous for all $i \in \mathcal{I}$, see [11, Proposition 2.1]. Without considering the existence of a solution of problem (1.1.8), the calmness can also be defined by using the perturbation function $V(u)$ given by

$$V(u) := \begin{cases} +\infty, & \text{if } \{x : x \in M(u)\} = \emptyset, \\ \min\{f(x) : x \in M(u)\}, & \text{otherwise.} \end{cases}$$

Problem (1.1.1) is said to be calm at x^* ([11]) if

$$\liminf_{u \to 0} \frac{V(u) - V(0)}{\|u\|} > -\infty. \tag{1.1.9}$$

The fact that the calmness implies the existence of an exact penalty parameter was established by Clarke [34, Proposition 6.4.3]. However, the reverse implication and the precision of this correspondence was first established by Burke ([11, Theorem 1.1], also see [12, Theorem 2.1]). Therefore, the notion of the calmness is in a sense equivalent to the notion of the exact penalization. The calmness hypothesis is quite weak and in many situations is easily verified.

Unfortunately, the ℓ_1-penalty function is nonsmooth and nondifferentiable, many effective algorithms such as quasi-Newton methods [103] cannot be adequately used. And general techniques for nondifferentiable optimization such as bundle methods [86], are also not efficient, as they do not take account of the special nature of the nondifferentiabilities. Even through the ℓ_1-penalty function is nondifferentiable, it has a directional derivative $D(P_1(x, \rho); d)$ along any direction $d \in \mathbb{R}^n$ given by

$$D(P_1(x, \rho); d) := \lim_{\epsilon \to 0^+} \frac{P_1(x + \epsilon d, \rho) - P_1(x, \rho)}{\epsilon}.$$

And the direction derivative of the function $P_1(x, \rho)$ at a feasible point x along a direction d can be easily written as

$$D(P_1(x, \rho); d) := \nabla f(x)^T d + \rho \sum_{i \in \mathcal{I}(x)} [\nabla c_i(x)^T d]_+.$$

A point $x^* \in \mathbb{R}^n$ is called a stationary point of the ℓ_1-penalty function if $D(P_1(x^*, \rho); d) \geq 0$ for all $d \in \mathbb{R}^n$. An important theorem [128, Theorem 17.4] states that if x^* is a stationary point of $P_1(x, \rho)$ for all ρ bigger than a certain threshold $\hat{\rho} > 0$ and $x^* \in \mathcal{F}$ then it satisfies the KKT conditions of problem (1.1.1). The existence of the the threshold $\hat{\rho} > 0$ can be guaranteed by the exact penalization of the ℓ_1-penalty function $P_1(x, \rho)$. Therefore, the function $P_1(x, \rho)$ can be used as a merit function in some iteration methods such as sequential quadratic programming (SQP) methods [63] to guarantee the global convergence of the iteration sequence by accepting or rejecting a trial step. The global convergence means that, under certain common assumptions, the iteration sequence converges to some KKT point of problem (1.1.1) from remote starting points. Fletcher [63] introduced a piecewise linear-quadratic model of $P_1(x, \rho)$ to compute an approximate descent direction d. The model is given by

$$q(d, \rho) := f(x) + \nabla f(x)^T d + \frac{1}{2} d^T H d + \rho \sum_{i \in \mathcal{I}} [c_i(x) + \nabla c_i(x)^T d]_+,$$

where H is a symmetric matrix approximating the Hessian of the Lagrange function (1.1.2). The model $q(d, \rho)$ is nonsmooth, but can be recast as a smooth quadratic programming problem by introducing artificial variables s_i as follows

$$\begin{aligned} \min_{d,s} \quad & f(x) + \nabla f(x)^T d + \frac{1}{2} d^T H d + \rho \sum_{i \in \mathcal{I}} s_i \\ \text{s.t.} \quad & c_i(x) + \nabla c_i(x)^T d \le s_i, \ i \in \mathcal{I}, \\ & s_i \ge 0, \ i \in \mathcal{I}. \end{aligned} \tag{1.1.10}$$

A standard sequential quadratic programming algorithm can be used to solve problem (1.1.10). Once the solution d is found, a line search such as Armijo line search or Wolfe line search is performed in the direction d to ensure that a sufficient decrease in the ℓ_1-exact penalty function $P_1(x, \rho)$ is achieved at the new iterate. The iteration method stated above is referred to as the sequential ℓ_1 quadratic programming (Sℓ_1QP, for short) which was proposed by Fletcher [63] and fully investigated in [128]. The Sℓ_1QP approach not only overcomes the difficulties posed by inconsistent constraint linearizations [63] but also can solve certain class of problems in which standard constraint qualifications such as the LICQ and the MFCQ are not satisfied [3, 102]. Further, there is no requirement for matrix H to be positive definite. However, this approach may fail to converge rapidly because it rejects steps that make good progress toward a solution. This undesirable phenomenon is called the Maratos effect, which was observed by Maratos [114]. A well-known example was constructed by Powell [144] to verify the Maratos effect. A great deal of effort has been made to overcoming this phenomenon, leading to the development of the so called watchdog (or nonmonotone) techniques [21, 77] and second-order correction techniques [38, 61, 62].

The choice of the penalty parameter ρ plays an important role in the efficiency of the Sℓ_1QP method. Examples [20] were given to show that if the penalty parameter ρ is too small, the ℓ_1-penalty function may be unbounded below, and the iterates diverge unless the value of ρ is corrected in time; if ρ is too big, the efficiency of the penalty approach may be impaired. Existing strategies [74, 119] for updating the penalty parameter ρ adaptively are based on tracking the size of the Lagrange multipliers or checking the optimality conditions for the ℓ_1-penalty function $P_1(x, \rho)$. As pointed out by Fletcher and Leyffer [64], these strategies are not without problems.

A breakthrough in updating the penalty parameter ρ for Sℓ_1QP methods with line search was the introduction of steering rules [17, 20] that adjust the penalty parameter dynamically at every iteration to ensure sufficient progress in linear feasibility and to promote acceptance of the step. In order to adjust the penalty parameter, an auxiliary linear programming problem must be solved and the quadratic programming problem (1.1.10) must be computed one or more times using big values of the penalty parameter. This extra cost may not be significant to small- or medium-scale problems because warm starts can be employed in the solution of these additional quadratic programming problems. However, these extra costs may be potentially expensive to large-scale problems.

Recently, many researchers are interested in a new type of nonlinear exact penalty functions called the $\ell_{\frac{1}{p}} (p > 1)$ (or lower order)-penalty function

$$P_{\frac{1}{p}}(x, \rho) := f(x) + \rho \sum_{i \in \mathcal{I}} ([c_i(x)]_+)^{\frac{1}{p}}, \tag{1.1.11}$$

which also can be obtained from $P(x, \rho)$ by setting $Q(t) = t$ and replacing the norm $\| \cdot \|$ with the nonlinear operator $\|z\|_{\frac{1}{p}}^{\frac{1}{p}} = \sum_{k=1}^{m} |z_k|^{\frac{1}{p}}$. This type of penalty function has been employed in the study of mathematical programs with equilibrium constraints and error bounds; see, e.g., Luo et al. [109] and Pang [134]. Necessary and sufficient conditions for the exact penalization of the $\ell_{\frac{1}{p}}$-penalty function have been established in [152, 153] by virtue of the following generalized calmness condition

$$\liminf_{u \to 0} \frac{V(u) - V(0)}{\|u\|^{\frac{1}{p}}} > -\infty, \tag{1.1.12}$$

which is weaker than the calmness condition (1.1.9). Therefore, one advantage of the $\ell_{\frac{1}{p}}$-penalty function is that it requires, in general, weaker conditions than the ℓ_1-penalty function for the exact penalization representation, see Huang and Yang [89]. It also was shown in [152, 154] that the smallest exact penalty parameter corresponding to the $\ell_{\frac{1}{p}}$-penalty function is substantially smaller than that of the ℓ_1-exact penalty function. Although the penalty parameter can be adjusted dynamically by using the steering rules, a smaller exact penalty parameter plays an important role in the efficiency in the numerical implementation.

However, the function $P_{\frac{1}{p}}(x, \rho)$ is referred to as a non-Lipschitzian function because it may be not locally Lipschitz at the point where $c_i(x) = 0$. It is well known that the ℓ_1-exact penalty function implies the KKT conditions of problem (1.1.1). For the $\ell_{\frac{1}{p}}$-exact penalty function, this implication does not hold in general. For example, consider the simple problem of minimizing x subject to $x^2 \leq 0$, for which the KKT conditions do not hold at the local solution $x = 0$. The ℓ_1-penalty function for this problem is not exact at $x = 0$, but the $\ell_{\frac{1}{p}}$-penalty function with $p = 2$ is exact at $x = 0$. Therefore, not every $\ell_{\frac{1}{p}}$-exact penalty function can be qualified for detecting the KKT conditions.

A breakthrough in establishing the existence of Lagrange multipliers for problem (1.1.1) by virtue of the $\ell_{\frac{1}{p}}$-exact penalty function was done by Yang and Meng [181] by introducing a type of conditions in terms of first-order and (generalized) second-order derivatives of the constraints. Furthermore, an example was given to show that these conditions with $p = 2$ do not imply the weakest GCQ, and vice versa. Meng and Yang [117] extended the work of Yang and Meng [181] by studying the theory of deriving optimality conditions for problem (1.1.1) from very general exact penalty functions, and developed a unified theory from a modern perspective of variational analysis popularized by Rockafellar and Wets' book [151].

The $\ell_{\frac{1}{p}}$-penalty function shares a greater chance to be exact than the ℓ_1-penalty function, and its exactness implies the KKT conditions under mild conditions. However, the $\ell_{\frac{1}{p}}$-penalty function is nonsmooth, nonconvex and non-Lipschitz when

$p > 1$. Many well known optimization algorithms lack effectiveness and efficiency in dealing with nonsmooth and nonconvex objective functions. Furthermore, for non-Lipschitz continuous functions, the Clarke generalized gradients [34] cannot be used directly in the analysis. Thus the minimization of the $\ell_{\frac{1}{p}}$-penalty function is not an easy task. It has been shown that the smoothing approximate techniques are efficient methods for solving certain specially structured nonsmooth problems, see [32, 115, 118, 127, 177, 179, 180, 182].

Yang et al. [182] proposed a smoothing method for the $\ell_{\frac{1}{p}}$-exact penalty function. They presented an algorithm for problem (1.1.1) based on the smoothed $\ell_{\frac{1}{p}}$-penalty function and proved that the limiting point of the sequence for minimizing the smoothed penalty function satisfies the KKT conditions as the smoothing parameter goes to zero. Other smoothing methods have been proposed to smooth the $\ell_{\frac{1}{p}}$-penalty function in [115, 118, 176, 177]. A great challenge for the smoothing methods is how to set the value of the smoothing parameter. It is well known that the solutions of minimizing the smoothed penalty problem are unstable as the smoothing parameter is sufficiently small.

1.1.3 Interior-Point Penalty Methods

Interior-point methods have been proved to be successful for nonlinear optimization, and are currently considered the most powerful algorithms for large-scale nonlinear programming problems. The interior-point method (also called barrier method) was proposed by Firsch [66] to solve convex programming problems. We define the log-barrier function

$$B(x, \mu) := f(x) - \mu \sum_{i \in \mathcal{I}} \log c_i(x), \tag{1.1.13}$$

where $\mu > 0$ is the barrier parameter and $\log(\cdot)$ denotes the natural logarithm function. The classical interior-point method consists of finding (approximate) solutions of minimizing the barrier function (1.1.13) for a sequence of positive barrier parameter μ that converges to zero. A first challenge of this method is how to find a strict feasible initial point x^0 with respect to the inequality constraints $c_i(x)$, $i \in \mathcal{I}$. A second challenge is that, in general, the Hessian matrix of the barrier function (1.1.13) becomes increasing ill-conditioned as the solution is approached and is singular in the limit, see, e.g., [108, 124].

In order to overcome the above drawbacks, Polyak [142] proposed the modified barrier method, which minimizes the modified barrier function $B_M : \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^1_+ \to \mathbb{R}^1$ by the formula

$$B_M(x,\mu,\rho) := \begin{cases} f(x) - \rho^{-1} \sum\limits_{i \in \mathcal{I}} \mu_i \log(1-\rho c_i(x)), & \text{if } x \in \text{int}\Omega_\rho, \\ +\infty, & \text{if } x \notin \text{int}\Omega_\rho, \end{cases} \tag{1.1.14}$$

where $\rho > 0$, $\Omega_\rho := \{x \in \mathbb{R}^n \mid 1 - \rho c_i(x) \geq 0, \ i \in \mathcal{I}\}$, and $c_i(x)$, $i \in \mathcal{I}$, are convex functions. It is clear that the modified barrier method can handle infeasibility naturally. As pointed out by Curtis [42] that the modified barrier method, which essentially incorporates the Lagrange multiplier estimates to play the role of penalty parameters within a logarithmic barrier term, could be seen as the first kind of interior-point penalty methods. Under the assumption of the LICQ at the solution and other standard assumptions, it was shown in [142] that the iteration sequence converges to some KKT point. Moreover, a superlinear rate of convergence was established by virtue of the Newton method.

Modern interior-point methods [128] are well known as infeasible interior-point methods [19] which do not enforce satisfaction of the inequality constraints at each iteration. They typically make use of slack variables to transform problem (1.1.1) into the equivalent problem

$$\begin{aligned} \min_{x,s} \quad & f(x) \\ \text{s.t.} \quad & c_i(x) + s_i = 0, \ i \in \mathcal{I}, \\ & s_i \geq 0, \ i \in \mathcal{I}. \end{aligned} \tag{1.1.15}$$

Yamashita [178] proposed the interior-point (barrier) ℓ_1-penalty method for problem (1.1.15) in the following two steps. First, problem (1.1.15) is reformulated as a logarithmic barrier subproblem

$$\begin{aligned} \min_{x,s} \quad & f(x) - \mu \sum_{i \in \mathcal{I}} \log s_i \\ \text{s.t.} \quad & c_i(x) + s_i = 0, \ i \in \mathcal{I}, \end{aligned} \tag{1.1.16}$$

which is an approximation to problem (1.1.15). Then, the Newton method is used to solve the KKT conditions of problem (1.1.16). Under the assumption of the LICQ for active constraints at the solution, the global convergence of this method using the Armijo line search was established by employing the barrier-penalty function

$$\phi_1(x,s,\rho,\mu) := f(x) - \mu \sum_{i\in\mathcal{I}} \log s_i + \rho \sum_{i\in\mathcal{I}} |c_i(x) + s_i|. \tag{1.1.17}$$

It was shown in [178] that this method keeps the advantage of interior-point methods for large-scale problems and overcomes the inevitable numerical difficulties occurred at the final stage of iterations for the classical interior-point methods.

Wächter and Biegler [166] constructed a well-posed analytical example to illustrate the failure of global convergence for a class of line search interior point methods [48, 165, 178] when starting from certain points. More examples were given by Byrd et al. [18]. Careful examination shows that the main difficulty stems from the possible rank deficiency of the Jacobian matrix for active inequality constraints at the infeasible not-stationary point. This difficulty can be readily avoided in inequality constrained problems by adding slack variables and employing certain feasibility control strategies; see, e.g., [14, 105].

Conn et al. [36] proposed a new interior-point ℓ_1-penalty method for problem (1.1.1), which makes use of the ℓ_1-exact penalty function (1.1.7). It is well known that the minimization of the ℓ_1-penalty function $P_1(x,\rho)$ can be reformulated as a smooth problem [74, 76]

$$\begin{aligned} \min_{x,s} \quad & P_1^S(x,s,\rho) := f(x) + \rho \sum_{i\in\mathcal{I}} s_i \\ \text{s.t.} \quad & s_i - c_i(x) \geq 0, \quad i \in \mathcal{I}, \\ & s_i \geq 0, \quad i \in \mathcal{I}. \end{aligned} \tag{1.1.18}$$

The point (x,s) is strictly feasible for problem (1.1.18) if the artificial variables are sufficiently large. The interior-point method places the inequality constraints in a barrier term leading to the following interior-point ℓ_1-penalty problem

$$\min_{x,s} \; P_1^B(x,s,\rho,\mu) := P_1^S(x,s,t,\rho) - \mu \sum_{i\in\mathcal{I}} \Big(\log(s_i - c_i(x)) + \log s_i\Big) \tag{1.1.19}$$

where $\mu > 0$ is the barrier parameter. Compared with the nondifferentiable merit function $\phi_1(x,s,\rho,\mu)$ given in (1.1.7), the function $P_1^B(x,s,\rho,\mu)$ is twice continuously differentiable under the assumptions of problem (1.1.1). They employed the trust region method that incorporates exact second-order derivative information of the function $P_1^B(x, s, t, \rho, \mu)$ to approximately solve problem (1.1.19). It is surprising that the MFCQ holds at every feasible point of problem (1.1.18), regardless of any constraint qualification being satisfied or not for problem (1.1.1). Then there always exist bounded Lagrange multipliers for the KKT conditions of problem (1.1.18). It was shown in [76] that the iteration sequence converges to some KKT point of problem (1.1.1) if there exists a threshold $\hat{\rho} > 0$ such that for all $\rho^i \leq \hat{\rho}$, where $\{\rho^i\}$ is the sequence of the penalty parameter used to produce the iteration sequence. On the other hand, the iteration sequence converges to some FJ point if the penalty parameter ρ goes to infinite and the MFCQ fails to hold at the limit point. Furthermore, the local Q-superlinear convergence was established under more restrictive assumptions such as the LICQ holds for the active inequality constraints at the solutions of problem (1.1.1).

Combining the regularization effects on the constraints from the ℓ_1-penalty function and the efficiency of Newton-like methods in large-scale optimization problems from interior-point methods, Curtis [42] introduced an interior-point ℓ_1-penalty method for problem (1.1.1). A common challenge in the implementation of both penalty methods and interior-point methods is the design of an effective strategy for updating the penalty and barrier parameters. Curtis presented an algorithm with novel feature on updating them.

Hyrd et al. [14] introduced an interior-point ℓ_2-penalty method, in which the merit function was constructed only with primal variables and can be written as

$$\phi_2(x, s, \rho, \mu) := f(x) - \mu \sum_{i \in \mathcal{I}} \log s_i + \rho \|c(x) + s\|. \tag{1.1.20}$$

This method applies the sequential quadratic programming techniques to a sequence of barrier problems (1.1.15), and uses the trust-region to ensure the robustness of the iteration and to allow the direct use of the second-order derivatives. The convergence to KKT points was established by assuming the LICQ for active constraints at the local solutions. Numerical experiments with an implementation of this method have been performed in [15] and showed that this approach holds much more promise. Furthermore, the superlinear convergence of this method was established in [16] under suitable assumptions. Tseng [163] further studied this method and established the convergence to KKT points under a relaxed MFCQ which is weaker than the LICQ employed in [14]. Moreover, Tseng established convergence to second-order stationary points under the LICQ, see [163, Corollary 6.2].

Using the same merit function $\phi_2(x, s, \rho, \mu)$, Liu and Sun [104, 106] introduced an interior-point ℓ_2-penalty method which has the theoretical properties of trust-region methods, but works entirely by the line search. Instead of introducing additional trust-region constraints, this method uses refined line search rules to generate a new iterate in a decomposed SQP framework. The search direction is determined by either a Newton-type step or a Cauchy-type step with the choice being made with reference to a penalty parameter in the merit function. Global convergence properties were derived without assuming regularity conditions, but a steepest descent approach would be used whenever the Newton direction fails to be a descent direction. Doing so guarantees the global convergence theoretically, but would greatly increase the iteration count within an implementation. However, unlike the trust-region rules used in [14], their method did not have the flexibility to allow the direct use of indefinite second-order derivatives. This method was improved by Liu and Yuan [107] by using the null-space techniques. Their proposed method approximately solves a sequence of subproblems (1.1.15) by computing a range-space step and a null-space step in every iteration. Under very mild conditions on range-space steps and approximate Hessian matrix, without assuming any regularity, the same convergent results were established in [105]. Furthermore, they analyzed the local convergence properties, and proved that by suitably controlling the exactness of range-space steps and selecting the barrier parameter and Hessian approximation, the approach generates a superlinear or quadratically convergent step.

Instead of using the SQP method and trust-region rules to solve barrier subproblem (1.1.15) in [14, 105, 163], Chen and Goldfarb [31] introduced another interior-point ℓ_2-penalty method which applies a modified Newton method [6, 78, 167] to approximately solve the KKT conditions of a sequence of barrier subproblems (1.1.15). This method can be seen as the line search regularized Newton method that takes the advantage of exact penalization of the ℓ_2-merit function defined in (1.1.20). Under mild assumptions, this method enjoys strong global convergence properties and fast local convergence after a slight modification.

1.2 Complementarity Problems

The general complementarity problem (GCP, for short) is to find a vector $x \in \mathbb{R}^n$ such that

$$F(x) \leq 0,\ G(x) \leq 0,\ F(x)^T G(x) = 0,$$

where the functions $F, G : \mathbb{R}^n \to \mathbb{R}^n$ are assumed to be continuously differentiable. It is well known that the GCP can be derived from the general variational inequality, which is a powerful tool to prove the existence of a solution to the GCP and has been fully studied in [129, 130, 137, 149] and the references therein. Some efficient methods such as the projection equation and trust region methods [92, 187] have been proposed for solving the GCP. As pointed out in [2] that the GCP can be reformulated as a mixed complementarity problem (MiCP, for short), which is equivalent to a variational inequality, see [50, Chapter 1]. In addition to optimization problems, many problems in real world can be cast as MiCPs, such as Nash equilibrium problems [125, 126], Oligopolistic electricity models [170], traffic equilibrium models [138], frictional contact problems [140], nonlinear obstacle problems [131] and pricing American options [172]. Overviews of how this is accomplished are given in [45, 50, 54, 82, 155].

In particular, if $G(x) \equiv x$, the GCP reduces to a nonlinear complementarity problem (NCP, for short) which was introduced by Cottle [39] for finding stationary points for nonlinear programming problems. Specifically, the NCP is to find a vector $x \in \mathbb{R}^n$ such that

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

Moreover, if F is an affine function, i.e., $F(x) = Ax - b$ with a matrix $A \in \mathbb{R}^{n \times n}$ and a vector $b \in \mathbb{R}^n$, then the NCP reduces to a linear complementarity problem (LCP, for short), which in turn contains linear and quadratic programming problems as special cases. A comprehensive investigation in the complementarity problems from the basic theoretical results to numerical methods can be found in monographs [40, 50, 51] and the vast references therein.

1.2.1 Equation-Based Methods

One type of the most powerful methods in solving the NCP is the equation-based methods, which are to reformulate the NCP as a system of nonlinear equations, or a minimization problem. A merit function whose global minima are the solutions of the NCP plays a vital role in these methods, which is defined, if not always, by a preliminary equation reformulation of the complementarity problem. Specifically, we define a system of equations $H(x) = 0$, whose solutions coincide with the solutions of the NCP, and then use the merit function $\Phi(x) := \|H(x)\|^2$ (or $\|H(x)\|$). There are several ways to construct the system of equations $H(x) = 0$. Mangasarian [111] introduced a class of smooth reformulations for $H(x)$, which have been further explored in [68, 94, 164]. A common drawback of the smooth reformulations as merit functions is that differentiable merit functions often fail to provide a sound basis for the development of fast local methods for degenerate problems, see [51, Proposition 9.1.1].

In the last two decades, the nonsmooth reformulations for $H(x)$ have received great attention [29, 43, 49, 52, 55, 58, 81, 83, 93, 95], since that they not only allow to define superlinearly convergent algorithms for degenerate problens but also the subproblems to be solved at each iteration tend to be more numerically stable. However, a price to pay is that the globalization becomes more complex since the merit function $\|H(x)\|^2$ is once but not twice continuously differentiable. Fortunately, these merit functions are B-differentiability [132, 133] or even semismoothness [146] that is a stronger analytical property than the B-differentiability, so using the recent powerful theory for solving B-differentiable equations [150] and semismooth equations [136, 145, 146], a fast local algorithm that only requires the solution of one linear system at each iteration can be developed to solve the NCP.

To construct the merit function for the NCP, a class of functions named NCP-functions plays a significant role. A function $\phi : \mathbb{R}^2 \to \mathbb{R}$ is called NCP-function if it satisfies $\phi(a, b) = 0$ if and only if $a \leq 0,\ b \leq 0,\ ab = 0$. By using the NCP-function ϕ, the merit function $\Phi : \mathbb{R}^n \to \mathbb{R}$ for the NCP can be constructed by

$$\Phi(x) := \begin{cases} \sum_{i=1}^{n} \phi(x_i, F_i(x)), & \text{if } \phi \text{ is nonnegative on } \mathbb{R}^2, \\ \sum_{i=1}^{n} \phi(x_i, F_i(x))^2, & \text{otherwise.} \end{cases}$$

There are many functions that belong to the class of NCP-functions. The up-to-date reviews on the NCP-functions can be found in [60, 69, 71, 160]. We recall three kinds of NCP-functions which have been well studied in the literature.

(a) $\phi_{NR}(a,b) := \max\{a,b\}$;

(b) $\phi_{FB}(a,b) := \sqrt{a^2+b^2}+a+b$;

(c) $\phi_{MS}(a,b) := ab + \frac{1}{2\alpha}\big(\max\{0,\alpha b-a\}^2 - a^2 + \max\{0,\alpha a-b\}^2 - b^2\big),\ \alpha>1.$

The function ϕ_{NR} can be rewritten as $\phi_{NR}(a,b) = \max\{b-a,0\}+a$, which is due to Wierzbicki [171]. The merit function Φ using the function ϕ_{NR} is well-known as the natural residual, which have been fully investigated to design efficient algorithms such as NE/SQP method [135] and smoothing methods [23, 25]. The function ϕ_{FB} which was considered by Fischer [57] and attributed to Burmeister plays a central role in the development of efficient algorithms for the NCP and has been intensively studied in [49, 51, 52, 59, 60, 93, 95]. There are many variants of the ϕ_{FB} based on which efficient numerical methods can be designed, such as generalized Fischer-Burmeister (FB) functions [26, 28, 29, 91, 97, 162] and penalized FB functions [22, 24]. The function ϕ_{MS} is nonnegative on $\mathbb{R}^2$ and the merit function based on it is the implicit Lagrangian proposed by Mangasarian and Solodov [113].

Another key issue on the equation-based methods is the regular condition which is used to guarantee that every stationary point of the merit function is a solution of the NCP. Different regular conditions corresponding to different kinds of merit functions have been proposed in [94, 113, 122, 147, 148]. Before listing them, we let $\mathcal{J} := \{1,2,\ldots,n\}$ and define the following three index sets at a solution x^* of the NCP

$$\begin{aligned}
\alpha &:= \{i\in\mathcal{J} \mid x_i^* < 0\};\\
\beta &:= \{i\in\mathcal{J} \mid x_i^* = 0 = F_i(x^*)\};\\
\gamma &:= \{i\in\mathcal{J} \mid F_i(x^*) < 0\}.
\end{aligned}$$

The solution x^* is said to be nondegenerate if $\beta=\emptyset$. A matrix $M\in\mathbb{R}^{n\times n}$ is called a P-matrix if every of its principal minors is positive. We review two kinds of regular conditions which have been widely used in the analysis for the NCP. The solution x^* is said to be

(a) b-regular if, for every every index set δ such that $\alpha \subseteq \delta \subseteq \alpha \cup \beta$, the principal submatrix $\nabla F_{\delta\delta}(x^*)$ is nonsingular;

(b) R-regular if $\nabla F_{\alpha\alpha}(x^*)$ is nonsingular and the Schur complement of $\nabla F_{\alpha\alpha}(x^*)$ in

$$\begin{pmatrix} \nabla F_{\alpha\alpha}(x^*) & \nabla F_{\alpha\beta}(x^*) \\ \nabla F_{\beta\alpha}(x^*) & \nabla F_{\beta\beta}(x^*) \end{pmatrix}$$

is a P-matrix, where $\nabla F(x^*)$ denotes the Jacobian matrix of function F at x^*.

Note that R-regularity implies b-regularity [135, 147], while b-regularity guarantees local uniqueness of the solution x^* [52, 101]. Some latest reviews on regular conditions for the NCP can be found in [43, 52, 59]. Jiang and Qi [93] proposed a distinctive regular condition for the merit function Φ generated by ϕ_{FB}, which requires that the function F is a uniform P-function. A function $F : \mathbb{R}^n \to \mathbb{R}^n$ is called a uniform P-function [121] if there exists a constant $\alpha > 0$ such that

$$\max_{1 \le i \le n} (y_i - x_i)\big(F_i(y) - F_i(x)\big) \ge \alpha \|y - x\|^2, \text{ for all } x, y \in \mathbb{R}^n.$$

Under this regular condition, they proved that every stationary point of the unconstrained problem is a global solution; furthermore, the level sets of the merit function are bounded. Geiger and Kanzow [73] proved the former if the function F is monotone, and the latter if the function F is strongly monotone.

1.2.2 Power Penalty Method

Recently, the power penalty method has received a great deal of attention in solving complementarity problems. The general power penalty problem for the NCP is to transform it into the following nonlinear equations, which are to find a vector $x^\rho \in \mathbb{R}^n$ satisfying

$$F(x) + \rho[x]_+^{\frac{1}{p}} = 0, \tag{1.2.21}$$

where $\rho > 0$ is the penalty parameter, $p \ge 1$ 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 $i \in \mathcal{J}$. As $p = 1$, the power penalty method reduces to the classical ℓ_1-penalty method which was proposed by Bensoussan and Lions [7] for solving the continuous variational inequality. They proved that the solution x^ρ

converges to a solution x^* of the NCP at a rate of $\mathcal{O}(\rho^{-\frac{1}{2}})$, that is, there exists a constant $C > 0$ such that

$$\|x^\rho - x^*\| \leq C\rho^{-\frac{1}{2}}.$$

Furthermore, the ℓ_1-penalty method was widely used to solve the LCP arising from American options [44, 65, 143, 172, 188], the Hamilton-Jacobi-Bellman (HJB) equations in finance [173, 174] and obstacle problems [156]. This square root rate of convergence requires that ρ is sufficiently big so as to achieve a given accuracy of the approximate solution. However, researchers in [63, 188] pointed out that big values of the penalty parameter ρ result in poorly conditional algebraic problems in solving nonlinear equations (1.2.21).

As $p > 1$, the power penalty method becomes the $\ell_{\frac{1}{p}}$-penalty method which was proposed by Wang et al. [169] to solve the LCP arising from American options. They proved the solution x^ρ converges to x^* at a rate of $\mathcal{O}(\rho^{-\frac{p}{2}})$, which improves significantly the existing theoretical result of the square root rate of convergence mentioned above. Zhang applied the $\ell_{\frac{1}{p}}$-penalty method to solve more models from American options [185, 186]. Furthermore, Huang and Wang [87] extended the $\ell_{\frac{1}{p}}$-penalty method to solve the NCP and they showed that the convergence rate between the solution of penalized equations and that of the NCP is of order $\mathcal{O}(\rho^{-\frac{p}{\xi}})$, provided that the function F is continuous and ξ-monotonicity for a positive constant $\xi > 1$. The same convergence rate has been established in [88] for the $\ell_{\frac{1}{p}}$-penalty method in solving the MiCP.

The ℓ_1-penalized equations can be solved efficiently by nonsmooth Newton methods [132, 146]. However, it is unfortunate that 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 [169, 185]. A vital drawback of smoothing methods is that their solutions become unstable as the smoothing parameter is sufficiently small.

1.3 Notation

In this thesis, the notation is standard. The space of real vectors of length n is denoted by $\mathbb{R}^n$, while the space of real $m \times n$ matrices is denoted by $\mathbb{R}^{m\times n}$. We write $\mathbb{R}^n_+$ to denote the set of nonnegative real vectors of length n, while $\mathbb{R}^n_{++}$ to denote the set of positive real vectors of length n. Given a vector $x \in \mathbb{R}^n$, we use x_i to denote its i-th component. We invariably assume that x is a column vector, and its transpose is denoted by x^T which is a row vector. We write $x \geq 0$ to indicate componentwise nonnegativity, that is, $x_i \geq 0$ for all $i = 1, \ldots, n$, while $x > 0$ indicates that $x_i > 0$ for all $i = 1, \ldots, n$. We write $[x]_+$ to indicate a new vector with components $([x]_+)_i = \max\{x_i, 0\}$ for all $i = 1, \ldots, n$, while $[x]_+^\sigma$ indicates that $([x]_+^\sigma)_i = \max\{x_i, 0\}^\sigma$ for all $i = 1, \ldots, n$, with given real number $\sigma \geq 0$. We write $[x]_-$ to indicate a new vector with components $([x]_-)_i = \max\{-x_i, 0\}$ for all $i = 1, \ldots, n$. We write $X = \mathrm{diag}(x)$ to indicate a diagonal matrix $X \in \mathbb{R}^{n \times n}$ whose i-th diagonal element is x_i for all $i = 1, \ldots, n$. We write $\|x\|$, $\|x\|_1$ and $\|x\|_\infty$ to indicate its Euclidean norm (also called ℓ_2-norm), ℓ_1-norm and ℓ_∞-norm, respectively.

Given vectors $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$, the standard inner product is $x^T y = \sum_{i=1}^n x_i y_i$. We write $x \geq (>) y$ to indicate that $x_i \geq (>) y_i$ for all $i = 1, \ldots, n$. 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 (x, y) (or $\binom{x}{y}$) to indicate a vector in $\mathbb{R}^{2n}$, that is $(x, y) := (x^T \ y^T)^T$.

Given a matrix $A \in \mathbb{R}^{m \times n}$, we specify its components by double subscripts as A_{ij}, for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. The transpose of A is denoted by A^T, while A^{-1} denotes the inverse of matrix A if A is invertible. We write $\|A\|$ to indicates its Frobenius norm, that is,

$$\|A\| = \Big(\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2 \Big)^{1/2}.$$

The matrix A is said to be square if $m = n$. A square matrix A is positive definite ($A \succ 0$) if there exists a positive scalar $\alpha > 0$ such that

$$x^T A x \geq \alpha x^T x, \text{ for all } x \in \mathbb{R}^n.$$

It is positive semidefinite ($A \succeq 0$) if

$$x^T A x \geq 0, \text{ for all } x \in \mathbb{R}^n.$$

Assume A is a positive semidefinite and diagonal matrix, we write A^σ to indicate a diagonal matrix with components $(A^\sigma)_{ii} := (A_{ii})^\sigma$ for all $i = 1, \ldots, n$, with given real number $\sigma \geq 0$.

