An Interior-Point $\ell_{\frac{1}{2}}$-Penalty Method for Inequality Constrained Nonlinear Optimization

Boshi Tian

Business School, Hunan University, Changsha 410082, Hunan Province, P.R. China

Xiaoqi Yang

Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, P.R. China

Kaiwen Meng

School of Economics and Management, Southwest Jiaotong University, Chengdu 610031, P.R. China

Abstract: In this paper, we study inequality constrained nonlinear programming problems by virtue of an $\ell_{\frac{1}{2}}$-penalty function and a quadratic relaxation. Combining with an interior-point method, we propose an interior-point $\ell_{\frac{1}{2}}$-penalty method. We introduce different kinds of constraint qualifications to establish the first-order necessary conditions for the quadratically relaxed problem. We apply the modified Newton method to a sequence of logarithmic barrier problems, and design some reliable algorithms. Moreover, we establish the global convergence results of the proposed method. We carry out numerical experiments on 266 inequality constrained optimization problems. Our numerical results show that the proposed method is competitive with some existing interior-point ℓ_1-penalty methods in term of iteration numbers and better when comparing the values of the penalty parameter.

Keywords: Nonlinear programming, Lower-order penalty function, Quadratic relaxation, Constraint qualification, Primal-dual interior-point method.

1 Introduction

Consider the inequality constrained nonlinear programming problem

$$\begin{aligned} &\min f(x) \\ &\text{s.t. } c_i(x) \le 0, \ \forall\, i \in \mathcal{I}, \end{aligned} \tag{1}$$

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

To solve problem (1), many efficient methods were proposed, in particular, for example, interior-point methods (28; 30; 36; 37; 41) and penalty methods (3; 14; 29). In the last decade, researchers paid great attention to studying interior-point penalty methods such as interior-point ℓ_2-penalty methods (6; 7; 20; 21) and interior-point ℓ_1-penalty methods (1; 10; 16), which aimed at solving large-scale optimization problems by combining the regularization effects on the constraints from exact penalty functions and the Newton-like qualities from interior-point methods. The interior-point ℓ_1-penalty method is to apply an interior-point method to a linearly relaxed optimization problem of the ℓ_1-penalty problem of problem (1). It was shown in (16) that the linearly relaxed optimization problem of problem (1) always satisfies the Mangasarian-Fromovitz constraint qualification (MFCQ, for short) at any of its feasible solutions. The global and local convergence results of interior-point penalty methods were established in (6; 16). However, it is a great challenge for the ℓ_1-penalty method in updating the values of the penalty parameter in the numerical implementation, especially when the exact penalty parameter is very large, see, e.g.,(4) and (14, Chapter 12). Different strategies for updating the penalty parameter adaptively were proposed in (3; 4; 15; 27).

Nonconvex and non-Lipschitzian lower-order penalty functions have been studied in the literature, see, e.g., (19; 22; 32). It was shown in (32) that the existence of lower-order exact penalty functions requires weaker conditions than that of an ℓ_1-exact penalty function and that the exact penalty parameter of lower-order exact penalty functions is also smaller than that of an ℓ_1-exact penalty function. Lower-order exact penalty functions have also been used in the establishment of the first-order necessary conditions for problem (1). More specifically, under some second-order conditions and the existence of lower-order exact penalty functions, the first-order necessary conditions of problem (1) were established in (39; 25). Furthermore, examples were given in (39) to show that these conditions do not imply the weakest Guignard constraints qualification (17) and vice versa. Recently, the $\ell_{\frac{1}{2}}$-regularization has received great attention for studying the sparse modeling (particularly on compressed sensing), see (5; 8; 38). Numerical results in (5) show that the stronger sparsity-promoting property of $\ell_{\frac{1}{2}}$-regularization has been achieved over ℓ_1-regularization.

However, there are only a few numerical methods developed so far to solve the lower-order penalty problem, such as smoothing approximation methods, see, e.g., (23; 26; 40). It is known that the solutions of the smoothing approximate method may become unstable when the smoothing parameter is getting small.

Properties of lower-order penalty functions mentioned above suggest that they may be useful in designing efficient optimization methods where certain accuracy can be achieved with smaller values of the penalty parameter and in solving some optimization problems without requiring usual constraint qualifications. In this study, we introduce an interior-point $\ell_{\frac{1}{2}}$-penalty method for solving problem (1), in particular, when problem (1) is nonconvex. More specifically, we first quadratically relax the following $\ell_{\frac{1}{2}}$-penalty problem:

$$\min_{x} \ \phi_{P,\frac{1}{2}}(x,\rho) := f(x) + \rho \sum_{i\in\mathcal{I}} [c_i(x)]_{+}^{\frac{1}{2}}, \tag{2}$$

as follows

$$\min_{x,s} \phi_{S,\frac{1}{2}}(x,s;\rho) := f(x) + \rho \sum_{i\in\mathcal{I}} s_i \tag{3a}$$

$$\text{s.t. } c_i(x) \le s_i^2 \ \forall\, i \in \mathcal{I}, \text{and} \tag{3b}$$

$$s_i \ge 0, \ \forall\, i \in \mathcal{I}, \tag{3c}$$

where $[a]_+ = \max\{a, 0\}$ for all $a \in \mathbb{R}$, $\rho > 0$ is the penalty parameter and $s = (s_i) \in \mathbb{R}^m_+$ are artificial variables. Let $(\hat{x}, \hat{s}) \in \mathbb{R}^{n+m}$ be a local solution of problem (3). Then we apply a primal-dual interior-point method to solve problem (3), i.e., solving the following interior-point penalty problem

$$\begin{aligned} &\min_{x,s} \phi_{B,\frac{1}{2}}(x,s;\rho,\mu) := \phi_{S,\frac{1}{2}}(x,s;\rho) - \mu^2 \sum_{i\in\mathcal{I}} \log\bigl(s_i^2 - c_i(x)\bigr) - \mu \sum_{i\in\mathcal{I}} \log s_i \\ &\text{s.t. } s_i^2 - c_i(x) > 0 \text{ and } s_i > 0, \ \forall\, i \in \mathcal{I}, \end{aligned} \tag{4}$$

where μ^2 and μ are the barrier parameter for constraints (3b) and (3c) respectively, both converging to zero from above. The further motivation for μ^2 being used for the term $\sum_{i\in\mathcal{I}} \log\bigl(s_i^2 - c_i(x)\bigr)$ will be provided in Remark 3.1.

Optimization problem (3) can be viewed as a quadratically relaxed problem of problem (1). Using characterizations in terms of the gradients and Hessians of constraints $c_i(x)$ with $i \in \mathcal{I}$, we introduce several kinds of constraint qualifications (CQs, for short) for the quadratically relaxed problem (3), under which we establish the first-order necessary conditions of problem (3). Especially, we present a new type of CQs (see Lemma 2.3(e) for its definition) which is strictly weaker than the classical MFCQ.

To solve problem (4), we employ a modified Newton's method (2) with an inexact line search to first-order necessary conditions of problem (4). Due to the quadratic relaxation, we add a condition on the Lagrange multipliers of original inequality constraints and that of inequality constraints of the quadratically relaxed problem (3) in order to guarantee the positive definiteness of the Hessian matrix for the interior-point penalty function in problem

(4). We detail our strategies for updating the Lagrange multipliers. Moreover, we describe three specific algorithms. The first algorithm is to solve the barrier problem (4) with a fixed penalty parameter ρ and a fixed barrier parameter μ, the second one is to solve a sequence of the quadratically relaxed problems (3) when ρ is fixed and μ goes to zero and the third one is to solve the penalty problem (2) when the penalty parameter ρ goes to infinite. Finally, under mild conditions, we establish the global convergence results of the proposed interior-point $\ell_{\frac{1}{2}}$-penalty method. Specifically, we prove that the iteration sequence converges to some KKT (or FJ) point of problem (1).

We carry out numerical experiments on 266 inequality constrained optimization problems from CUTEr collection, COPS, MITT and Global test sets. We compare the performance of our method with two existing interior-point ℓ_1-penalty methods introduced in (10) in term of the number of iterations and the values of the penalty parameter.

This paper is organized as follows. In Section 2, we study the first-order necessary conditions of the quadratically relaxed problem (3). In Section 3, we propose an interior-point $\ell_{\frac{1}{2}}$-penalty method and present analysis on a modified Newton method and its global convergence. In Section 4, we present our numerical results.