We write e_i to indicate a vector with i-th component 1 and 0 otherwise. We write $e = e_1 + \cdots + e_n$ to indicate a vector whose all components are 1. The identity matrix, denoted by E, is the square diagonal matrix whose diagonal components are all 1.

Given a point $x \in \mathbb{R}^n$, we call $\mathcal{N} \in \mathbb{R}^n$ a neighborhood of x if it is an open set containing x. We write $\mathbb{B}(x, \epsilon)$ to indicate a open ball of radius ϵ around x, that is,

$$\mathbb{B}(x, \epsilon) := \{y \in \mathbb{R}^n \mid \|y - x\| \leq \epsilon\},$$

while $\mathbb{B}$ denotes the unit ball centered at origin.

Considering the function $f : \mathcal{D} \to \mathbb{R}^m$ where $\mathcal{D} \subset \mathbb{R}^n$ for general m and n. The function f is said to be Lipschitz continuous on some set $\mathcal{N} \subset \mathcal{D}$ if there exists a constant $L > 0$ such that

$$\|f(x) - f(y)\| \leq L\|x - y\|, \text{ for all } x, y \in \mathcal{N}.$$

The function f is called a real-valued function if $m = 1$ and is called vector-valued function if $m > 1$. For a twice continuously differentiable real-valued function f, we write $\nabla f(x)$ to denote its gradient vector of f at x, while $\nabla^2 f(x)$ to indicate its Hessian matrix of f at x. For a continuously differentiable vector-valued function f, we write $\nabla f(x)$ to denote its Jacobian matrix of f at x. The Dini upper-directional derivative [181] and subderivative [151] of the real-valued function f at x in the direction $u \in \mathbb{R}^n$ are defined, respectively, by

$$\begin{aligned} D_+ f(x)(u) &:= \limsup_{t \to 0^+} \frac{f(x + tu) - f(x)}{t}, \\ df(x)(u) &:= \liminf_{t \to 0^+,\, u' \to u} \frac{f(x + tu') - f(x)}{t}. \end{aligned}$$

We say that function $\eta : \mathbb{R} \to \mathbb{R}$ converges to 0 at a rate of $\mathcal{O}(\nu)$ if there exists a constant $C > 0$ such that $|\eta(\nu)| \leq C|\nu|$, when η is sufficiently small.

1.4 Motivation and Outline of the Thesis

The $\ell_{\frac{1}{p}}$ $(p > 1)$-penalty method is becoming a powerful tool to solve some fundamental mathematical models such as constrained nonlinear programming problems and complementarity problems. For constrained nonlinear programming problems, it was shown in [152] that the existence of an $\ell_{\frac{1}{p}}$-exact penalty function requires weaker conditions than that of the ℓ_1-exact penalty function and that the smallest exact penalty parameter of the $\ell_{\frac{1}{p}}$-exact penalty function is also smaller than that of the ℓ_1-exact penalty function. Furthermore, the $\ell_{\frac{1}{p}}$-exact penalty function has been used in the establishment of first-order optimality conditions. Specifically, under some second order conditions and the existence of the $\ell_{\frac{1}{p}}$-exact penalty function, first-order optimality conditions of constrained nonlinear programming problems were established and examples were given to show that these conditions do not imply the weakest GCQ and vice versa in [117, 181].

However, the $\ell_{\frac{1}{p}}$-penalty function is locally nonconvex and non-Lipschitzian. These features make many well-known optimization methods such as quasi-Newton methods [128] and gradient sampling methods [13] lack the effectiveness and the efficiency in solving the minimization of the $\ell_{\frac{1}{p}}$-penalty function directly. Smoothing methods [115, 118, 179, 182] seem to be the only choice in dealing with the the $\ell_{\frac{1}{p}}$-penalty function. Nevertheless, it is well known that the solutions of minimizing the smoothed $\ell_{\frac{1}{p}}$-penalty function are unstable as the smoothing parameter is sufficiently small. In this thesis, motivated by the interior-point ℓ_2-penalty methods [30, 105] and interior-point ℓ_1-penalty methods [5, 76], we propose an interior-point $\ell_{\frac{1}{p}}$-penalty method to solve inequality constrained nonlinear programming problems in Chapter 2.

The $\ell_{\frac{1}{p}}$-penalty method was introduced to solve a LCP arising from the American option valuation in [169]. Under mild conditions, their convergence rate is faster than that of the ℓ_1-penalty method proposed by Bensoussan and Lions [7]. More specifically, the solution x^ρ of $\ell_{\frac{1}{p}}$-penalized equations converges to a solution x^* of the complementarity problem in the speed of $\mathcal{O}(\rho^{-\frac{p}{2}})$, that is, there exists a constant $C > 0$ such that $\|x^\rho - x^*\| \leq C\rho^{-\frac{p}{2}}$. However, the convergence rate of the ℓ_1-penalty method is only of $\mathcal{O}(\rho^{-\frac{1}{2}})$. The penalty parameter ρ, which is vital to keep the stability of solution of the penalized equations, used for $\ell_{\frac{1}{p}}$-penalty method is smaller than that used for the ℓ_1-penalty method in order to achieve a given accuracy. The same order of convergence rate has been proved for the LCP [168] under the assumption of a M-matrix. Furthermore, the convergent rate of $\mathcal{O}(\rho^{-\frac{p}{\xi}})$ has been proved to the NCP and MiCP under the assumptions of the continuity and the ξ-monotonicity with $\xi \in (1, 2]$ in [87, 88]. In Chapter 3, we propose a box-constrained differentiable penalty method for solving nonlinear complementarity problems, which not only shares the convergence rate of the existing $\ell_{\frac{1}{p}}$-penalty method but also overcomes the drawback of the non-Lipschitzianness corresponding to the $\ell_{\frac{1}{p}}$-penalized equations. Furthermore, we introduce an unconstrained differentiable penalty method to solve general complementarity problems in Chapter 4.

The outline of the thesis is as follows.

In Chapter 2, we aim at designing algorithms that solve the inequality constrained nonlinear programming problems efficiently by virtue of the $\ell_{\frac{1}{p}}$-penalty function. In Section 2.2, we introduce a technique of p-order relaxation to relax the nonconvex and non-Lipschitzian $\ell_{\frac{1}{p}}$-penalty problem into an equivalent constrained problem which shares the same differentiable property as the original problem. Combining with an interior-point method, we propose an interior-point $\ell_{\frac{1}{p}}$-penalty method. Then, we introduce different kinds of constraint qualifications to establish first-order necessary conditions for the relaxed problem. Combining with an interior-point method, in Section 2.3, we propose an interior-point $\ell_{\frac{1}{p}}$-penalty method. We employ the modified Newton method to solve a sequence of logarithmic barrier subproblems and detail three numerical algorithms which constitute the interior-point $\ell_{\frac{1}{p}}$-penalty method. Furthermore, under mild conditions, we prove that the iteration sequence converges to a KKT (or FJ) point of the original problem. In Section 2.4, we conduct our numerical experiments on three test problems sets: small- to medium-scale problems, large-scale problems and problems with degenerate constraints. We use the first test set to compare the performance of the interior-point $\ell_{\frac{1}{p}}$-penalty method with different values of the power p. Then we compare the performance of the interior-point $\ell_{\frac{1}{2}}$-penalty method with existing interior-point ℓ_1-penalty methods.

In Chapter 3, we propose a box-constrained differentiable penalty method by virtue of the $\ell_{\frac{1}{p}}$-penalty method for the NCP. In Section 3.2, we introduce a new definition for the function F named a uniform ξ-P-function which is weaker than the ξ-monotonicity and reduces to the P-function if function F is linear. Then, we propose a box-constrained differentiable penalty method which not only inherits the convergence rate of the $\ell_{\frac{1}{p}}$-penalty method but also can be implemented efficiently by classical iteration methods. Specifically, we prove that the solution of the boxed-constrained penalized equations converges to a solution of the NCP at a rate of $\mathcal{O}(\rho^{-\frac{p}{\xi}})$ if the function F is a uniform ξ-P-function. Instead of solving box-constrained penalized equations directly, in Section 3.3, we solve a least squares problem with box constraints by use of a trust-region Gauss-Newton method [123]. In Section 3.4, we carry out our numerical experiments on the test problems from MCPLIB [45]. We first set $p = 2$ and compare the performances of our method with the smoothed $\ell_{\frac{1}{2}}$-penalty method [87] and the ℓ_1-penalty method [7] in terms of the number of function evaluations and the values of the penalty parameter. Then different values of the power p are chosen to test the efficiency of our method. Furthermore, we compare the performance of our method with the smooth approximation method [23] and the nonsmooth equations method [93] in terms of the number of function evaluations.

In Chapter 4, we propose an unconstrained differentiable penalty method for the GCP. In Section 4.2, we establish the convergence rate of the order $\mathcal{O}(\rho^{-\frac{p}{\xi}})$ between the solution of penalized equations and that of the original problem, under the assumption of a uniform ξ-P-function. In Section 4.3, we carry out our numerical experiments on the same test problems used in Chapter 3. We first set $p = 2$ to the proposed method to compare its performance with the box-constrained differentiable penalty method with $p = 2$ and the ℓ_1-penalty method [7] in terms of the number of function evaluations and the values of the penalty parameter. Using the same terms, we test the performances of the new method with different values of power p. Finally, we compare the performance of the new method with two well known methods in terms of the number of function evaluations.

In Chapter 5, we conclude the thesis and provide directions for future research work.

Chapter 2

An Interior-Point $\ell_{\frac{1}{p}}$-Penalty Method for Nonlinear Optimization

2.1 Introduction

In this chapter, we consider the inequality constrained nonlinear programming problem

$$\begin{aligned} &\min\ f(x) \\ &\text{s.t. } c_i(x) \le 0,\ i \in \mathcal{I}, \end{aligned} \tag{2.1.1}$$

where the functions f and $c_i : \mathbb{R}^n \to \mathbb{R}$ are assumed to be twice continuously differentiable and $\mathcal{I} = \{1, 2, \ldots, m\}$.

Motivated by interior-point ℓ_1-penalty methods [5, 42, 76], we introduce a technique of the p-order relaxation to the nonsmooth and non-Lipschitzian $\ell_{\frac{1}{p}}$-penalty problem to transform it into an equivalent problem which shares the same differentiable property as problem (2.1.1). We introduce different kinds of constraint qualifications to establish first-order necessary conditions for the relaxed problem. Combining the interior-point method, we propose an interior-point $\ell_{\frac{1}{p}}$-penalty method for problem (2.1.1).

We employ a modified Newton's method with an inexact line search to solve the first-order necessary conditions of the barrier problem. Due to the p-order relaxation, we present a condition on the Lagrange multipliers of original inequality constraints and that of inequality constraints of the relaxed problem in order to guarantee the positive definiteness of the Jacobian matrix of the first-order necessary conditions. We describe three specific algorithms. The first algorithm is to solve the barrier problem with a fixed penalty parameter ρ and a fixed barrier parameter μ, the second one is to solve a sequence of relaxed problems when ρ is fixed and μ goes to zero and the third one is to solve the penalty problem when the penalty parameter ρ goes to infinite. Finally, under mild conditions, we prove that the iteration sequence converges to some KKT (or FJ) point of problem (2.1.1).

We carry out numerical experiments on three test problems sets. The first one contains 266 small-scale and medium-scale test problems from CUTEr collection, COPS, MITT and Global test sets, the second one contains 26 large-scale test problems from COPS and MITT and the last one contains 37 test problems with degenerate constraints from DEGEN_collection and one degenerate problem from [117]. We compare our method with two existing interior-point ℓ_1-penalty methods: PIPAL-a and PIPAL-c in [42].

This chapter is organized as follows. In Section 2.2, we introduce a p-order relaxation scheme to the $\ell_{\frac{1}{p}}$-penalty problem and investigate its optimality conditions under different constraint qualifications. In Section 2.3, we propose an interior-point $\ell_{\frac{1}{p}}$-penalty method and present its analysis on a modified Newton method and corresponding algorithms, moreover on global convergence. In the last section, we present the numerical results.

2.2 p-Order Relaxation of the $\ell_{\frac{1}{p}}$-Penalty Problem

In this section, a technique of p-order relaxation is introduced to recast the minimization of the $\ell_{\frac{1}{p}}$-penalty function as an equivalent constrained problem that shares the same differentiability as problem (2.1.1). Specifically, we relax the following $\ell_{\frac{1}{p}}$-penalty problem

$$\min_x \ \phi_{P,\frac{1}{p}}(x,\rho) := f(x) + \rho \sum_{i\in\mathcal{I}} [c_i(x)]_+^{\frac{1}{p}} \tag{2.2.1}$$

as follows

$$\begin{aligned} \min_{x,s} \ & \phi_{S,\frac{1}{p}}(x,s;\rho) := f(x) + \rho \sum_{i\in\mathcal{I}} s_i \\ \text{s.t. } & c_i(x) \le s_i^p \text{ and } \ s_i \ge 0, \ i \in \mathcal{I}, \end{aligned} \tag{2.2.2}$$

where $\rho > 0$ is the penalty parameter, $p \geq 1$ is the power, $[a]_+ = \max\{a, 0\}$ for any $a \in \mathbb{R}$ and $s = (s_i) \in \mathbb{R}^m_+$ are artificial variables. As $p = 1$, the p-order relaxation is known as the linear relaxation which plays an important role in the interior-point ℓ_1-penalty methods [5, 76]. In this chapter, we mainly focus on the case of $p > 1$. Let $(\hat{x}, \hat{s}) \in \mathbb{R}^{n+m}$ be a local solution of problem (2.2.2).

Throughout this chapter, we define the index sets at $x \in \mathbb{R}^n$ as follows

$$\begin{aligned} \mathcal{I}^-(x) &:= \{i \in \mathcal{I} \mid c_i(x) < 0\}; \\ \mathcal{I}^0(x) &:= \{i \in \mathcal{I} \mid c_i(x) = 0\}; \\ \mathcal{I}^+(x) &:= \{i \in \mathcal{I} \mid c_i(x) > 0\}. \end{aligned}$$

We introduce the following index sets for $(x, s) \in \mathbb{R}^{n+m}$

$$\begin{aligned} S^0(x,s) &:= \{i \in \mathcal{I} \mid s_i = 0 \text{ and } c_i(x) \le 0\}; \\ S^+(x,s) &:= \{i \in \mathcal{I} \mid s_i > 0 \text{ and } c_i(x) \le s_i^p\}; \\ S^=(x,s) &:= \{i \in S^+(x,s) \mid c_i(x) = s_i^p\}; \\ CS^0(x,s) &:= \{i \in S^0(x,s) \mid c_i(x) = 0\}. \end{aligned}$$

We define the feasible set $\widehat{\mathcal{F}}$ for problem (2.2.2) by

$$\widehat{\mathcal{F}} := \{(x,s) \in \mathbb{R}^{n+m} \mid c_i(x) \le s_i^p, \ s_i \ge 0, \ \forall\, i \in \mathcal{I}\}.$$

The following proposition concludes that the $\ell_{\frac{1}{p}}$-penalty problem (2.2.1) and its p-order relaxed problem (2.2.2) are equivalent in the sense that they have the same local solution.

Proposition 2.2.1. *Given the penalty parameter $\rho > 0$, a point $\hat{x}$ solves problem (2.2.1) locally if and only if the point $(\hat{x}, \hat{s})$ solves problem (2.2.2) locally with $\hat{s}_i = [c_i(\hat{x})]_+^{\frac{1}{p}}$ for all $i \in \mathcal{I}$.*

Proof. We prove this proposition by considering two cases.

Case 1. We assume $\hat{x} \in \mathcal{F}$. In this case, we have $\hat{s} = 0$. Suppose that $\hat{x}$ solves problem (2.2.1). Take $\hat{s} = 0$, and then $(\hat{x}, 0)$ solves problem (2.2.2) locally. Conversely, let $(\hat{x}, 0)$ solve problem (2.2.2) locally. Assume to the contrary that $\hat{x}$ does not solve problem (2.2.1) locally. Thus there exists a sequence $\{x^k\} \to \hat{x}$ such that

$$f(x^k) + \rho \sum_{i\in\mathcal{I}} [c_i(x^k)]_+^{\frac{1}{p}} < f(\hat{x}) + \rho \sum_{i\in\mathcal{I}} [c_i(\hat{x})]_+^{\frac{1}{p}} = f(\hat{x}). \tag{2.2.3}$$

Since $(\hat{x}, 0)$ solves problem (2.2.2) locally, it follows that there exists a neighborhood $\mathcal{N}(\hat{x}, 0)$ such that for all points $(x, s) \in \mathcal{N}(\hat{x}, 0)$, it holds

$$\begin{aligned} &f(\hat{x}) \leq f(x) + \rho \sum_{i\in\mathcal{I}} s_i, \\ &c_i(x) \leq s_i^p \text{ and } -s_i \leq 0,\ \forall\, i \in \mathcal{I}. \end{aligned} \tag{2.2.4}$$

By $x^k \to \hat{x}$ and $c(\hat{x}) \leq 0$, we have $c_i(x^k) \to 0$. Letting $s_i^k = \sqrt[p]{[c_i(x^k)]_+}$, we have $s_i^k \to 0$ as $k \to \infty$ for all $i \in \mathcal{I}$. Therefore, we see that $(x^k, s^k) \in \mathcal{N}(\hat{x}, 0)$ as $k \to \infty$. By (2.2.4), we have

$$f(\hat{x}) \leq f(x^k) + \rho \sum_{i\in\mathcal{I}} s_i^k. \tag{2.2.5}$$

Combining (2.2.3), (2.2.5) and $s_i^k = \sqrt[p]{[c_i(x_k)]_+}$, we achieve a contradiction. We have shown that $\hat{x}$ solves problem (2.2.1) locally.

Case 2. We assume $\hat{x} \notin \mathcal{F}$. In this case, we have $\hat{s} \neq 0$. Let $(\hat{x}, \hat{s})$ solve problem (2.2.2) locally. Assume to the contrary that $\hat{x}$ does not solve problem (2.2.1) locally. Thus there exists a sequence $\{x^k\} \to \hat{x}$ such that

$$f(x^k) + \rho \sum_{i\in\mathcal{I}} [c_i(x^k)]_+^{\frac{1}{p}} < f(\hat{x}) + \rho \sum_{i\in\mathcal{I}} [c_i(\hat{x})]_+^{\frac{1}{p}}. \tag{2.2.6}$$

Since $(\hat{x}, \hat{s})$ solves problem (2.2.2) locally, it follows that there exists a neighborhood

$\mathcal{N}(\hat{x},\hat{s})$ such that for all points $(x,s)\in\mathcal{N}(\hat{x},\hat{s})$, it holds

$$\begin{aligned} f(\hat{x})+\rho\sum_{i\in\mathcal{I}}\hat{s}_i &\le f(x)+\rho\sum_{i\in\mathcal{I}}s_i, \\ c_i(x)\le s_i^p &\text{ and } -s_i\le 0,\ \forall\, i\in\mathcal{I}. \end{aligned} \tag{2.2.7}$$

By the continuity of $c_i(x)$ and $x^k\to\hat{x}$, we have $c_i(x^k)\to c_i(\hat{x})$ as $k\to\infty$ for all $i\in\mathcal{I}$. Letting $s_i^k=[c_i(x^k)]_+^{\frac{1}{p}}$ for all $i\in\mathcal{I}$. If the set $S^0(\hat{x},\hat{s})$ is nonempty, then we have $c_i(x^k)\to c_i(\hat{x})\le 0$ and $s_i^k\to\hat{s}_i=0$ for all $i\in S^0(\hat{x},\hat{s})$ as $k\to\infty$. Since $\hat{s}\neq 0$, we see that the set $S^+(\hat{x},\hat{s})$ is nonempty. We have $c_i(x^k)\to c_i(\hat{x})>0$, and $s_i^k\to[c_i(\hat{x})]_+^{\frac{1}{p}}=\hat{s}_i$ as $k\to\infty$ for all $i\in S^+(\hat{x},\hat{s})$. Consequently, we obtain that $(x^k,s^k)\in\mathcal{N}(\hat{x},\hat{s})$ as $k\to\infty$. Substituting (x^k,s^k) into (2.2.7), we have

$$\begin{aligned} f(\hat{x})+\rho\sum_{i\in\mathcal{I}}\hat{s}_i &\le f(x^k)+\rho\sum_{i\in\mathcal{I}}s_i^k=f(x^k)+\rho\sum_{i\in\mathcal{I}}[c_i(x^k)]_+^{\frac{1}{p}} \\ &< f(\hat{x})+\rho\sum_{i\in\mathcal{I}}[c_i(\hat{x})]_+^{\frac{1}{p}}=f(\hat{x})+\rho\sum_{i\in\mathcal{I}}\hat{s}_i. \end{aligned}$$

We reach a contradiction. Therefore, we have shown that $\hat{x}$ solves problem (2.2.1) locally.

Conversely, assume $\hat{x}$ solves problem (2.2.1) locally. Taking $\hat{s}_i=[c_i(\hat{x})]_+^{\frac{1}{p}}$ for all $i\in\mathcal{I}$, we have that $(\hat{x},\hat{s})$ solves problem (2.2.2) locally.

Summarizing the above two cases, we proved this proposition. □

2.2.1 Exact Penalization

Next we consider the $\ell_{\frac{1}{p}}$-penalty problem for the p-order relaxed problem (2.2.2) as follows

$$\min_{x,s}\Phi(x,s,\rho;\pi):=f(x)+\rho\sum_{i\in\mathcal{I}}s_i+\pi\Big(\sum_{i\in\mathcal{I}}[c_i(x)-s_i^p]_+^{\frac{1}{p}}+\sum_{i\in\mathcal{I}}[-s_i]_+^{\frac{1}{p}}\Big), \tag{2.2.8}$$

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

Lemma 2.2.1. *For any $a,\ b\in\mathbb{R}$ satisfying $a\ge b\ge 0$ and $\tau\in\mathbb{R}$ with $0\le\tau\le 1$, we* *have*

$$(a-b)^\tau \geq a^\tau - b^\tau. \tag{2.2.9}$$

Proof. It is trivial to prove this lemma if $b = 0$, or $\tau = 0$ and $\tau = 1$. In the following, we prove other cases. Let the function $g : [0,1] \to \mathbb{R}$ be defined by $g(x) := (1-x)^\tau + x^\tau - 1$. Then the function g is monotonically increasing on $[0, \frac{1}{2}]$ and monotonically decreasing on $[\frac{1}{2}, 1]$; moreover, $g(0) = g(1) = 0$. Therefore, we conclude that $g(x) \geq 0$ for all $x \in [0,1]$. Taking $x = \frac{b}{a}$ with $a \geq b > 0$, we have $g(\frac{b}{a}) = (1 - \frac{b}{a})^\tau + (\frac{b}{a})^\tau - 1 \geq 0$, which implies that the inequality (2.2.9) holds. The proof is complete. $\square$

Lemma 2.2.2. *For any $a,\ b \in \mathbb{R}$ and $p \geq 1$, we have*

$$\sqrt[p]{[a-b]_+} \geq \sqrt[p]{[a]_+} - \sqrt[p]{[b]_+}. \tag{2.2.10}$$

Proof. We consider the following cases:

(i) If $a \geq b \geq 0$, by Lemma 2.2.1, we have $\sqrt[p]{a-b} \geq \sqrt[p]{a} - \sqrt[p]{b}$, i.e., (2.2.10) holds;
(ii) If $a \geq 0 \geq b$, we have $a - b \geq a \geq 0$, i.e., $\sqrt[p]{[a-b]_+} \geq \sqrt[p]{[a]_+}$, i.e., (2.2.10) holds;
(iii) If $0 \geq a \geq b$, we have $\sqrt[p]{[a-b]_+} \geq 0$, i.e., (2.2.10) holds;
(iv) If $a < b \leq 0$, $a \leq 0 \leq b$ or $0 \leq a < b$, it is trivial that (2.2.10) holds.

By (i) – (iv), we have shown this lemma. $\square$

Using the above lemma, we prove that the $\ell_{\frac{1}{p}}$-penalty function in problem (2.2.8) is exact for any $\pi \geq 1$.

Proposition 2.2.2. *Let $\rho > 0$ be fixed. If $(\hat{x}, \hat{s})$ solves problem (2.2.2) locally, then there exists a real number $\hat{\pi} > 0$ such that for all $\pi \geq \hat{\pi}$, $(\hat{x}, \hat{s})$ solves problem (2.2.8) locally.*

Proof. Since $(\hat{x}, \hat{s})$ solves problem (2.2.2) locally, by Proposition 2.2.1, we see that $\hat{x}$ solves problem (2.2.1) locally and $\hat{s}_i = [c_i(\hat{x})]_+^{\frac{1}{p}}$ for all $i \in \mathcal{I}$. Thus, there is a neighborhood $\mathcal{N}(\hat{x})$ such that

$$f(x) + \rho \sum_{i \in \mathcal{I}} [c_i(x)]_+^{\frac{1}{p}} \geq f(\hat{x}) + \rho \sum_{i \in \mathcal{I}} [c_i(\hat{x})]_+^{\frac{1}{p}},\ \forall\, x \in \mathcal{N}(\hat{x}).$$

Let $\hat{\pi} = 1$. By Lemma 2.2.2, we have for $\pi \geq \hat{\pi}$,

$$\begin{aligned}
& f(x) + \rho \sum_{i \in \mathcal{I}} s_i + \pi \Big(\sum_{i \in \mathcal{I}} [c_i(x) - s_i^p]_+^{\frac{1}{p}} + \sum_{i \in \mathcal{I}} [-s_i]_+^{\frac{1}{p}} \Big) \\
\geq \;& f(x) + \rho \sum_{i \in \mathcal{I}} s_i + \Big(\sum_{i \in \mathcal{I}} [c_i(x) - s_i^p]_+^{\frac{1}{p}} + \sum_{i \in \mathcal{I}} [-s_i]_+^{\frac{1}{p}} \Big) \\
\geq \;& f(x) + \rho \sum_{i \in \mathcal{I}} s_i + \Big(\sum_{i \in \mathcal{I}} [c_i(x)]_+^{\frac{1}{p}} - \sum_{i \in \mathcal{I}} [s_i^p]_+^{\frac{1}{p}} + \sum_{i \in \mathcal{I}} [-s_i]_+^{\frac{1}{p}} \Big) \\
\geq \;& f(x) + \rho \sum_{i \in \mathcal{I}} [c_i(x)]_+^{\frac{1}{p}} + \Big(\sum_{i \in \mathcal{I}} s_i - \sum_{i \in \mathcal{I}} |s_i| + \sum_{i \in \mathcal{I}} [-s_i]_+^{\frac{1}{p}} \Big) \\
= \;& f(\hat{x}) + \rho \sum_{i \in \mathcal{I}} [c_i(\hat{x})]_+^{\frac{1}{p}} + \sum_{i \in \mathcal{I}} \Big(s_i - |s_i| + [-s_i]_+^{\frac{1}{p}} \Big) \\
\geq \;& f(\hat{x}) + \rho \sum_{i \in \mathcal{I}} \hat{s}_i,
\end{aligned}$$

for all $x \in \mathcal{N}(\hat{x})$ and $s_i \geq -2^{-\frac{p}{p-1}}$ for all $i \in \mathcal{I}$. The last inequality is derived from

$$s_i - |s_i| + [-s_i]_+^{\frac{1}{p}} \geq 0, \text{ for all } s_i \geq -2^{-\frac{p}{p-1}} \text{ and } i \in \mathcal{I}.$$

The proof is complete. □

2.2.2 First-Order Necessary Conditions

Throughout this subsection, we assume that $\rho > 0$ is fixed and that $(\hat{x}, \hat{s})$ is a local solution of the p-order relaxed problem (2.2.2). It is well-known that under some suitable regularity condition (also known as constraint qualification), the first-order necessary conditions hold at $(\hat{x}, \hat{s})$ for the p-order relaxed problem (2.2.2), i.e., there exist vectors $y, u \in \mathbb{R}^m$ such that

$$\nabla f(\hat{x}) + A(\hat{x}) y = 0, \tag{2.2.11a}$$
$$\rho e - pY \hat{s}^{p-1} - u = 0, \tag{2.2.11b}$$
$$Y \big(c(\hat{x}) - \hat{s}^p \big) = 0, \tag{2.2.11c}$$
$$U \hat{s} = 0, \tag{2.2.11d}$$
$$\hat{s}^p - c(\hat{x}) \geq 0, \tag{2.2.11e}$$
$$\hat{s}, y, u \geq 0, \tag{2.2.11f}$$

where the vectors $y, u \in \mathbb{R}^m_+$ are called Lagrange multipliers, $Y = \text{diag}(y)$ and $U = \text{diag}(u)$ are diagonal matrices. Since $(\hat{x}, \hat{s})$ is assumed to be a local solution of the problem (2.2.2), we have $\hat{s}_i = \sqrt{\max\{c_i(\hat{x}), 0\}}$ for all $i \in \mathcal{I}$, and thus there is no $i \in \mathcal{I}$ such that $c_i(\hat{x}) < \hat{s}_i^2$ and $\hat{s}_i > 0$, implying that $S^=(\hat{x}, \hat{s}) = S^+(\hat{x}, \hat{s})$ and $\mathcal{I} = S^=(\hat{x}, \hat{s}) \cup S^0(\hat{x}, \hat{s})$. By using the index sets above, we can reformulate (2.2.11) as

$$
\begin{aligned}
&\nabla f(\hat{x}) + \sum_{i \in \mathcal{I}} y_i \nabla c_i(\hat{x}) = 0, \\
&y_i = \frac{\rho}{p\hat{s}_i^{p-1}}, \ \forall\, i \in S^=(\hat{x}, \hat{s}), \quad y_i \geq 0, \ \forall\, i \in CS^0(\hat{x}, \hat{s}), \\
&y_i = 0, \ \forall\, i \in S^0(\hat{x}, \hat{s}) \backslash CS^0(\hat{x}, \hat{s}), \\
&u_i = 0, \ \forall\, i \in S^=(\hat{x}, \hat{s}), \quad u_i = \rho, \ \forall\, i \in S^0(\hat{x}, \hat{s}), \\
&\hat{s}^p - c(\hat{x}) \geq 0, \quad \hat{s} \geq 0.
\end{aligned}
\tag{2.2.12}
$$

If $\hat{x}$ is feasible to problem (2.1.1), we have $\hat{s} = 0$ and $S^=(\hat{x}, \hat{s}) = \emptyset$, and moreover, the first-order necessary conditions (2.2.11) or (2.2.12) recover the first-order necessary conditions at $\hat{x}$ for problem (2.1.1).

If $\hat{s} \in \mathbb{R}^m_{++} := \{x \mid x_i > 0, \ \forall\, i \in \mathcal{I}\}$, the p-order relaxed problem (2.2.2) only has the inequalities $c_i(x) - s_i^p \leq 0$ with $i \in \mathcal{I}$ being active at $(\hat{x}, \hat{s})$, and the Jacobian matrix $(A(\hat{x})^T, -p\text{diag}(\hat{s}^{p-1}))$ of $c(x) - s^p$ at $(\hat{x}, \hat{s})$ has full rank, implying that the LICQ holds at $(\hat{x}, \hat{s})$. In this case, the first-order necessary conditions (2.2.11) hold automatically.

In the remainder of this subsection, we assume that $\hat{s} \in \mathbb{R}^m_+ \backslash \mathbb{R}^m_{++}$ and shall give some CQs for p-order relaxed problem (2.2.2) to possess the first-order necessary conditions (2.2.11). To begin with, we show in the following lemma that the LICQ (resp. the MFCQ) holds at $(\hat{x}, \hat{s})$ for the p-order relaxed problem (2.2.2) if and only if the LICQ (resp. the MFCQ) holds at $\hat{x}$ for the inequality system

$$
c_i(x) \leq 0, \quad \forall\, i \in CS^0(\hat{x}, \hat{s}). \tag{2.2.13}
$$

Lemma 2.2.3. *Assume that $\hat{s} \in \mathbb{R}^m_+ \backslash \mathbb{R}^m_{++}$. Consider the following CQs.*

(a) *The* LICQ *holds at $\hat{x}$ for the inequality system (2.2.13), i.e., the vectors $\nabla c_i(\hat{x})$ with $i \in CS^0(\hat{x}, \hat{s})$ are linearly independent.*

(b) *The* MFCQ *holds at* $\hat{x}$ *for the inequality system (2.2.13), i.e., there exists some* $d \in \mathbb{R}^n$ *such that*

$$\nabla c_i(\hat{x})^T d < 0, \quad \forall\, i \in CS^0(\hat{x}, \hat{s}),$$

or in other words,

$$\sum_{i \in CS^0(\hat{x},\hat{s})} y_i \nabla c_i(\hat{x}) = 0, \quad y_i \geq 0,\ \forall\, i \in CS^0(\hat{x}, \hat{s}) \Longrightarrow y_i = 0,\ \forall\, i \in CS^0(\hat{x}, \hat{s}). \tag{2.2.14}$$

Then (a) *holds if and only if the* LICQ *holds at* $(\hat{x}, \hat{s})$ *for the p-order relaxed problem (2.2.2), while* (b) *holds if and only if the* MFCQ *holds at* $(\hat{x}, \hat{s})$ *for the p-order relaxed problem (2.2.2).*

Proof. By definition, the MFCQ holds at $(\hat{x}, \hat{s})$ for the p-order relaxed problem (2.2.2) if,

$$\left.\begin{array}{l} \displaystyle\sum_{i \in S^=(\hat{x},\hat{s}) \cup CS^0(\hat{x},\hat{s})} y_i \nabla c_i(\hat{x}) = 0 \\ -p\hat{s}_i^{p-1} y_i = 0,\ \forall\, i \in S^=(\hat{x}, \hat{s}), \\ u_i = 0,\ \forall\, i \in S^0(\hat{x}, \hat{s}), \\ y_i \geq 0,\ \forall\, i \in S^=(\hat{x}, \hat{s}) \cup CS^0(\hat{x}, \hat{s}), \\ u_i \geq 0,\ \forall\, i \in S^0(\hat{x}, \hat{s}) \end{array}\right\} \Longrightarrow \left\{\begin{array}{l} y_i = 0,\ \forall\, i \in S^=(\hat{x}, \hat{s}) \cup CS^0(\hat{x}, \hat{s}), \\ \\ u_i = 0,\ \forall\, i \in S^0(\hat{x}, \hat{s}). \end{array}\right. \tag{2.2.15}$$

Observing that $\hat{s}_i > 0$ for all $i \in S^=(\hat{x}, \hat{s})$, the equivalence of (2.2.14) and (2.2.15) follows immediately. The case for the LICQ can be proved in a similar way. □

Remark 2.2.1. *It is well-known in the field of the nonlinear programming that the MFCQ amounts to the boundedness of Lagrange multipliers. Thus, in the case of* $\hat{s} \in \mathbb{R}^m_+ \backslash \mathbb{R}^m_{++}$*, the p-order relaxed problem (2.2.2) has bounded Lagrange multipliers* (y, u) *as defined by (2.2.11) if and only if Lemma 2.2.3* (b) *is fulfilled. If the MFCQ holds at a feasible point* $x_0 \in \mathcal{F}$ *for problem (2.1.1), then for all* $(\hat{x}, \hat{s})$ *with* $\hat{x}$ *near* x_0*, the quadratically relaxed problem (2.2.2) has bounded Lagrange multipliers at* $(\hat{x}, \hat{s})$ *provided that it is a local solution of problem (2.2.2).*

Besides having the CQs in Lemma 2.2.3 for the first-order necessary conditions (2.2.11), we can use the techniques in [116, 117, 181] to derive some other CQs, some of which turn out to be strictly weaker than the ones in Lemma 2.2.3. Three cases, $p = 2$, $1 \leq p < 2$ and $p > 2$, are considered, respectively. Because the case $p = 2$ is typical, we shall consider this case first. In this case, the p-order relaxation reduces to the quadratical relaxation.