2 Notations and Necessary Conditions

2.1 Notations and Definitions

For a vector $s \in \mathbb{R}^n$, we invariably assume that s is a column vector and use s_i to denote its i-th component. We write $S := \mathrm{diag}(s)$ to denote a diagonal matrix whose i-th diagonal element is s_i and s^2 to denote a vector in $\mathbb{R}^n$ whose i-th element is s_i^2. Given a matrix $\mathcal{V} \in \mathbb{R}^{n\times n}$, the transpose of $\mathcal{V}$ is denoted as $\mathcal{V}^T$, while $\mathcal{V}^{-1}$ denotes the inverse of matrix $\mathcal{V}$ if $\mathcal{V}$ is invertible. Given two vectors $u, v \in \mathbb{R}^n$, we say $u \geq (>) v$ if and only if $u_i \geq (>) v_i$, for all $i = 1, 2, \ldots, n$. For two symmetric matrices $\mathcal{U}, \mathcal{V} \in \mathbb{R}^{n\times n}$, we write $\mathcal{U} \succ (\succeq) \mathcal{V}$ to mean the matrix $\mathcal{U} - \mathcal{V}$ is positive definite (positive semi-definite). Let $e \in \mathbb{R}^n$ denote a vector whose all components are 1 and E denote an identity matrix. Throughout this paper, we use $\|\cdot\|$ to indicate the Euclidean norm. Given a vector valued function $c : \mathbb{R}^n \to \mathbb{R}^m$, we write $C(x) := \mathrm{diag}(c(x))$ to denote the diagonal matrix whose i-th diagonal element is $c_i(x)$, furthermore, we write $A(x)$ to indicate the transpose of the Jacobian matrix of $c(x)$, i.e., $A(x) := [\nabla c_1(x), \ldots, \nabla c_m(x)]$. Given a real value function $f : \mathbb{R}^n \to \mathbb{R}$, we write $\nabla f(x)$ to denote its gradient vector. The

Dini upper-directional derivative (39) and subderivative (31) of f at x in the direction $u \in \mathbb{R}^n$ are defined, respectively, by

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

Throughout this paper, 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 define the feasible sets of problem (1) and the quadratically relaxed problem (3), respectively, by

$$\begin{aligned} \mathcal{F} &:= \{x \in \mathbb{R}^n \mid c_i(x) \le 0,\ \forall\, i \in \mathcal{I}\}; \\ \widehat{\mathcal{F}} &:= \{(x,s) \in \mathbb{R}^{n+m} \mid c_i(x) \le s_i^2,\ s_i \ge 0,\ \forall\, i \in \mathcal{I}\}. \end{aligned}$$

Definition 2.1 *We say that $x^* \in \mathbb{R}^n$ is a Fritz-John (FJ, for short) point of problem (1) if there exist $\lambda_0 \ge 0$ and $\lambda \in \mathbb{R}^m_+$ satisfying the following conditions*

$$\begin{aligned} \lambda_0 \nabla f(x^*) + A(x^*)\lambda &= 0, \\ C(x^*)\lambda &= 0, \\ -c(x^*) &\ge 0. \end{aligned}$$

Definition 2.2 *We say that $x^* \in \mathbb{R}^n$ is a Karush-Kuhn-Tucker (KKT, for short) point of problem (1) if there exists $\lambda \in \mathbb{R}^m_+$ satisfying the following conditions*

$$\nabla f(x^*) + A(x^*)\lambda = 0, \tag{5a}$$

$$C(x^*)\lambda = 0, \tag{5b}$$

$$-c(x^*) \ge 0. \tag{5c}$$

Definition 2.3 *A vector $x^* \in \mathbb{R}^n$ is called a local solution of problem (1) if $x^* \in \mathcal{F}$ and there is a neighborhood $\mathcal{N}$ of x^* such that $f(x^*) \le f(x)$ for all $x \in \mathcal{N} \cap \mathcal{F}$.*

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 quadratically relaxed problem (3). In the next proposition we conclude that the $\ell_{\frac{1}{2}}$-penalty problem (2) and its quadratically relaxed problem (3) are equivalent in the sense that they have the same local solution. Its proof is easy and the details are omitted.

Proposition 2.1 *Let $\rho > 0$ be fixed. Then $\hat{x} \in \mathbb{R}^n$ is a local solution of problem (2) if and only if $(\hat{x}, \hat{s}) \in \mathbb{R}^{n+m}$ is a local solution of problem (3) with $\hat{s}_i = [c_i(\hat{x})]_+^{\frac{1}{2}}$ for all $i \in \mathcal{I}$.*

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 quadratically relaxed problem (3), i.e., there exist vectors $y, u \in \mathbb{R}^m$ such that

$$\nabla f(\hat{x}) + A(\hat{x})y = 0, \tag{6a}$$

$$\rho e - 2Y\hat{s} - u = 0, \tag{6b}$$

$$Y\big(c(\hat{x}) - \hat{s}^2\big) = 0, \tag{6c}$$

$$U\hat{s} = 0, \tag{6d}$$

$$\hat{s}^2 - c(\hat{x}) \geq 0, \tag{6e}$$

$$\hat{s}, y, u \geq 0, \tag{6f}$$

where the vectors $y, u \in \mathbb{R}^m_+$ are called Lagrange multipliers, $Y = \mathrm{diag}(y)$ and $U = \mathrm{diag}(u)$ are diagonal matrices. To investigate regularity conditions under which (6) can be fulfilled, we introduce a few index sets defined for $(\hat{x}, \hat{s}) \in \mathbb{R}^{n+m}$ as follows:

$$\begin{aligned}
S^0(\hat{x}, \hat{s}) &:= \{i \in \mathcal{I} \mid \hat{s}_i = 0 \text{ and } c_i(\hat{x}) \leq 0\}; \\
S^+(\hat{x}, \hat{s}) &:= \{i \in \mathcal{I} \mid \hat{s}_i > 0 \text{ and } c_i(\hat{x}) \leq \hat{s}_i^2\}; \\
S^=(\hat{x}, \hat{s}) &:= \{i \in S^+(\hat{x}, \hat{s}) \mid c_i(\hat{x}) = \hat{s}_i^2\}; \\
CS^0(\hat{x}, \hat{s}) &:= \{i \in S^0(\hat{x}, \hat{s}) \mid c_i(\hat{x}) = 0\}.
\end{aligned}$$

Since $(\hat{x}, \hat{s})$ is assumed to be a local solution of the quadratically relaxed problem (3), 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 (6) as

$$
\begin{aligned}
& \nabla f(\hat{x}) + \sum_{i \in \mathcal{I}} y_i \nabla c_i(\hat{x}) = 0, \\
& y_i = \frac{\rho}{2\hat{s}_i}, \ \forall\, i \in S^{=}(\hat{x}, \hat{s}), \ y_i \geq 0, \ \forall\, i \in CS^0(\hat{x}, \hat{s}), \\
& y_i = 0, \ \forall\, i \in S^0(\hat{x}, \hat{s}) \backslash CS^0(\hat{x}, \hat{s}), \\
& u_i = 0, \ \forall\, i \in S^{=}(\hat{x}, \hat{s}), \ u_i = \rho, \ \forall\, i \in S^0(\hat{x}, \hat{s}), \\
& \hat{s}^2 - c(\hat{x}) \geq 0, \ \hat{s} \geq 0.
\end{aligned} \tag{7}
$$

If $\hat{x}$ is feasible to problem (1), we have $\hat{s} = 0$ and $S^{=}(\hat{x}, \hat{s}) = \emptyset$, and moreover, the first-order necessary conditions (6) or (7) recover the first-order necessary conditions at $\hat{x}$ for problem (1).

If $\hat{s} \in \mathbb{R}^m_{++} := \{x \mid x_i > 0, \ \forall\, i \in \mathcal{I}\}$, the quadratically relaxed problem (3) only has the inequalities $c_i(x) - s_i^2 \leq 0$ with $i \in \mathcal{I}$ being active at $(\hat{x}, \hat{s})$, and the Jacobian matrix $(A(\hat{x})^T, -2\mathrm{diag}(\hat{s}))$ of $c(x) - s^2$ at $(\hat{x}, \hat{s})$ has full rank, implying that the linearly independent constraint qualification (LICQ, for short) holds at $(\hat{x}, \hat{s})$. In this case, the first-order necessary conditions (6) 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 the quadratically relaxed problem (3) to possess the first-order necessary conditions (6). To begin with, we show in the following lemma that the LICQ (resp. the MFCQ) holds at $(\hat{x}, \hat{s})$ for the quadratically relaxed problem (3) if and only if the LICQ (resp. the MFCQ) holds at $\hat{x}$ for the inequality system

$$
c_i(x) \leq 0, \quad \forall\, i \in CS^0(\hat{x}, \hat{s}). \tag{8}
$$

Lemma 2.1 *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 (8), 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 (8), 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,

$$
\sum_{i \in CS^0(\hat{x}, \hat{s})} y_i \nabla c_i(\hat{x}) = 0, \quad y_i \geq 0, \ \forall\, i \in CS^0(\hat{x}, \hat{s}) \Longrightarrow y_i = 0, \ \forall\, i \in CS^0(\hat{x}, \hat{s}). \tag{9}
$$

Then (a) *holds if and only if the* LICQ *holds at* $(\hat{x}, \hat{s})$ *for the quadratically relaxed problem (3), while* (b) *holds if and only if the* MFCQ *holds at* $(\hat{x}, \hat{s})$ *for the quadratically relaxed problem (3).*