Case $p = 2$.

We conduct the analysis in the next lemma below by showing that the linearized tangent cone

$$L_{\widehat{\mathcal{F}}}(\hat{x}, \hat{s}) := \left\{ (w, \beta) \in \mathbb{R}^n \times \mathbb{R}^m \;\middle|\; \begin{array}{r} \langle \nabla c_i(\hat{x}), w \rangle - 2\hat{s}_i \beta_i \leq 0, \quad \forall\, i \in S^{=}(\hat{x}, \hat{s}) \\ \langle \nabla c_i(\hat{x}), w \rangle \leq 0, \quad \forall\, i \in CS^0(\hat{x}, \hat{s}) \\ -\beta_i \leq 0, \quad \forall\, i \in S^0(\hat{x}, \hat{s}) \end{array} \right\} \tag{2.2.16}$$

to the feasible set $\widehat{\mathcal{F}}$ of problem (2.2.2) with $p = 2$ at $(\hat{x}, \hat{s})$ coincides with the kernel of the subderivative (or Dini upper directional derivative) of the penalty term

$$\phi(x, s) := \sum_{i \in \mathcal{I}} \sqrt{\max\{c_i(x) - s_i^2, 0\}} + \sum_{i \in \mathcal{I}} \sqrt{\max\{-s_i, 0\}}. \tag{2.2.17}$$

Next we give characterizations in terms of the gradients and the Hessians of the functions c_i with $i \in \mathcal{I}$ for two equalities

$$L_{\widehat{\mathcal{F}}}(\hat{x}, \hat{s}) = \{(w, \beta) \in \mathbb{R}^{n+m} \mid D_+\phi(\hat{x}, \hat{s})(w, \beta) = 0\} \tag{2.2.18}$$

and

$$L_{\widehat{\mathcal{F}}}(\hat{x}, \hat{s}) = \{(w, \beta) \in \mathbb{R}^{n+m} \mid d\phi(\hat{x}, \hat{s})(w, \beta) = 0\}. \tag{2.2.19}$$

Lemma 2.2.4. *Assume that* $\hat{s} \in \mathbb{R}^m_+ \backslash \mathbb{R}^m_{++}$. *Let*

$$\Omega := \{w \in \mathbb{R}^n \mid \langle \nabla c_i(\hat{x}), w \rangle \leq 0,\ \forall\, i \in CS^0(\hat{x}, \hat{s})\}.$$

Consider the following CQs:

(a) *The equality (2.2.18) holds.*

(b) *For each* $w \in \Omega$ *and* $i \in S^{=}(\hat{x}, \hat{s})$, *it follows that*

$$2\hat{s}_i^2 \langle w, \nabla^2 c_i(\hat{x}) w \rangle \leq \langle \nabla c_i(\hat{x}), w \rangle^2,$$

and for each $w \in \Omega$ *and* $i \in CS^0(\hat{x}, \hat{s})$ *with* $\langle \nabla c_i(\hat{x}), w \rangle = 0$, *it follows that*

$$\langle w, \nabla^2 c_i(\hat{x}) w \rangle \le 0.$$

(c) *For each* $w \in \Omega$ *and* $i \in CS^0(\hat{x}, \hat{s})$ *with* $\langle \nabla c_i(\hat{x}), w \rangle = 0$, *it follows that*

$$\langle w, \nabla^2 c_i(\hat{x}) w \rangle \le 0.$$

(d) *For each* $w \in \Omega$ *and* $i \in CS^0(\hat{x}, \hat{s})$ *with* $\langle \nabla c_i(\hat{x}), w \rangle = 0$, *there exists some* $z \in \mathbb{R}^n$ *such that*

$$\langle \nabla c_i(\hat{x}), z \rangle + \langle w, \nabla^2 c_i(\hat{x}) w \rangle \le 0.$$

(e) *For each* $w \in \Omega$, *it follows that*

$$\max \left\{ \sum_{i \in CS^0(\hat{x},\hat{s})} \lambda_i \langle w, \nabla^2 c_i(\hat{x}) w \rangle \,\middle|\, \sum_{i \in CS^0(\hat{x},\hat{s})} \lambda_i \nabla c_i(\hat{x}) = 0,\ \lambda_i \ge 0,\ \forall\, i \in CS^0(\hat{x}, \hat{s}) \right\} = 0.$$

(f) *The equality (2.2.19) holds.*

Then we have

$$\text{(a)} \Longleftrightarrow \text{(b)} \Longrightarrow \text{(c)} \Longrightarrow \text{(d)} \Longleftrightarrow \text{(e)} \Longleftrightarrow \text{(f)}.$$

Proof. The implications (b) $\Longrightarrow$ (c) $\Longrightarrow$ (d) hold trivially. By a nonhomogeneous Farkas' Lemma [159, Lemma 4.2], it is straightforward to verify that (d) $\Longleftrightarrow$ (e). To show (e) $\Longleftrightarrow$ (f), we introduce another square root penalty term for the quadratically relaxed problem (2.2.2) as follows:

$$\tilde{\phi}(x, s) := \sqrt{\sum_{i \in \mathcal{I}} \max\{c_i(x) - s_i^2, 0\} + \sum_{i \in \mathcal{I}} \max\{-s_i, 0\}}.$$

According to [89, Lemma 4.1], we have $\tilde{\phi} \le \phi \le 2m\tilde{\phi}$ and hence

$$\{(w, \beta) \mid d\tilde{\phi}(\hat{x}, \hat{s})(w, \beta) = 0\} = \{(w, \beta) \mid d\phi(\hat{x}, \hat{s})(w, \beta) = 0\}. \tag{2.2.20}$$

Applying [116, Proposition 2.1], we have the equality

$$L_{\widehat{\mathcal{F}}}(\hat{x}, \hat{s}) = \{(w, \beta) \mid d\tilde{\phi}(\hat{x}, \hat{s})(w, \beta) = 0\} \tag{2.2.21}$$

if and only if for all $(w, \beta) \in L_{\widehat{\mathcal{F}}}(\hat{x}, \hat{s})$,

$$\max \left\{ \sum_{i \in CS^0(\hat{x},\hat{s})} \lambda_i [\langle w, \nabla^2 c_i(\hat{x}) w\rangle - 2\beta_i^2] \,\middle|\, \sum_{i \in CS^0(\hat{x},\hat{s})} \lambda_i \nabla c_i(\hat{x}) = 0,\ \lambda_i \geq 0,\ \forall\, i \in CS^0(\hat{x}, \hat{s}) \right\} = 0.$$

The latter condition holds if and only if for all $w \in \Omega$ and $\beta \in \mathbb{R}^m$ with $\beta_i \geq 0$ for all $i \in CS^0(\hat{x}, \hat{s})$,

$$\max \left\{ \sum_{i \in CS^0(\hat{x},\hat{s})} \lambda_i [\langle w, \nabla^2 c_i(\hat{x}) w\rangle - 2\beta_i^2] \,\middle|\, \sum_{i \in CS^0(\hat{x},\hat{s})} \lambda_i \nabla c_i(\hat{x}) = 0,\ \lambda_i \geq 0,\ \forall i \in CS^0(\hat{x}, \hat{s}) \right\} = 0,$$

because $(w, \beta) \in L_{\widehat{\mathcal{F}}}(\hat{x}, \hat{s})$ amounts to that $w \in \Omega$, $\beta_i \geq \langle \frac{\nabla c_i(\hat{x})}{2\hat{s}_i}, w\rangle$ for all $i \in S^{=}(\hat{x}, \hat{s})$ and $\beta_i \geq 0$ for all $i \in CS^0(\hat{x}, \hat{s})$. Since $\lambda_i[\langle w, \nabla^2 c_i(\hat{x}) w\rangle - 2\beta_i^2] \leq \lambda_i \langle w, \nabla^2 c_i(\hat{x}) w\rangle$ whenever $\lambda_i \geq 0$, the equality (2.2.21) holds if and only if (e) holds. In view of (2.2.20), we have (e) $\Longleftrightarrow$ (f).

By [181, Lemma 2.3] or [117, Remark 2.2], (a) holds if and only if, for each $i \in S^{=}(\hat{x}, \hat{s})$ and $(w, \beta) \in L_{\widehat{\mathcal{F}}}(\hat{x}, \hat{s})$ with $\langle \nabla c_i(\hat{x}), w\rangle - 2\hat{s}_i \beta_i = 0$, it follows that

$$\langle w, \nabla^2 c_i(\hat{x}) w\rangle - 2\beta_i^2 \leq 0 \quad \text{or} \quad 2\hat{s}_i^2 \langle w, \nabla^2 c_i(\hat{x}) w\rangle \leq \langle \nabla c_i(\hat{x}), w\rangle^2,$$

and for each $i \in CS^0(\hat{x}, \hat{s})$ and $(w, \beta) \in L_{\widehat{\mathcal{F}}}(\hat{x}, \hat{s})$ with $\langle \nabla c_i(\hat{x}), w\rangle = 0$ (or in other words, for each $i \in CS^0(\hat{x}, \hat{s})$ and $w \in \Omega$ with $\langle \nabla c_i(\hat{x}), w\rangle = 0$), it follows that

$$\langle w, \nabla^2 c_i(\hat{x}) w\rangle \leq 0.$$

That is, we have (a) $\Longleftrightarrow$ (b). This completes the proof. □

Remark 2.2.2. *It is clear to see that the CQ given by Lemma 2.2.4* (e) *is implied by the CQ given by Lemma 2.2.3* (b)*. But the converse may not hold as can be seen from [116, Example 2.3] in the case of* $\hat{s} = 0$.

In view of Lemma 2.2.2 and [116, Theorem 2.1], we now confirm that the first-order necessary conditions (2.2.11) hold at $(\hat{x}, \hat{s})$ for the quadratically relaxed problem (2.2.2) provided that one of the CQs in Lemmas 2.2.3 and 2.2.4 is fulfilled. To be precise, we now summarize what we have discussed so far on the first-order necessary conditions for the quadratically relaxed problem (2.2.2) in the following theorem.

Theorem 2.2.1. *Let $\rho > 0$ and $p = 2$. Assume that $(\hat{x}, \hat{s})$ is a local solution of the problem (2.2.2). Then the first-order necessary conditions (2.2.11) hold at $(\hat{x}, \hat{s})$ if either $\hat{s} \in \mathbb{R}^m_{++}$ or $\hat{s} \in \mathbb{R}^m_{+} \backslash \mathbb{R}^m_{++}$ with one of the CQs in Lemmas 2.2.3 and 2.2.4 being fulfilled.*

Case $1 \leq p < 2$.

Similar to [181, Theorem 2.2], we have the next theorem.

Theorem 2.2.2. *Let $\rho > 0$ and $1 \leq p < 2$. Assume that $(\hat{x}, \hat{s})$ is a local solution of the problem (2.2.2) and c_i $(i \in \mathcal{I})$ are twice continuously differentiable. Then the first-order necessary conditions (2.2.11) hold at $(\hat{x}, \hat{s})$.*

Proof. The conclusion follows from Proposition 2.2.2, [181, Lemmas 2.2 and 2.4] and a homogeneous Farkas' Lemma [159, Lemma 4.1]. □

Case $p > 2$.

Using Proposition 2.2.2, [181, Lemmas 2.2 and 2.5] and a homogeneous Farkas' Lemma [159, Lemma 4.1], we derive the next theorem.

Theorem 2.2.3. *Let $\rho > 0$ and $p > 2$. Assume that $(\hat{x}, \hat{s})$ is a local solution of the problem (2.2.2) and c_i $(i \in \mathcal{I})$ are twice continuously differentiable. In addition, assume that for each $w \in \Omega$ and $i \in S^{=}(\hat{x}, \hat{s})$, it follows that*

$$p\hat{s}_i^p \langle w, \nabla^2 c_i(\hat{x}) w \rangle < (p-1) \langle \nabla c_i(\hat{x}), w \rangle^2,$$

and for each $w \in \Omega$ and $i \in CS^0(\hat{x}, \hat{s})$ with $\langle \nabla c_i(\hat{x}), w \rangle = 0$, it follows that

$$\langle w, \nabla^2 c_i(\hat{x}) w \rangle < 0.$$

Then the first-order necessary conditions (2.2.11) hold at $(\hat{x}, \hat{s})$.

2.3 Interior-Point $\ell_{\frac{1}{p}}$-Penalty Method

In this section, we introduce an interior-point $\ell_{\frac{1}{p}}$-penalty method. Then we establish the global convergence results of the proposed method under mild conditions.

2.3.1 A Basic Interior-Point Method

A primal-dual interior-point method is used to solve problem (2.2.2). Specifically, we minimize a sequence of logarithmic barrier functions

$$\begin{aligned} &\min_{x,s} \phi_{B,\frac{1}{2}}(x,s;\rho,\mu) := \phi_{S,\frac{1}{p}}(x,s;\rho) - \mu^p \sum_{i\in\mathcal{I}} \log\big(s_i^p - c_i(x)\big) - \mu \sum_{i\in\mathcal{I}} \log s_i \\ &\text{s.t. } s_i^p - c_i(x) > 0 \text{ and } s_i > 0,\ i \in \mathcal{I}, \end{aligned} \tag{2.3.1}$$

where $\mu > 0$ is the barrier parameter. Let (x, s) be a local solution of problem (2.3.1). Then the first-order necessary conditions of problem (2.3.1) are

$$\nabla f(x) + A(x)y = 0, \tag{2.3.2a}$$

$$\rho e - pYs^{p-1} - u = 0, \tag{2.3.2b}$$

$$Y\big(s^p - c(x)\big) - \mu^p e = 0, \tag{2.3.2c}$$

$$Us - \mu e = 0, \tag{2.3.2d}$$

where vectors y,$u \in \mathbb{R}^m_{++}$ are the Lagrange multipliers, $Y = \text{diag}(y)$ and $U = \text{diag}(u)$ are the diagonal matrices.

Remark 2.3.1. *Here we note that it is reasonable to choose μ^p as the barrier parameter for the term $\sum_{i\in\mathcal{I}} \log\big(s_i^p - c_i(x)\big)$ in (2.3.1a). Indeed, suppose that the Lagrange multiplier y is bounded. We obtain from (2.3.2b) that the Lagrange multiplier $u \to \rho e$ as $s \to 0^+$. From (2.3.2d), we have $\mu = O(\|s\|)$, which can be guaranteed by setting the barrier parameter μ^p for the term $\sum_{i\in\mathcal{I}} \log\big(s_i^p - c_i(x)\big)$ in (2.3.1).*

Applying a modified Newton's method (see [6]) to the nonlinear system (2.3.2) in variables x, s, y and u, we obtain

$$\Omega(x,y,s,u,H)\begin{pmatrix}\triangle x\\ \triangle s\\ \triangle y\\ \triangle u\end{pmatrix} = -\begin{pmatrix}\nabla f(x)+A(x)y\\ \rho e - pYs^{p-1}-u\\ Y\big(s^p - c(x)\big)-\mu^p e\\ Us-\mu e\end{pmatrix} \tag{2.3.3}$$

where

$$\Omega(x,y,s,u,H) := \begin{pmatrix} H(x,y) & 0 & A(x) & 0\\ 0 & -p(p-1)YS^{p-2} & -pS^{p-1} & -E\\ -YA(x)^T & pYS^{p-1} & S^p - C(x) & 0\\ 0 & U & 0 & S\end{pmatrix},$$

and

$$H(x,y) := \nabla^2 f(x) + \sum_{i\in\mathcal{I}} y_i \nabla^2 c_i(x). \tag{2.3.4}$$

Noting that $Ys = Sy$ and $Us = Su$, we rewrite (2.3.3) as follows

$$H(x,y)\triangle x + A(x)(y+\triangle y) = -\nabla f(x), \tag{2.3.5a}$$

$$pS^{p-1}(y+\triangle y) + E(u+\triangle u) + p(p-1)S^{p-2}Y\triangle s = \rho e, \tag{2.3.5b}$$

$$(S^p - C(x))(y+\triangle y) + pYS^{p-1}\triangle s - YA(x)^T\triangle x = \mu^p e, \tag{2.3.5c}$$

$$U\triangle s + S(u+\triangle u) = \mu e. \tag{2.3.5d}$$

Solving $\hat{y} := y + \triangle y$ and $\hat{u} := u + \triangle u$ from (2.3.5c) and (2.3.5d), we get

$$\hat{y} = \big(S^p - C(x)\big)^{-1}\big(\mu^p e - pYS^{p-1}\triangle s + YA(x)^T\triangle x\big), \tag{2.3.6a}$$

$$\hat{u} = S^{-1}\big(\mu e - U\triangle s\big). \tag{2.3.6b}$$

Substituting (2.3.6a) and (2.3.6b) into (2.3.5a) and (2.3.5b), we obtain

$$\mathcal{M}\begin{pmatrix}\triangle x\\ \triangle s\end{pmatrix} = \begin{pmatrix} -\rho\nabla f(x) - \mu^p A(x)\big(S^p - C(x)\big)^{-1}e\\ p\mu^p S^{p-1}\big(S^p - C(x)\big)^{-1}e + \mu S^{-1}e - \rho e\end{pmatrix} \tag{2.3.7}$$

where

$$\mathcal{M} := \begin{pmatrix} \widehat{H}(x,s,y) & -pA(x)\mathcal{N}S^{p-1} \\ -p\mathcal{N}S^{p-1}A(x)^T & \Xi \end{pmatrix} \tag{2.3.8}$$

with $\mathcal{N} := \big(S^p - C(x)\big)^{-1}Y$, $\widehat{H}(x,s,y) := H(x,y) + A(x)\mathcal{N}A(x)^T$ and $\Xi := p^2S^{p-1}\mathcal{N}S^{p-1} + S^{-1}U - p(p-1)S^{p-2}Y$.

In order to establish the global convergence of the interior-point method, we need to ensure that the matrix $\mathcal{M}$ is sufficiently positive definite [47, 48]. Assume that

$$u - p(p-1)Ys^{p-1} \geq 0. \tag{2.3.9}$$

Since $\mathcal{N} \succ 0$ and $S \succ 0$, it follows from the above assumption, we obtain $\Xi \succ 0$. To guarantee $\mathcal{M} \succ 0$, by the Schur complement, we need to ensure

$$\widehat{H}(x,s,y) - \Big(p\mathcal{N}S^{p-1}A(x)^T\Big)\Big(\Xi\Big)^{-1}\Big(pA(x)\mathcal{N}S^{p-1}\Big) \succ 0.$$

Substituting $\widehat{H}(x,s,y)$ into the above inequality, we attain

$$H(x,y) + A(x)\Big\{\mathcal{N} - p\mathcal{N}S^{p-1}\Big(\Xi\Big)^{-1}pS^{p-1}\mathcal{N}\Big\}A(x)^T \succ 0. \tag{2.3.10}$$

However, inequality (2.3.10) may not always hold in general. We can modify $H(x,y)$ by adding a term of the form δE where δ is chosen to be large enough to ensure that it holds, that is, we can replace $H(x,y)$ by $H(x,y)+\delta E$ with a suitable δ so that (2.3.10) holds [6, 157, 165].

Remark 2.3.2. *In order to use the Schur complement to matrix $\mathcal{M}$, we force (2.3.9) to hold in every iteration (see (2.3.14) and (2.3.15)). Here, we note that this assumption is reasonable. Indeed, as $s \to 0^+$, assume that multiplier y is bounded above, it follows from (2.3.2b) that $u \to \rho e$ and (2.3.9) holds automatically.*

At the k-th iteration (x^k, s^k), we can get $(\triangle x^k, \triangle s^k)$ by solving (2.3.7). Then we let

$$x^{k+1} := x^k + \alpha_P^k \triangle x^k, \tag{2.3.11a}$$

$$s^{k+1} := s^k + \alpha_P^k \triangle s^k, \tag{2.3.11b}$$

where $\alpha_P^k := \max\{\ \bar{\rho}^j \mid j = 0, 1, 2, \dots\}$ with $\bar{\rho} \in (0, 1)$ is a step length, which satisfies the following conditions:

$$(s^{k+1})^p - c(x^{k+1}) > 0, \tag{2.3.12a}$$

$$s^{k+1} > 0, \tag{2.3.12b}$$

$$\begin{aligned}\phi_{B,\frac{1}{p}}(x^{k+1}, s^{k+1}; \rho, \mu) - \phi_{B,\frac{1}{p}}(x^k, s^k; \rho, \mu) \le \tau_1 \alpha_P^k \Big(&\nabla_x \phi_{B,\frac{1}{p}}(x^k, s^k; \rho, \mu)^T \triangle x^k + \\ &\nabla_s \phi_{B,\frac{1}{p}}(x^k, s^k; \rho, \mu)^T \triangle s^k \Big),\end{aligned} \tag{2.3.12c}$$

for some $\tau_1 \in (0, \frac{1}{2})$, where the last inequality is a standard Armijo line search condition in [175] on the decrease of the barrier objective function in problem (2.3.1).

Remark 2.3.3. *In practice, the parameter τ_1 is chosen to be quite small. In this chapter, following [42], $\tau_1 = 10^{-8}$ is set, see Table 2.1 in Section 2.4.*

2.3.2 Updating the Lagrange Multipliers

Two steps are used to update the Lagrange multipliers (y^k, u^k) at the k-th iteration. We first use the strategy introduced in [5, 30, 36] to update them as follows, $\forall\ i \in \mathcal{I}$,

$$\tilde{y}_i^{k+1} := \begin{cases} \min\{\gamma_{min} y_i^k, \frac{\mu^p}{(s_i^k)^p - c_i(x^k)}\}, & \text{if} \quad \hat{y}_i^{k+1} < \min\{\gamma_{min} y_i^k, \frac{\mu^p}{(s_i^k)^p - c_i(x^k)}\}, \\ \frac{\mu^p \gamma_{max}}{(s_i^k)^p - c_i(x^k)}, & \text{if} \quad \hat{y}_i^{k+1} > \frac{\mu^p \gamma_{max}}{(s_i^k)^p - c_i(x^k)}, \\ \hat{y}_i^{k+1}, & \text{otherwise}, \end{cases} \tag{2.3.13a}$$

$$\tilde{u}_i^{k+1} := \begin{cases} \min\{\gamma_{min} u_i^k, \frac{\mu}{s_i^k}\}, & \text{if} \quad \hat{u}_i^{k+1} < \min\{\gamma_{min} u_i^k, \frac{\mu}{s_i^k}\}, \\ \frac{\mu \gamma_{max}}{s_i^k}, & \text{if} \quad \hat{u}_i^{k+1} > \frac{\mu \gamma_{max}}{s_i^k}, \\ \hat{u}_i^{k+1}, & \text{otherwise}, \end{cases} \tag{2.3.13b}$$

where the parameters γ_{min} and γ_{max} satisfy $0 < \gamma_{min} < 1 < \gamma_{max}$.

The second step is to guarantee the new Lagrange multipliers (y^{k+1}, u^{k+1}) satisfying the assumption (2.3.9). Specifically, if $(\tilde{y}^{k+1}, \tilde{u}^{k+1})$ satisfies (2.3.9), we let $(y^{k+1}, u^{k+1}) := (\tilde{y}^{k+1}, \tilde{u}^{k+1})$ as the new Lagrange multiplier vector. otherwise, we set

$$y^{k+1} := \gamma_1 \tilde{y}^{k+1},\ u^{k+1} := \gamma_2 \tilde{u}^{k+1}, \tag{2.3.14}$$

where $\gamma_1 \in (0, 1]$ and $\gamma_2 \geq 1$ satisfy

$$\frac{\gamma_2}{\gamma_1} \geq \max_{i \in \mathcal{I}} \left\{ \frac{p(p-1)(s_i^{k+1})^{p-1} \tilde{y}_i^{k+1}}{\tilde{u}_i^{k+1}} \right\}. \tag{2.3.15}$$

Remark 2.3.4. *Here we note that, to guarantee the dual multipliers (y^k, u^k) being bounded, the sequences $\{(\hat{y}^k, \hat{u}^k)\}$ is truncated in (2.3.13) through choosing a proper γ_{max}. In practice, γ_{max} should be very large, for example, $\gamma_{max} = 10^{20}$ was used in [36]. In this chapter, $\gamma_{max} = 10^{23}$ is chosen; see Table 2.1 in Section 2.4.*

Rather than solving the barrier subproblem (2.3.1) accurately, our iteration continues until the conditions (2.3.2) are satisfied within a tolerance ϵ_μ for the current barrier parameter μ, that is

$$\mathrm{Res}(x, s, \hat{y}, \hat{u}; \rho, \mu) := \left\| \begin{array}{c} \nabla f(x) + A(x)\hat{y} \\ \rho e - p\widehat{Y} s^{p-1} - \hat{u} \\ \widehat{Y}(s^p - c(x)) - \mu^p e \\ \widehat{U} s - \mu e \end{array} \right\| < \epsilon_\mu, \tag{2.3.16a}$$

$$(\hat{y}, \hat{u}) \succeq -\epsilon_\mu(e, e), \tag{2.3.16b}$$

where $\epsilon_\mu > 0$ is a μ-related tolerance parameter, which satisfies $\epsilon_\mu \downarrow 0$ as $\mu \to 0$.

2.3.3 Specific Algorithms

In this subsection, we describe three specific algorithms to solve problem (2.1.1) by virtue of the $\ell_{\frac{1}{p}}$-penalty function. More implementation details will be stated in Section 2.4. The first algorithm gives a description of the approximate solution of problem (2.3.1) with the fixed penalty parameter $\rho > 0$ and barrier parameter $\mu > 0$.

Algorithm 2.1:	Inner algorithm for solving problem (2.3.1).
Step 0	Initialization. Set τ_1, γ_{min} and $\gamma_1 \in (0,1)$, γ_{max} and $\gamma_2 > 1$. Let $k := 0$;
Step 1	If (2.3.16) holds at point $(x^k, s^k, \hat{y}^k, \hat{u}^k)$, stop;
Step 2	If (2.3.10) dose not hold then replace $H(x^k, y^k)$ by $H(x^k, y^k) + \delta E$ with a proper $\delta > 0$ such that inequality (2.3.10) holds;
Step 3	Computing $(\triangle x^k, \triangle s^k)$ from (2.3.7) and $(\hat{y}^{k+1}, \hat{u}^{k+1})$ from (2.3.6); we compute the primal step length α_P^k such that it satisfies (2.3.12) and compute $(x^{k+1},\ s^{k+1})$ from (2.3.11); based on (2.3.13)-(2.3.14), we update the dual multipliers to obtain (y^{k+1}, u^{k+1});
Step 4	Let $k := k + 1$, go to Step 1.

In order to solve the relaxed problem (2.2.2), we need to solve a series of barrier subproblems (2.3.1) for decreasing the values of μ with a fixed penalty parameter $\rho > 0$.

Algorithm 2.2:	Inner algorithm for solving problem (2.2.2).
Step 0	Initialization. Set $\mu^0 > 0$, $\epsilon_{\mu^0} > 0$ and $\gamma \in (0,1)$. Let $j := 0$;
Step 1	If $\mathrm{Res}(x^j, s^j, \hat{y}^j, \hat{u}^j; \rho, 0) \le \bar{\epsilon}$ and $(\hat{y}^j, \hat{u}^j) \ge 0$, stop;
Step 2	Starting from $(x^j, s^j, \hat{y}^j, \hat{u}^j)$, we apply Algorithm 2.1 to solve problem (2.3.1) with the barrier parameter μ^j and the stopping tolerance ϵ_{μ^j}. Let the solution be $(x^{j+1}, s^{j+1}, \hat{y}^{j+1}, \hat{u}^{j+1})$;
Step 3	Set $\mu^{j+1} := \gamma\mu^j$, $\epsilon_{\mu^{j+1}} := \gamma\epsilon_{\mu^j}$ and let $j := j + 1$, go to Step 1.

If $\|s^j\|$ is sufficiently small at point (x^j, s^j), we declare that point x^j as a KKT or FJ point of problem (2.1.1). Otherwise, we increase the penalty parameter ρ and solve the relaxed problem (2.2.2) again. A formal description of algorithm to solve problem (2.1.1) is given as follows.

Algorithm 2.3:	Outer algorithm for solving problem (2.1.1).
Step 0	Initialization. Set $p \geq 1$, $x^0 \in R^n$, $\rho^0 > 0$, $y^0 = \hat{y}^0 > 0$, $u^0 = \hat{u}^0 > 0$, $\nu > 1$, $\bar{\epsilon} > 0$ and $s_l^0 \geq \sqrt[p]{\max\{c_l(x^0), 0\}} + \frac{1}{2}$ for all $l \in \mathcal{I}$. Let $i := 0$;
Step 1	If $\|s^i\| \leq \bar{\epsilon}$, stop;
Step 2	Starting from point $(x^i, s^i, \hat{y}^i, \hat{u}^i)$, we apply Algorithm 2.2 to solve problem (2.2.2) with the penalty parameter ρ^i. Let the solution be $(x^{i+1}, s^{i+1}, \hat{y}^{i+1}, \hat{u}^{i+1})$;
Step 3	Set $\rho^{i+1} := \nu\rho^i$ and let $i := i + 1$, go to Step 1.

2.3.4 Convergence Analysis

In this subsection, we establish the global convergence of the interior-point $\ell_{\frac{1}{p}}$-penalty method. The following assumptions are needed.

Assumption 1: The feasible set $\mathcal{F}$ is nonempty.

Assumption 2: The functions $f(x)$ and $c_i(x)$, for all $i \in \mathcal{I}$ are twice continuously differentiable on $\mathbb{R}^n$.

Assumption 3: The primal iterate sequence $\{x^k\}$ lies in a bounded set.

Assumption 4: The Hessian matrix sequence $\{H^k\} := \{H(x^k, y^k; \rho)\}$ lies in a bounded set.

Let the strictly feasible set of problem (2.2.2) be defined by

$$\widehat{\mathcal{F}}^+ := \{(x, s) \in \mathbb{R}^{n+m} \mid c_i(x) < s_i^p,\ s_i > 0,\ i \in \mathcal{I}\}.$$

Lemma 2.3.1. *The set $\widehat{\mathcal{F}}^+$ is nonempty.*

Proof. Let $\tilde{x} \in \mathbb{R}^n$ and $\tilde{s}_i > \sqrt[p]{\max\{c_i(\tilde{x}), 0\}}$, for all $i \in \mathcal{I}$. Doing so ensures that $\tilde{s}_i^p - c_i(\tilde{x}) > 0$ and $\tilde{s}_i > 0$ for all $i \in \mathcal{I}$. Therefore, the point $(\tilde{x}, \tilde{s})$ lies in the interior of the feasible region of problem (2.2.2). This proves that the strictly feasible set $\widehat{\mathcal{F}}^+$ is nonempty. $\square$

The next lemma shows that the sequence $\{(\triangle x^k, \triangle s^k)\}$ generated by Algorithm I is a descent direction of the merit function $\phi_{B,\frac{1}{p}}(x^k, s^k; \rho, \mu)$ provided $\mathcal{M}^k \succ 0$ or has been modified to be so.

Lemma 2.3.2. *Let the penalty parameter* $\rho > 0$ *and the barrier parameter* $\mu > 0$ *be fixed. Suppose that Assumptions* 2-4 *hold and, at the* k*-th iteration of Algorithm* 2.1, *the linear system (2.3.5) has a solution* $(\triangle x^k, \triangle s^k, \hat{y}^{k+1}, \hat{u}^{k+1})$. *Then we have*

$$\phi_{B,\frac{1}{p}}{}'(x^k, s^k; \rho, \mu; \triangle x^k, \triangle s^k) \leq -(\triangle x^k, \triangle s^k)^T \mathcal{M}_k (\triangle x^k, \triangle s^k), \tag{2.3.17}$$

where $\phi_{B,\frac{1}{p}}{}'(x^k, s^k; \rho, \mu; \triangle x^k, \triangle s^k)$ *denotes the directional derivative of the function* $\phi_{B,\frac{1}{p}}(x, s; \rho, \mu)$ *at point* (x^k, s^k) *in the direction* $(\triangle x^k, \triangle s^k)$.

Proof. Since the merit function $\phi_{B,\frac{1}{p}}(x, s; \rho, \mu)$ is continuously differentiable, it follows that

$$\nabla_x \phi_{B,\frac{1}{p}}(x^k, s^k; \rho, \mu) = \nabla f(x^k) + \mu^p A(x^k) \big((S^k)^p - C(x^k)\big)^{-1} e, \tag{2.3.18a}$$

$$\nabla_s \phi_{B,\frac{1}{p}}(x^k, s^k; \rho, \mu) = \rho e - p\mu^p (S^k)^{p-1} \big((S^k)^p - C(x^k)\big)^{-1} e - \mu (S^k)^{-1} e, \tag{2.3.18b}$$

$$\phi_{B,\frac{1}{p}}{}'(x^k, s^k; \rho, \mu; \triangle x^k, \triangle s^k) = \nabla_x \phi_{B,\frac{1}{p}}(x^k, s^k; \rho, \mu)^T \triangle x^k + \nabla_s \phi_{B,\frac{1}{p}}(x^k, s^k; \rho, \mu)^T \triangle s^k. \tag{2.3.18c}$$

Substituting (2.3.18a) and (2.3.18b) into (2.3.18c) and combining (2.3.3) and (2.3.7), we can reach inequality (2.3.17). $\square$

In spite of the descent property of the sequence $\{(\triangle x^k, \triangle s^k)\}$, we cannot conclude its tendency to zero. A possible case is that instead of the search direction, the line search steplength may tend to zero. The following two lemmas prove that the line search steplength is sufficiently positive.

Lemma 2.3.3. *Let the penalty parameter* $\rho > 0$ *and the barrier parameter* $\mu > 0$ *be fixed. Suppose that Assumptions* 2-4 *hold and Algorithm* 2.1 *does not terminate at Step* 1 *in the* $(k+1)$*-th iteration. Then we have* $(\triangle x^k, \triangle s^k) \neq 0$.

Proof. Assume to the contrary that $(\triangle x^k, \triangle s^k) = 0$. From (2.3.6a) and (2.3.6b), we have that

$$\begin{aligned} \hat{y}^{k+1} &= \big((S^k)^p - C(x^k)\big)^{-1}\mu^p e, \\ \hat{u}^{k+1} &= (S^k)^{-1}\mu e. \end{aligned} \tag{2.3.19}$$

By line search (2.3.12a) and (2.3.12b), we see that $(\hat{y}^{k+1}, \hat{u}^{k+1}) > 0$. It follows from inequality (2.3.10) we have the matrix $\mathcal{M}^k$ is positive definite. Combining (2.3.7), we have

$$\begin{aligned} -\nabla f(x^k) - \mu^p A(x^k)\big((S^k)^p - C(x^k)\big)^{-1} e = 0, \\ p\mu^p (S^k)^{p-1}\big((S^k)^p - C(x^k)\big)^{-1} e + \mu(S^k)^{-1}e - \rho e = 0. \end{aligned} \tag{2.3.20}$$

By (2.3.19) and (2.3.20), we conclude that the point $(x^{k+1}, s^{k+1}, \hat{y}^{k+1}, \hat{u}^{k+1})$ satisfies the termination condition (2.3.16). Then the Algorithm 2.1 will terminate at the $(k+1)$-th iteration, which contradicts the assumption. □

Lemma 2.3.4. *Let the penalty parameter $\rho > 0$ and the barrier parameter $\mu > 0$ be fixed. Suppose that Assumptions* 2-4 *hold and Algorithm* 2.1 *does not terminate at Step* 1 *in the $(k+1)$-th iteration. Then there exists a constant $\bar{\alpha}_P^k \in (0, 1]$ such that line search condition (2.3.12) holds for all $\alpha_P^{k,j} \in (0, \bar{\alpha}_P^k]$.*

Proof. Let the function $R(x, s) : \mathbb{R}^n \times \mathbb{R}^m_+ \to \mathbb{R}^m$ be defined as $R(x, s) = s^p - c(x)$. Then we have the function $R(x, s)$ is continuous and strictly positive at point (x^k, s^k). Therefore, there exists a constant $\tilde{\alpha}_P^k > 0$ such that condition (2.3.12a) holds for all $\alpha_P^k \in (0, \tilde{\alpha}_P^k]$. By $s^k > 0$, there exists a constant $\hat{\alpha}_P^k > 0$ such that condition (2.3.12b) holds for all $\alpha_P^k \in (0, \hat{\alpha}_P^k]$. By Lemma 2.3.3, we have $(\triangle x^k, \triangle s^k) \neq 0$, and it follows from (2.3.17) that $\phi_{B,\frac{1}{p}}{}'(x^k, s^k, \rho, \mu; \triangle x^k, \triangle s^k) < 0$. Hence, we conclude that there exists a $\breve{\alpha}_P^k > 0$ such that condition (2.3.12c) holds for all $\alpha_P^k \in (0, \breve{\alpha}_P^k]$. Letting $\bar{\alpha}_P^k = \min\{\tilde{\alpha}_P^k, \hat{\alpha}_P^k, \breve{\alpha}_P^k\}$, we prove this lemma. □

Lemma 2.3.5. *Let the penalty parameter $\rho > 0$ and the barrier parameter $\mu > 0$ be fixed. Suppose that Assumptions* 2-4 *hold. Then the sequences $\{(s^k)^p - c(x^k)\}$ and $\{s^k\}$ generated by Algorithm* 2.1 *are bounded from above and componentwise bounded away from zero, so is the sequence $\{(y^k, u^k)\}$ generated by our update strategy (2.3.13)-(2.3.14).*

Proof. Since the sequence $\{(x^k, s^k)\}$ is generated by a descent line search method, it follows that $\phi_{B,\frac{1}{p}}(x^k, s^k; \rho, \mu) \le \phi_{B,\frac{1}{p}}(x^0, s^0; \rho, \mu)$ for all $k \ge 1$. Specifically, we have

$$f(x^k) + \rho \sum_{i\in\mathcal{I}} s_i^k - \mu^p \sum_{i\in\mathcal{I}} \log\big((s_i^k)^p - c_i(x^k)\big) - \mu \sum_{i\in\mathcal{I}} \log s_i^k \le \phi_{B,\frac{1}{p}}(x^0, s^0; \rho, \mu), \quad (2.3.21)$$

for all $k \ge 1$. Assume to the contrary that the sequence $\{s^k\}$ is unbounded. Then we have (taking a subsequence of the sequence $\{s^k\}$ if necessary) $\lim\limits_{k\to\infty} \sum\limits_{i\in\mathcal{I}} s_i^k = +\infty$, as $s_i^k \ge 0$, for all $i \in \mathcal{I}$ and $k \ge 1$. Since the sequence $\{x^k\}$ lies in a bounded set, there exists a vector $x^* \in \mathbb{R}^n$ (taking a subsequence if necessary) such that $\lim\limits_{k\to\infty} x^k = x^*$. By the continuity of the functions f and c_i, $i \in \mathcal{I}$, we have $\lim\limits_{k\to\infty} f(x^k) = f(x^*)$ and $\lim\limits_{k\to\infty} c_i(x^k) = c_i(x^*)$, $i \in \mathcal{I}$. Dividing on both sides of inequality (2.3.21) by $\sum\limits_{i\in\mathcal{I}} s_i^k$ and taking the limit as $k \to \infty$, we have $1 \le 0$ as the facts $\lim\limits_{k\to\infty} \frac{\mu^p \sum_{i\in\mathcal{I}} \log\big((s_i^k)^p - c_i(x^k)\big)}{\sum_{i\in\mathcal{I}} s_i^k} = 0$, $\lim\limits_{k\to\infty} \frac{\mu \sum_{i\in\mathcal{I}} \log s_i^k}{\sum_{i\in\mathcal{I}} s_i^k} = 0$ and the right hand side of inequality (2.3.21) is bounded. Therefore, we prove that the sequence $\{s^k\}$ is bounded above, so is the sequence $\{(s^k)^p - c(x^k)\}$. There exists a vector $s^* \in \mathbb{R}^m$ (taking a subsequence if necessary) such that $\lim\limits_{k\to\infty} s^k = s^*$. Similarly, we can prove that $\lim\limits_{k\to\infty} (s^k)^p - c(x^k) = (s^*)^p - c(x^*) > 0$ and $s^* > 0$. The last part can be proved by virtue of the rules (2.3.13)-(2.3.14) for updating the dual multipliers. Here, the details are omitted. □

Lemma 2.3.6. *Let the penalty parameter $\rho > 0$ and barrier parameter $\mu > 0$ be fixed. Suppose that Assumptions* 2-4 *hold. Then the sequence $\{(\hat{y}^k, \hat{u}^k)\}$ generated by Algorithm* 2.1 *is bounded.*

Proof. Assume to the contrary that the sequence $\{(\hat{y}^k, \hat{u}^k)\}$ is unbounded. Then we have (taking a subsequence if necessary) that $\|(\hat{y}^k, \hat{u}^k)\| \to \infty$ as $k \to \infty$. By Assumptions 3 and 4, there exist a vector x^* and a matrix H^* such that $\lim\limits_{k\to\infty} x^k = x^*$ and $\lim\limits_{k\to\infty} H^k = H^*$. By Assumption 2, we have that

$$\lim_{k\to\infty} \nabla f(x^k) = \nabla f(x^*), \ \lim_{k\to\infty} c(x^k) = c(x^*), \ \lim_{k\to\infty} A(x^k) = A(x^*).$$

It follows from inequality (2.3.10) there exists a positive definite matrix $\mathcal{M}^*$ such that $\lim\limits_{k\to\infty} \mathcal{M}^k = \mathcal{M}^*$. By Lemma 2.3.5, there exist vectors $s^* > 0$, $(y^*, u^*) > 0$ and a constant

$M > 0$ such that $\lim_{k\to\infty} s^k = s^*$, $\lim_{k\to\infty} (y^k, u^k) = (y^*, u^*)$ and

$$(s^*)^p - c(x^*) > 0, \ \|s^*\| \le M, \ \|(s^*)^p - c(x^*)\| \le M, \ \|(y^*, u^*)\| \le M.$$

It follows from equation (2.3.6) we have

$$\hat{y}^k = \big((S^k)^p - C(x^k)\big)^{-1}\big(\mu^p e - pY^k(S^k)^{p-1}\triangle s^k + Y^k A(x^k)^T \triangle x^k\big)$$
$$\hat{u}^k = (S^k)^{-1}\big(\mu e - U^k \triangle s^k\big).$$

Taking limit as $k \to \infty$ on both sides of the above two equations, we conclude that $\lim_{k\to\infty} \|(\triangle x^k, \triangle s^k)\| = \infty$. By equation (2.3.7), we have

$$\mathcal{M}^k \begin{pmatrix} \triangle x^k \\ \triangle x^k \end{pmatrix} = \begin{pmatrix} -\nabla f(x^k) - \mu^p A(x^k)\big((S^k)^p - C(x^k)\big)^{-1} e \\ p\mu^p (S^k)^{p-1}\big((S^k)^p - C(x^k)\big)^{-1} e + \mu (S^k)^{-1} e - \rho e \end{pmatrix}$$

Taking limit as $k \to \infty$ on both sides of the last equation, we conclude that $\lim_{k\to\infty} \|\mathcal{M}^k\| = \|\mathcal{M}^*\| = 0$, which contradicts the fact that the matrix $\mathcal{M}^*$ is positive definite. We prove this lemma. $\square$

Similar to the proof of [30, Lemma 4.11], we can prove the next lemma. Here the details are omitted.

Lemma 2.3.7. *Let the penalty parameter $\rho > 0$ and barrier parameter $\mu > 0$ be fixed. Suppose that Assumptions $2-4$ hold. Then the sequence $\{(\triangle x^k, \triangle s^k)\}$ generated by Algorithm* 2.1 *is bounded from above and $\|(\triangle x^k, \triangle s^k)\| \to 0$ as $k \to \infty$.*

We prove that the sequence $\{(x^k, s^k)\}$ generated by Algorithm 2.1 converges to a KKT point of problem (2.3.1).

Theorem 2.3.1. *Let the penalty parameter $\rho > 0$ and the barrier parameter $\mu > 0$ be fixed. Suppose that Assumptions* 2-4 *hold. Then the sequence $\{(x^k, s^k)\}$ generated by Algorithm* 2.1 *converges to a KKT point of problem (2.3.1).*

Proof. By Assumption 3, Lemmas 2.3.5 and 2.3.6, we have the sequence $\{(x^k, s^k, \hat{y}^k, \hat{u}^k)\}$ lies in a bounded set. Then there exists a vector (x^*, s^*, y^*, u^*) such that

$\lim_{k\to\infty}(x^k, s^k, \hat{y}^k, \hat{u}^k) = (x^*, s^*, y^*, u^*)$ (taking a subsequence if necessary). By Assumption 4, there exists a matrix H^* such that $\lim_{k\to\infty} H^k = H^*$. By Assumption 2, we have that

$$\lim_{k\to\infty} \nabla f(x^k) = \nabla f(x^*),\ \lim_{k\to\infty} c(x^k) = c(x^*),\ \lim_{k\to\infty} A(x^k) = A(x^*).$$

By Lemma 2.3.5, there exist a vector $(y^{**}, u^{**}) > 0$ and a constant $M > 0$ such that $\lim_{k\to\infty}(y^k, u^k) = (y^{**}, u^{**})$ and

$$(s^*)^p - c(x^*) > 0,\ \|s^*\| \le M,\ \|(s^*)^p - c(x^*)\| \le M,\ \|(y^{**}, u^{**})\| \le M.$$

By Lemma 2.3.7, we have $\|(\triangle x^k, \triangle s^k)\| \to 0$ as $k \to \infty$. At the k-th iteration, by (2.3.5), we have

$$\begin{aligned}
\nabla f(x^k) + H(x^k, y^k; \rho)\triangle x^k + A(x^k)(y^k + \triangle y^k) &= 0,\\
p(S^k)^{p-1}(y^k + \triangle y^k) + E(u^k + \triangle u^k) + p(p-1)(S^k)^{p-2}Y^k\triangle s^k &= \rho e,\\
((S^k)^p - C(x^k))(y^k + \triangle y^k) + p(Y^k)(S^k)^{p-1}\triangle s^k - (Y^k)A(x^k)^T\triangle x^k &= \mu^p e,\\
(U^k)\triangle s^k + (S^k)(u^k + \triangle u^k) &= \mu e.
\end{aligned}$$

Taking limit as $k \to \infty$ on both sides of the above equations, we have

$$\begin{aligned}
\nabla f(x^*) + (x^*)y^* &= 0,\\
p(S^*)^{p-1}y^* + u^* &= \rho e,\\
((S^*)^p - C(x^*))y^* &= \mu^p e,\\
S^*u^* &= \mu e.
\end{aligned}$$

Therefore, we prove that the sequence $\{(x^k, s^k)\}$ converges to a KKT point of problem (2.3.1) $\square$

We establish the convergent results of the sequence $\{(x^j, s^j)\}$ generated by Algorithm 2.2.

Theorem 2.3.2. *Let the penalty parameter $\rho > 0$ be fixed. Suppose that Assumptions* 2-4 *hold and that the sequence $\{(x^j, s^j, \hat{y}^j, \hat{u}^j)\}$ is generated by Algorithm* 2.2. *Then we conclude that* *(i) If the sequence $\{(\hat{y}^j, \hat{u}^j)\}$ is unbounded, then the sequence $\{(x^j, s^j)\}$ converges to a FJ point of problem (2.2.2);*

(ii) If the sequence $\{(\hat{y}^j, \hat{u}^j)\}$ is bounded, then the sequence $\{(x^j, s^j)\}$ converges to a KKT point of problem (2.2.2).

Proof. We first suppose that the sequence $\{(\hat{y}^j, \hat{u}^j)\}$ is unbounded. By Assumptions 2 and 3, we have (taking a subsequence if necessary) that there exists a vector x^* such that $\lim_{j\to\infty} x^j = x^*$, $\lim_{j\to\infty} f(x^j) = f(x^*)$, $\lim_{j\to\infty} c(x^j) = c(x^*)$, $\lim_{j\to\infty} \nabla f(x^j) = \nabla f(x^*)$, $\lim_{j\to\infty} A(x^j) = A(x^*)$. By Lemma 2.3.7, there exist a vector $s^* \geq 0$ and a constant $M > 0$ such that $(s^j)^p - c(x^j) \to (s^*)^p - c(x^*) \geq 0$ and $s^j \to s^* \geq 0$ as $j \to \infty$; moreover, $\|(s^*)^p - c(x^*)\| \leq M$ and $\|s^*\| \leq M$. Let $\varpi^j := \max\{\|\hat{y}^j\|, \|\hat{u}^j\|, 1\}$, $\bar{y}^j = (\varpi^j)^{-1}\hat{y}^j$ and $\bar{u}^j = (\varpi^j)^{-1}\hat{u}^j$. We have the sequence $\{(\bar{y}^j, \bar{u}^j)\}$ is bounded. Then we have (taking a subsequence if necessary) there exists a vector $(\bar{y}, \bar{u})$ such that $(\bar{y}^j, \bar{u}^j) \to (\bar{y}, \bar{u})$ as $j \to \infty$; furthermore, $\|(\bar{y}, \bar{u})\| = 1$.

At the j-th iteration, dividing on both sides of inequalities (2.3.16a) and (2.3.16b) by ϖ^j and taking limit as $j \to \infty$, we reach that

$$\begin{aligned} A(x^*)\bar{y} &= 0, \\ p(S^*)^{p-1}\bar{y} + \bar{u} &= 0, \\ ((S^*)^p - C(x^*))\bar{y} &= 0, \\ S^*\bar{u} &= 0, \end{aligned}$$

and $(\bar{y}, \bar{u}) \geq 0$. Consequently, we conclude that the limit point (x^*, s^*) is a FJ point of problem (2.2.2).

We then consider the case when the sequence $\{(\hat{y}^j, \hat{u}^j)\}$ is bounded. Since the sequences $\{x^j\}$ and $\{s^j\}$ are all bounded, there exists a vector (x^*, s^*, y^*, u^*) such that $(x^j, s^j, \hat{y}^j, \hat{u}^j) \to (x^*, s^*, y^*, u^*)$ as $j \to \infty$ (taking a subsequence if necessary). Algorithm 2.2 implies that $\epsilon_{\mu^j} \to 0$ as $j \to \infty$. By (2.3.16a), we conclude that

$\lim_{j\to\infty} \operatorname{Res}(x^j, s^j, \hat{y}^j, \hat{u}^j; \rho, \mu^j) = \operatorname{Res}(x^*, s^*, y^*, u^*; \rho, 0) = 0$. Specifically, we have

$$\begin{aligned}
\nabla f(x^*) + A(x^*)y^* &= 0, \\
\rho e - p(S^*)^{p-1}y^* - u^* &= 0, \\
((S^*)^p - C(x^*))y^* &= 0, \\
S^*u^* &= 0.
\end{aligned}$$

By (2.3.16b), we have $(y^*, u^*) \geq 0$. Combining $(s^*)^p - c(x^*) \geq 0$ and $s^* \geq 0$, we prove that (x^*, s^*) is a KKT point of problem (2.2.2). $\square$

We are now ready to prove the globally convergent results of Algorithm 2.3.

Theorem 2.3.3. *Suppose that Assumptions* 1-4 *hold and the sequence* $\{(x^i, s^i, \hat{y}^i, \hat{u}^i)\}$ *is generated by Algorithm* 2.3. *Then we conclude that*

(i) There exists a constant $\hat{\rho} > 0$ such that the penalty parameter $\rho^i \leq \hat{\rho}$ for all $i \geq 1$, and the sequence $\{(\hat{y}^i, \hat{u}^i)\}$ is bounded, then the sequence $\{x^i\}$ converges to a KKT point of problem (2.1.1);

(ii) There exists a constant $\hat{\rho} > 0$ such that the penalty parameter $\rho^i \leq \hat{\rho}$ for all $i \geq 1$, and the sequence $\{(\hat{y}^i, \hat{u}^i)\}$ is unbounded, then the sequence $\{x^i\}$ converges to a FJ point of problem (2.1.1);

(iii) The penalty parameter ρ^i goes to infinite, then the sequence $\{x^i\}$ converges to a FJ point of problem (2.1.1).

Proof. We consider the following two cases.

Case 1. Assume that there exists a constant $\hat{\rho} > 0$ such that $\rho^i \leq \hat{\rho}$ for all $i \geq 1$. Then the penalty parameter updates in a finite number of times before the termination condition $\|s^i\| \leq \bar{\epsilon}$ is satisfied. If the sequence $\{(\hat{y}^i, \hat{u}^i)\}$ is bounded, by Theorem 2.3.2, the sequence $\{(x^i, s^i, \hat{y}^i, \hat{u}^i)\}$ satisfies the conditions as follows

$$\begin{aligned}
\nabla f(x^i) + A(x^i)\hat{y}^i &= 0, \\
\rho^i e - p(S^i)^{p-1}\hat{y}^i - \hat{u}^i &= 0, \\
((S^i)^p - C(x^i))\hat{y}^i &= 0, \\
S^i \hat{u}^i &= 0, \\
(s^i)^p - c(x^i) &\geq 0, \\
s^i &\geq 0, \\
(\hat{y}^i, \hat{u}^i) &\geq 0,
\end{aligned} \tag{2.3.25}$$

which reduces to the KKT conditions of problem (2.1.1) as $\|s^i\|$ approaches zero. Therefore, we prove the statement (i).

Assume that the sequence $\{(\hat{y}^i, \hat{u}^i)\}$ is unbounded. Let $\varpi^i := \max\{\|\hat{y}^i\|, \|\hat{u}^i\|, 1\}$, $\bar{y}^i := (\varpi^i)^{-1}\hat{y}^i$ and $\bar{u}^i := (\varpi^i)^{-1}\hat{u}^i$. By Theorem 2.3.2, we have that the sequence $\{(x^i, s^i, \bar{y}^i, \bar{u}^i)\}$ satisfies the conditions

$$\begin{aligned}
A(x^i)\bar{y}^i &= 0, \\
p(S^i)^{p-1}\bar{y}^i + \bar{u}^i &= 0, \\
((S^i)^p - C(x^i))\bar{y}^i &= 0, \\
S^i \bar{u}^i &= 0, \\
(s^i)^p - c(x^i) &\geq 0, \\
s^i &\geq 0, \\
(\bar{y}^i, \bar{u}^i) &\geq 0,
\end{aligned} \tag{2.3.26}$$

which reduces to the FJ conditions of problem (2.1.1) as $\|s^i\|$ approaches to zero. Consequently, we prove the statement (ii).

Case 2. By Algorithm 3, we have the sequence $\{(x^{i+1}, s^{i+1}, \hat{y}^{i+1}, \hat{u}^{i+1}, \rho^i)\}$ satisfying

$$\text{Res}(x^{i+1}, s^{i+1}, \hat{y}^{i+1}, \hat{u}^{i+1}; \rho^i, 0) \leq \bar{\epsilon}, \;\; (\hat{y}^{i+1}, \hat{u}^{i+1}) \succeq 0.$$

Therefore, we have the sequence $\{(\hat{y}^i, \hat{u}^i)\}$ is unbounded above as ρ^i goes to infinite. Let $\varpi^i := \max\{\rho^i, \|\hat{y}^{i+1}\|, \|\hat{u}^{i+1}\|, 1\}$, $\bar{\rho}^i := (\varpi^i)^{-1}\rho^i$, $\bar{y}^{i+1} := (\varpi^i)^{-1}\hat{y}^{i+1}$ and $\bar{u}^{i+1} :=$

$(\varpi^i)^{-1}\hat{u}^{i+1}$ for all $i = 0, 1, \ldots$. Since the sequence $\{(x^i, s^i)\}$ and $\{\bar{\rho}^i\}$ are all bounded, there exists a vector $(x^*, s^*, y^*, u^*, \bar{\rho})$ such that $(x^i, s^i, \bar{y}^i, \hat{u}^i, \bar{\rho}^i) \to (x^*, s^*, y^*, u^*, \bar{\rho})$ as $i \to \infty$ (taking a subsequence if necessary). After the i-th iteration, dividing on both sides of inequalities (2.3.16a) and (2.3.16b) by ϖ^i and taking limit as $i \to \infty$, we reach that x^* is a FJ point of problem (2.1.1). Therefore, we have proved the statement (iii).

$\square$

2.4 Numerical Experiments

In this section, we present numerical results of our algorithms described in Section 2.3.3 using MATLAB 7.10.0. We conduct numerical testing on Ubuntu 9.04 with 1.689GB of main memory and Intel(R) Core(TM) 2 Duo 3.0GHz processors.

We refer to the implementation of Algorithms 2.1, 2.2 and 2.3 as the IPLOP method, which stands for the Interior-Point Lower-Order Penalty method. We carry out the numerical experiment on three sets of optimization problems: small- to medium-scale problems, large-scale problems and problems with degenerate constraints. In order to show the robustness of the IPLOP method, we compare its numerical performance with two existing interior-point ℓ_1-penalty methods PIPAL-a and PIPAL-c in [42] in terms of the number of iterations and the relative error.

Before presenting the numerical results, we illustrate the implementation details of our method as follows.

In the implementation, we use the default initial point $x^0 \in \mathbb{R}^n$ as the one provided for every test problem from the test problem collections and set $s_i^0 = \sqrt[p]{\max\{c_i(x^0), 0\}} + \frac{1}{2}$ for all $i \in \mathcal{I}$ unless specified otherwise. We set MaxiterI=1000 as the maximum number of iterations for Algorithm 2.1, and similarly we set MaxiterII=5000 and MaxiterIII=5000 for Algorithm 2.2 and Algorithm 2.3, respectively.

Next, we illustrate our strategy for choosing δ large enough such that the matrix $\mathcal{M}$ (see (2.3.8)) with $\widehat{H}(x, s, y)$ being replaced by $\widehat{H}(x, s, y) + \delta E$ is sufficiently positive definite. However, we would like to keep it as small as possible in order to make this algorithm work more efficiently in practice, as large values of δ make the algorithm behave like a steepest descent method, which is not desirable. Since matrix $\mathcal{M}$ is symmetric and the matrix $4S\mathcal{N}S + S^{-1}U - 2Y$ is diagonal and positive definite, it follows from the LDL^T factorization for a symmetric indefinite matrix in [165, 167] that we can find a sufficiently small δ such that $\mathcal{M}$ is positive definite. In our implementation, we use the factorization routine MA57 in MATLAB 7.10.0 for this purpose.

Having computed search directions from (2.3.7), the step size $\alpha_p^k \in (0, 1]$ has to be determined in order to obtain the next iterate by (2.3.11). In our implementation, we first obtain $\bar{\alpha}_P^k := \max\{\ \bar{\rho}_1^j \mid j = 0, 1, 2, \dots\}$ with $\bar{\rho}_1 \in (0, 1)$ such that (2.3.12c) holds. Then, we let $\alpha_P^k := \max\{\bar{\alpha}_p^k \bar{\rho}_2{}^j \mid j = 0, 1, 2, \dots\}$ with $\bar{\rho}_2 \in (0, 1)$ satisfying

$$(s^{k+1})^p - c(x^{k+1}) \geq (1 - \hat{\eta})\Big((s^k)^p - c(x^k)\Big), \tag{2.4.27a}$$

$$s^{k+1} \geq (1 - \hat{\eta})s^k, \tag{2.4.27b}$$

where $\hat{\eta} = \max\{0.99, 1 - \mu\}$ in our implementation. We see that (2.4.27a) and (2.4.27b) imply (2.3.12a) and (2.3.12b), respectively. The modification (2.3.12a) as (2.4.27a) is due to the nonlinearity of $(s^{k+1})^2$ in (2.3.12a), in which case the classical fraction-to-boundary rule cannot be used anymore. The above strategy of computing stepsize α_P^k is shown to be efficient in our numerical experiments. In Algorithm 2.2, we set $\epsilon_{\mu^j} = \mu^j$ and $\epsilon_{\mu^{j+1}} = \max\{\gamma\epsilon_{\mu^j}, 10^{-7}\}$ at the j-th iteration.

The default settings for different parameters are listed in Table 2.1 below.

Table 2.1: Input parameter values for the IPLOP method.

Parameter	Value	Parameter	Value
ρ^0	0.1	ν	5
μ^0	0.1	γ	0.1
γ_{min}	0.5	γ_{max}	10^{23}
γ_1	1	$\bar{\rho}_2$	0.1
τ_1	10^{-8}	$\bar{\rho}_1$	0.5
$\bar{\epsilon}$	10^{-6}		

2.4.1 Experiments with the Different Power p

In this subsection, we select our first test problems set, a total of 266 inequality constrained optimization problems, from the CUTEr collection[2], COPS[3], MITT[4] and GLOBAL Library[5] test sets, see Table 2.2. We first use this test set to test the performances of the IPLOP method with different values of the power p in term of the number of iterations and the values of the penalty parameter ρ. Some values of the power p are chosen as $p = 1.0,\ 4/3,\ 3/2,\ 2,\ 4,\ 5$. For convenience, we write the IPLOP method with different values p as the IPLOP_1, $\text{IPLOP}_{3/4}$, $\text{IPLOP}_{2/3}$, $\text{IPLOP}_{1/2}$, $\text{IPLOP}_{1/4}$ and $\text{IPLOP}_{1/5}$ methods, respectively.

Table 2.2: Problem names for the first test set.

Problem	**Problem**	**Problem**	**Problem**	**Problem**
3pk	allinit	avgasa	avgasb	bearing_50_100
bearing_50_50	bearing_50_70	biggsb1	biggsc4	bqpgabim
bqpgasim	camel6	camshape_100	cantilvr	cb2
cb3	chaconn1	chaconn2	circle	congigmz
coshfun	deer	demymalo	dipigri	eg1
eigena	emfl_vareps	esfl_socp	ex14_1_2m	ex14_1_4
ex14_1_5m	ex14_1_8	ex14_1_9	ex14_2_1m	ex14_2_2m
ex14_2_3m	ex14_2_4m	ex14_2_4m	ex14_2_5m	ex14_2_7m
ex14_2_8m	ex14_2_9m	ex2_1_1	ex2_1_10	ex2_1_3
ex2_1_4	ex2_1_5	ex2_1_6	ex2_1_7	ex3_1_2
ex3_1_3	ex3_1_4	ex4_1_5	ex4_1_9	ex7_2_1
ex7_2_5	ex7_2_6	ex7_3_1	ex8_1_1	ex8_6_2
expfita	expfitb	expquad	fekete	fekete2
fekete3	fir_convex	fir_linear	fir_socp	gpp
hadamals	haifam	haifas	haldmads	hart6
hatfldc	himmelp1	himmelp2	himmelp5	himmelp6
hs001	hs002	hs003	hs004	hs005
hs010	hs011	hs012	hs015	hs016
hs017	hs018	hs020	hs021	hs022
hs023	hs024	hs029	hs030	hs031
hs033	hs034	hs035	hs036	hs037
hs038	hs043	hs044	hs045	hs059
hs064	hs065	hs066	hs076	hs083
hs084	hs086	hs088	hs093	hs095
hs096	hs097	hs098	hs100	hs100mod
hs108	hs110	hs113	hs117	hs118

[2]http://orfe.princeton.edu/~rvdb/ampl/nlmodels/.
[3]http://www.mcs.anl.gov/~more/cops/.
[4]http://plato.asu.edu/ftp/ampl_files/lukvl_ampl/lukvl/.
[5]http://www.gamsworld.org/global/globallib.htm.

Table 2.2: Problem names for the first test set (continued).

Problem	Problem	Problem	Problem	Problem
hubfit	jbearing100	jbearing25	jbearing50	jbearing75
kiwcresc	least	logcheb	lootsma	lowpass
madsen	madsschj	makela1	makela3	matrix2
median_exp	median_nonconvex	mifflin1	mifflin2	minmaxrb
minsurf_50_100	minsurf_50_50	minsurf_50_75	mistake	oet7
optprloc	pacman	palmer1	palmer1a	palmer1b
palmer2	palmer2a	palmer2b	palmer3	palmer3a
palmer3b	palmer4	palmer4a	palmer4b	palmer5a
palmer5b	palmer5e	palmer6a	palmer6e	palmer7a
palmer7e	palmer8a	palmer8e	pentagon	polak4
polygon_100	polygon_50	polygon25	polygon75	prolog
pspdoc	qr3d	qr3dbd	qr3dls	qrtquad
rbrock	s222	s223	s224	s225
s226	s227	s228	s229	s230
s231	s232	s233	s234	s236
s237	s238	s239	s242	s244
s249	s250	s251	s253	s257
s259	s264	s268	s270	s277
s278	s279	s280	s284	s285
s315	s323	s324	s326	s330
s331	s337	s339	s340	s341
s343	s346	s354	s356	s357
s359	s360	s361	s365	s365mod
s366	s368	s384	s385	s387
s388	s389	sineali	spiral	springs
springs_nonconvex	stancmin	synthes1	torsion_50_50	turtle
twobars	weeds	yfit	zecevic3	zecevic4
zy2				

Using the performance profiles of Dolan and Moré in [46], we plot Figure 2.1, where the plots $\pi_s(\tau)$ denote the scaled performance profile

$$\pi_s(\tau) := \frac{\text{no. of problems } \hat{p} \text{ where } \log_2(r_{\hat{p},s}) \leq \tau}{\text{total no. of problems}}, \ \tau \geq 0,$$

where $\log_2(r_{\hat{p},s})$ is the scaled performance ratio between the iteration number to solve problem $\hat{p}$ by solver s over the fewest iteration number required by the solvers of the IPLOP method with different p. 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 [46] for more details regarding the performance profiles.

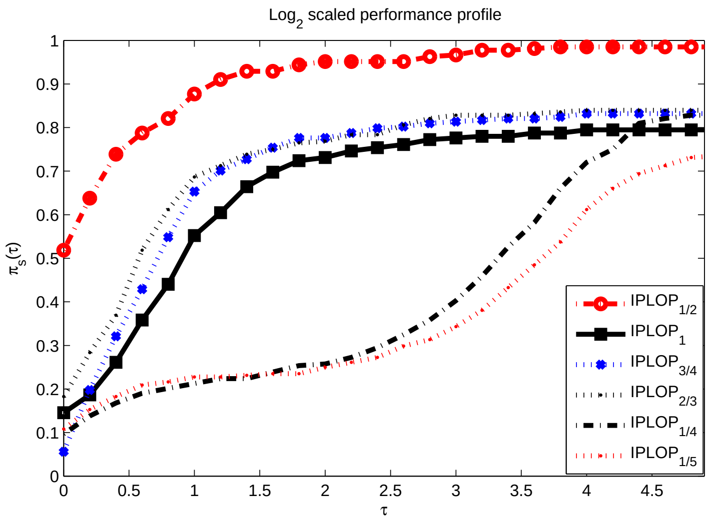

Figure 2.1: Performance profiles based on the number of iterations for the IPLOP method with the different p.