Proof. By definition, the MFCQ holds at $(\hat{x}, \hat{s})$ for problem (3) if,

$$\left.\begin{array}{l} \displaystyle\sum_{i \in S^{=}(\hat{x},\hat{s}) \cup CS^0(\hat{x},\hat{s})} y_i \nabla c_i(\hat{x}) = 0 \\ -2\hat{s}_i y_i = 0, \ \forall\, i \in S^{=}(\hat{x}, \hat{s}), \\ u_i = 0, \ \forall\, i \in S^0(\hat{x}, \hat{s}), \\ y_i \geq 0, \ \forall\, i \in S^{=}(\hat{x}, \hat{s}) \cup CS^0(\hat{x}, \hat{s}), \\ u_i \geq 0, \ \forall\, i \in S^0(\hat{x}, \hat{s}) \end{array}\right\} \Longrightarrow \left\{\begin{array}{l} y_i = 0, \ \forall\, i \in S^{=}(\hat{x}, \hat{s}) \cup CS^0(\hat{x}, \hat{s}), \\ \\ u_i = 0, \ \forall\, i \in S^0(\hat{x}, \hat{s}) \end{array}\right. \tag{10}$$

Observing that $\hat{s}_i > 0$ for all $i \in S^{=}(\hat{x}, \hat{s})$, the equivalence of (9) and (10) follows immediately. The case for the LICQ can be proved in a similar way. ■

Remark 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 quadratically relaxed problem (3) has bounded Lagrange multipliers* (y, u) *as defined by (6) if and only if Lemma 2.1* (b) *is fulfilled. If the MFCQ holds at a feasible point* $x_0 \in \mathcal{F}$ *for problem (1), then for all* $(\hat{x}, \hat{s})$ *with* $\hat{x}$ *near* x_0*, the quadratically relaxed problem (3) has bounded Lagrange multipliers at* $(\hat{x}, \hat{s})$ *provided that it is a local solution of problem (3).*

Besides having the CQs in Lemma 2.1 for the first-order necessary conditions (6), we can use the techniques in (24; 25; 39) to derive some other CQs, some of which turn out to be strictly weaker than the ones in Lemma 2.1. We conduct the analysis in two lemmas below by first showing the exactness of an $\ell_{\frac{1}{2}}$-penalty function for the quadratically relaxed problem (3) defined by

$$f(x) + \rho \sum_{i \in \mathcal{I}} s_i + \pi \left(\sum_{i \in \mathcal{I}} \sqrt{\max\{c_i(x) - s_i^2, 0\}} + \sum_{i \in \mathcal{I}} \sqrt{\max\{-s_i, 0\}} \right), \tag{11}$$

and then requiring that the linearized tangent cone

$$L_{\widehat{\mathcal{F}}}(\hat{x}, \hat{s}) := \left\{ (w, \beta) \in \mathbb{R}^n \times \mathbb{R}^m \;\middle|\; \begin{array}{rl} \langle \nabla c_i(\hat{x}), w \rangle - 2\hat{s}_i \beta_i \leq 0, & \forall\, i \in S^{=}(\hat{x}, \hat{s}) \\ \langle \nabla c_i(\hat{x}), w \rangle \leq 0, & \forall\, i \in CS^0(\hat{x}, \hat{s}) \\ -\beta_i \leq 0, & \forall\, i \in S^0(\hat{x}, \hat{s}) \end{array} \right\} \tag{12}$$

to the feasible set $\widehat{\mathcal{F}}$ of problem (3) at $(\hat{x}, \hat{s})$ coincides with the kernel of the subderivative (or Dini upper directional derivative) of the penalty term

$$\phi(x,s) := \sum_{i\in\mathcal{I}} \sqrt{\max\{c_i(x) - s_i^2, 0\}} + \sum_{i\in\mathcal{I}} \sqrt{\max\{-s_i, 0\}}. \tag{13}$$

First we show that the penalty function (11) is exact at $(\hat{x}, \hat{s})$ in the sense that it admits a local minimum at $(\hat{x}, \hat{s})$ when the penalty parameter π is greater than 1.

Lemma 2.2 *Assume that $\hat{s} \in \mathbb{R}^m_+ \backslash \mathbb{R}^m_{++}$. Then $(\hat{x}, \hat{s})$ is a local minimizer of the function (11) whenever $\pi \geq 1$.*

Proof. To begin with, we show that for any two real numbers a and b,

$$\sqrt{\max\{a-b, 0\}} \geq \sqrt{\max\{a, 0\}} - \sqrt{\max\{b, 0\}}. \tag{14}$$

In fact, if $a - b \leq 0$, we have $\sqrt{\max\{a, 0\}} \leq \sqrt{\max\{b, 0\}}$ and hence (14) follows. If $a - b > 0$ and $b \geq 0$, (14) follows immediately from (18, Lemma 4.1). And if $a - b > 0$ and $b < 0$, we have $a - b > a$ and hence

$$\sqrt{\max\{a-b, 0\}} \geq \sqrt{\max\{a, 0\}} = \sqrt{\max\{a, 0\}} - \sqrt{\max\{b, 0\}}.$$

That is, (14) holds in all cases.

Since $(\hat{x}, \hat{s})$ is a local solution of problem (3). It follows from Proposition 2.1, we have $\hat{s}_i = [c_i(\hat{x})]_+^{\frac{1}{2}}$ for all $i \in \mathcal{I}$, and that $\hat{x}$ is a local solution of problem (2), or in other words, the following inequality holds for all x in some neighborhood V of $\hat{x}$:

$$f(x) + \rho \sum_{i\in\mathcal{I}} [c_i(x)]_+^{\frac{1}{2}} \geq f(\hat{x}) + \rho \sum_{i\in\mathcal{I}} [c_i(\hat{x})]_+^{\frac{1}{2}} = f(\hat{x}) + \rho \sum_{i\in\mathcal{I}} \hat{s}_i.$$

Then for all $x \in V$ and all $s \in \mathbb{R}^m$ with $s_i \geq -\frac{1}{4}$ for all $i \in \mathcal{I}$, it follows that

$$\begin{aligned}
& f(x) + \rho \sum_{i\in\mathcal{I}} s_i + \sum_{i\in\mathcal{I}} \sqrt{\max\{c_i(x) - s_i^2, 0\}} + \sum_{i\in\mathcal{I}} \sqrt{\max\{-s_i, 0\}} \\
\geq\ & f(x) + \rho \sum_{i\in\mathcal{I}} s_i + \sum_{i\in\mathcal{I}} \sqrt{\max\{c_i(x), 0\}} - \sum_{i\in\mathcal{I}} \sqrt{\max\{s_i^2, 0\}} + \sum_{i\in\mathcal{I}} \sqrt{\max\{-s_i, 0\}} \\
=\ & f(x) + \rho \sum_{i\in\mathcal{I}} [c_i(x)]_+^{\frac{1}{2}} + \sum_{i\in\mathcal{I}} \left(s_i - |s_i| + \sqrt{\max\{-s_i, 0\}} \right) \\
\geq\ & f(x) + \rho \sum_{i\in\mathcal{I}} [c_i(x)]_+^{\frac{1}{2}} \\
\geq\ & f(\hat{x}) + \rho \sum_{i\in\mathcal{I}} \hat{s}_i,
\end{aligned}$$

where the first inequality follows from (14) and the second inequality follows from the fact that $\alpha - |\alpha| - \sqrt{\max\{-\alpha, 0\}} \geq 0$ whenever $\alpha \geq -\frac{1}{4}$. This shows that $(\hat{x}, \hat{s})$ is a local minimizer of the function (11) with $\pi = 1$, and hence with any $\pi \geq 1$. This completes the proof. ■

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

$$L_{\widehat{\mathcal{F}}}(\hat{x}, \hat{s}) = \{(w, \beta) \in \mathbb{R}^{n+m} \mid D_+\phi(\hat{x}, \hat{s})(w, \beta) = 0\} \tag{15}$$

and

$$L_{\widehat{\mathcal{F}}}(\hat{x}, \hat{s}) = \{(w, \beta) \in \mathbb{R}^{n+m} \mid d\phi(\hat{x}, \hat{s})(w, \beta) = 0\}, \tag{16}$$

where $L_{\widehat{\mathcal{F}}}(\hat{x}, \hat{s})$ and ϕ are given by (12) and (13), respectively.

Lemma 2.3 *Assume that $\bar{s} \in \mathbb{R}^m_+ \backslash \mathbb{R}^m_{++}$. 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 (15) 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

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

(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*

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

(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*

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

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

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

(f) *The equality (16) holds.*

Then we have

$$\text{(a)} \iff \text{(b)} \implies \text{(c)} \implies \text{(d)} \iff \text{(e)} \iff \text{(f)}.$$

Proof. The implications (b) $\implies$ (c) $\implies$ (d) hold trivially. By a nonhomogeneous Farkas' Lemma (34, Lemma 4.2), it is straightforward to verify that (d) $\iff$ (e). To show (e) $\iff$ (f), we introduce another square root penalty term for the quadratically relaxed problem (3) 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 (18, Lemma 4.1), we have $\tilde{\phi} \le \phi \le 2m\tilde{\phi}$ and hence

$$\{(w,\beta) \mid d\tilde{\phi}(\hat{x},\hat{s})(w,\beta) = 0\} = \{(w,\beta) \mid d\phi(\hat{x},\hat{s})(w,\beta) = 0\}. \tag{17}$$

Applying (24, Proposition 2.1), we have the equality

$$L_{\widehat{\mathcal{F}}}(\hat{x},\hat{s}) = \{(w,\beta) \mid d\tilde{\phi}(\hat{x},\hat{s})(w,\beta) = 0\} \tag{18}$$

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

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

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

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

because $(w,\beta) \in L_{\widehat{\mathcal{F}}}(\hat{x},\hat{s})$ amounts to that $w \in \Omega$, $\beta_i \ge \langle \frac{\nabla c_i(\hat{x})}{2\bar{s}_i}, w\rangle$ for all $i \in S^{=}(\hat{x},\hat{s})$ and $\beta_i \ge 0$ for all $i \in S^{=}(\hat{x},\hat{s}) \cup CS^0(\hat{x},\hat{s})$. Since $\lambda_i[\langle w, \nabla^2 c_i(\hat{x}) w\rangle - 2\beta_i^2] \le \lambda_i \langle w, \nabla^2 c_i(\hat{x}) w\rangle$ whenever $\lambda_i \ge 0$ and $\beta_i \ge 0$, the equality (18) holds if and only if (e) holds. In view of (17), we have (e) $\iff$ (f).

By (39, Lemma 2.3) or (25, 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

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

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 *It is clear to see that the CQ given by Lemma 2.3* (e) *is implied by the CQ given by Lemma 2.1* (b)*. But the converse may not hold as can be seen from (24, Example 2.3) in the case of* $\hat{s} = 0$*.*

In view of Lemma 2.2 and (24, Theorem 2.1), we now confirm that the first-order necessary conditions (6) hold at $(\hat{x}, \hat{s})$ for the quadratically relaxed problem (3) provided that one of the CQs in Lemmas 2.1 and 2.3 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 (3) in the following theorem.

Theorem 2.1 *Let* $\rho > 0$ *be fixed and let* $(\hat{x}, \hat{s})$ *be a local solution of the quadratically relaxed problem (3). Then the first-order necessary conditions (6) 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 2.1 and 2.3 being fulfilled.*

3 Interior-Point $\ell_{\frac{1}{2}}$-Penalty Method

In this section, we introduce an interior-point $\ell_{\frac{1}{2}}$-penalty method for problem (1). Then we establish the global convergence results for the proposed method under mild conditions.

3.1 A Primal-Dual Interior-Point Method

Throughout this subsection, we let the penalty parameter $\rho > 0$ be fixed and apply the primal-dual interior-point method to solve the quadratically relaxed problem (3). Let (x, s) be a local solution of problem (4). Then the first-order necessary conditions of problem (4) can be written as

$$\nabla f(x) + A(x) y = 0, \tag{19a}$$

$$\rho e - 2Ys - u = 0, \tag{19b}$$

$$Y\big(s^2 - c(x)\big) - \mu^2 e = 0, \tag{19c}$$

$$Us - \mu e = 0, \tag{19d}$$

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

Remark 3.1 *We note that it is reasonable to choose μ^2 as the barrier parameter for the term $\sum_{i\in\mathcal{I}} \log\big(s_i^2 - c_i(x)\big)$ in problem (4). Indeed, suppose that the Lagrange multiplier y is bounded. It follows from (19b) that the Lagrange multiplier $u \to \rho e$ as $s \to 0^+$. By (19d), we have $\mu = O(\|s\|)$. Thus $\mu^2 = O(\|s^2\|)$, which can be guaranteed by setting μ^2 for the term $\sum_{i\in\mathcal{I}} \log\big(s_i^2 - c_i(x)\big)$ in problem (4).*

Applying a modified Newton's method (2) to the nonlinear system (19) in variables x, s, y and u, we obtain

$$\Omega(x, y, s, u, H) \begin{pmatrix} \triangle x \\ \triangle s \\ \triangle y \\ \triangle u \end{pmatrix} = - \begin{pmatrix} \nabla f(x) + A(x)y \\ \rho e - 2Ys - u \\ Y\big(s^2 - c(x)\big) - \mu^2 e \\ Us - \mu e \end{pmatrix} \tag{20}$$

where

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

and

$$H(x, y) := \nabla^2 f(x) + \sum_{i\in\mathcal{I}} y_i \nabla^2 c_i(x). \tag{21}$$

Observing that $Ys = Sy$ and $Us = Su$, we can reformulate (20) as follows

$$H(x, y)\triangle x + A(x)(y + \triangle y) = -\nabla f(x), \tag{22a}$$
$$2S(y + \triangle y) + E(u + \triangle u) + 2Y\triangle s = \rho e, \tag{22b}$$
$$(S^2 - C(x))(y + \triangle y) + 2YS\triangle s - YA(x)^T\triangle x = \mu^2 e, \tag{22c}$$
$$U\triangle s + S(u + \triangle u) = \mu e. \tag{22d}$$

Solving $\hat{y} := y + \triangle y$ and $\hat{u} := u + \triangle u$ from (22c) and (22d), we get

$$\hat{y} = \big(S^2 - C(x)\big)^{-1}\big(\mu^2 e - 2YS\triangle s + YA(x)^T\triangle x\big), \tag{23a}$$
$$\hat{u} = S^{-1}\big(\mu e - U\triangle s\big). \tag{23b}$$

Substituting (23a) and (23b) into (22a) and (22b), we obtain

$$\mathcal{M}\begin{pmatrix} \triangle x \\ \triangle s \end{pmatrix} = \begin{pmatrix} -\nabla f(x) - \mu^2 A(x)\big(S^2 - C(x)\big)^{-1} e \\ 2\mu^2 S\big(S^2 - C(x)\big)^{-1} e + \mu S^{-1} e - \rho e \end{pmatrix} \tag{24}$$

where

$$\mathcal{M} := \begin{pmatrix} \widehat{H}(x,s,y) & -2A(x)\mathcal{N}S \\ -2\mathcal{N}SA(x)^T & 4S\mathcal{N}S + S^{-1}U - 2Y \end{pmatrix} \tag{25}$$

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

In order to establish the global convergence of the interior-point method, we need to ensure the matrix $\mathcal{M}$ is sufficiently positive definite (13; 12). Assume that

$$u - 2Ys \geq 0. \tag{26}$$

Since $\mathcal{N} \succ 0$ and $S \succ 0$, it follows from the assumption above, we have $4S\mathcal{N}S + S^{-1}U - 2Y \succ 0$. To guarantee $\mathcal{M} \succ 0$, by the Schur complement, we need to ensure

$$\widehat{H}(x,s,y) - \Big(2\mathcal{N}SA(x)^T\Big)\Big(4S\mathcal{N}S + S^{-1}U - 2Y\Big)^{-1}\Big(2A(x)\mathcal{N}S\Big) \succ 0.$$

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

$$H(x,y) + A(x)\Big\{\mathcal{N} - 2\mathcal{N}S\Big(4S\mathcal{N}S + S^{-1}U - 2Y\Big)^{-1} 2S\mathcal{N}\Big\}A(x)^T \succ 0. \tag{27}$$

However, inequality (27) 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 (27) holds (2; 33; 35).

Remark 3.2 *In order to use the Schur complement to matrix $\mathcal{M}$, we force (26) to hold in every iteration (see (31) and (32)). Here, we note that the assumption (26) is reasonable. Indeed, as $s \to 0^+$, assume that multiplier y is bounded above, it follows from (19b) that $u \to \rho e$ and (26) holds automatically.*

At the k-th iteration (x^k, s^k), we get $(\triangle x^k, \triangle s^k)$ by solving (24). Then we let

$$x^{k+1} = x^k + \alpha_P^k \triangle x^k, \tag{28a}$$

$$s^{k+1} = s^k + \alpha_P^k \triangle s^k, \tag{28b}$$

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

$$(s^{k+1})^2 - c(x^{k+1}) > 0, \tag{29a}$$

$$s^{k+1} > 0, \tag{29b}$$

$$\phi_{B,\frac{1}{2}}(x^{k+1}, s^{k+1}; \rho, \mu) - \phi_{B,\frac{1}{2}}(x^k, s^k; \rho, \mu) \le \tau_1 \alpha_P^k \Big(\nabla_x \phi_{B,\frac{1}{2}}(x^k, s^k; \rho, \mu)^T \triangle x^k + \nabla_s \phi_{B,\frac{1}{2}}(x^k, s^k; \rho, \mu)^T \triangle s^k \Big), \tag{29c}$$

for some $\tau_1 \in (0, \frac{1}{2})$, where the last inequality is a standard Armijo line search condition in (37) on the decrease of the barrier objective function in problem (4).