Figure 2.1 shows that on this test set the $\mathrm{IPLOP}_{1/2}$ method is the most efficient among all the six methods as the performance profile for the $\mathrm{IPLOP}_{1/2}$ method lies above all others for all performance ratios. Moreover, the $\mathrm{IPLOP}_{1/2}$ method uses the least number of iterations on approximately 52% of test problems, and solves the most problems (about 97%) successfully. The $\mathrm{IPLOP}_{3/4}$ and $\mathrm{IPLOP}_{2/3}$ methods share the nearly same performance and are more efficient than the IPLOP_1 method. The robustness of the $\mathrm{IPLOP}_{1/4}$ method is almost identical with the $\mathrm{IPLOP}_{3/4}$ method, but the $\mathrm{IPLOP}_{1/4}$ method is less efficient than the $\mathrm{IPLOP}_{3/4}$ method and even the IPLOP_1 method. Furthermore, the $\mathrm{IPLOP}_{1/5}$ method is the weakest solver among them as its performance profile lies below all the others.

We use the values of ρ to plot Figure 2.2, which shows that the $\mathrm{IPLOP}_{1/2}$ method uses the smallest values of penalty parameter ρ on approximately 93% of test problems. Furthermore, Figure 2.2, to some extent, verifies the theorem in [152, Theorem 7.2], which states that the smallest exact penalty parameter for the $\ell_{\frac{1}{p}}(p > 1)$-exact penalty function is smaller than that of the ℓ_1-exact penalty function.

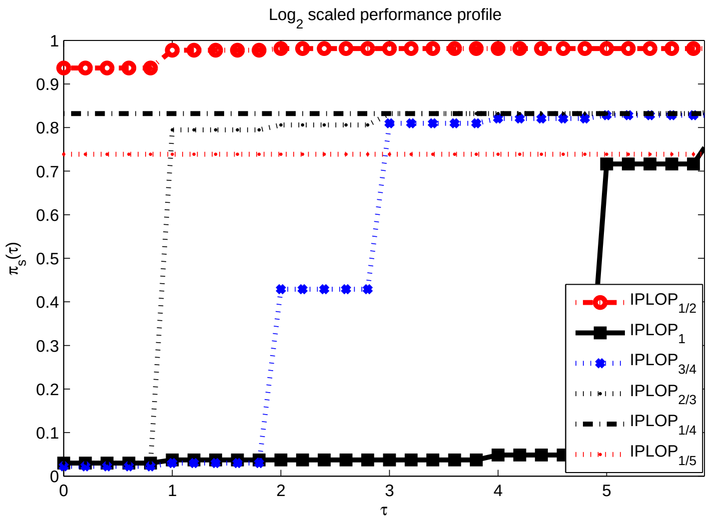

Figure 2.2: Performance profiles based on the values of the penalty parameter for the IPLOP method with the different p.

2.4.2 Experiments with Small-Scale and Medium-Scale Problems

In this subsection, using the first test set, we compare the performance of the IPLOP$_{1/2}$ method with the interior-point ℓ_1-penalty methods PIPAL-a and PIPAL-c implemented in PIPAL1.0[1] by Curtis [42].

We plot Figures 2.3-2.4, which describe the performance of these solvers in the number of iterations and the values of the penalty parameter, respectively. Figure 2.3 shows that the IPLOP$_{1/2}$ method uses the least number of iterations on approximately 58% of test problems and shares the nearly same robustness with other two solvers.

[1]http://coral.ie.lehigh.edu/ frankecurtis/software.

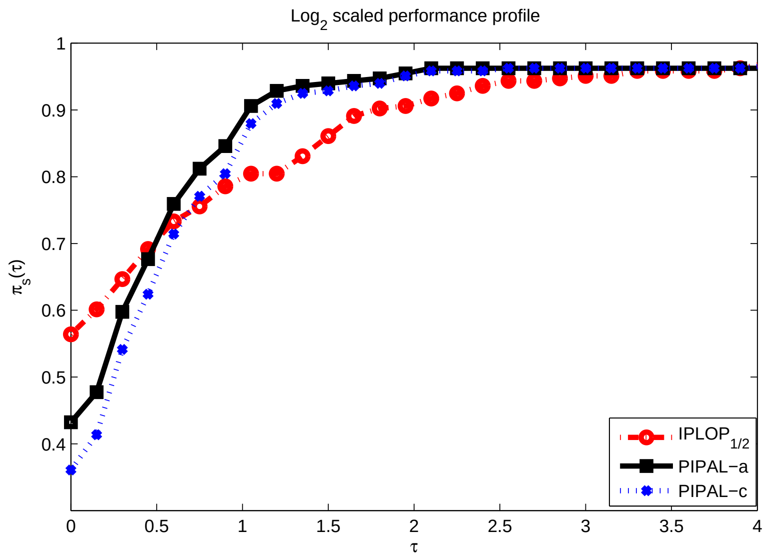

Figure 2.3: Performance profiles based on the number of iterations for the IPLOP$_{1/2}$, PIPAL-a and PIPAL-c methods.

Figure 2.4 is plotted by the values of ρ, which shows the IPLOP$_{1/2}$ method uses smaller values of the penalty parameter than that of the PIPAL-c method which employs the same strategy for updating the penalty parameter as the IPLOP$_{1/2}$ method.

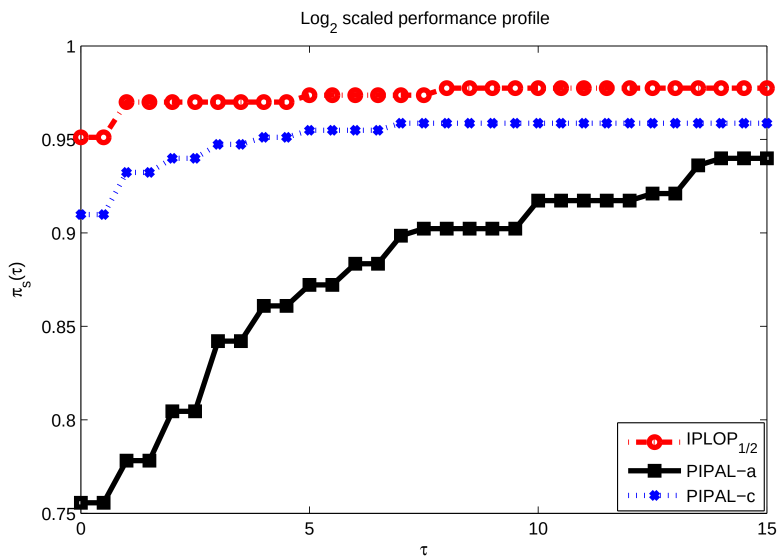

Figure 2.4: Performance profiles based on the values of the penalty parameter for the IPLOP$_{1/2}$, PIPAL-a and PIPAL-c methods.

We compare the performance of the IPLOP$_{1/2}$ method with that of the PIPAL-a and PIPAL-c methods in terms of the relative error. The relative error is defined as

$\frac{|f(x^*)-f^*|}{|f^*|+\varepsilon}$, where f^* denotes the known local minimum of object function, $f(x^*)$ denotes the computed local minimum by a solver with the same starting point, positive constant ε is very small to guarantee the relative error making sense as $f^* = 0$. In our first test set, there are 250 problems that we know their best local minima with given starting points. Based on the relative error, we plot Figure 2.5, which shows that the IPLOP$_{1/2}$ method can solve about 90% of test problems with the smallest relative error.

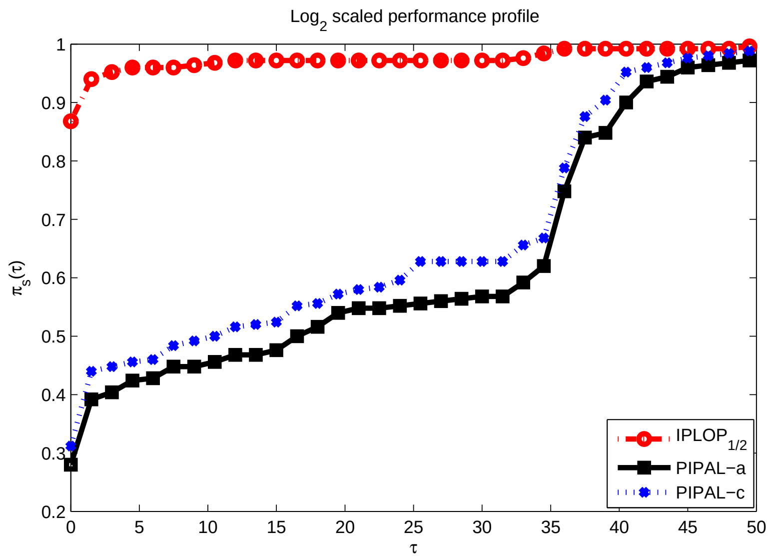

Figure 2.5: Performance profiles based on the relative error for the IPLOP$_{1/2}$, PIPAL-a and PIPAL-c methods.

2.4.3 Experiments with Large-Scale Problems

In this subsection, we choose 26 large-scale inequality constrained optimization problems from COPS and MITT test sets as the second test set, These problems cannot be solved by either the PIPAL-a method or the PIPAL-c method. We show the test problem data and the numerical performance of the IPLOP$_{1/2}$ method for solving these large-scale problems in Table 2.4, whose abbreviations are illustrated in Table 2.3. Table 2.4 shows that the IPLOP$_{1/2}$ method can successfully solve very large-scale problems.

Table 2.3: Abbreviations on the experiments for large scale problems.

Problem	name of test problem
♯var	number of variables (not including the slack variables)
♯ineq	number of inequality or range constraints (including the bounded constraints)
♯Iter1	number of iterations of Algorithm 2.1
♯Iter2	number of iterations of Algorithm 2.2
♯Iter3	number of iterations of Algorithm 2.3
CPU [s]	CPU time in seconds
$f(x^*)$	computed objective function value

Table 2.4: Performance of the IPLOP$_{1/2}$ method to large-scale problems.

Problem	**♯var**	**♯ineq**	**♯Iter1**	**♯Iter2**	**♯Iter3**	**CPU [s]**	$f(x^*)$
cvxbqp1	10000	20000	12	6	1	8.18784e+00	2.25023e+06
expquad	120	20	30	5	1	1.33528e-01	-3.62460e+06
lukvli2	50000	49993	43	6	1	2.34645e+02	1.32666e+06
lukvli3	50000	2	12	6	1	6.40883e+00	1.15775e+01
lukvli3_100000	100000	2	11	5	1	1.34654e+01	1.15775e+01
lukvli6	49999	24999	48	5	1	9.01176e+01	3.14423e+06
lukvli7	50000	4	24	5	1	1.33679e+01	-1.86339e+04
lukvli9	50000	6	192	6	1	5.31867e+01	4.99467e+03
lukvli11	49997	33330	19	5	1	4.02668e+01	1.26468e-03
lukvli12	49997	37497	21	5	1	2.58388e+01	2.97710e-05
lukvli16	49997	37497	25	5	1	3.18699e+01	4.62037e-02
lukvli17	49997	37497	25	5	1	3.02674e+01	1.05490e-02
lukvli17_149990	149990	112492	27	5	1	1.37633e+02	3.54839e-03
lukvli18	49997	37497	15	5	1	1.38716e+01	3.01438e-02
sinrosnb	1000	2000	9	5	1	2.10994e-01	-9.99010e+04
svanberg	5000	15000	154	6	2	4.61292e+01	8.36142e+03
bearing_200_200	40000	40000	74	6	1	1.53298e+02	-1.54829e-01
torsion_50_75	3750	7500	53	6	2	6.11683e+00	-4.18199e-01
torsion_50_100	5000	10000	56	6	2	8.00671e+00	-4.18239e-01
torsion_200_200	40000	80000	46	5	1	1.27469e+02	-4.18469e-01
torsion_400_400	160000	320000	77	6	1	9.65081e+02	-4.18488e-01
polygon_100	198	5444	78	5	1	2.75431e+01	8.07387e-10
polygon_200	398	20894	120	6	2	1.76088e+02	7.45598e-10
duct12	6906	18875	9	5	1	1.92118e+01	2.23076e+04
duct15	2895	8671	9	5	1	5.89805e+00	1.04951e+04
hook	1200	4071	10	5	1	2.38681e+00	6.05735e+03

2.4.4 Experiments with Degenerate Problems

It is a great challenge to design efficient algorithms for solving optimization problems with degenerate constraints (OPDC, for short). We select 37 degenerate test problems from the DEGEN_collection[6] and one degenerate problem from [117, Example 2.3] as our third test set shown in Table 2.6. All these 38 problems are only with inequality constraints and have the unique minimizer. We use the classification rules in [90] for naming these problems, that is, using the form T-DD-NN as test identifiers. Their meanings are explained in Table 2.5 below.

Table 2.5: Classification rules for degenerate test problems.

T	problem type	
	0	problems satisfying LICQ but violating strict complementarity
	1	problems satisfying MFCQ but violating LICQ
	2	problems violating MFCQ
	3	MPCCs(problems with complementarity constraints)
	4	MPVCs(problems with vanishing constraints)
	5	problems satisfying lower-order exact penalty but violating GCQ
DD	number of variables	
NN	number in the T-DD group	

Table 2.6: Problem names for the third test set.

Problem	**Problem**	**Problem**	**Problem**	**Problem**
00201	00302	00501	10202	10203
10204	10205	10207	10401	20103
20104	20105	20209	20211	20213
20216	20221	20222	20226	20227
20304	30201	30203	30205	30206
30210	30301	40201	40202	40203
40204	40205	40206	40207	40208
40210	40401	50201		

For each test problem, we perform 100 runs from randomly generated starting points by a uniform distribution in $[-10, 10]$. Using the relative error, we plot Figure 2.6 to

[6]http://www.impa.br/ optim/solodov.html.

qualify the ability of finding the minimum of the objective function for each method. Figure 2.6 shows that the IPLOP$_{1/2}$ method reaches the smallest relative error on approximately 90% of test problems and is more reliable and robust than the PIPAL-a and PIPAL-c methods as its performance profile lies above them for all performance ratios.

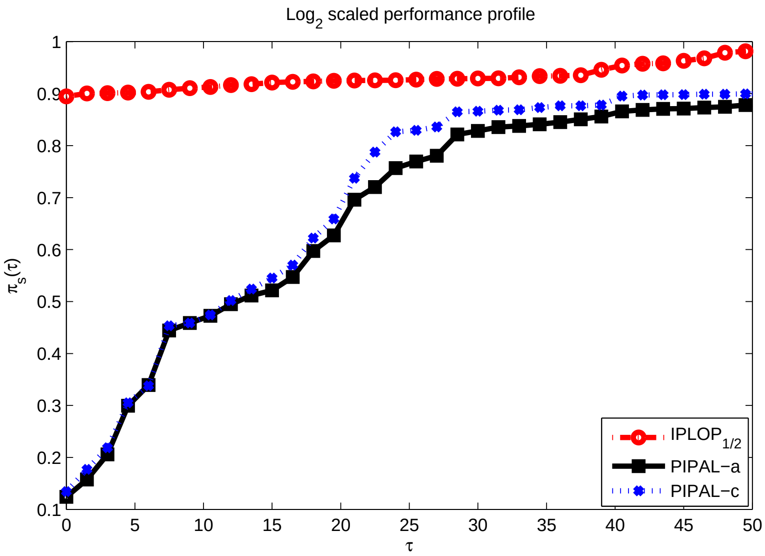

Figure 2.6: Performance profiles based on the relative error of degenerate test problems for the IPLOP$_{1/2}$, PIPAL-a and PIPAL-c methods.

Chapter 3

A Box-Constrained Differentiable Penalty Method for Nonlinear Complementarity Problems

3.1 Introduction

In this chapter, we consider the following NCP of finding a vector $x \in \mathbb{R}^n$ satisfying the following conditions

$$x \le 0,\ F(x) \le 0,\ x^T F(x) = 0, \tag{3.1.1}$$

where the function $F : \mathbb{R}^n \to \mathbb{R}^n$ is assumed to be continuously differentiable. Throughout this chapter, we use X^* to denote the solution set of problem (3.1.1) and $\mathcal{J} = \{1, 2, \ldots, n\}$. When the function F is linear, i.e., $F(x) = Ax - b$ with a given matrix $A \in \mathbb{R}^{n\times n}$ and a vector $b \in \mathbb{R}^n$, problem (3.1.1) is reduced to a LCP. Complementarity problems play an important role in operations research, option pricing, economic equilibrium models and the engineering sciences; see, e.g., [53, 54, 82].

We introduce a definition for the function F named a uniform ξ-P-function, which is weaker than the ξ-monotonicity and coincides with a uniform P-function (or a P-function) when the function F is linear. Under the assumption of a uniform ξ-P-function, we show that problem (3.1.1) has a unique solution, and moreover the penalized equation (1.2.21) has a unique solution for any $\rho > 0$. Then we introduce a box-constrained differentiable penalty method for solving problem (3.1.1), which not only inherits the convergence rate of the existing $\ell_{\frac{1}{p}}$-penalty method but also mitigates its drawback. Specifically, we consider a differentiable system of 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. Instead of solving the above system directly, we consider a corresponding least squares problem, which can be solved by a trust-region Gauss-Newton method introduced by Macconi et al. [123].

We carry out numerical experiments on test problems from MCPLIB [9]. We first set $p = 2$ and compare the performance of our method with the smoothed $\ell_{\frac{1}{2}}$-penalty method [87] and the ℓ_1-penalty method [7] in terms of the number of function evaluations and the values of the penalty parameter. Numerical results show that our method is more efficient and robust than other two methods. Then different values of the power $p = 1, 2, 100, 1000, 5000, 10000$ are chosen to test the efficiency of our method. Furthermore, we compare the performance of our method with the smooth approximation method [23] and the nonsmooth equations method [93] in terms of the number of function evaluations.

This chapter is organized as follows. In Section 3.2, we propose a differentiable penalty method for problem (3.1.1). 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 (3.1.1) in Section 3.3. In the last section, preliminary numerical experiments are shown.

3.2 Box-Constrained Differentiable Penalty Method

In this section, we first introduce a new definition for the function F named a uniform ξ-P-function with $\xi \in (1, 2]$. Then we propose a box-constrained differentiable penalty method and establish its convergence rate theorem.

3.2.1 Uniform ξ-P-function

To begin, we first recall some useful definitions on function F.

Definition 3.2.1 ([50]). *A function $G : \mathbb{R}^n \to \mathbb{R}^n$ is said to be ξ-monotone for some $\xi \in (1, 2]$, if there exists a constant $\alpha > 0$ such that*

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

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

Definition 3.2.2 ([50]). *A function $G : \mathbb{R}^n \to \mathbb{R}^n$ 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*

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

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

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

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

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

Definition 3.2.3 ([50]). *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 definition for the function F called a uniform ξ-P-function, which is stronger than a P-function.

Definition 3.2.4. *A function $G : \mathbb{R}^n \to \mathbb{R}^n$ is said to be a uniform ξ-P-function for some $\xi \in (1, 2]$, if there exists constant $\alpha > 0$ such that for all pairs of vectors x and y in $\mathbb{R}^n$,*

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

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 [87] to establish the convergence rate of $\mathcal{O}(\rho^{-\frac{p}{\xi}})$ by which the solution of problem (1.2.21) converges to that of problem (3.1.1). If $A \in \mathbb{R}^{n\times n}$ is a P-matrix, then the function $G(x) = Ax$ becomes a uniform P-function (also a P-function). The following propositions are useful to investigate properties of the uniform ξ-P-function.

Proposition 3.2.1 ([50])**.** *Let $G : D \subset \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable P_0-function on the open set D. Then $\nabla G(x)$ is a P_0-matrix for each $x \in D$.*

Corollary 3.2.1. *Let $G : D \subset \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable ξ-P-function on the open set D. Then $\nabla G(x)$ is a P_0-matrix for each $x \in D$.*

Proposition 3.2.2 ([50])**.** *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 $x_i\big(Ax\big)_i \geq 0$.*

Proposition 3.2.3 ([50])**.** *Let the linear function $G : \mathbb{R}^n \to \mathbb{R}^n$ be $G(x) = Ax - b$ with a given matrix $A \in \mathbb{R}^{n\times n}$ and a vector $b \in R^n$. Then*

(a) G is ξ-monotone if and only if matrix A is positive definite;

(b) G is a (uniform) P-function if and only if A is a P-matrix.

Corollary 3.2.2. *Let the linear function $G : \mathbb{R}^n \to \mathbb{R}^n$ be $G(x) = Ax - b$ with a given matrix $A \in \mathbb{R}^{n\times n}$ and a vector $b \in \mathbb{R}^n$. Then G is a uniform ξ-P-function if and only if A is a P-matrix.*

Proposition 3.2.4 ([50])**.** *Let $G : D \subset \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable on the open convex set D. Then G is 2-monotone on D if and only if its Jacobian matrix $\nabla G(x)$ is uniformly positive definite for all x in D, i.e., there exists a constant $c' > 0$ such that*

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

for all $x \in D$.

We present an example from [40, Example 3.3.2] below, which shows that the uniform ξ-P-function is weaker than the ξ-monotonicity.

Example 3.2.1. *Let $F(x) = Ax - b$ 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 3.2.3 that we know function $F(x)$ is a uniform ξ-P-function, but not ξ-monotone.

Next we describe a nonlinear example to show that the uniform P-function is weaker than the 2-monotonicity.

Example 3.2.2. *Consider function $G : \mathbb{R}^2 \to \mathbb{R}^2$ as*

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

where $F(x)$ is a linear function defined in Example 3.2.1. 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 3.2.1 shows that the matrix* $\nabla G(0)$ *is not positive definite. Therefore, by Proposition 3.2.4, we conclude that the function* $G(x)$ *is not* 2*-monotone. Since the function* $F(x)$ *is a uniform* P*-function, it follows that there exists constant* $\alpha > 0$ *such that for all pairs of vectors* x *and* y *in* $\mathbb{R}^2$ *the inequality* $\max\limits_{1\le\kappa\le 2}(x_\kappa - y_\kappa)(F_\kappa(x) - F_\kappa(y)) \ge \alpha\|x-y\|^2$ *holds. We 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*

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

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 (1.2.21) is unique. Before doing this, we first prove an auxiliary proposition.

Proposition 3.2.5. *Assume that the function* $F : \mathbb{R}^n \to \mathbb{R}^n$ *is a continuous uniform* ξ*-*P*-function. Then problem (3.1.1) has a unique solution.*

Proof. It follows from [50, Proposition 1.1.3] that problem (3.1.1) is equivalent to the following variational inequality problem: Find a vector $x \in \mathcal{K}$ such that for all vectors $y \in \mathcal{K}$

$$(y-x)^T F(x) \ge 0, \tag{3.2.2}$$

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

Using the [50, Proposition 3.5.1 (a)], we need to prove that there exists a vector $x^{\text{ref}} \in \mathcal{K}$ such that the set

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

is nonempty and bounded. Let $x^{\text{ref}} \in \mathcal{K}$ and $\|x^{\text{ref}}\| \ne 0$. By the continuity of the 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 the contrary that the set $L'_{\leq}$ is unbounded. There exists a sequence $\{x^k\} \subset \mathcal{K}$ such that for all k,

$$F_\nu(x^k)(x^k_\nu - x^{\text{ref}}_\nu) \leq 0, \ \forall\, \nu \in \mathcal{J} \text{ such that } x^k_\nu \neq x^{\text{ref}}_\nu \tag{3.2.3}$$

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

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 $x^k_\nu \neq x^{\text{ref}}_\nu$ such that

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

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

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

which contradicts with inequality (3.2.3). Therefore, the set $L'_{\leq}$ is bounded for any given $x^{\text{ref}} \in \mathcal{K}$. By [50, Proposition 3.5.1 (c)], we conclude that the variational inequality problem (3.2.2) has a solution. Since a uniform ξ-P-function is a P-function, it follows from [50, Proposition 3.5.10 (a)] that the variational inequality problem (3.2.2) has at most one solution. Therefore, we have proved that the solution of variational inequality problem (3.2.2) is unique. We concluded that problem (3.1.1) has a unique solution. □

Proposition 3.2.6. *Assume the function $F : \mathbb{R}^n \to \mathbb{R}^n$ is a uniform ξ-P-function. Then the penalized nonlinear equations (1.2.21) 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

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

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

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

where the last inequality is from Definition 3.2.4. Therefore, the function $\phi(x,\rho)$ is a uniform ξ-P-function for any $\rho \geq 0$ and 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 3.2.5. We proved that the penalized equations (??) have a unique solution. $\square$

3.2.2 Box-Constrained Differentiable Penalty Method

In this subsection, we introduce a box-constrained differentiable penalty method for solving problem (3.1.1), 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:

$$\mathcal{F}(x,\rho) := \begin{pmatrix} x_1F_1(x) & + & \rho[F_1(x)]_+^q \\ x_2F_2(x) & + & \rho[F_2(x)]_+^q \\ \vdots & \vdots & \vdots \\ x_nF_n(x) & + & \rho[F_n(x)]_+^q \end{pmatrix} = 0, \quad x \in \Omega, \tag{3.2.4}$$

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 $q > 1$. We see that the function $\mathcal{F}(\cdot,\rho) : \mathbb{R}^n \to \mathbb{R}^n$ is first order continuously differentiable for any given ρ. The system (3.2.4) can be solved efficiently by algorithms [98, 123].

Remark 3.2.1. *Alternately, we can consider another system of constrained equations for problem (3.1.1) as follows*

$$\begin{pmatrix} x_1F_1(x) & + & \rho[x_1]_+^q \\ x_2F_2(x) & + & \rho[x_2]_+^q \\ \vdots & \vdots & \vdots \\ x_nF_n(x) & + & \rho[x_n]_+^q \end{pmatrix} = 0, \quad x \in \widehat{\Omega}, \tag{3.2.5}$$

where $\widehat{\Omega} = \{x \in \mathbb{R}^n \mid F(x) \leq 0\}$. However, the feasible set $\widehat{\Omega}$ in general is not convex in general. It is not easy to solve the system (3.2.5) when the function F is nonlinear.

Proposition 3.2.7. *Let $x^* \in \mathbb{R}^n$ be a solution of problem (3.1.1). Then x^* solves $\mathcal{F}(x, \rho) = 0$ for any given $\rho > 0$.*

We present an example that shows the converse of Proposition 3.2.7 is not true.

Example 3.2.3. *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 (3.1.1) when $x^* > 0$.*

Remark 3.2.2. *Example 3.2.3 indicates that the constraint set Ω in the system (3.2.4) is vital to the box-constrained differentiable penalty method for problem (3.1.1).*

Given the penalty parameter ρ and the power p. The solution of the system (3.2.4) in general is not unique even if problem (3.1.1) has a unique solution, which is verified by the next example.

Example 3.2.4. *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 $x(x+1) + \rho[x+1]_+^2 = 0$ with $x \leq 0$. The constrained equation has two solutions, one is $\bar{x}^\rho = -1$ and the other one is $\hat{x}^\rho = -\frac{\rho}{\rho+1}$.*

3.2.3 Convergence Rate Analysis

In this subsection, we establish that the solution x^ρ of system (3.2.4) converges to a solution x^* of problem (3.1.1) at a rate of $\mathcal{O}(\rho^{-\frac{p}{\xi}})$ provided that function F is a uniform ξ-P-function. We first show some useful lemmas as follows.

Lemma 3.2.1. *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 (3.2.4). 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 (3.2.4), it follows that $x_i^\rho F_i(x^\rho) + \rho F_i(x^\rho)[F_i(x^\rho)]_+^{\frac{1}{p}} = 0$, which means $x_i^\rho F_i(x^\rho) \leq 0$, for all $i \in \mathcal{J}$.

By the uniform ξ-P-function of the 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 have proved this lemma with $M_1 = \sqrt[\xi-1]{\frac{1}{\alpha}\|F(0)\|_\infty}$. $\square$

Lemma 3.2.1 implies that, for any $\rho > 0$, the solution of system (3.2.4) always lies in a bounded closed set. Assuming further the continuity of function F, we have that there exists a positive constant L, independent of x^ρ, ρ and p, such that $\|F(x^\rho)\| \leq L$, for all $\rho > 0$.

Lemma 3.2.2. *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 (3.2.4). Then there exists a positive C_1, independent of x^ρ and ρ, such that*

$$\|[F(x^\rho)]_+\| \leq C_1\rho^{-p}.$$

Proof. Since x^ρ is a solution of system (3.2.4), it follows that $\rho[F_i(x^\rho)]_+^q = -F_i(x^\rho)x_i^\rho \leq \|F(x^\rho)\|_\infty\|x^\rho\|_\infty$ for all index $i \in \mathcal{J}$. Therefore, we have $\|[F(x^\rho)]_+\|_\infty \leq \rho^{-p}\|x^\rho\|_\infty^p$. By the fact that all norms in $\mathbb{R}^n$ are equivalent, there exists constant $\widetilde{C} > 0$ such that $\|[F(x^\rho)]_+\| \leq \widetilde{C}\|[F(x^\rho)]_+\|_\infty$. By Lemma 3.2.1, we have there exists a constant C_1 such that

$$\|[F(x^\rho)]_+\| \leq C_1\rho^{-p}$$

with $C_1 = \widetilde{C}M_1^p$. $\square$

Now, we establish our main convergence rate theorem.

Theorem 3.2.1. *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 (3.1.1) and system (3.2.4), respectively. Then there exist constants $\widehat{C} > 0$ and $\xi > 1$, independent of x^ρ and ρ, such that*

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

Proof. Since x^ρ is a solution of system (3.2.4), the index set at x^ρ can be split into the following two sets:

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

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

$$\Big(x_i^* - x_i^\rho\Big)\Big(F_i(x^*) - F_i(x^\rho) + [F_i(x^\rho)]_+\Big) = \Big(x_i^* - x_i^\rho\Big)\Big(F_i(x^*) + [F_i(x^\rho)]_-\Big) \le 0 \quad (3.2.6)$$

where $[a]_- := \max\{-a, 0\}$ for all $a \in \mathbb{R}$. Note that x^* is a solution of problem (3.1.1), the following two cases are considered.

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

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

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

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

Therefore, we proved that the inequality (3.2.6) 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

$$\begin{aligned}
\alpha\|x^* - x^\rho\|^\xi &\le \max_{1\le i\le n} \big(x_i^* - x_i^\rho\big)\big(F_i(x^*) - F_i(x^\rho)\big)\\
&\le \max_{1\le i\le n} \big(-[F_i(x^\rho)]_+(x_i^* - x_i^\rho)\big)\\
&\le C_1\rho^{-p}\|x^* - x^\rho\|_\infty\\
&\le 2C_1M_1\rho^{-p},
\end{aligned}$$

where the second inequality is from inequality (3.2.6), the third one is from Lemma 3.2.2 and the last one is from Lemma 3.2.1. Therefore, we proved this theorem with $\widehat{C} = \sqrt[\xi]{\frac{2C_1M_1}{\alpha}}$. □

Similar to the proof of Theorem 3.2.1, 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 matrix A to the LCP), which is weaker than that of the ξ-monotonicity for the function F (or a positive definite matrix for the matrix A to the LCP) used in [87]. Here, the details are omitted.

Theorem 3.2.2. *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 (3.1.1) and system (1.2.21), respectively. Then there exist constants $\widehat{C} > 0$ and $\xi > 1$, independent of x^ρ and ρ, such that*

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

Corollary 3.2.3. *For each $\rho > 0$, assume that the linear function $F : \mathbb{R}^n \to \mathbb{R}^n$ is $F(x) = Ax$ with the matrix $A \in \mathbb{R}^{n\times n}$ being a P-matrix. Let vectors x^* and x^ρ in $\mathbb{R}^n$ be the solutions of problem (3.1.1) and system (1.2.21), respectively. Then, there exists a constant $\widehat{C} > 0$, independent of x^ρ and ρ, such that*

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

Remark 3.2.3. *We note that the assumption of a P-matrix for matrix A is weaker than the assumption of a M-matrix used in [168] and the assumption of positive definiteness used in [87].*

Remark 3.2.4. *It has been proved in [121] 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 [40, Example 3.3.10], which verifies the conclusions in Remarks 3.2.3 and 3.2.4.

Example 3.2.5. *Let matrix*

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

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.

3.3 Numerical Algorithms

In this section, we describe specific algorithms to solve system (3.2.4). Instead of solving the box-constrained system (3.2.4) directly, we consider the corresponding least squares problem

$$\min_{x \in \Omega} \; \Psi(x, \rho) := \frac{1}{2}\|\mathcal{F}(x, \rho)\|^2. \tag{3.3.7}$$

The first-order necessary condition for a vector $x^\rho \in \Omega$ to be a local solution of problem (3.3.7) for given $\rho > 0$ is stated in the following proposition.

Proposition 3.3.1 ([8]). *For each $\rho > 0$, assume that x^ρ is a local solution of problem (3.3.7). Then*

$$\nabla\Psi(x^\rho, \rho)^T(x - x^\rho) \geq 0, \;\; \forall\, x \in \Omega, \tag{3.3.8}$$

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

Theorem 3.3.1 ([35]). *For each $\rho > 0$, a vector $x^\rho \in \mathbb{R}^n$ satisfies inequality (3.3.8) if and only if x^ρ is the solution of the following nonlinear system*

$$D(x)\nabla\Psi(x, \rho) = D(x)\nabla\mathcal{F}(x, \rho)^T\mathcal{F}(x, \rho) = 0,$$

where $\nabla\mathcal{F}$ is the Jacobian matrix of function $\mathcal{F}$ and D is the scaling matrix $D(x) := diag(|d_1(x)|, \ldots, |d_n(x)|)$ with

$$d_i(x) := \begin{cases} x_i, & \text{if } (\nabla\Psi(x, \rho))_i < 0, \\ 1, & \text{otherwise.} \end{cases}$$

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

$$\nabla\mathcal{F}(x, \rho) := \Theta(x) + \Pi(x, \rho)\nabla F(x), \tag{3.3.9}$$

where

$$\Theta(x) := \mathrm{diag}(F_1(x), \ldots, F_n(x)),\ \Pi(x,\rho) := \mathrm{diag}(G_1(x,\rho), \ldots, G_n(x,\rho))$$

are diagonal matrices, $\nabla F(x)$ is the Jacobian matrix of function F and $G_i(x,\rho) := x_i + \rho(1+\frac{1}{p})[F_i(x)]_+^{\frac{1}{p}}$ for all $i \in \mathcal{J}$.