Remark 3.3 *In practice, τ_1 is chosen to be quite small. In this paper, following (10), $\tau_1 = 10^{-8}$ is set, see Table 1 in Section 4.*

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 (1; 6; 9) to update them as follows,

$$\tilde{y}_i^{k+1} := \begin{cases} \min\{\gamma_{min} y_i^k, \frac{\mu^2}{(s_i^k)^2 - c_i(x^k)}\}, & \text{if} \quad \hat{y}_i^{k+1} < \min\{\gamma_{min} y_i^k, \frac{\mu^2}{(s_i^k)^2 - c_i(x^k)}\}, \\ \frac{\mu^2 \gamma_{max}}{(s_i^k)^2 - c_i(x^k)}, & \text{if} \quad \hat{y}_i^{k+1} > \frac{\mu^2 \gamma_{max}}{(s_i^k)^2 - c_i(x^k)}, \\ \hat{y}_i^{k+1}, & \text{otherwise}, \end{cases} \tag{30a}$$

$$\tilde{u}_i^{k+1} := \begin{cases} \min\{\gamma_{min} u_i^k, \frac{\mu}{s_i^k}\}, & \text{if} \quad \hat{u}_i^{k+1} < \min\{\gamma_{min} u_i^k, \frac{\mu}{s_i^k}\}, \\ \frac{\mu \gamma_{max}}{s_i^k}, & \text{if} \quad \hat{u}_i^{k+1} > \frac{\mu \gamma_{max}}{s_i^k}, \\ \hat{u}_i^{k+1}, & \text{otherwise}, \end{cases} \tag{30b}$$

where 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 (26). Specifically, if $(\tilde{y}^{k+1}, \tilde{u}^{k+1})$ satisfies (26), we let $(y^{k+1}, u^{k+1}) := (\tilde{y}^{k+1}, \tilde{u}^{k+1})$ as the new Lagrange multipliers, otherwise, we set

$$y^{k+1} := \gamma_1 \tilde{y}^{k+1}, \ u^{k+1} := \gamma_2 \tilde{u}^{k+1}, \tag{31}$$

where $0 < \gamma_1 \le 1$ and $\gamma_2 \ge 1$ satisfy

$$\frac{\gamma_2}{\gamma_1} \ge \max_{i \in \mathcal{I}} \Big\{ \frac{2 s_i^{k+1} \tilde{y}_i^{k+1}}{\tilde{u}_i^{k+1}} \Big\}. \tag{32}$$

Remark 3.4 *Here note that, to guarantee the dual multipliers* (y^k, u^k) *being bounded, the sequences* $\{(\hat{y}^k, \hat{u}^k)\}$ *is truncated in (30) through choosing a proper* γ_{max}. *In practice,* γ_{max} *should be very large, for example,* $\gamma_{max} = 10^{20}$ *was used in (9). In this paper,* $\gamma_{max} = 10^{23}$ *is chosen, see Table 1 in Section 4.*

Rather than solving the barrier subproblem (4) accurately, our iteration continues until the condition (19) is satisfied within a tolerance ϵ_μ for the current barrier parameter μ, that is,

$$\mathrm{Res}(x, s, \hat{y}, \hat{u}; \rho, \mu) := \left\| \begin{array}{c} \nabla f(x) + A(x)\hat{y} \\ \rho e - 2\widehat{Y}s - \hat{u} \\ \widehat{Y}(s^2 - c(x)) - \mu^2 e \\ \widehat{U}s - \mu e \end{array} \right\| < \epsilon_\mu, \tag{33a}$$

$$(\hat{y}, \hat{u}) \succeq -\epsilon_\mu(e, e), \tag{33b}$$

where ϵ_μ is a μ-related tolerance parameter, which goes to zero from above as $\mu \to 0$.

3.3 Specific Algorithms

In this subsection, we describe three specific algorithms to solve problem (1) by virtue of the $\ell_{\frac{1}{2}}$-penalty function. More implementation details will be stated in Section 4. The first algorithm gives a description of the approximate solution of problem (4) with a fixed penalty parameter $\rho > 0$ and a barrier parameter $\mu > 0$.

Algorithm 1:	Inner algorithm for solving problem (4).
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 (33) holds at $(x^k, s^k, \hat{y}^k, \hat{u}^k)$, stop;
Step 2	If (27) 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 it is positive definite;
Step 3	Compute $(\triangle x^k, \triangle s^k)$ from (24) and $(\hat{y}^{k+1}, \hat{u}^{k+1})$ from (23); we compute the primal step length α_P^k such that (29) is satisfied and compute (x^{k+1}, s^{k+1}) from (28); based on (30)-(31) 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 quadratically relaxed problem (3), we need to solve a series of barrier subproblems (4) for decreasing the values of the barrier parameter μ with a fixed penalty parameter $\rho > 0$.

Algorithm 2:	Inner algorithm for solving problem (3).
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 $(\hat{y}^j, \hat{u}^j) \succeq 0$, stop;
Step 2	Starting from $(x^j, s^j, \hat{y}^j, \hat{u}^j)$, we apply Algorithm 1 to solve problem (4) 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 $\mu^{j+1} := \gamma\mu^j$ and $\epsilon_{\mu^{j+1}} := \gamma\epsilon_{\mu^j}$, let $j := j + 1$ and go to Step 1.

If $\|s^j\|$ is sufficiently small at (x^j, s^j), we declare that point x^j as a KKT or FJ point of problem (1). Otherwise, we increase the penalty parameter ρ and solve the quadratically relaxed problem (3) again. A formal description of the algorithm to solve problem (1) is given as follows.

Algorithm 3:	Outer algorithm for solving problem (1).
Step 0	Initialization. Set $x^0 \in R^n$, $\rho^0 > 0$, $y^0 = \hat{y}^0 > 0$, $u^0 = \hat{u}^0 > 0$, $\nu > 1$, $\bar{i} > 1$, $\bar{\epsilon} \ge 0$ and $s_l^0 \ge \sqrt{\max\{c_l(x^0), 0\} + \frac{1}{2}}$ for all $l \in \mathcal{I}$. Let $i := 0$;
Step 1	If $\|s^i\| \le \bar{\epsilon}$, stop;
Step 2	Starting from $(x^i, s^i, \hat{y}^i, \hat{u}^i)$, we apply Algorithm 2 to solve problem (3) 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 $i := i + 1$, go to Step 1.

Remark 3.5 *Generally, we set $\bar{\epsilon} = 0$ to establish the convergence analysis for the sequence $\{x^i\}$ to some stationary point of problem (1). In numerical experiments, $\bar{\epsilon}$ should be set a sufficiently small constant. We let $\bar{\epsilon} = 10^{-6}$ in our numerical experiments, please see Table 1.*

3.4 Convergence Analysis

In this subsection, we establish the global convergence results for the interior-point $\ell_{\frac{1}{2}}$-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 $\{x^k\}$ generated by Algorithm 1 lies in a bounded set.

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

Remark 3.6 *Assumptions* 3 *and* 4 *are standard assumptions in establishing the global convergence of the interior-point method, see (6, Assumptions* 3 *and* 4*).*

We define the strictly feasible set of problem (3) as follows

$$\widehat{\mathcal{F}}^+ := \{(x,s) \in \mathbb{R}^{n+m} \mid c_i(x) < s_i^2,\ s_i > 0,\ \forall\, i \in \mathcal{I}\}.$$

Lemma 3.1 *The set $\widehat{\mathcal{F}}^+$ is nonempty.*

Proof. Let $\tilde{x} \in \mathbb{R}^n$ and $\tilde{s}_i > \sqrt[2]{\max\{c_i(\tilde{x}), 0\}}$, for all $i \in \mathcal{I}$. Doing so ensures that $\tilde{s}_i^2 - c_i(\tilde{x}) > 0$ 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 (3). This proves that the strictly feasible set $\widehat{\mathcal{F}}^+$ is nonempty. ■

The next lemma shows that the sequence $\{(\triangle x^k, \triangle s^k)\}$ generated by Algorithm 1 is a descent direction of the merit function $\phi_{B,\frac{1}{2}}(x^k, s^k; \rho, \mu)$ provided that $\mathcal{M}^k \succ 0$ or has been modified to be so.

Lemma 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* 1*, the linear system (22) has a solution $(\triangle x^k, \triangle s^k, \hat{y}^{k+1}, \hat{u}^{k+1})$. Then we have*

$$\phi'_{B,\frac{1}{2}}(x^k, s^k; \rho, \mu; \triangle x^k, \triangle s^k) \leq -(\triangle x^k, \triangle s^k)^T \mathcal{M}_k (\triangle x^k, \triangle s^k), \tag{34}$$

where $\phi'_{B,\frac{1}{2}}(x^k, s^k; \rho, \mu; \triangle x^k, \triangle s^k)$ denotes the directional derivative of the function $\phi_{B,\frac{1}{2}}(x, s; \rho, \mu)$ at point (x^k, s^k) in the direction $(\triangle x^k, \triangle s^k)$.