In the following, we apply a trust-region Gauss-Newton method to solve the least squares problem (3.3.7) for given $\rho > 0$; more details can be found in [37, 110, 128]. At the k-th iteration, we consider a quadratic model $m^k(d,\rho)$ for $\Psi(x,\rho)$ at $x^k \in \Omega$ and replace the problem (3.3.7) by a trust-region problem

$$\min \quad m^k(d,\rho) \quad \text{s.t. } \|d\| \leq \Delta^k, \tag{3.3.10}$$

with the objective function

$$m^k(d,\rho) := \frac{1}{2}\|\mathcal{F}(x^k,\rho) + \nabla\mathcal{F}(x^k,\rho)d\|^2, \tag{3.3.11}$$

where Δ^k is the trust-region radius.

A formal description of the trust-region Gauss-Newton method for problem (3.3.7) for given $\rho > 0$ is presented as follows.

Algorithm 3.1: Trust-region Gauss-Newton method.

1 Input: $x^0 \in \Omega,\ \bar{\nu} > 0,\ \bar{\beta},\ \bar{\sigma},\ \bar{\gamma} \in (0,1),\ \Delta_0 \geq \Delta_{min} > 0,\ \beta_1,\ \beta_2,\ \hat{\mu} \in (0,1),\ \mu \geq 0,\ \rho,\ \epsilon_\rho > 0,$ $\hat{\epsilon} > 0$ and let $k := 0$;

2 **if** $\min\{\|D(x^k)\nabla\Psi(x^k,\rho)\|, \|P_\Omega(x^k - \nabla\Psi(x^k,\rho)) - x^k\|\} \leq \hat{\epsilon}\sqrt{n}$ *or* $\|\mathcal{F}(x^k,\rho)\| \leq \epsilon_\rho$ **then**

3   Stop;

4 **else**

5   *Compute the minimum-length solution* d_N^k *to the problem* $\min_d m^k(d,\rho)$. *Compute the generalized Cauchy step* d_C^k *as follows*:

$$d_C^k = \operatorname*{argmin}_{d \in \text{span}\{p^k\}} m^k(d,\rho) \quad \text{s.t. } \|d\| \leq \Delta^k,\ x^k + d \in \Omega.$$

    if $\|d_N^k\| \leq \Delta^k$ **then**

6   Set $d_{tr} = d_N^k$;

7   **else**

8   Find the dogleg step d_{tr} for (3.3.10).

9   **end**

10   *Let* $\bar{d}_{tr} = P_\Omega(x^k + d_{tr}) - x^k$.

11   **if**

$$\zeta_c(\bar{d}_{tr}) := \frac{m^k(0,\rho) - m^k(\bar{d}_{tr},\rho)}{m^k(0,\rho) - m^k(d_C^k,\rho)} \geq \beta_1$$

    then

12   Set $d^k = \bar{d}_{tr}^k$;

13   **else**

14   Find some scale $t \in (0,1]$ such that $d^k = td_C^k + (1-t)\bar{d}_{tr}$ satisfies $\zeta_c(d^k) \geq \beta_1$.

15   **end**

16   **if**

$$\zeta_\Psi(d^k) := \frac{\Psi(x^k,\rho) - \Psi(x^k + d^k,\rho)}{m^k(0,\rho) - m^k(d^k,\rho)} \geq \beta_2 \tag{3.3.12}$$

    then

17   Set $x^{k+1} = x^k + d^k$, if $\zeta_\Psi(d^k) \geq 0.75$ we set $\Delta^{k+1} = \max\{\Delta^k, 2\|d^k\|, \Delta_{min}\}$, otherwise we let $\Delta^{k+1} = \max\{\Delta^k, \Delta_{min}\}$, let $k := k+1$ and go to 2;

18   **else**

19   Set $\Delta^k = \min\{\frac{\Delta^k}{4}, \frac{\|d^k\|}{2}\}$ and go to 6;

20   **end**

21 **end**

In Algorithm 3.1, the function $\mathrm{P}_\Omega : \mathbb{R}^n \to \mathbb{R}$ denotes the projection operator defined by

$$\mathrm{P}_\Omega(x) := \{y \in \mathbb{R}^n \mid \|y - x\| = \min_{z \in \Omega} \|z - x\|\}.$$

We present a box-constrained differentiable penalty algorithm for problem (3.1.1). Before doing this, we define the termination criterion for this algorithm as follows

$$\text{Termination}(x) := \min\{\|[x]_+\|,\ \|[F(x)]_+\|, \|F(x) \circ x\|\} \le \epsilon, \tag{3.3.13}$$

where $\epsilon > 0$ is the tolerance parameter, which should be small enough, $F(x) \circ x$ denotes a vector with components $(F(x) \circ x)_i = F_i(x)x_i$, for all $i \in \mathcal{J}$. Now, a formal description of our algorithm for problem (3.1.1) is given as follows.

Algorithm 3.2: Box-constrained differentiable penalty method for the NCP.

Initializing $\rho^0 > 0,\ \rho^{min};\ \sigma > 1,\ \epsilon > 0$ and an initial point x^0 and let $i := 0$;
while $\rho^i > \rho^{min}$ **do**
  if $\text{Termination}(x^i) \le \epsilon$ **then**
    Stop;
  else
    Using Algorithm 3.1 *to solve problem (3.3.7) with starting point* x^i*, termination tolerance* ϵ_{ρ^i} *and penalty parameter* ρ^i*, we obtain* x^{i+1}*;*
  end
  Letting $\rho^{i+1} := \sigma\rho^i$ and $i := i + 1$;
end

3.3.1 Convergence Analysis

In this subsection, we establish the connection between solutions of the least squares problem (3.3.7) and solutions of problem (3.1.1).

Theorem 3.3.2. *Suppose that vector* $x^i \in \Omega$ *is the exact global solution of problem (3.3.7), and that* $\rho^i \to \infty$*. Then every limit point of the sequence* $\{x^i\}$ *is a solution of problem (3.1.1).*

Proof. Let x^* be the solution of problem (3.1.1). Then $\Psi(x^*, \rho) = 0$ for any $\rho > 0$. Since x^i is the exact global solution of problem (3.3.7) for given $\rho^i > 0$, we have

$\Psi(x^i, \rho^i) \leq \Psi(x^*, \rho^i)$, which means that $\Psi(x^i, \rho^i) = 0$. Specifically, we have

$$\frac{1}{2}\sum_{l=1}^{n}(x_l^i F_l(x^i))^2 + \rho^i\Big(\sum_{l=1}^{n} x_l^i [F_l(x^i)]_+^{q+1} + \frac{\rho^i}{2}\sum_{l=1}^{n}[F_l(x^i)]_+^{2q}\Big) = 0. \tag{3.3.14}$$

By rearranging this expression, we obtain

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

which means that

$$\sum_{l=1}^{n}[F_l(x^i)]_+^{2q} \leq -\frac{2}{\rho^i}\sum_{l=1}^{n} x_l^i[F_l(x^i)]_+^{q+1}. \tag{3.3.15}$$

Suppose that $\bar{x}$ is a limit point of the sequence $\{x^i\}$, so there is an infinite subsequence $\mathcal{K}$ such that $\bar{x} = \lim_{i\in\mathcal{K}} x^i \leq 0$, which implies $\bar{x} \in \Omega$. By taking the limit as $i \to \infty$, $i \in \mathcal{K}$, on both sides of (3.3.15), we have

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

where the last equality follows from $\rho^i \to \infty$. Therefore, we have $F_l(\bar{x}) \leq 0$ for all $l \in \mathcal{J}$. Moreover, by taking the limit as $i \to \infty$ for $i \in \mathcal{K}$ in (3.3.14), we have

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

where the last equality follows from (3.2.4) that $\rho^i x_l^i[F_l(x^i)]_+^{q+1} = -(x_l^i)^2[F_l(x^i)]_+^2$ for all $l \in \mathcal{J}$.

Therefore, we have proved that $\bar{x} \leq 0$, $F(\bar{x}) \leq 0$ and $\sum_{l=1}^{n}(\bar{x}_l F_l(\bar{x}))^2 = 0$, that is, $\bar{x}$ is a solution of problem (3.1.1). $\square$

Theorem 3.3.3. *Suppose that F is a uniform ξ-P-function and the set X^* is nonempty. Moreover, assume that x^ρ is a local solution of problem (3.3.7) for given $\rho > 0$ and* *satisfies $F(x^\rho) \le 0$. Then x^ρ is the solution of problem (3.1.1).*

Proof. Applying Proposition 3.3.1 at x^ρ for given penalty parameter $\rho > 0$, 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 as an explicit form via equality (3.3.9) as follows

$$\begin{cases} \big(\Theta(x^\rho)\mathcal{F}(x^\rho,\rho) + \nabla F(x^\rho)^T \Pi(x^\rho,\rho)\mathcal{F}(x^\rho,\rho)\big)_i = 0, & \text{if } x_i^\rho < 0, \\ \big(\Theta(x^\rho)\mathcal{F}(x^\rho,\rho) + \nabla F(x^\rho)^T \Pi(x^\rho,\rho)\mathcal{F}(x^\rho,\rho)\big)_i \le 0, & \text{if } x_i^\rho = 0. \end{cases} \tag{3.3.16}$$

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

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 $\mathcal{F}_1(x^\rho,\rho) = x_1^\rho F_1(x^\rho)$, we see that $F_1(x^\rho) \ne 0$ and $x_1^\rho \ne 0$. It follows from (3.3.16) that

$$\Big(\rho\Theta(x^\rho)\mathcal{F}(x^\rho,\rho) + \nabla F(x^\rho)^T \Pi(x^\rho,\rho)\mathcal{F}(x^\rho,\rho)\Big)_1 = 0. \tag{3.3.17}$$

Thus,

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

It follows from equality (3.3.17) that

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

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 have proved that $\mathcal{F}(x^\rho,\rho) = 0$, which means that x^ρ is the solution of problem (3.1.1). □

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 3.3.4. *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

$$\Psi(x^k,\rho) \leq \hat{\alpha}. \tag{3.3.18}$$

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

$$\begin{aligned} \alpha \sum_{i\in\mathcal{T}} (x_i^k)^\xi &= \alpha \|x^k - z^k\|^\xi \\ &\leq (x_\nu^k - z_\nu^k)(F_\nu(x^k) - F_\nu(z^k)) \\ &\leq \max_{i\in\mathcal{T}} x_i^k (F_i(x^k) - F_i(z^k)) \\ &= x_j^k (F_j(x^k) - F_j(z^k)) \\ &\leq |x_j^k||F_j(x^k) - F_j(z^k)|, \end{aligned} \tag{3.3.19}$$

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

$$\{|x_j^k|\} \to \infty. \tag{3.3.20}$$

Dividing by $|x_j^k|$ on both sides of inequality (3.3.19), we have

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

this, in turn, implies

$$\{|F_j(x^k)|\} \to \infty, \tag{3.3.21}$$

since $F_j(z^k)$ is bounded. However, (3.3.20) and (3.3.21) imply that $\{|\mathcal{F}_j(x^k,\rho)|\} \to \infty$, which contradicts with (3.3.18). $\square$

3.4 Numerical Experiments

In this section, we present numerical results of our Algorithms described in Section 3.3 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 [9].

We refer to the implementation of Algorithm 3.2 as the CDLOP method, which s-tands for the Constrained Differentiable Lower Order Penalty method. For convenience, we write the CDLOP method with $p = 2$ and 100 as the $\text{CDLOP}_{1/2}$ and $\text{CDLOP}_{1/100}$ methods, respectively. We first compare the performances of the $\text{DLOPP}_{1/2}$ method with the $\ell_{\frac{1}{2}}$-penalty method [169] and ℓ_1-penalty method [7] in terms of the number of function evaluations and the values of the penalty parameter ρ. Using the same terms, we compare the performances of the CDLOP method with different values of the power $p = 1, 2, 100, 1000, 5000, 10000$. Finally, based on the number of function evaluations, we compare the performance of our method with some well known approaches, such as the smooth approximation method [23, 25] and the nonsmooth equations method [93].

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 [169] 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 $\text{SLOP}_{1/2}$ method. The ℓ_1-penalty method employs the semismooth Newton method [146] 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. The implementation of Algorithm 3.1 is by virtue of a Matlab solver TRESNEI[1], which is a trust-region Gauss-Newton method developed by Morini and Porcelli [123] for bound-constrained (or unconstrained) nonlinear least squares problems. Furthermore, the solver TRESNEI is used to solve the corresponding least squares problems for the $\text{SLOP}_{1/2}$ and SSOOP_1 methods.

Throughout the experiments, we set parameters $\rho^0 = 1$, $\rho^{min} = 10^{16}$, $\sigma = 0.1$ and $\epsilon = 1.0e-6$ in Algorithm 3.2. We use $\hat{\epsilon} = 10^{-5}$ for the value of smoothing factor in

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

the $\mathrm{SLOP}_{1/2}$ method. We follow all default parameters in the solver TRESNEI. For example, we terminate the Algorithm 3.1 when the number of iteration or the number of function evaluations is over 1000. Other details can be found in [123]. We employ the performance profile introduced by Dolan and Moré [46] to present our numerical results. See Section 2.4 or [46] for more details regarding the performance profiles.

We select 22 test problems from MCPLIB shown in Table 3.1. 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 3.1: 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 3.1, the **Problem** denotes the name of test problem, the **Dim** denotes the dimension of problem (3.1.1) and the **Interval** denotes the interval in which a starting point is generated by a uniform distribution.

We present the numerical results as follows. We plot Figures 3.1-3.2 to compare the performance of the $\mathrm{CDLOP}_{1/2}$ method with the $\mathrm{SLOP}_{1/2}$ and SSOOP_1 methods in terms of the number of function evaluations and the values of the penalty parameter.

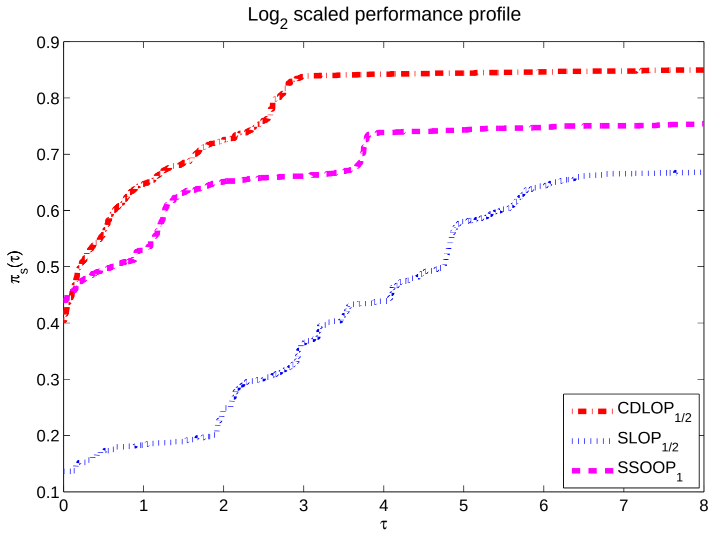

Figure 3.1: Performance profiles based on the number of function evaluations for the $\text{CDLOP}_{1/2}$, $\text{SLOP}_{1/2}$ and SSOOP_1 methods.

Figure 3.1 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 of the test problems successfully. The $\text{SLOP}_{1/2}$ method is the weakest solver among them.

The performance profiles in Figure 3.2 are plotted by the values of ρ. Figure 3.2 indicates the $\text{CDLOP}_{1/2}$ method can solve about 68% of the test problems with the smallest values of penalty parameter ρ. A small portion of the test problems can be solved by the SSOOP_1 method with the smallest penalty parameter ρ. However, the SSOOP_1 method is more robust than the $\text{SLLOP}_{1/2}$ method.

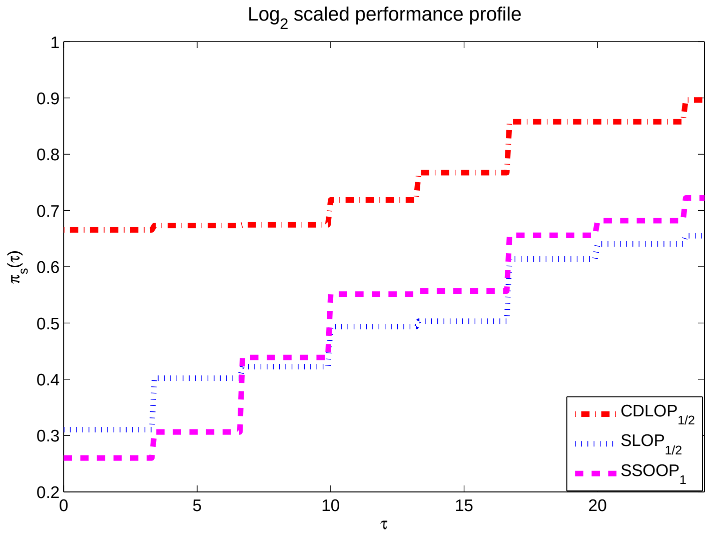

Figure 3.2: Performance profiles based on the values of the penalty parameter for the CDLOP$_{1/2}$, SLOP$_{1/2}$ and SSOOP$_1$ methods.

We plot Figures 3.3 and 3.4 to compare performance of the CDLOP method with different values of p in term of the number of function evaluations and the values of the penalty parameter.

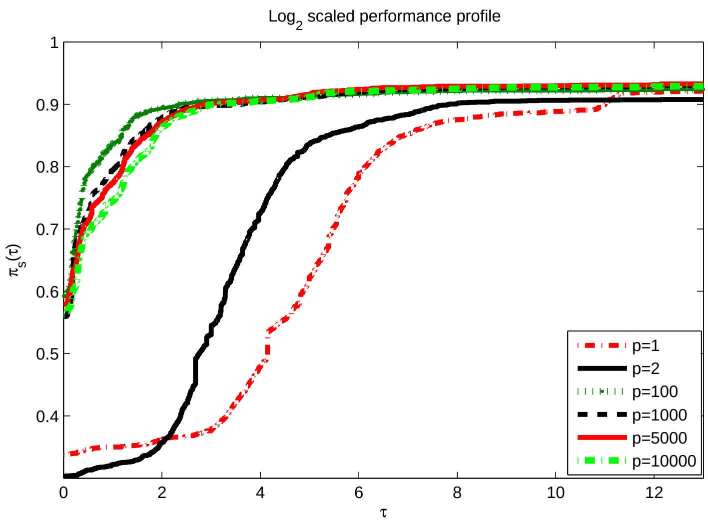

Figure 3.3: Performance profiles based on the number of function evaluations for the CDLOP method with the different p.

Figure 3.3 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 will 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.

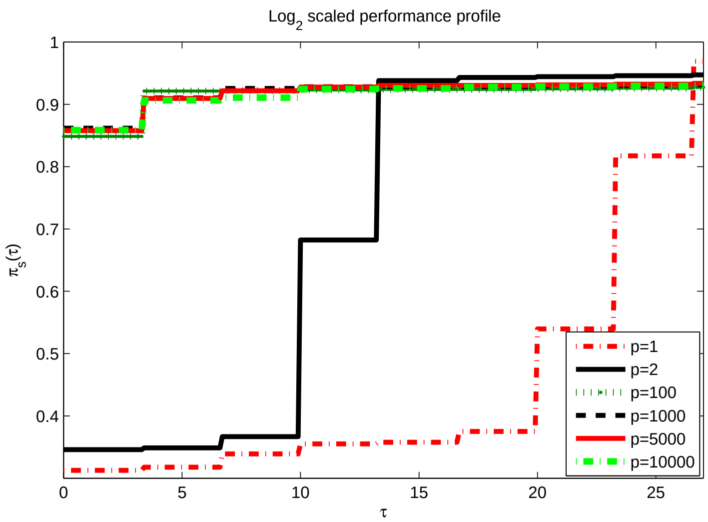

Figure 3.4: Performance profiles based on the values of the penalty parameter for the CDLOP method with the different p.

The performance profiles in Figure 3.4 are plotted by the values of ρ. Figure 3.4 indicates that the CDLOP method with $p = 1$ uses the smallest values of penalty parameter. smaller 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 3.2.1.

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 [183] is used in the smooth approximation method to smooth its normal equations. The nonsmooth equations method employs the semismooth Newton method [146] 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.

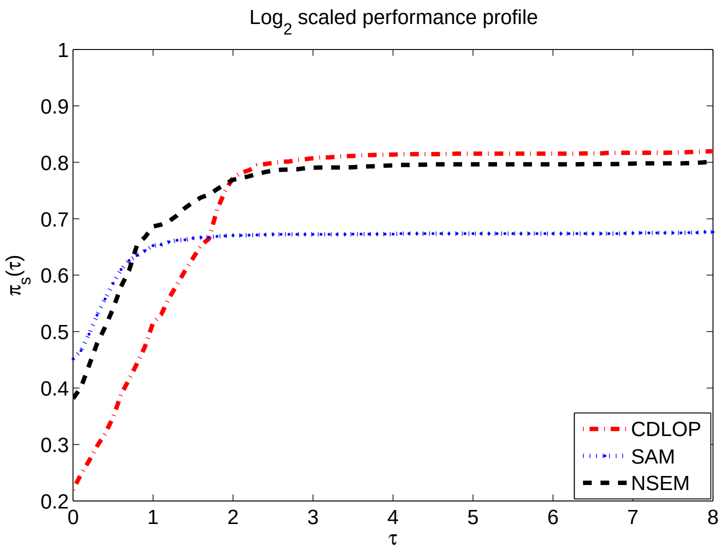

Figure 3.5: Performance profiles based on the number of function evaluations for the CDLOP method with $p = 100$, the SAM and NSEM methods.

Figure 3.5 indicates that the SAM method 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 method is more efficient than the SAM method. The CDLOP method with $p = 100$ can successfully solve the most test problems among them.

Next, we compare the performance of the proposed method with projection-type methods that have been studied widely for solving monotone linear and nonlinear variational inequalities, see [51, Chapter 12] and [84, 85, 158]. The extra-gradient method with modifications 1 and 2 in [84] (EGA$_{12}$, for short) is used to compare the performance of the CDLOP method with $p = 100$ in term of the number of function evaluations and the CPU time on the next test example. We take

$$F(x) = D(x) + Mx + q,$$

where $D(x)$ and $Mx+q$ are the nonlinear part and the linear part of $F(x)$, respectively. The matrix $M = A^T A + B$ where A is an $n \times n$ matrix whose entries are randomly generated in the interval $(-5, 5)$ and a skew-symmetric matrix B is generated in the same way. The vector q is generated from a uniform distribution in the interval $(-500, 500)$. In $D(x)$, the nonlinear part of $F(x)$, the components are $d_j * \arctan(x_j)$, where d_j is a random variable in $(0, 1)$. It is easy to see that the Jacobian matrix of F is positive semidefinite (not necessarily symmetric) and hence the problem is monotone. We test problems with dimension $n = 100, 200, 300$. All methods started at the same x^0 generated from a uniform distribution in the interval $(0, 10)$. To obtain more stable results, we run each test case 5 times. The average numbers of function evaluations and the computation times of these methods for problem with different sizes are given in the following table, where Dim denotes the dimension of problem, NF denotes the number of function evaluations and CPU denotes the CPU time.

Table 3.2: Numerical results for methods of EGA_{12} and CDLOP.

	EGA_{12}		CDLOP	
Dim	NF	CPU	NF	CPU
100	640	0.044	26	0.101
200	870	0.113	26	0.375
300	1017	0.525	30	1.376

Table 3.2 shows that more number of function evaluations is used by the projection-type method than that of the proposed method. However, the proposed method uses much CPU time. This is due to the fact that the CDLOP method needs to solve some linear equations of high dimensions, while the EGA_{12} method does not need to. However, the EGA_{12} method cannot be used to solve the complementarity problems without monotonicity, which can be solved efficiently by the proposed method CDLOP if they satisfy the assumption of the uniform ξ-P-function.

We plot the following Figures 3.6-3.9 in terms of the number of function evaluations to illustrate the sensitivity of the proposed algorithms' performance on the starting penalty parameter ρ^0, the rules of adjusting the penalty parameter ρ^i and the accuracy of solving the subproblems. Figure 3.6 describes the performance of the proposed method using different values of the starting penalty parameter $\rho^0 = 10^{-1}, 10^0, 10^1$ and 10^3, which indicates that the starting $\rho^0 = 1$ and $\rho^0 = 10$ make the proposed method more efficient and robust.

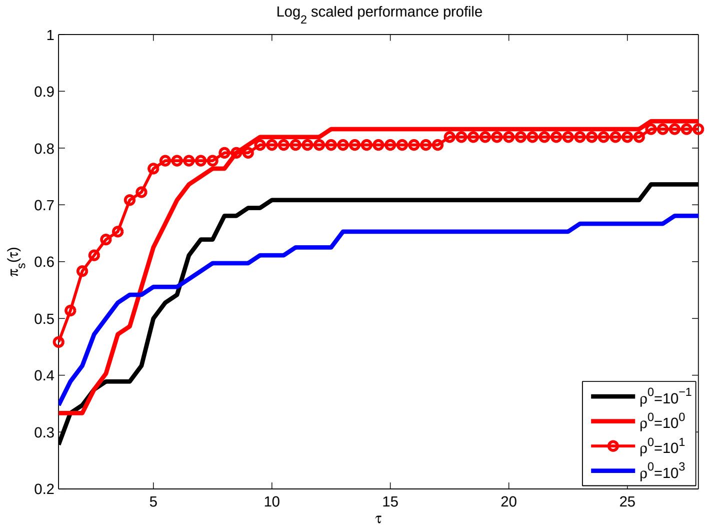

Figure 3.6: Performance profiles based on different values of the starting penalty parameter for the CDLOP method with $p = 100$.

Figure 3.7 is plotted by use of different values of $\sigma = 1/5, 1/10, 1/15$ and $1/25$ in Algorithm 3.2, which implies that the proposed method with the adjusting parameter $\sigma = 0.1$ is more efficient.

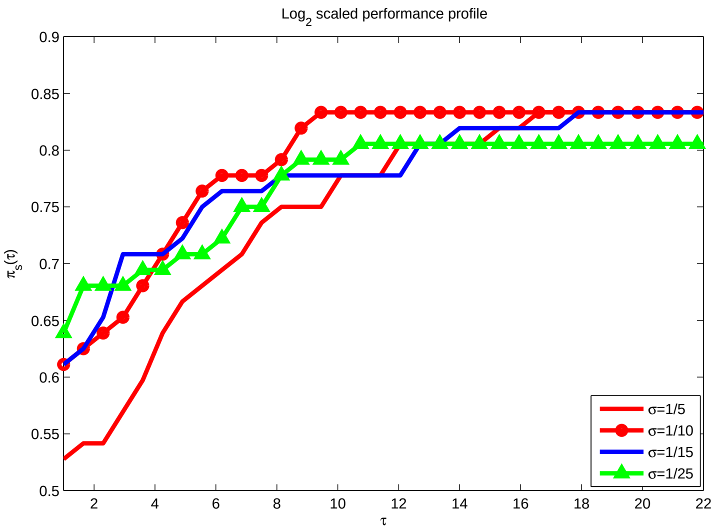

Figure 3.7: Performance profiles based on different rules of adjusting the penalty parameter for the CDLOP method with $p = 100$.

In order to test the sensitivity of the proposed method on the accuracy of solving the subproblems, we let $\epsilon_\rho := \max\{\nu\frac{1}{\rho}, 10^{-6}\}$ in Algorithm 3.1, where $\nu \geq 0$ is a parameter that determines the accuracy of solving the subproblems. We plot Figures 3.8 and 3.9 using different values of $\nu = 0, 0.1, 0.5$ and 1. Figure 3.8 shows that the proposed method with $p = 2$ use less number of function evaluations and is more robust if its subproblems are solved by some inexact rules. However, the performance profiles of Figure 3.9 show that the least number of function evaluations is used by the proposed method with $p = 100$ if the subproblems can be solved more accurately.

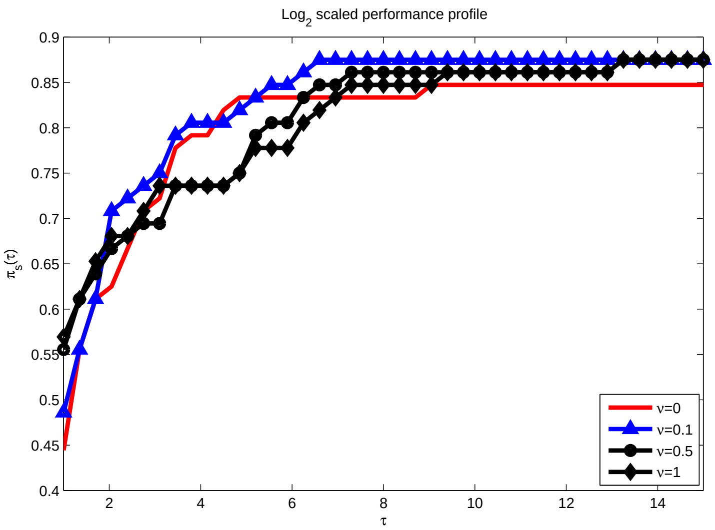

Figure 3.8: Performance profiles based on different accuracy of solving the subproblems for the CDLOP method with $p = 2$.

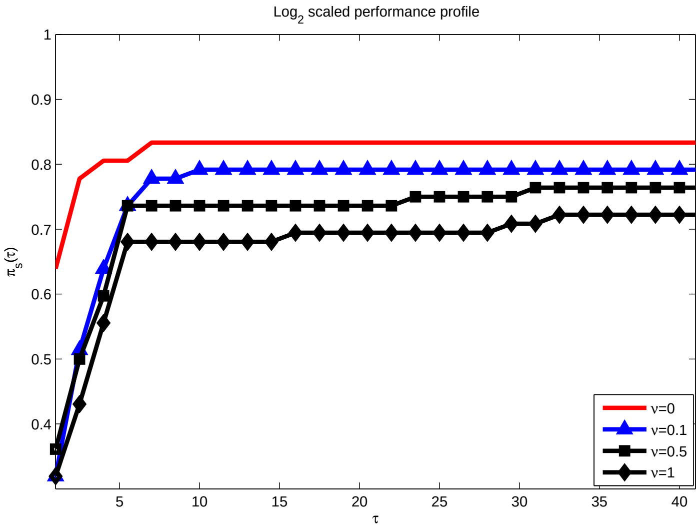

Figure 3.9: Performance profiles based on different accuracy of solving the subproblems for the CDLOP method with $p = 100$.

Chapter 4

An Unconstrained Differentiable Penalty Method for General Complementarity Problems

4.1 Introduction

In this chapter, we consider the GCP, which is to find a vector $x \in \mathbb{R}^n$ satisfying the following conditions,

$$H(x) \leq 0, \ F(x) \leq 0, \ H(x)^T F(x) = 0, \tag{4.1.1}$$

where the functions $H, \ F : \mathbb{R}^n \to \mathbb{R}^n$ are continuously differentiable. Throughout this chapter, we write $\mathcal{J} = \{1, 2, \ldots, n\}$ and use X^* to denote the solution set of problem (4.1.1), which is assumed to be nonempty. When the function $H(x) := x$, problem (4.1.1) reduces to the NCP studied in Chapter 3. Furthermore, problem (4.1.1) becomes a LCP if the function $H(x) := x$ and function $F(x)$ is linear, i.e., $F(x) := Ax - b$ for a given matrix $A \in \mathbb{R}^{n \times n}$ and a vector $b \in \mathbb{R}^n$.

An unconstrained minimization formulation for problem (4.1.1) was studied by

Tseng et al. [164] by virtue of the Mangasarian and Solodov's implicit Lagrangian function [113]. Kanzow and Fukushima [96] employed the Fischer's function [57] to transform problem (4.1.1) into an unconstrained minimization formulation. They presented mild conditions to guarantee that the global minima of the unconstrained problem coincide with the solutions of problem (4.1.1). The last unconstrained formulation was further investigated by Jiang et al. [92] and they proposed a trust region method for solving problem (4.1.1). The global convergence and local Q-superlinear convergence were established under a nonsingularity assumption.

In Chapter 3, we introduced a box-constrained differentiable penalty method for the NCP. However, it is not efficient to use this method to solve problem (4.1.1) directly as the corresponding constrained differentiable penalty problem has nonlinear constraints as the functions H and F are nonlinear. It is well known that optimization with nonlinear constraints are munch harder to solve in general than optimization problems with box constraints. We note the fact that problem (4.1.1) can be reformulated as a MiCP by virtue of artificial variables, which can be solved by the existing $\ell_{\frac{1}{p}}$-penalty method [88]. We can use the box-constrained differentiable penalty method to solve problem (4.1.1) by introducing artificial variables, which however doubles the number of nonlinear equations.

In this chapter, we introduce an unconstrained differentiable penalty method for problem (4.1.1) without introducing any artificial variables. Specifically, we will consider the system of unconstrained equations as follows:

$$\mathcal{G}(x,\rho) := H(x) \circ F(x) + \rho\big([H(x)]_+^{1+\frac{1}{p}} + [F(x)]_+^{1+\frac{1}{p}}\big) = 0, \tag{4.1.2}$$

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 $i \in \mathcal{J}$, 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 two vectors $H(x)$ and $F(x)$ with components $\big(H(x) \circ F(x)\big)_i = H_i(x)F_i(x)$, for all $i \in \mathcal{J}$. We note that the function $\mathcal{G}(x,\rho)$ is continuously differentiable for any $\rho > 0$ and $p \in [1,\infty)$. We establish that the solution x^ρ of system (4.1.2) converges to a solution x^* of problem (4.1.1) at a rate of $\mathcal{O}(\rho^{-\frac{p}{\xi}})$ under the assumption of a uniform ξ-P-function on functions

H and G, that is, there exists a constant $C > 0$ such that,

$$\|x^\rho - x^*\| \leq C\rho^{-\frac{p}{\xi}}. \tag{4.1.3}$$

Instead of solving the unconstrained equations (4.1.2) directly, we consider a unconstrained minimization problem that is solved by a trust-region Gauss-Newton method.

We carry out our numerical experiments on the same test problems used in Chapter 3. We set $p = 2$ in the unconstrained differentiable penalty method to compare its performance with the box-constrained differentiable penalty method with $p = 2$ and the ℓ_1-penalty method [7] 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 test the efficiency of the proposed method. Finally, we compare the performance of the proposed method with box-constrained differentiable penalty method in Chapter 3, the smooth approximation method [23], and the nonsmooth equations method [93] in terms of the function evaluations.

This chapter is organized as follows. In Section 4.2, we introduce an unconstrained differentiable penalty method for problem (4.1.1) and establish the convergence rate theorem for this method under the assumption of a uniform ξ-P-function. In the last section, we present a numerical implementation of the proposed method and detail our numerical results.

4.2 Unconstrained Differentiable Penalty Method

In this section, we establish that the solution of system (4.1.2) converges to a solution of problem (4.1.1) in the order of $\mathcal{O}(\rho^{-\frac{p}{\xi}})$ under the assumption of a uniform ξ-P-function on functions H and F.

Now, we reformulate problem (4.1.1) as a mixed complementarity problem by virtue of artificial variables as follows:

$$\begin{aligned} T(x,y) &= 0, \\ G(x,y) &\leq 0, \\ y^T G(x,y) &= 0, \\ y &\leq 0, \end{aligned} \tag{4.2.4}$$

where $T(x,y) := H(x) - y$ and $G(x,y) := F(x)$.

Proposition 4.2.1. *The vector $x^* \in \mathbb{R}^n$ is a solution of problem (4.1.1) if and only if there exists a vector $y^* \in \mathbb{R}^n$ satisfying $y^* = H(x^*)$ such that the vector $\binom{x^*}{y^*} \in \mathbb{R}^{2n}$ is a solution of problem (4.2.4).*

We consider a system of unconstrained equations as follows:

$$\mathcal{F}(x,y,\rho) := \begin{pmatrix} T(x,y) \\ G(x,y) \circ y \end{pmatrix} + \rho \begin{pmatrix} 0 \\ [y]_+^{1+\frac{1}{p}} + [G(x,y)]_+^{1+\frac{1}{p}} \end{pmatrix} = 0, \tag{4.2.5}$$

where $G(x,y) \circ y$ denotes a vector with components $\big(G(x,y) \circ y\big)_i = G_i(x,y)y_i$, for all $i \in \mathcal{J}$.

Proposition 4.2.2. *Given $\rho > 0$, the vector $x^\rho \in \mathbb{R}^n$ is a solution of system (4.1.2) if and only if there exists a vector $y^\rho \in \mathbb{R}^n$ satisfying $y^\rho = H(x^\rho)$ such that the vector $\binom{x^\rho}{y^\rho} \in \mathbb{R}^{2n}$ is a solution of system (4.2.5).*

Proof. We first assume that the vector x^ρ is a solution of system (4.1.2). Letting $y^\rho = H(x^\rho)$, we have $T(x^\rho, y^\rho) = H(x^\rho) - y^\rho = 0$. Consequently, we conclude that the vector $\binom{x^\rho}{y^\rho}$ is a solution of system (4.2.5). Now, we assume that the vector $\binom{x^\rho}{y^\rho}$ is a solution of system (4.2.5). Since $T(x^\rho, y^\rho) = 0$, we have $y^\rho = H(x^\rho)$. Therefore, the vector x^ρ is a solution of system (4.1.2). $\square$

The next example shows that the solution of system (4.1.2) is not unique in general, even if problem (4.1.1) has a unique solution.

Example 4.2.1. *Let $F(x) = x + 1$ and $H(x) = x$ with $x \in \mathbb{R}$. It is clear that $x^* = -1$ is the unique solution of problem (4.1.1). Take $p = 1$. Its unconstrained differentiable penalized equation is $x(x+1) + \rho\big([x]_+^2 + [x+1]_+^2\big) = 0$. Then $\bar{x}^\rho = -1$ and $\hat{x}^\rho = -\frac{\rho}{\rho+1}$ are two solutions of the last equation.*

Next we define the function $Z : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ by $Z(x,y) = \binom{T(x,y)}{G(x,y)}$. We prove that the solution $\binom{x^\rho}{y^\rho}$ of system (4.2.5) converges to a solution $\binom{x^*}{y^*}$ of problem (4.2.4) at a rate of $\mathcal{O}(\rho^{-\frac{p}{\xi}})$ under the assumption of a uniform ξ-P-function on function Z. Before doing this, we first show some lemmas.

Lemma 4.2.1. *For each $\rho > 0$, assume that the function Z is a uniform ξ-P-function and let the vector $\binom{x^\rho}{y^\rho}$ be a solution of system (4.2.5). Then there exists a constant $M > 0$, independent of $\binom{x^\rho}{y^\rho}$, ρ and p such that*

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

Proof. Since $\binom{x^\rho}{y^\rho}$ is a solution of system (4.2.5), it follows that $T(x^\rho, y^\rho) = 0$, and moreover $G_i(x^\rho, y^\rho) y_i^\rho + \rho\big([y_i^\rho]^{1+\frac{1}{p}} + [G_i(x^\rho, y^\rho)]_+^{1+\frac{1}{p}}\big) = 0$ for all $i \in \mathcal{J}$. Therefore, we have $T_i(x^\rho, y^\rho) x_i^\rho = 0$ and $G_i(x^\rho, y^\rho) y_i^\rho \leq 0$ for all $i \in \mathcal{J}$. By the uniform ξ-P-function assumption on function Z, there exist constants $\alpha > 0$ and $\xi > 1$ such that

$$\begin{aligned}
\alpha \left\| \begin{pmatrix} x^\rho \\ y^\rho \end{pmatrix} \right\|^\xi &\leq \max_{i \in \mathcal{J}} \begin{pmatrix} x_i^\rho (T_i(x^\rho, y^\rho) - T_i(0,0)) \\ y_i^\rho (G_i(x^\rho, y^\rho) - G_i(0,0)) \end{pmatrix} \\
&\leq \max_{i \in \mathcal{J}} \begin{pmatrix} -x_i^\rho T_i(0,0) \\ -y_i^\rho G_i(0,0) \end{pmatrix} \\
&\leq \left\| \begin{pmatrix} x^\rho \\ y^\rho \end{pmatrix} \right\| \|Z(0,0)\|_\infty.
\end{aligned}$$

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

Lemma 4.2.1 implies that the solution of problem (4.2.5) always lies in a bounded closed set for any $\rho > 0$. Assuming the continuity of function Z, we have that there exists a positive constant L, independent of $\binom{x^\rho}{y^\rho}$, ρ and p such that

$$\|Z(x^\rho, y^\rho)\| \leq L. \tag{4.2.6}$$

Lemma 4.2.2. *For each $\rho > 0$, assume that function Z is a uniform ξ-P-function and let the vector $\binom{x^\rho}{y^\rho}$ be a solution of system (4.2.5). Then there exist constants $C_1 > 0$ and $C_2 > 0$, independent of $\binom{x^\rho}{y^\rho}$ and ρ, such that*

$$\|[y^\rho]_+\| \leq C_1 \rho^{-p} \textit{ and } \|[G(x^\rho, y^\rho)]_+\| \leq C_2 \rho^{-p}.$$

Proof. Since $\binom{x^\rho}{y^\rho}$ is a solution of system (4.2.5), it follows that $T(x^\rho, y^\rho) = 0$, and moreover $G_i(x^\rho, y^\rho)y_i^\rho + \rho\big([y_i^\rho]^{1+\frac{1}{p}} + [G_i(x^\rho, y^\rho)]_+^{1+\frac{1}{p}}\big) = 0$ for all $i \in \mathcal{J}$. Then

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

for all $i \in \mathcal{J}$. We have $\|[y^\rho]_+\|_\infty \leq \rho^{-p}\|G(x^\rho, y^\rho)\|_\infty^p$. By the fact that all norms in $\mathbb{R}^n$ are equivalent, there exists a constant $\widetilde{C} > 0$ such that $\|[y^\rho]_+\| \leq \widetilde{C}\|[y^\rho]_+\|_\infty$. Combining inequality (4.2.6), we have that there exists a constant C_1 such that $\|[y^\rho]_+\| \leq C_1\rho^{-p}$ with $C_1 = \widetilde{C}L^p$. Similarly, there exists a constant C_2 such that $\|[G(x^\rho, y^\rho)]_+\| \leq C_2\rho^{-p}$ with $C_2 = \widetilde{C}M^p$. $\square$

Theorem 4.2.1. *For each $\rho > 0$, assume that function Z is a uniform ξ-P-function and let $\binom{x^*}{y^*}$ and $\binom{x^\rho}{y^\rho}$ be the solutions of problem (4.2.5) and system (4.1.2), respectively. 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 point $\binom{x^\rho}{y^\rho}$ as follows

$$\begin{aligned} y_a^\rho &= \{i \in \mathcal{J} \mid y_i^\rho = 0,\ G_i(x^\rho, y^\rho) > 0\}; \\ y_b^\rho &= \{i \in \mathcal{J} \mid y_i^\rho = 0,\ G_i(x^\rho, y^\rho) = 0\}; \\ y_c^\rho &= \{i \in \mathcal{J} \mid y_i^\rho = 0,\ G_i(x^\rho, y^\rho) < 0\}; \\ y_d^\rho &= \{i \in \mathcal{J} \mid y_i^\rho > 0,\ G_i(x^\rho, y^\rho) > 0\}; \\ y_e^\rho &= \{i \in \mathcal{J} \mid y_i^\rho > 0,\ G_i(x^\rho, y^\rho) = 0\}; \\ y_f^\rho &= \{i \in \mathcal{J} \mid y_i^\rho > 0,\ G_i(x^\rho, y^\rho) < 0\}; \\ y_g^\rho &= \{i \in \mathcal{J} \mid y_i^\rho < 0,\ G_i(x^\rho, y^\rho) > 0\}; \\ y_h^\rho &= \{i \in \mathcal{J} \mid y_i^\rho < 0,\ G_i(x^\rho, y^\rho) = 0\}; \\ y_s^\rho &= \{i \in \mathcal{J} \mid y_i^\rho < 0,\ G_i(x^\rho, y^\rho) < 0\}. \end{aligned}$$

Since $\binom{x^\rho}{y^\rho}$ is the solution of system (4.1.2), it follows that the sets y_a^ρ, y_d^ρ, y_e^ρ and y_s^ρ are empty. Let $\Lambda := y_b^\rho \cup y_c^\rho \cup y_f^\rho$ and $\Gamma := y_g^\rho \cup y_s^\rho$. Then $\mathcal{J} = \Lambda \cup \Gamma$. In the following, we first prove the inequality

$$\Big(y_i^* - y_i^\rho + [y_i^\rho]_+\Big)\Big(G_i(x^*,y^*) - G_i(x^\rho,y^\rho)\Big) = \Big(y_i^* + [y_i^\rho]_-\Big)\Big(G_i(x^*,y^*) - G_i(x^\rho,y^\rho)\Big) \le 0 \tag{4.2.7}$$

holds for $i \in \Lambda$.

(I) Let $i \in y_b^\rho$. Then

$$\Big(y_i^* + [y_i^\rho]_-\Big)\Big(G_i(x^*,y^*) - G_i(x^\rho,y^\rho)\Big) = y_i^* G_i(x^*,y^*) \le 0.$$

(II) Let $i \in y_c^\rho$. Then

$$\begin{aligned} & \Big(y_i^* + [y_i^\rho]_-\Big)\Big(G_i(x^*,y^*) - G_i(x^\rho,y^\rho)\Big) \\ = \;& y_i^* G_i(x^*,y^*) - y_i^* G_i(x^\rho,y^\rho) + [y_i^\rho]_- G_i(x^*,y^*) - [y_i^\rho]_- G_i(x^\rho,y^\rho) \\ = \;& -y_i^* G_i(x^\rho,y^\rho) \le 0. \end{aligned}$$

(III) Let $i \in y_f^\rho$. Then

$$\begin{aligned} & \Big(y_i^* + [y_i^\rho]_-\Big)\Big(G_i(x^*,y^*) - G_i(x^\rho,y^\rho)\Big) \\ = \;& y_i^* G_i(x^*,y^*) - y_i^* G_i(x^\rho,y^\rho) + [y_i^\rho]_- G_i(x^*,y^*) - [y_i^\rho]_- G_i(x^\rho,y^\rho) \\ = \;& -y_i^* G_i(x^\rho,y^\rho) \le 0. \end{aligned}$$

In the next, we prove that the inequality

$$\Big(y_i^* - y_i^\rho\Big)\Big(F_i(x^*,y^*) - G_i(x^\rho,y^\rho) + [G_i(x^\rho,y^\rho)]_+\Big) \le 0 \tag{4.2.8}$$

holds for all $i \in \Gamma$.

(I) Let $i \in y_g^\rho$. Then

$$\begin{aligned} & \Big(y_i^* - y_i^\rho\Big)\Big(G_i(x^*,y^*) + [G_i(x^\rho,y^\rho)]_-\Big) \\ = \;& y_i^* G_i(x^*,y^*) + y_i^*[G_i(x^\rho,y^\rho)]_- - y_i^\rho G_i(x^*,y^*) - y_i^\rho [G_i(x^\rho,y^\rho)]_- \\ = \;& -y_i^\rho G_i(x^*,y^*) \le 0. \end{aligned}$$

(II) Let $i \in y_h^\rho$. Then

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

Since $\binom{x^*}{y^*}$ solves problem (4.2.5) and $\binom{x^\rho}{y^\rho}$ is a solution of system (4.1.2), we have $T_i(x^*, y^*) = 0$ and $T_i(x^\rho, y^\rho) = 0$ for all $i \in \mathcal{J}$. Therefore, we have

$$
\begin{aligned}
& \max_{i\in\Lambda} \begin{pmatrix} (x_i^* - x_i^\rho)(T_i(x^*, y^*) - T_i(x^\rho, y^\rho)) \\ (y_i^* - y_i^\rho)(G_i(x^*, y^*) - G_i(x^\rho, y^\rho)) \end{pmatrix} \\
= \;& \max_{i\in\Lambda} \begin{pmatrix} 0 \\ (y_i^* - y_i^\rho)(G_i(x^*, y^*) - G_i(x^\rho - y^\rho)) \end{pmatrix} \\
\leq \;& \max_{i\in\Lambda} \begin{pmatrix} 0 \\ -[y_i^\rho]_+(G_i(x^*, y^*) - G_i(x^\rho - y^\rho)) \end{pmatrix} \\
\leq \;& \|[y^\rho]_+\| \|\big(G(x^*, y^*) - G(x^\rho, y^\rho)\big)\|_\infty \\
\leq \;& C_1 \rho^{-p} \|\big(G(x^*, y^*) - G(x^\rho, y^\rho)\big)\|_\infty \\
\leq \;& 2C_1 L \rho^{-p},
\end{aligned}
$$

where the first inequality comes from inequality (4.2.7) and the third inequality is from Lemma 4.2.2.

Furthermore, we have

$$
\begin{aligned}
& \max_{i\in\Gamma} \begin{pmatrix} (x_i^* - x_i^\rho)(T_i(x^*, y^*) - T_i(x^\rho, y^\rho)) \\ (y_i^* - y_i^\rho)(G_i(x^*, y^*) - G_i(x^\rho, y^\rho)) \end{pmatrix} \\
= \;& \max_{i\in\Gamma} \begin{pmatrix} 0 \\ (y_i^* - y_i^\rho)(G_i(x^*, y^*) - G_i(x^\rho - y^\rho)) \end{pmatrix} \\
\leq \;& \max_{i\in\Gamma} \begin{pmatrix} 0 \\ -[G_i(x^\rho, y^\rho)]_+(y_i^* - y_i^\rho) \end{pmatrix} \\
\leq \;& \|[G(x^\rho, y^\rho)]_+\| \|y^* - y^\rho\|_\infty \\
\leq \;& C_2 \rho^{-p} \|y^* - y^\rho\|_\infty \\
\leq \;& 2C_2 M_1 \rho^{-p},
\end{aligned}
$$

where the first inequality is from inequality (4.2.8) and the third inequality comes from Lemma 4.2.2.

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

$$
\begin{aligned}
& \alpha \left\| \begin{pmatrix} x^* \\ y^* \end{pmatrix} - \begin{pmatrix} x^\rho \\ y^\rho \end{pmatrix} \right\|^\xi \\
\leq \; & \max_{i \in \mathcal{J}} \begin{pmatrix} (x_i^* - x_i^\rho)(T_i(x^*, y^*) - T_i(x^\rho, y^\rho)) \\ (y_i^* - y_i^\rho)(G_i(x^*, y^*) - G_i(x^\rho, y^\rho)) \end{pmatrix} \\
= \; & \max_{i \in \Lambda \cup \Gamma} \begin{pmatrix} (x_i^* - x_i^\rho)(T_i(x^*, y^*) - T_i(x^\rho, y^\rho)) \\ (y_i^* - y_i^\rho)(G_i(x^*, y^*) - G_i(x^\rho, y^\rho)) \end{pmatrix} \\
\leq \; & \widehat{C} \rho^{-p}.
\end{aligned}
$$

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

Theorem 4.2.2. *For each $\rho > 0$, assume that functions H and F are uniform ξ-P-functions. Let x^* and x^ρ be the solutions of problem (4.1.1) and system (4.1.2), respectively. 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. Since x^* is a solution of problem (4.1.1), it follows from Proposition 4.2.1 that there exists y^* such that $\begin{pmatrix} x^* \\ y^* \end{pmatrix}$ is a solution of problem (4.2.4). Since x^ρ is a solution of system (4.1.2), it follows from Proposition 4.2.2 that there exists y^ρ such that $\begin{pmatrix} x^\rho \\ y^\rho \end{pmatrix}$ is a solution of system (4.2.5). Using Theorem 3.2.2, 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 $\widetilde{C}_1 = \max\left\{ \sqrt[\xi]{\frac{2C_1 L}{\alpha}}, \sqrt[\xi]{\frac{2C_2 M_1}{\alpha}} \right\}$. □

We consider a system of box-constrained nonlinear equations for problem (4.1.1) as follows:

$$
\mathcal{E}(x, y, \rho) := \begin{pmatrix} T(x, y) \\ G(x, y) \circ y \end{pmatrix} + \rho \begin{pmatrix} 0 \\ [G(x, y)]_+^{1+\frac{1}{p}} \end{pmatrix} = 0, \; y \in \Omega, \tag{4.2.9}
$$

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

Similar to the proof of Theorem 3.2.1, we establish the following convergence rate theorem for problem (4.1.1). Here, the details are omitted.

Theorem 4.2.3. *For each $\rho > 0$, assume that functions H and F are uniform ξ-P-functions. Let x^* and $\begin{pmatrix} x^\rho \\ y^\rho \end{pmatrix}$ be solutions of problem (4.1.1) and system (4.2.9), respectively. Then there exists a constant $\widehat{C} > 0$, independent of $\begin{pmatrix} x^\rho \\ y^\rho \end{pmatrix}$ and ρ, such* *that*

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

4.3 Numerical Algorithms and Experiments

In this section, we first present a numerical algorithm for problem (4.1.1) by virtue of the unconstrained differentiable penalty method. Then we use the same test problems described in Chapter 3 to compare the performance of our method with some existing methods in terms of the number of function evaluations and the values of the penalty parameter.

Instead of solving the penalized equations (4.2.5) directly, we consider the corresponding unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} \Psi(x, \rho) := \frac{1}{2}\|\mathcal{G}(x, \rho)\|^2. \tag{4.3.10}$$

For each $\rho > 0$, assume that $x^\rho \in \mathbb{R}^n$ is a local solution of problem (4.3.10). Then we have that x^ρ satisfies the next equations

$$\nabla\mathcal{G}(x, \rho)^T \mathcal{G}(x, \rho) = 0,$$

where $\nabla\mathcal{G}(x, \rho)$ is the Jacobian matrix of the function $\mathcal{G}(x, \rho)$, 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 functions $F(x)$ and $H(x)$, respectively, $\Theta(x, \rho) := \mathrm{diag}(G_1(x, \rho), \ldots, G_n(x, \rho))$ and $\Pi(x, \rho) := \mathrm{diag}(Q_1(x, \rho), \ldots, Q_n(x, \rho))$ are diagonal matrices with for all $i \in \mathcal{J}$,

$$G_i(x, \rho) := F_i(x) + \rho(1 + \frac{1}{p})[H_i(x)]_+^{\frac{1}{p}} \text{ and } Q_i(x, \rho) := H_i(x) + \rho(1 + \frac{1}{p})[F_i(x)]_+^{\frac{1}{p}}.$$

4.3.1 Convergence Analysis

In this subsection, we establish the connection between solutions of the unconstrained optimization problem (4.3.10) and that of problem (4.1.1).

Theorem 4.3.1. *Suppose that $x^i \in \mathbb{R}^n$ is a global solution of problem (4.3.10) for each $\rho^i > 0$ and that $\rho^i \to \infty$. Then every limit point of the sequence $\{x^i\}$ is a solution of problem (4.1.1).*

Proof. Let x^* be a solution of problem (4.1.1). Then we have $\Psi(x^*, \rho) = 0$ for each $\rho > 0$. Therefore, we have $\Psi(x^i, \rho^i) \leq \Psi(x^*, \rho^i) = 0$, which means that $\Psi(x^i, \rho^i) = 0$. Specifically, we have

$$\begin{aligned}&\frac{1}{2}\sum_{l=1}^{n}\big(H_l(x^i)^2F_l(x^i)^2 + (\rho^i)^2[H_l(x^i)]_+^{2+\frac{2}{p}} + (\rho^i)^2[F_l(x^i)]_+^{2+\frac{2}{p}}\big)\\ &+\sum_{l=1}^{n}\big(\rho^iF_l(x^i)[H_l(x^i)]_+^{2+\frac{1}{p}} + \rho^iH_l(x^i)[F_l(x^i)]_+^{2+\frac{1}{p}}\big)\\ &+\sum_{l=1}^{n}(\rho^i)^2[H_l(x^i)]_+^{1+\frac{1}{p}}[F_l(x^i)]_+^{1+\frac{1}{p}} = 0.\end{aligned} \tag{4.3.11}$$

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 above equation, we have

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

Therefore, we conclude that $F(\bar{x}) \leq 0$ and $H(\bar{x}) \leq 0$. It follows from (4.3.11) and take the limit as $i \xrightarrow{\mathcal{K}} \infty$ we have

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

where the second equality follows from (4.1.2) that, for all $l \in \mathcal{J}$,

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

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 (4.3.10) without the assumption of the convexity for the objective function $\Psi(x,\rho)$ for each $\rho > 0$. We mainly focus on the local solution of problem (4.3.10) in practice. In the next theorem, we prove that the local solution of problem (4.3.10) solves the general complementarity problem under the assumption of the uniform P-function on functions F and H that is strictly weaker than the assumption of the convexity for the objective function $\Psi(x,\rho)$ for each $\rho > 0$.

Theorem 4.3.2. *Suppose that the functions F and $H : \mathbb{R}^n \to \mathbb{R}^n$ are uniform P-functions. Moreover, assume that x^ρ is a local solution of problem (4.3.10) for each $\rho > 0$ and satisfies $F(x^\rho) \le 0$ and $H(x^\rho) \le 0$. Then x^ρ is a solution of problem (4.1.1).*

Proof. Since x^ρ is a local solution of problem (4.3.10) for given $\rho > 0$, we have $\big(\nabla\mathcal{G}(x,\rho)^T\mathcal{G}(x,\rho)\big)_i = 0$ for all $i \in \mathcal{J}$. Specifically, we have, for all $i \in \mathcal{J}$,

$$\big(\nabla H(x^\rho)^T\Theta(x^\rho,\rho)\mathcal{G}(x^\rho,\rho) + \nabla F(x^\rho)^T\Pi(x^\rho,\rho)\mathcal{G}(x^\rho,\rho)\big)_i = 0. \tag{4.3.12}$$

Since x^ρ satisfies $F(x^\rho) \le 0$ and $H(x^\rho) \le 0$, it follows that

$$\begin{aligned}
\mathcal{G}(x^\rho,\rho) &= (H_1(x^\rho)F_1(x^\rho),\ldots,H_n(x^\rho)F_n(x^\rho))^T,\\
\Theta(x^\rho,\rho) &= \mathrm{diag}(F_1(x^\rho),\ldots,F_n(x^\rho)),\\
\Pi(x^\rho,\rho) &= \mathrm{diag}(H_1(x^\rho),\ldots,H_n(x^\rho)).
\end{aligned}\tag{4.3.13}$$

We first prove that $\mathcal{G}(x^\rho,\rho) = 0$. Assume on the contrary that $\mathcal{G}(x^\rho,\rho) \neq 0$. Then there exists at least one index $i \in \mathcal{J}$ such that $\mathcal{G}_i(x^\rho,\rho) \neq 0$. Without loss of generality, we assume $\mathcal{G}_1(x^\rho,\rho) \neq 0$ and $\mathcal{G}_i(x^\rho,\rho) = 0$ for all $i = 2,\ldots,n$. It follows from $\mathcal{G}_1(x^\rho,\rho) \neq 0$ that we have $F_1(x^\rho) \neq 0$ and $H_1(x^\rho) \neq 0$. It follows from (4.3.13) that we have $\big(\Theta(x^\rho,\rho)\mathcal{G}(x^\rho,\rho)\big)_1 = H_1(x^\rho)F_1(x^\rho)^2 < 0$, $\big(\Pi(x^\rho,\rho)\mathcal{G}(x^\rho,\rho)\big)_1 = F_1(x^\rho)H_1(x^\rho)^2 < 0$, $\big(\Theta(x^\rho,\rho)\mathcal{G}(x^\rho,\rho)\big)_i = 0$ and $\big(\Pi(x^\rho,\rho)\mathcal{G}(x^\rho,\rho)\big)_i = 0$ for all $i = 2,\ldots,n$. Since the function F is a uniform P-function, it follows that there exists a constant $c > 0$ such that

$$\big(\Pi(x^\rho,\rho)\mathcal{G}(x^\rho,\rho)\big)_1\big(\nabla F(x^\rho)^T\Pi(x^\rho,\rho)\mathcal{G}(x^\rho,\rho)\big)_1 \geq c\|\Pi(x^\rho,\rho)\mathcal{G}(x^\rho,\rho)\|^2 > 0.$$

By (4.3.12), we have

$$\begin{aligned} & \big(\Theta(x^\rho,\rho)\mathcal{G}(x^\rho,\rho)\big)_1\big(\nabla H(x^\rho)^T\Theta(x^\rho,\rho)\mathcal{G}(x^\rho,\rho)\big)_1 \\ = & -\big(\Theta(x^\rho,\rho)\mathcal{G}(x^\rho,\rho)\big)_1\big(\nabla F(x^\rho)^T\Pi(x^\rho,\rho)\mathcal{G}(x^\rho,\rho)\big)_1 \\ = & -\frac{F_1(x^\rho)}{H_1(x^\rho)}\big(\Pi(x^\rho,\rho)\mathcal{G}(x^\rho,\rho)\big)_1\big(\nabla F(x^\rho)^T\Pi(x^\rho,\rho)\mathcal{G}(x^\rho,\rho)\big)_1 < 0. \end{aligned}$$

Therefore, we conclude that

$$\max_{1\leq i\leq n}\big(\Theta(x^\rho,\rho)\mathcal{G}(x^\rho,\rho)\big)_i\big(\nabla H(x^\rho)^T\Theta(x^\rho,\rho)\mathcal{G}(x^\rho,\rho)\big)_i = 0,$$

which contradicts the fact that $\max\limits_{1\leq i\leq n} z_i\big(\nabla H(x^\rho)^T z\big)_i \geq \bar{c}\|z\|^2$ for some constant $\bar{c} > 0$ and for all $z \in \mathbb{R}^n$ as the function H is a uniform P-function. Thus, we have proved that $\mathcal{G}(x^\rho,\rho) = 0$. Since $F(x^\rho) \leq 0$ and $H(x^\rho) \leq 0$, we conclude that x^ρ is a solution of problem (4.1.1). The proof is complete. $\square$

4.3.2 Numerical Algorithms

We apply a trust-region Gauss-Newton method to solve the unconstrained least squares problem (4.3.10) for each $\rho > 0$, see Algorithm 3.1 in Chapter 3. Before presenting our unconstrained differentiable penalty method for solving problem (4.1.1), we define the termination criterion for it as follows

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

where $\epsilon > 0$ is the tolerance parameter, which is set to be small enough, $F(x)\circ H(x)$ denotes a vector with components $(F(x)\circ H(x))_i = F_i(x)H_i(x)$, for all $i \in \mathcal{J}$. Now, a formal description of our algorithm for problem (4.1.1) is given as follows.

Algorithm 4.1: Unconstrained differentiable penalty method for the GCP.

Initializing $\rho^0 > 0,\ \rho^{min},\ \sigma > 1,\ \epsilon > 0$ and an initial point x^0 and let $i := 0$;
while $\rho^i > \rho^{min}$ **do**
    if Termination$(x^i) \leq \epsilon$ **then**
        Stop;
    else
        Using Algorithm 3.1 *to solve the unconstrained problem (4.3.10) with starting point* x^i*, termination tolerance* ϵ_{ρ^i} *and penalty parameter* ρ^i*, we obtain* x^{i+1}*;*
    end
    Letting $\rho^{i+1} := \sigma\rho^i$ and $i := i + 1$;
end

4.3.3 Numerical Experiments

In this subsection, we implement the Algorithm 4.1 with our code in MATLAB R2011b for the same test problems described in Table 3.1. We conduct 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 4.1 as the UDLOP method, which stands for the Unconstrained Differentiable Lower Order Penalty method. The same abbreviations for the existing methods used in Chapter 3 are rewritten in Table 4.1.

Table 4.1: Abbreviations for some existing methods.

CDLOP_1	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	semismooth one order penalty method
SAM	smooth approximate method
NSEM	nonsmooth equations 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. Throughout the experiments, all parameters are set the same as that in Chapter 3. The solver TRESNEI [123] is used to solve the corresponding least squares problem of every method. We employ the performance profile introduced by Dolan and Moré [46] to present our numerical results.

In the following, 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 penalty parameter.

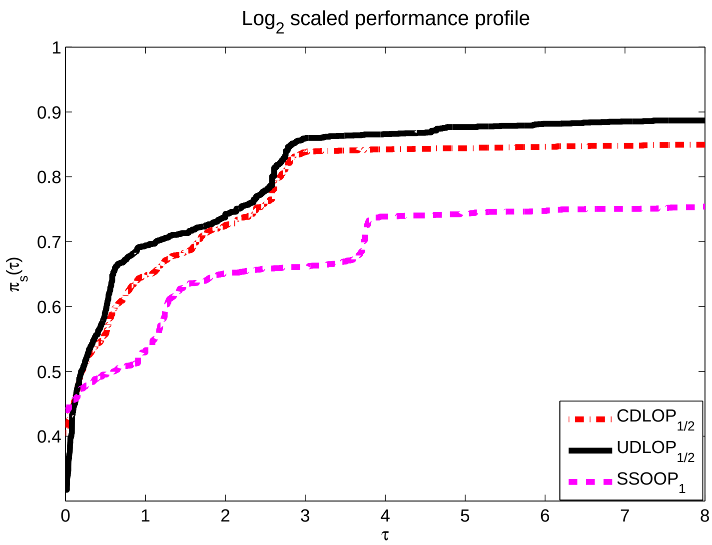

Figure 4.1: Performance profiles based on the number of function evaluations for the $\text{CDLOP}_{1/2}$, $\text{UDLOP}_{1/2}$ and SSOOP_1 methods.

Figure 4.1 indicates that the SSOOP_1 method solves about 47% test problems with the least number of function evaluations but this method is the weakest solver as it only can solve 80% test problems. The $\text{UDLOP}_{1/2}$ method is the most robust and can solve about 93% test problems.

We use the values of ρ to plot Figure 4.2, which shows that the SSOOP_1 method employs bigger values of penalty parameter than that of the $\text{CDLOP}_{1/2}$ method in order to achieve an approximate solution within the given accuracy.

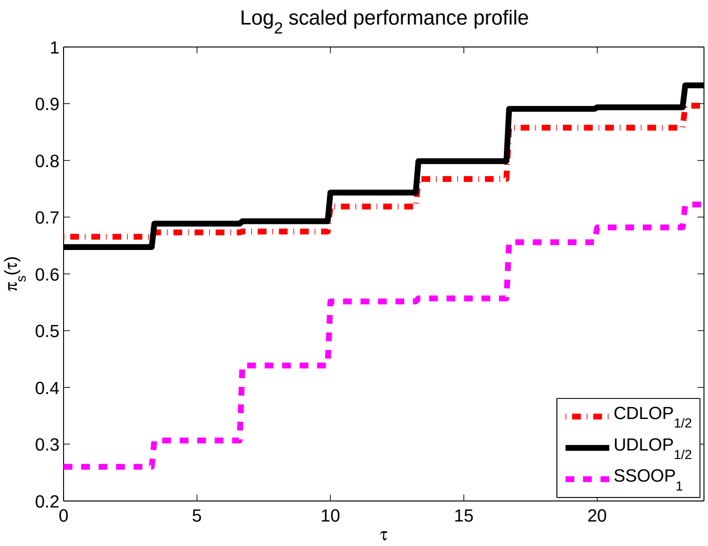

Figure 4.2: Performance profiles based on the values of the penalty parameter for the CDLOP$_{1/2}$, UDLOP$_{1/2}$ and SSOOP$_1$ methods.

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

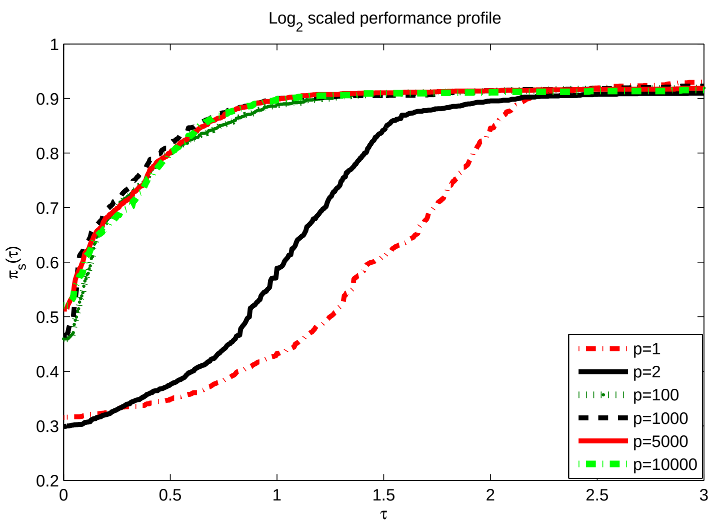

Figure 4.3: Performance profiles based on the number of function evaluations for the UDLOP method with different p.

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. Furthermore, the UDLOP method shares the almost same robustness for different power p.

We use the values of ρ to plot Figure 4.4, which indicates that the UDLOP_1 method is the weakest solver among them.