Proof. Since the merit function $\phi_{B,\frac{1}{2}}(x,s;\rho,\mu)$ is continuously differentiable, it follows that

$$\nabla_x\phi_{B,\frac{1}{2}}(x^k,s^k;\rho,\mu) = \nabla f(x^k) + \mu^2 A(x^k)\big((S^k)^2 - C(x^k)\big)^{-1}e, \tag{35a}$$

$$\nabla_s\phi_{B,\frac{1}{2}}(x^k,s^k;\rho,\mu) = \rho e - 2\mu^2 S^k\big((S^k)^2 - C(x^k)\big)^{-1}e - \mu(S^k)^{-1}e, \tag{35b}$$

$$\phi_{B,\frac{1}{2}}{}'(x^k,s^k;\rho,\mu;\triangle x^k,\triangle s^k) = \nabla_x\phi_{B,\frac{1}{2}}(x^k,s^k;\rho,\mu)^T\triangle x^k + \nabla_s\phi_{B,\frac{1}{2}}(x^k,s^k;\rho,\mu)^T\triangle s^k. \tag{35c}$$

Substituting (35a) and (35b) into (35c) and combining (20) and (24), we obtain the inequality (34). ■

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 3.3 *Let the penalty parameter $\rho > 0$ and the barrier parameter $\mu > 0$ be fixed. Suppose that Assumptions 2-4 hold and Algorithm 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 (23a) and (23b), we have that

$$\begin{aligned} \hat{y}^{k+1} &= \mu^2\big((S^k)^2 - C(x^k)\big)^{-1}e,\\ \hat{u}^{k+1} &= \mu(S^k)^{-1}e. \end{aligned} \tag{36}$$

By (29a) and (29b), we see that $(\hat{y}^{k+1},\hat{u}^{k+1}) > 0$. It follows from inequality (27) we have the matrix $\mathcal{M}^k$ is positive definite. Combining (24), we have

$$\begin{aligned} -\nabla f(x^k) - \mu^2 A(x^k)\big((S^k)^2 - C(x^k)\big)^{-1}e = 0,\\ 2\mu^2 S^k\big((S^k)^2 - C(x^k)\big)^{-1}e + \mu(S^k)^{-1}e - \rho e = 0. \end{aligned} \tag{37}$$

By (36) and (37), we conclude that the point $(x^{k+1},s^{k+1},\hat{y}^{k+1},\hat{u}^{k+1})$ satisfies the termination condition (33). Then the Algorithm 1 will terminate at the $(k+1)$-th iteration, which contradicts the assumption. ■

Lemma 3.4 *Let the penalty parameter $\rho > 0$ and the barrier parameter $\mu > 0$ be fixed. Suppose that Assumptions 2-4 hold and Algorithm 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 (29) holds for all $\alpha_P^k \in (0,\bar{\alpha}_P^k]$.*

Proof. Let the function $R(x,s):\mathbb{R}^n\times\mathbb{R}^m_+\to\mathbb{R}^m$ be defined as $R(x,s)=s^2-c(x)$. 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 (29a) 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 (29b) holds for all $\alpha_P^k\in(0,\hat{\alpha}_P^k]$. By Lemma 3.3, we have $(\triangle x^k,\triangle s^k)\neq 0$, and it follows from (34) that $\phi_{B,\frac{1}{2}}{}'(x^k,s^k;\rho,\mu;\triangle x^k,\triangle s^k)<0$. Hence, we conclude that there exists a $\breve{\alpha}_P^k>0$ such that condition (29c) 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 have proved this lemma. ■

Lemma 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 $\{(s^k)^2-c(x^k)\}$ and $\{s^k\}$ generated by Algorithm 1 are bounded from above and componentwise bounded away from zero, so is the sequence $\{(y^k,u^k)\}$ generated by our update strategy (30)-(31).*

Proof. Since the sequence $\{(x^k,s^k)\}$ is generated by a descent line search method, it follows that $\phi_{B,\frac{1}{2}}(x^k,s^k;\rho,\mu)\le\phi_{B,\frac{1}{2}}(x^0,s^0;\rho,\mu)$ for all $k\ge 1$. Specifically, we have

$$f(x^k)+\rho\sum_{i\in\mathcal{I}}s_i^k-\mu^2\sum_{i\in\mathcal{I}}\log\big((s_i^k)^2-c_i(x^k)\big)-\mu\sum_{i\in\mathcal{I}}\log s_i^k\le\phi_{B,\frac{1}{2}}(x^0,s^0;\rho,\mu),\tag{38}$$

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 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 (38) 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 $\lim\limits_{k\to\infty}\frac{\mu^2\sum_{i\in\mathcal{I}}\log\left((s_i^k)^2-c_i(x^k)\right)}{\sum\limits_{i\in\mathcal{I}}s_i^k}=0$, $\lim\limits_{k\to\infty}\frac{\mu\sum_{i\in\mathcal{I}}\log s_i^k}{\sum\limits_{i\in\mathcal{I}}s_i^k}=0$ and the right hand side of inequality (38) is bounded. Therefore, we have proved that the sequence $\{s^k\}$ is bounded above, so is the sequence $\{(s^k)^2-c(x^k)\}$. There exists a vector $s^*\in\mathbb{R}^m$ (taking a subsequence of the sequence $\{s^k\}$ if necessary) such that $\lim\limits_{k\to\infty}s^k=s^*$. Similarly, we can prove that $\lim\limits_{k\to\infty}(s^k)^2-c(x^k)=(s^*)^2-c(x^*)>0$ and $s^*>0$. The last part can be proved by virtue of the rules (30)-(31) for updating the dual multipliers. Here, the details are omitted. ■

Lemma 3.6 *Let the penalty parameter $\rho>0$ and barrier parameter $\mu>0$ be fixed. Assume that Assumptions 2-4 hold. Then the sequence $\{(\hat{y}^k,\hat{u}^k)\}$ generated by Algorithm 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 $\lim_{k\to\infty} x^k = x^*$ and $\lim_{k\to\infty} H^k = H^*$. By Assumption 2, we have that

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

It follows from inequality (27) there exists a positive definite matrix $\mathcal{M}^*$ such that $\lim_{k\to\infty} \mathcal{M}^k = \mathcal{M}^*$. By Lemma 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

$$(s^*)^2 - c(x^*) > 0,\ \|s^*\| \le M,\ \|(s^*)^2 - c(x^*)\| \le M,\ \|(y^*, u^*)\| \le M.$$

It follows from equation (23) that we have

$$\begin{aligned}
\hat{y}^k &= \big((S^k)^2 - C(x^k)\big)^{-1}\big(\mu^2 e - 2Y^k S^k \triangle s^k + Y^k A(x^k)^T \triangle x^k\big) \\
\hat{u}^k &= (S^k)^{-1}\big(\mu e - U^k \triangle s^k\big).
\end{aligned}$$

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 (24), we have

$$\mathcal{M}^k \begin{pmatrix} \triangle x^k \\ \triangle x^k \end{pmatrix} = \begin{pmatrix} -\nabla f(x^k) - \mu^2 A(x^k)\big((S^k)^2 - C(x^k)\big)^{-1} e \\ 2\mu^2 S^k \big((S^k)^2 - C(x^k)\big)^{-1} e + \mu (S^k)^{-1} e - \rho e \end{pmatrix}$$

Taking limit as $k \to \infty$ on both sides of the above 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 have proved this lemma. ■

Similar to the proof of (6, Lemma 4.11), we can prove the next lemma. Here the details are omitted.

Lemma 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 1 is bounded from above and $\|(\triangle x^k, \triangle s^k)\| \to 0$ as $k \to \infty$.*

Next we prove that the sequence $\{(x^k, s^k)\}$ generated by Algorithm 1 converges to an approximate KKT point of problem (4).

Theorem 3.1 *Let the penalty parameter $\rho > 0$ and the barrier parameter $\mu > 0$ be fixed. Suppose that Assumptions 2-4. Then the sequence $\{(x^k, s^k)\}$ generated by Algorithm 1 converges to a KKT point of problem (4).*

Proof. By Assumption 3, Lemmas 3.5 and 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 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

$$(s^*)^2 - c(x^*) > 0,\ \|s^*\| \le M,\ \|(s^*)^2 - c(x^*)\| \le M,\ \|(y^{**}, u^{**})\| \le M.$$

By Lemma 3.7, we have $\|(\triangle x^k, \triangle s^k)\| \to 0$ as $k \to \infty$. At the k-th iteration, by (22), we have

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

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

$$\begin{aligned}
\nabla f(x^*) + A(x^*)y^* &= 0,\\
2S^* y^* + u^* &= \rho e,\\
((S^*)^2 - C(x^*))y^* &= \mu^2 e,\\
S^* u^* &= \mu e.
\end{aligned}$$

Therefore, we have proved that the sequence $\{(x^k, s^k)\}$ converges to a KKT point of problem (4). ∎

We establish the convergence results of the sequence $\{(x^j, s^j)\}$ generated by Algorithm 2 in the next theorem.

Theorem 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. 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 (3);

(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 (3).

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 3.7, there exist a vector $s^* \geq 0$ and a constant $M > 0$ such that $(s^j)^2 - c(x^j) \to (s^*)^2 - c(x^*) \geq 0$ and $s^j \to s^* \geq 0$ as $j \to \infty$; moreover, $\|(s^*)^2 - c(x^*)\| \leq M$ 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 (33a) and (33b) by ϖ^j and taking limit as $j \to \infty$, we reach that

$$\begin{aligned} A(x^*)\bar{y} &= 0, \\ 2S^*\bar{y} + \bar{u} &= 0, \\ ((S^*)^2 - C(x^*))\bar{y} &= 0, \\ S^*\bar{u} &= 0, \end{aligned}$$

and $(\bar{y}, \bar{u}) \geq 0$. Consequently, we conclude that the limit point (x^*, s^*) is a FJ point of problem (3).

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 implies that $\epsilon_{\mu^j} \to 0$ as $j \to \infty$. By (33a), we conclude that $\lim_{j\to\infty} \mathrm{Res}(x^j, s^j, \hat{y}^j, \hat{u}^j; \rho, \mu^j) = \mathrm{Res}(x^*, s^*, y^*, u^*; \rho, 0) = 0$. Specifically, we have

$$\begin{aligned} \nabla f(x^*) + A(x^*)y^* &= 0, \\ \rho e - 2S^*y^* - u^* &= 0, \\ ((S^*)^2 - C(x^*))y^* &= 0, \\ S^*u^* &= 0. \end{aligned}$$

By (33b), we have $(y^*, u^*) \geq 0$. Combining $(s^*)^2 - c(x^*) \geq 0$ and $s^* \geq 0$, we have proved that (x^*, s^*) is a KKT point of problem (3). ■

We are now ready to prove the global convergence results of Algorithm 3.

Theorem 3.3 *Suppose that Assumptions* 1-4 *hold and that the sequence* $\{(x^i, s^i, \hat{y}^i, \hat{u}^i)\}$ *is generated by Algorithm* 3*. Moreover, we assume that the sequence* $\{s^i\}$ *is bounded above. Then we conclude that:*

(i) If 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 (1);

(ii) If the penalty parameter ρ^i goes to infinite, then the sequence $\{x^i\}$ converges to a FJ point of problem (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 $\|s^i\| = 0$ is satisfied. If the sequence $\{(\hat{y}^i, \hat{u}^i)\}$ is bounded, by Theorem 3.2, the sequence $\{(x^i, s^i, \hat{y}^i, \hat{u}^i)\}$ satisfies the conditions as follows

$$\begin{aligned} \nabla f(x^i) + A(x^i)\hat{y}^i &= 0, \\ \rho^i e - 2S^i\hat{y}^i - \hat{u}^i &= 0, \\ ((S^i)^2 - C(x^i))\hat{y}^i &= 0, \\ S^i\hat{u}^i &= 0, \\ (s^i)^2 - c(x^i) &\geq 0, \\ s^i &\geq 0, \\ (\hat{y}^i, \hat{u}^i) &\geq 0, \end{aligned} \tag{42}$$

which reduces to the KKT conditions of problem (1) since $\|s^i\| = 0$ at the final iteration. Therefore, we have proved the statement (i).

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

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

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 (33a) and (33b) by ϖ^i and taking limit as $i \to \infty$, we reach that x^* is a FJ point of problem (1). Therefore, we have proved the statement (ii). ■

4 Numerical Experiments

In this section, we present numerical results for the proposed method 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 1-3 as the IPLOP method, which stands for the Interior-Point Lower-Order Penalty method. We carry out the numerical experiments on 266 inequality constrained optimization problems from the CUTEr collection[1], COPS[2], MITT[3] and GLOBAL Library[4] test sets. See Table 2. 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 implemented in PIPAL1.0[5] in (10) in terms of the number of iterations and the values of the penalty parameter.

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

In the implementation, we use the same initial point $x^0 \in \mathbb{R}^n$ as the one provided for every test problem from the test problem collections and set $s_i^0 = \sqrt{\max\{c_i(x^0), 0\}} + \frac{1}{2}$ for all $i \in \mathcal{I}$ unless specified otherwise. We set MaxiterI=1000, that is, the maximum number of iterations for Algorithm 1 is 1000, and similarly we also set MaxiterII=1000 and MaxiterIII=1000 for Algorithm 2 and Algorithm 3, respectively.

Next, we illustrate our strategy for choosing δ large enough such that the matrix $\mathcal{M}$ (see (25)) with $\widehat{H}(x, s, y)$ being replaced by $\widehat{H}(x, s, y) + \delta E$ is sufficiently positive definite. However, we would also like to keep it as small as possible as well in order to make our method work more efficiently in practice, as large values of δ will make the algorithm behave like a steepest descent method, which is not desirable. Since $\mathcal{M}$ 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 (35; 36) 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 (24), the steplength $\alpha_P^k \in (0, 1]$ has to be determined in order to obtain the next iterate by (28). In our implementation, we first obtain $\bar{\alpha}_P^k := \max\{\ \bar{\rho}_1^j \mid j = 0, 1, 2, \ldots\}$ with $\bar{\rho}_1 \in (0, 1)$ such that (29c) holds. Then, we

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

let $\alpha_P^k := \max\{\bar{\alpha}_P^k \bar{\rho}_2^j \mid j = 0, 1, 2, \ldots\}$ with $\bar{\rho}_2 \in (0, 1)$ satisfying

$$(s^{k+1})^2 - c(x^{k+1}) \geq (1 - \hat{\eta})\Big((s^k)^2 - c(x^k)\Big), \tag{43a}$$

$$s^{k+1} \geq (1 - \hat{\eta}) s^k, \tag{43b}$$

where $\hat{\eta} = \max\{0.99, 1 - \mu\}$ in our implementation. Obviously, (43a)-(43b) imply (29a)-(29b). The modification (29a) as (43a) is due to the nonlinearity of $(s^{k+1})^2$ in (29a), in which case the classical fraction-to-boundary rule cannot be used anymore. The above strategy of computing steplength α_P^k is efficient in our numerical experiments. In Algorithm 2, we set $\epsilon_{\mu^j} = \mu^j$ and $\epsilon_{\mu^{j+1}} = \max\{\gamma \epsilon_{\mu^j}, 10^{-7}\}$.

The default values of input parameters are listed in Table 1 below.

Table 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}		

Table 2: Names of the inequality constrained optimization problems

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

Table 2: (continued)

Problem	Problem	Problem	Problem	Problem
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
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				

Our results show that there are 6, 10 and 9 test problems which cannot be solved successfully by the IPLOP, PIPAL-a and PIPAL-c methods, respectively. So we plot only the corresponding results for solved ones. Using the performance profiles of Dolan and Moré in (11), we plot Figure 1, where the plots $\pi_s(\tau)$ denote the scaled performance profile

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

where $\log_2(r_{\hat{p},s})$ is the scaled performance ratio between the iteration number to solve problem $\hat{p}$ by solver s over the fewest iteration number required by the solvers IPLOP, PIPAL-a and PIPAL-c. 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 (11) for more details regarding the performance profiles.

Figure 1 shows that the IPLOP method uses the least number of iterations on approximate 58% of test problems and shares the nearly same robustness with other two solvers.

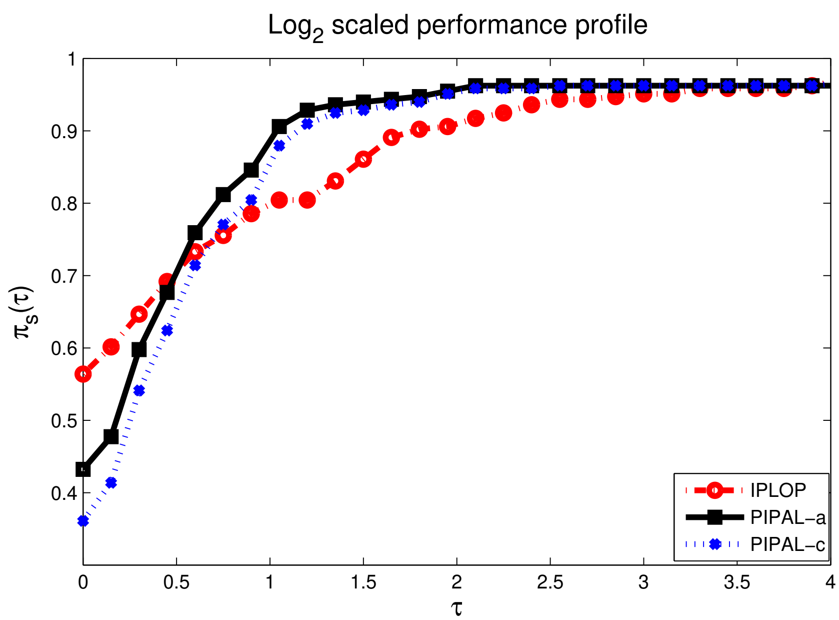

Figure 1: Performance profiles based on the number of iterations for the IPLOP, PIPAL-a and PIPAL-c methods.

Figure 2 is plotted by the values of the penalty parameter ρ, which shows that the IPLOP 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.

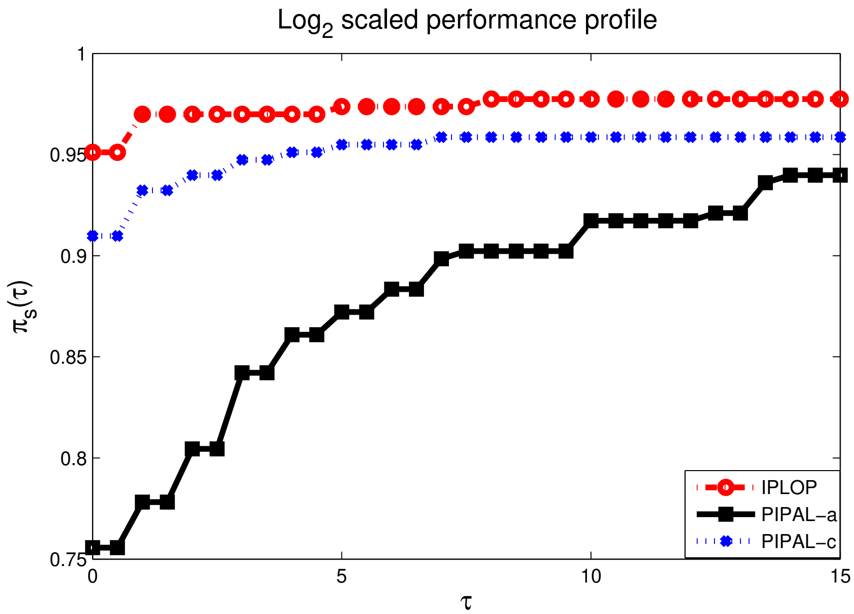

Figure 2: Performance profiles based on the values of the penalty parameter for the IPLOP, PIPAL-a and PIPAL-c methods.