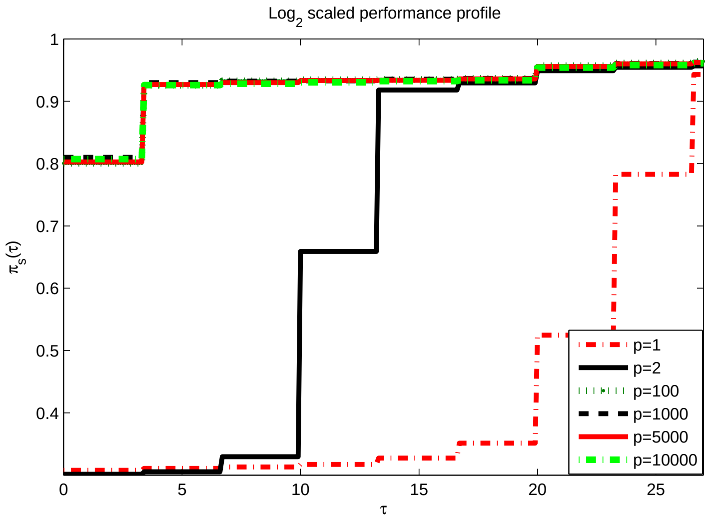

Figure 4.4: Performance profiles based on the values of the penalty parameter for the UDLOP method with different p.

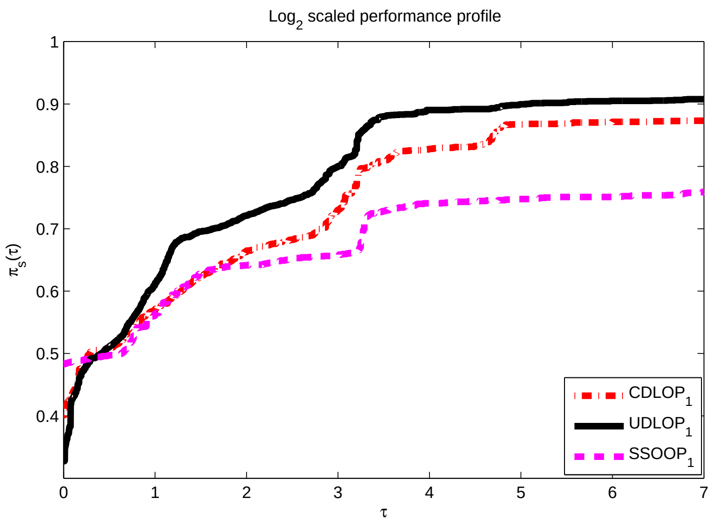

Figure 4.5: Performance profiles based on the number of function evaluations for the CDLOP_1, UDLOP_1 and SSOOP_1 methods.

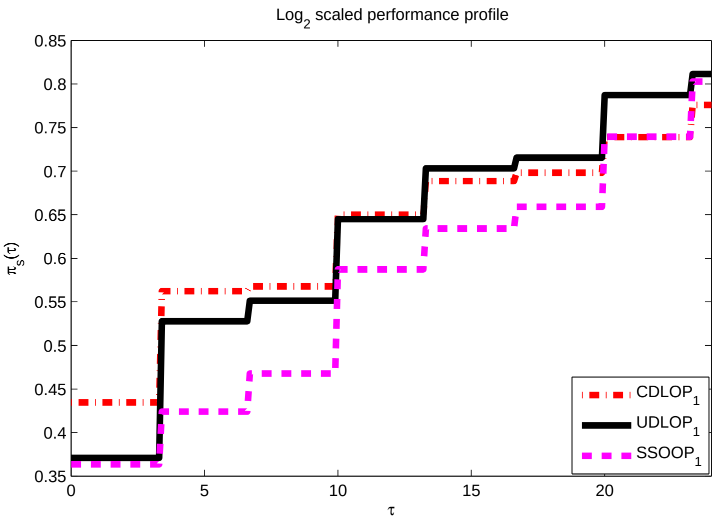

Figure 4.6: Performance profiles based on the values of the penalty parameter for the CDLOP_1, UDLOP_1 and SSOOP_1 methods.

Figures 4.5 and 4.6 indicate that the UDLOP_1 method performs better than the CDLOP_1 method and the SSOOP_1 method is the weakest solver among them.

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 smooth approximation method and the nonsmooth equations method.

Figure 4.7 indicates that the SAM method 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.

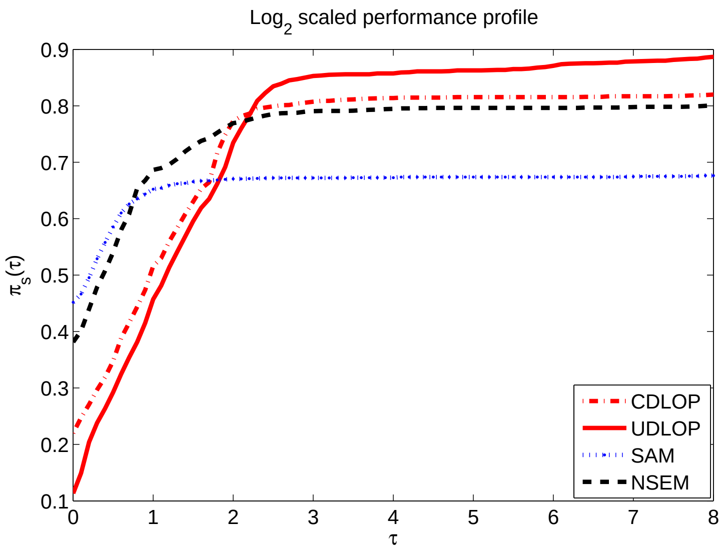

Figure 4.7: Performance profiles based on the number of function evaluations for the CDLOP$_{1/100}$, UDLOP$_{1/100}$, SAM and NSEM methods.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

In this thesis, we aimed at designing efficient algorithms for inequality constrained nonlinear programming problems and complementarity problems by virtue of $\ell_{\frac{1}{p}}(p > 1)$-penalty functions. A technique of the p-order relaxation was used to the nonconvex and non-Lipschitzian $\ell_{\frac{1}{p}}$-penalty function. Combining with an interior-point method, we proposed an interior-point $\ell_{\frac{1}{p}}$-penalty method to solve inequality constrained nonlinear programming problems. We introduce different kinds of constraint qualifications to establish first-order necessary conditions for the relaxed problem. We employed the modified Newton method to solve a sequence of logarithmic barrier subproblems and detailed three reliable algorithms which constitute the interior-point $\ell_{\frac{1}{p}}$-penalty method and established the global convergence of the proposed method under mild conditions. Specifically, we proved that the iteration sequence generated by the interior-point $\ell_{\frac{1}{p}}$-penalty method converges to some KKT (or FJ) point of original problem. Preliminary numerical experiments have been done, which show that the interior-point $\ell_{\frac{1}{2}}$-penalty method is competitive with other existing interior-point ℓ_1-penalty method in terms of iteration numbers and better when comparing the number of updating the penalty parameter and the relative error.

Furthermore, we proposed box-constrained and unconstrained differentiable penalty methods for complementarity problems and established their convergence rate between the solution of original problem and that of differentiable penalized equations under the assumption of a unform ξ-P-function. Our methods not only inherit the convergence rate of the existing $\ell_{\frac{1}{p}}$-penalty method but also overcome the shortcoming of the non-Lipschitzianness of the $\ell_{\frac{1}{p}}$-penalized term. Instead of solving differentiable penalized equations directly, we solved a corresponding least squares problem by the trust-region Gauss-Newton method. Numerical experiments were carried out on the test problems from MCPLIB, and numerical results showed that the differentiable $\ell_{\frac{1}{2}}$-penalty methods are more efficient than both the smoothing $\ell_{\frac{1}{2}}$-penalty method and the ℓ_1-penalty method in terms of the number of function evaluations and the values of the penalty parameter.

5.2 Future Work

We believe that our methods proposed in this thesis open a leaf of window to examine the non-Lipschitzian $\ell_{\frac{1}{p}}$-penalty function from the point of view of numerical implementation. However, there are many other issues that are needed to deal with in the future work. We summarize them as follows.

(I) As pointed out by Fletcher [63] that the strategy of updating the penalty parameter plays a central role in the numerical implementation for penalty methods, some adaptive strategies have been introduced in [17, 20] to update the penalty parameter for the ℓ_1-penalty method. It is well-known that the smallest exact penalty parameter of the $\ell_{\frac{1}{p}}$-exact penalty function is smaller than that of the ℓ_1-exact penalty function. However, a precise criterion for adjustment of the penalty parameter in the numerical implementation of the $\ell_{\frac{1}{p}}$-penalty method has not been studied in the thesis.

(II) We have run both the interior-point $\ell_{\frac{1}{2}}$-penalty method and two interior-point ℓ_1-penalty methods developed by Curtis [42] with the same stopping criterion on the set of 38 test problems with degenerate constraints and with the same starting point. Our numerical results showed that the interior-point $\ell_{\frac{1}{2}}$-penalty method can find a local minimum more accurately than that of the interior-point ℓ_1-penalty methods. However, our numerical findings are lack of the theoretical justification.

(III) Our interior-point $\ell_{\frac{1}{p}}$-penalty methods are only efficient to solve inequality constrained optimization problems. It is possible that we can utilize artificial variables to transform all equality constraints into inequality constraints to convert the optimization problem with equality and inequality constraints into an optimization problem with only inequality constraints, which can be solved by interior-point $\ell_{\frac{1}{p}}$-penalty methods. Following the procedure above, we have conducted numerical experiments, whose results show that the interior-point $\ell_{\frac{1}{p}}$-penalty method lack efficiency for optimization problems with equality constraints. We will combine the techniques of augmented Lagrangian and interior-point $\ell_{\frac{1}{p}}$-penalization to tackle the equality and inequality constraints, respectively.

(IV) Recently, second-order cone complementarity problems [27, 70, 161] have received a great deal of attention. However, there are few numerical algorithms that can solve these problems efficiently, especially for large scale problems. We will apply our differentiable penalty methods to design efficient numerical algorithms for solving second order cone complementarity problems.

Bibliography

[1] J. Abadie. On the Kuhn-Tucker theorem. *In: J. Abadie, eds., Nonlinear Programming (North-Holland, Amsterdam)*, pages 19–36, 1966.

[2] R. Andreani, A. Friedlander, and S. A. Santos. On the resolution of the generalized nonlinear complementarity problem. *SIAM Journal on Optimization*, 12(2):303–321, 2002.

[3] M. Anitescu. Global convergence of an elastic mode approach for a class of mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 16(1):120–145, 2005.

[4] L. Armijo. Minimization of functions having Lipschitz continuous first partial derivatives. *Pacific Journal of Mathematics*, 16(1):1–3, 1966.

[5] H. Y. Benson, A. Sen, and D. F. Shanno. *Interior-point methods for nonconvex nonlinear programming: Convergence analysis and computational performance.* http://rutcor.rutgers.edu/~shanno/converge5.pdf, 2009.

[6] H. Y. Benson, D. F. Shanno, and R. J. Vanderbei. Interior-point methods for nonconvex nonlinear programming: Jamming and numerical testing. *Mathematical Programming*, 99(1):35–48, 2004.

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

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

[9] 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.

[10] J. M. Borwein. Stability and regular points of inequality systems. *Journal of Optimization Theory and Applications*, 48(1):9–52, 1986.

[11] J. V. Burke. Calmness and exact penalization. *SIAM Journal on Control and Optimization*, 29(2):493–497, 1991.

[12] J. V. Burke. An exact penalization viewpoint of constrained optimization. *SIAM Journal on Control and Optimization*, 29(4):968–998, 1991.

[13] J. V. Burke, A. S. Lewis, and M. L. Overton. A robust gradient sampling algorithm for nonsmooth, nonconvex optimization. *SIAM Journal on Optimization*, 15(3):751–779, 2005.

[14] R. H. Byrd, J. C. Gilbert, and J. Nocedal. A trust region method based on interior point techniques for nonlinear programming. *Mathematical Programming*, 89(1):149–185, 2000.

[15] R. H. Byrd, M. E. Hribar, and J. Nocedal. An interior point algorithm for large-scale nonlinear programming. *SIAM Journal on Optimization*, 9(4):877–900, 1999.

[16] R. H. Byrd, G. H. Liu, and J. Nocedal. On the local behavior of an interior point method for nonlinear programming. *In: D.F. Griffiiths and D. J. Higham, eds., Numerical Analysis (Addison-Wesley Longman)*, pages 37–56, 1997.

[17] R. H. Byrd, G. Lopez-Calva, and J. Nocedal. A line search exact penalty method using steering rules. *Mathematical Programming*, 133(1-2):39–73, 2012.

[18] R. H. Byrd, M. Marazzi, and J. Nocedal. On the convergence of Newton iterations to non-stationary points. *Mathematical Programming*, 99(1):127–148, 2004.

[19] R. H. Byrd, J. Nocedal, and R. A. Waltz. Feasible interior methods using slacks for nonlinear optimization. *Computational Optimization and Applications*, 26(1):35–61, 2003.

[20] R. H. Byrd, J. Nocedal, and R. A. Waltz. Steering exact penalty methods for nonlinear programming. *Optimization Methods and Software*, 23(2):197–213, 2008.

[21] R. M. Chamberlain, M. J. D. Powell, C. Lemarechal, and H. C. Pedersen. The watchdog technique for forcing convergence in algorithms for constrained optimization. *Algorithms for Constrained Minimization of Smooth Nonlinear Functions*, 16:1–17, 1982.

[22] B. Chen, X. Chen, and C. Kanzow. A penalized Fischer-Burmeister NCP-function. *Mathematical Programming*, 88(1):211–216, 2000.

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

[24] B. T. Chen, X. J. Chen, and C. Kanzow. *A penalized Fischer-Burmeister NCP-function: Theoretical investigation and numerical results.* Technical Report, Internaltional Symposium on Mathematical Programming in Lausanne, Switzerland, 1997.

[25] 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.

[26] J. S. Chen. On some NCP-functions based on the generalized Fischer–Burmeister function. *Asia-Pacific Journal of Operational Research*, 24(3):401–420, 2007.

[27] J. S. Chen, X. Chen, and P. Tseng. Analysis of nonsmooth vector-valued functions associated with second-order cones. *Mathematical Programming*, 101(1):95–117, 2004.

[28] J. S. Chen, Z. H. Huang, and C. Y. She. A new class of penalized NCP-functions and its properties. *Computational Optimization and Applications*, 50(1):49–73, 2011.

[29] J. S. Chen and S. Pan. A family of NCP functions and a descent method for the nonlinear complementarity problem. *Computational Optimization and Applications*, 40(3):389–404, 2008.

[30] L. Chen and D. Goldfarb. Interior-point ℓ_2-penalty methods for nonlinear programming with strong global convergence properties. *Mathematical Programming*, 108(1):1–36, 2006.

[31] L. Chen and D. Goldfarb. An interior-point piecewise linear penalty method for nonlinear programming. *Mathematical Programming*, 128(1):73–122, 2011.

[32] X. J. Chen. Smoothing methods for nonsmooth, nonconvex minimization. *Mathematical Programming*, 134(1):71–99, 2012.

[33] F. H. Clarke. A new approach to Lagrange multipliers. *Mathematics of Operations Research*, 1(2):165–174, 1976.

[34] F. H. Clarke. *Optimization and Nonsmooth Analysis*. SIAM, 1990.

[35] T. F. Coleman and Y.Y. Li. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization*, 6(2):418–445, 1996.

[36] A. R. Conn, N. I. M. Gould, D. Orban, and P. L. Toint. A primal-dual trust-region algorithm for non-convex nonlinear programming. *Mathematical Programming*, 87(2):215–249, 2000.

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

[38] A. R. Conn and T. Pietrzykowski. A penalty function method converging directly to a constrained optimum. *SIAM Journal on Numerical Analysis*, 14(2):348–375, 1977.

[39] R. W. Cottle. Nonlinear programs with positively bounded Jacobians. *SIAM Journal on Applied Mathematics*, 14(1):147–158, 1966.

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

[41] R. Courant. Variational methods for the solution of problems of equilibrium and vibrations. *Bulletin of the American Mathematical Society*, 49(1):1–23, 1943.

[42] F. E. Curtis. A penalty-interior-point algorithm for nonlinear constrained optimization. *Mathematical Programming Computation*, 4(2):181–209, 2012.

[43] 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.

[44] Y. d′Halluin, P. A. Forsyth, and G. Labahn. A penalty method for American options with jump diffusion processes. *Numerische Mathematik*, 97(2):321–352, 2004.

[45] S. P. Dirkse and M. C. Ferris. MCPLIB: A collection of nonlinear mixed complementarity problems. *Optimization Methods and Software*, 5(4):319–345, 1995.

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

[47] C. Durazzi. On the Newton interior-point method for nonlinear programming problems. *Journal of Optimization Theory and Applications*, 104(1):73–90, 2000.

[48] A. S. El-Bakry, R. A. Tapia, T. Tsuchiya, and Y. Zhang. On the formulation and theory of the Newton interior-point method for nonlinear programming. *Journal of Optimization Theory and Applications*, 89(3):507–541, 1996.

[49] F. Facchinei and C. Kanzow. On unconstrained and constrained stationary points of the implicit Lagrangian. *Journal of Optimization Theory and Applications*, 92(1):99–115, 1997.

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

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

[52] F. Facchinei and J. Soares. A new merit function for nonlinear complementarity problems and a related algorithm. *SIAM Journal on Optimization*, 7(1):225–247, 1997.

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

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

[55] M. C. Ferris and D. Ralph. Projected gradient methods for nonlinear complementarity problems via normal maps. *In: D. Z. Du, L. Qi and R. S. Womersley, eds., Recent Advances in Nonsmooth Optimization(World Scientific Publishing)*, pages 57–87, 1995.

[56] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques.* SIAM, 1990.

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

[58] A. Fischer. Solution of monotone complementarity problems with locally Lipschitzian functions. *Mathematical Programming*, 76(3):513–532, 1997.

[59] A. Fischer. New constrained optimization reformulation of complementarity problems. *Journal of Optimization Theory and Applications*, 97(1):105–117, 1998.

[60] A. Fischer and H. Y. Jiang. Merit functions for complementarity and related problems: A survey. *Computational Optimization and Applications*, 17(2):159–182, 2000.

[61] R. Fletcher. A model algorithm for composite nondifferentiable optimization problems. *Nondifferential and Variational Techniques in Optimization*, 17:67–76, 1982.

[62] R. Fletcher. Second order corrections for non-differentiable optimization. *Numerical Analysis, Lecture Notes in Mathematics*, 912:85–114, 1982.

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

[64] R. Fletcher and S. Leyffer. Nonlinear programming without a penalty function. *Mathematical Programming*, 91(2):239–269, 2002.

[65] P. A. Forsyth and K. R. Vetzal. Quadratic convergence of a penalty method for valuing American options. *SIAM Journal on Scientific Computation*, 23(6):2095–2122, 2002.

[66] K. R. Frisch. *The logarithmic potential method of convex programming.* Unpublished manuscript, Instituation of Economics Memorandum, Oslo University, 1955.

[67] J. Fritz. Extremum problems with inequalities as side conditions. *Studies and Essays, Courant Anniversary Volume*, pages 187–204, 1948.

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

[69] M. Fukushima. Merit functions for variational inequality and complementarity problems. *In: G. DiPillo and F. Giannessi, eds., Nonlinear Optimization and Applications (Plenum Press, New York)*, pages 155–170, 1996.

[70] M. Fukushima, Z. Q. Luo, and P. Tseng. Smoothing functions for second-order-cone complementarity problems. *SIAM Journal on Optimization*, 12(2):436–460, 2002.

[71] A. Galántai. Properties and construction of NCP functions. *Computational Optimization and Applications*, 52(3):805–824, 2012.

[72] J. Gauvin. A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming. *Mathematical Programming*, 12(1):136–138, 1977.

[73] C. Geiger and C. Kanzow. On the resolution of monotone complementarity problems. *Computational Optimization and Applications*, 5(2):155–173, 1996.

[74] P. E. Gill, W. Murray, and M. H. Wright. *Practical Optimization.* Academic Press, 1981.

[75] F.J. Gould and J. W. Tolle. A necessary and sufficient qualification for constrained optimization. *SIAM Journal on Applied Mathematics*, 20(2):164–172, 1971.

[76] N. I. M. Gould, P. L. Toint, and D. Orban. *An interior-point ℓ_1-penalty method for nonlinear optimization.* Groupe d'études et de recherche en analyse des décisions, 2010.

[77] L. Grippo, F. Lampariello, and S. Lucidi. A nonmonotone line search technique for newton's method. *SIAM Journal on Numerical Analysis*, 23(4):707–716, 1986.

[78] I. Griva, D. F. Shanno, and R. J. Vanderbei. Convergence analysis of a primal-dual interior-point method for nonlinear programming. *Optimization Online*, 2004.

[79] M. Guignard. Generalized Kuhn-Tucker conditions for mathematical programming problems in a Banach space. *SIAM Journal on Control*, 7(2):232–241, 1969.

[80] S. P. Han and O. L. Mangasarian. Exact penalty functions in nonlinear programming. *Mathematical Programming*, 17(1):251–269, 1979.

[81] S. P. Han, J. S. Pang, and N. Rangaraj. Globally convergent Newton methods for nonsmooth equations. *Mathematics of Operations Research*, 17(3):586–607, 1992.

[82] 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.

[83] P. T. Harker and B. Xiao. Newton's method for the nonlinear complementarity problem: A B-differentiable equation approach. *Mathematical Programming*, 48(1):339–357, 1990.

[84] B. S. He, X. M. Yuan, and J. Z. Zhang. Comparison of two kinds of prediction-correction methods for monotone variational inequalities. *Computational Optimization and Applications*, 27(3):247–267, 2004.

[85] B.S. He and L.Z. Liao. Improvements of some projection methods for monotone nonlinear variational inequalities. *Journal of Optimization Theory and Applications*, 112(1):111–128, 2002.

[86] J. B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms: Part 1: Fundamentals*. Springer, 1996.

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

[88] 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.

[89] 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.

[90] A. F. Izmailov and M. V. Solodov. Examples of dual behaviour of Newton-type methods on optimization problems with degenerate constraints. *Computational Optimization and Applications*, 42(2):231–264, 2009.

[91] H. Y. Jiang. *Smoothed Fischer-Burmeister equation methods for the complementarity problem.* Department of Mathematics, The University of Melbourne, Australia, 1997.

[92] 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.

[93] 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.

[94] C. Kanzow. Some equation-based methods for the nonlinear complementarity problem. *Optimization Methods and Software*, 3(4):327–340, 1994.

[95] C. Kanzow. Nonlinear complementarity as unconstrained optimization. *Journal of Optimization Theory and Applications*, 88(1):139–155, 1996.

[96] 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.

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

[98] 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.

[99] W. Karush. *Minima of functions of several variables with inequalities as side constraints.* Master's thesis, Department of Mathematics, University of Chicago, 1939.

[100] H. W. Kuhn and A. W. Tucker. Nonlinear programming. *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, 5:481–492, 1951.

[101] J. Kyparisis. Uniqueness and differentiability of solutions of parametric nonlinear complementarity problems. *Mathematical Programming*, 36(1):105–113, 1986.

[102] S. Leyffer, G. López-Calva, and J. Nocedal. Interior methods for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 17(1):52–77, 2006.

[103] D. C. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1-3):503–528, 1989.

[104] X. W. Liu and J. Sun. Generalized stationary points and an interior-point method for mathematical programs with equilibrium constraints. *Mathematical Programming*, 101(1):231–261, 2004.

[105] X. W. Liu and J. Sun. A robust primal-dual interior-point algorithm for nonlinear programs. *SIAM Journal on Optimization*, 14(4):1163–1186, 2004.

[106] X. W. Liu and J. Sun. Global convergence analysis of line search interior-point methods for nonlinear programming without regularity assumptions. *Journal of Optimization Theory and Applications*, 125(3):609–628, 2005.

[107] X. W. Liu and Y. X. Yuan. A null-space primal-dual interior-point algorithm for nonlinear optimization with nice convergence properties. *Mathematical Programming*, 125(1):163–193, 2010.

[108] F. A. Lootsma. Hessian matrices of penalty functions for solving constrained optimization problems. *Philips Research Reports*, 24:322–331, 1969.

[109] Z. Q. Luo, J. S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, 1996.

[110] 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.

[111] O. L. Mangasarian. Equivalence of the complementarity problem to a system of nonlinear equations. *SIAM Journal on Applied Mathematics*, 31(1):89–92, 1976.

[112] O. L. Mangasarian and S. Fromovitz. The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *Journal of Mathematical Analysis and Applications*, 17(1):37–47, 1967.

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

[114] N. Maratos. *Exact Penalty Function Algorithms for Finite Dimensional and Control Optimization Problems.* PhD thesis, Imperial College London, 1978.

[115] K. W. Meng, S. J. Li, and X. Q. Yang. A robust SQP method based on a smoothing lower order penalty functions. *Optimization*, 58(1):23–38, 2009.

[116] K. W. Meng and X. Q. Yang. First- and second-order necessary conditions via exact penalty functions. *Submitted*.

[117] K. W. Meng and X. Q. Yang. Optimality conditions via exact penalty functions. *SIAM Journal on Optimization*, 20(6):3208–3231, 2010.

[118] Z. Q. Meng, C. Y. Dang, and X. Q. Yang. On the smoothing of the square-root exact penalty function for inequality constrained optimization. *Computational Optimization and Applications*, 35(3):375–398, 2006.

[119] M. Mongeau and A. Sartenaer. Automatic decrease of the penalty parameter in exact penalty function methods. *European Journal of Operational Research*, 83(3):686–699, 1995.

[120] B. S. Mordukhovich. *Variational Analysis and Generalized Differentiation. I: Basic Theory. II: Applications.* Springer, 2006.

[121] 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.

[122] J. J. Moré. Global methods for nonlinear complementarity problems. *Mathematics of Operations Research*, 21(3):589–614, 1996.

[123] 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.

[124] W. Murray. Analytical expressions for the eigenvalues and eigenvectors of the Hessian matrices of barrier and penalty functions. *Journal of Optimization Theory and Applications*, 7(3):189–196, 1971.

[125] J. F. Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49, 1950.

[126] J. F. Nash. Non-cooperative games. *The Annals of Mathematics*, 54(2):286–295, 1951.

[127] Y. Nesterov. Smooth minimization of non-smooth functions. *Mathematical Programming*, 103(1):127–152, 2005.

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

[129] A. M. Noor. General variational inequalities. *Applied Mathematics Letters*, 1(2):119–122, 1988.

[130] A. M. Noor. Some developments in general variational inequalities. *Applied Mathematics and Computation*, 152(1):199–277, 2004.

[131] P. D. Panagiotopoulos. *Inequality Problems in Mechanics and Applications: Convex and Nonconvex Energy Functions*. Springer, 1985.

[132] J. S. Pang. Newton's method for B-differentiable equations. *Mathematics of Operations Research*, 15(2):311–341, 1990.

[133] J. S. Pang. A B-differentiable equation-based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems. *Mathematical Programming*, 51(1-3):101–131, 1991.

[134] J. S. Pang. Error bounds in mathematical programming. *Mathematical Programming*, 79(1-3):299–332, 1997.

[135] J. S. Pang and S. A. Gabriel. NE/SQP: A robust algorithm for the nonlinear complementarity problem. *Mathematical Programming*, 60(1):295–337, 1993.

[136] J. S. Pang and L. Q. Qi. Nonsmooth equations: motivation and algorithms. *SIAM Journal on Optimization*, 3(3):443–465, 1993.

[137] J. S. Pang and J. C. Yao. On a generalization of a normal map and equation. *SIAM Journal on Control and Optimization*, 33(1):168–184, 1995.

[138] P. Patriksson. *The Traffic Assignment Problem: Models and Methods*. CRC Press, 1994.

[139] D. W. Peterson. A review of constraint qualifications in finite-dimensional spaces. *SIAM Review*, 15(3):639–654, 1973.

[140] F. Pfeiffer. *Multibody Dynamics with Unilateral Contacts*. Wiley, 1996.

[141] T. Pietrzykowski. An exact potential method for constrained maxima. *SIAM Journal on Numerical Analysis*, 6(2):299–304, 1969.

[142] R. Polyak. Modified barrier functions (theory and methods). *Mathematical Programming*, 54(1):177–222, 1992.

[143] D. M. Pooley, P.A. Forsyth, and K. R. Vetzal. Numerical convergence properties of option pricing PDEs with uncertain volatility. *IMA Journal of Numerical Analysis*, 23(2):241–267, 2003.

[144] M. J. D. Powell. Convergence properties of algorithms for nonlinear optimization. *SIAM Review*, 28(4):487–500, 1986.

[145] L. Q. Qi. Convergence analysis of some algorithms for solving nonsmooth equations. *Mathematics of Operations Research*, 18(1):227–244, 1993.

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

[147] S. M. Robinson. Strongly regular generalized equations. *Mathematics of Operations Research*, 5(1):43–62, 1980.

[148] S. M. Robinson. Generalized equations and their solutions, Part II: Applications to nonlinear programming. *Optimality and Stability in Mathematical Programming, Mathematical Programming Studies*, 19:200–221, 1982.

[149] S. M. Robinson. Normal maps induced by linear transformations. *Mathematics of Operations Research*, 17(3):691–714, 1992.

[150] S. M. Robinson. Newton's method for a class of nonsmooth functions. *Set-Valued Analysis*, 2(1-2):291–305, 1994.

[151] R. T. Rockafellar and R. J. B. Wets. *Variational Analysis*. Springer, 2011.

[152] A. M. Rubinov, B. M. Glover, and X. Q. Yang. Decreasing functions with applications to penalization. *SIAM Journal on Optimization*, 10(1):289–313, 1999.

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

[154] A. M. Rubinov, X. Q. Yang, and A. M. Bagirov. Penalty functions with a small penalty parameter. *Optimization Methods and Software*, 17(5):931–964, 2002.

[155] T. F Rutherford. Extension of GAMS for complementarity problems arising in applied economic analysis. *Journal of Economic Dynamics and Control*, 19(8):1299–1324, 1995.

[156] R. Scholz. Numerical solution of the obstacle problem by the penalty method. *Numerische Mathematik*, 49(2-3):255–268, 1986.

[157] D. F. Shanno and R. J. Vanderbei. Interior-point methods for nonconvex nonlinear programming: Orderings and higher-order methods. *Mathematical Programming*, 87(2):303–316, 2000.

[158] M. V. Solodov and P. Tseng. Modified projection-type methods for monotone variational inequalities. *SIAM Journal on Control and Optimization*, 34(5):1814–1830, 1996.

[159] G Still and M Streng. Optimality conditions in smooth nonlinear programming. *Journal of Optimization Theory and Applications*, 90(3):483–515, 1996.

[160] D. F. Sun and L. Q. Qi. On NCP-functions. *Computational Optimization and Applications*, 13(1-3):201–220, 1999.

[161] D.F. Sun and J. Sun. Strong semismoothness of the Fischer-Burmeister SDC and SOC complementarity functions. *Mathematical Programming*, 103(3):575–581, 2005.

[162] P. Tseng. Growth behavior of a class of merit functions for the nonlinear complementarity problem. *Journal of Optimization Theory and Applications*, 89(1):17–37, 1996.

[163] P. Tseng. Convergent infeasible interior-point trust-region methods for constrained minimization. *SIAM Journal on Optimization*, 13(2):432–469, 2002.

[164] 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.

[165] R. J. Vanderbei and D. F. Shanno. An interior-point algorithm for nonconvex nonlinear programming. *Computational Optimization and Applications*, 13(1):231–252, 1999.

[166] A. Wächter and L. T. Biegler. Failure of global convergence for a class of interior point methods for nonlinear programming. *Mathematical Programming*, 88(3):565–574, 2000.

[167] A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57, 2006.

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

[169] 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.

[170] J. Y. Wei and S. Yves. Spatial oligopolistic electricity models with cournot generators and regulated transmission prices. *Operations Research*, 47(1):102–112, 1999.

[171] A. P. Wierzbicki. Note on the equivalence of Kuhn-Tucker complementarity conditions to an equation. *Journal of Optimization Theory and Applications*, 37(3):401–405, 1982.

[172] P. Wilmott, J. Dewynne, and S. Howison. *Option Pricing: Mathematical Models and Computation*. Oxford Financial Press, 2000.

[173] J. H. Witte and C. Reisinger. A penalty method for the numerical solution of Hamilton-Jacobi-Bellman (HJB) equations in finance. *SIAM Journal on Numerical Analysis*, 49(1):213–231, 2011.

[174] J. H. Witte and C. Reisinger. Penalty methods for the solution of discrete HJB equations-continuous control and obstacle problems. *SIAM Journal on Numerical Analysis*, 50(2):595–625, 2012.

[175] S. J. Wright. *Primal-dual Interior-Point Methods*. SIAM, 1987.

[176] Z. Y. Wu, F. S. Bai, X. Q. Yang, and L. S. Zhang. An exact lower order penalty function and its smoothing in nonlinear programming. *Optimization*, 53(1):51–68, 2004.

[177] X. S. Xu, Z. Q. Meng, J. W. Sun, L. G. Huang, and R. Shen. A second-order smooth penalty function algorithm for constrained optimization problems. *Computational Optimization and Applications*, 55(1):155–172, 2013.

[178] H. Yamashita. A globally convergent primal-dual interior point method for constrained optimization. *Optimization Methods and Software*, 10(2):443–469, 1998.

[179] X. Q. Yang. Smoothing approximations to nonsmooth optimization problems. *Journal of the Australian Mathematical Society, Series B*, 36(3):274–285, 1995.

[180] X. Q. Yang. A comparative study of smoothing approximations. *Journal of the Australian Mathematical Society, Series B*, 38(2):194–200, 1996.

[181] X. Q. Yang and Z. Q. Meng. Lagrange multipliers and calmness conditions of order p. *Mathematics of Operations Research*, 32(1):95–101, 2007.

[182] X. Q. Yang, Z. Q. Meng, X. X. Huang, and G. T. Y. Pong. Smoothing nonlinear penalty functions for constrained optimization problems. *Numerical Functional Analysis and Optimization*, 24(3):351–364, 2003.

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

[184] W. I. Zangwill. Non-linear programming via penalty functions. *Management Science*, 13(5):344–358, 1967.

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

[186] 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.

[187] Y. Zhao and D. F. Sun. Alternative theorems for nonlinear projection equations and applications to generalized complementarity problems. *Nonlinear Analysis: Theory Methods and Applications*, 46(6):853–868, 2001.

[188] 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.