We plot Figure 3 using the CPU time for the IPLOP, PIPAL-a and PIPAL-c methods, which shows that the proposed method outperforms both the PIPAL-a and PIPAL-c methods in CPU time.

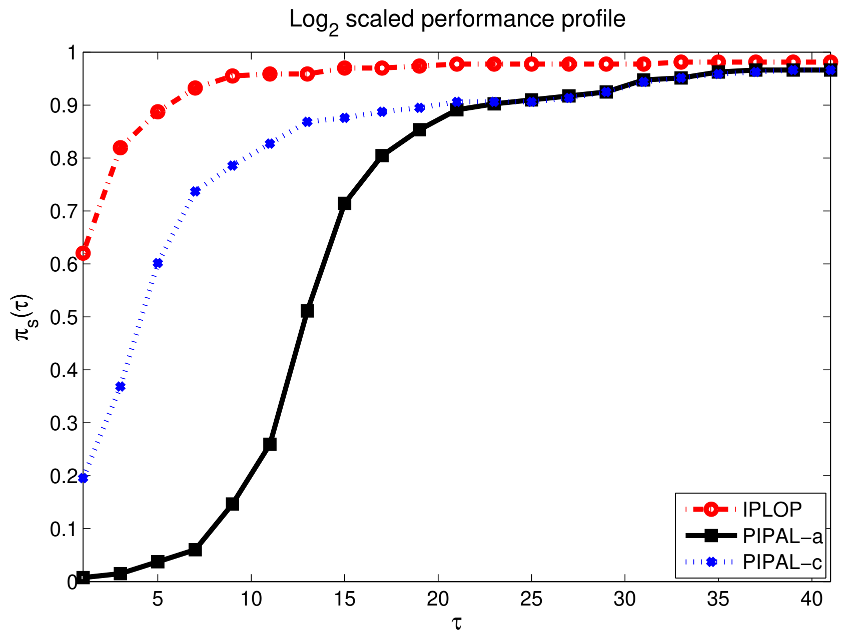

Figure 3: Performance profiles based on CPU time for the IPLOP, PIPAL-a and PIPAL-c methods.

5 Conclusion

We have proposed an interior-point $\ell_{\frac{1}{2}}$-penalty method for solving the inequality constrained nonlinear optimization by virtue of the $\ell_{\frac{1}{2}}$-penalty function. The quadratic relaxation was introduced and different kinds of constraint qualifications were investigated to establish the first-order necessary conditions of the quadratically relaxed problem. Moreover, we proved the global convergence of the proposed method under mild conditions. We conducted our numerical experiments on 266 inequality constrained optimization problems to compare the performance of the proposed method with existing interior-point ℓ_1-penalty methods in terms of the number of iterations, the values of the penalty parameter and the CPU time. Numerical results indicate that the $\ell_{\frac{1}{2}}$-penalty function is competitive with the ℓ_1-penalty function from the view of numerical implementation. However, there are many other issues that are needed to deal with in our future work. We summarize them as follows.

(I) As pointed out by Fletcher (14) 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 (3; 4) to update the penalty parameter for the ℓ_1-

penalty method. It is well-known that the smallest exact penalty parameter of the $\ell_{\frac{1}{2}}$-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 has not been studied in the paper.

(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 (10) with the same stopping criterion on 38 test problems with degenerate constraints and 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) In our further research, we will introduce an interior-point $\ell_{\frac{1}{p}}$-penalty method for solving the inequality constrained optimization problems by combining the interior-point method and extending the quadratic relaxation to the p-order relaxation for the $\ell_{\frac{1}{p}}(p > 1)$-penalty function.

Acknowledgments. The authors sincerely thank Professor Weijun Zhou from School of Mathematics and Computing Science of Changsha University of Science & Technology for his valuable comments on the draft version. Moreover, the authors sincerely thank the two anonymous referees for their careful reading of the paper and their suggestions and questions, all of which greatly improved this paper. The first author was supported by the Fundamental Research Funds for the Central Universities, the second author was supported by grants from the Research Grants Council of Hong Kong(PolyU 5295/12E) and the third author was supported by the NSF(11201383) of China.

References

[1] 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.

[2] H. Y. Benson, D. F. Shanno and R. J. Vanderbei, Interior-point methods for nonconvex nonlinear programming: Jamming and numerical testing, *Mathematical Programming*, **99** (2004), 35–48.

[3] R. H. Byrd, G. Lopez-Calva and J. Nocedal, A line search exact penalty method using steering rules, *Mathematical Programming*, **133** (2012), 39–73.

[4] R. H. Byrd, J. Nocedal and R. A. Waltz, Steering exact penalty methods for nonlinear programming, *Optimization Methods and Software*, **23** (2008), 197–213.

[5] R. Chartrand and V. Staneva, Restricted isometry properties and nonconvex compressive sensing, *Inverse Problems*, **24** (2008), 1–14.

[6] L. Chen and D. Goldfarb, Interior-point ℓ_2-penalty methods for nonlinear programming with strong global convergence properties, *Mathematical Programming*, **108** (2006), 1–36.

[7] L. Chen and D. Goldfarb, *On the fast local convergence of interior-point ℓ_2-penalty methods for nonlinear programming*, Technical report, Columbia University, 2006.

[8] X. J. Chen, Smoothing methods for nonsmooth, nonconvex minimization, *Mathematical Programming*, **134** (2012), 71–99.

[9] 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** (2000), 215–249.

[10] F. E. Curtis, A penalty-interior-point algorithm for nonlinear constrained optimization, *Mathematical Programming Computation*, **4** (2012), 181–209.

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

[12] C. Durazzi, On the Newton interior-point method for nonlinear programming problems, *Journal of Optimization Theory and Applications*, **104** (2000), 73–90.

[13] 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** (1996), 507–541.

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

[15] P. E. Gill, W. Murray and M. H. Wright, *Practical Optimization*, Academic Press, 1981.

[16] 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.

[17] M. Guignard, Generalized Kuhn-Tucker conditions for mathematical programming problems in a Banach space, *SIAM Journal on Control*, **7** (1969), 232–241.

[18] X. X. Huang and X. Q. Yang, Convergence analysis of a class of nonlinear penalization methods for constrained optimization via first-order necessary optimality conditions, *Journal of Optimization Theory and Applications*, **116** (2003), 311–332.

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

[20] X. W. Liu and J. Sun, A robust primal-dual interior-point algorithm for nonlinear programs, *SIAM Journal on Optimization*, **14** (2004), 1163–1186.

[21] 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** (2005), 609–628.

[22] Z. Q. Luo, J. S. Pang and D. Ralph, *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, 1996.

[23] K. W. Meng, S. J. Li and X. Q. Yang, A robust SQP method based on a smoothing lower order penalty functions, *Optimization*, **58** (2009), 23–38.

[24] K. W. Meng and X. Q. Yang, First- and second-order necessary conditions via exact penalty functions, *Submitted*.

[25] K. W. Meng and X. Q. Yang, Optimality conditions via exact penalty functions, *SIAM Journal on Optimization*, **20** (2010), 3208–3231.

[26] 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** (2006), 375–398.

[27] M. Mongeau and A. Sartenaer, Automatic decrease of the penalty parameter in exact penalty function methods, *European Journal of Operational Research*, **83** (1995), 686–699.

[28] A. S. Nemirovski and M. J. Todd, Interior-point methods for optimization, *Acta Numerica*, **17** (2008), 191–234.

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

[30] I. Pólik and T. Terlaky, Interior point methods for nonlinear optimization, *In: G. Di Pillo, F. Schoen, eds., Nonlinear Optimization*, 215–276.

[31] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*, Springer, 2011.

[32] A. M. Rubinov, B. M. Glover and X. Q. Yang, Decreasing functions with applications to penalization, *SIAM Journal on Optimization*, **10** (1999), 289–313.

[33] D. F. Shanno and R. J. Vanderbei, Interior-point methods for nonconvex nonlinear programming: Orderings and higher-order methods, *Mathematical Programming*, **87** (2000), 303–316.

[34] G. Still and M. Streng, Optimality conditions in smooth nonlinear programming, *Journal of Optimization Theory and Applications*, **90** (1996), 483–515.

[35] R. J. Vanderbei and D. F. Shanno, An interior-point algorithm for nonconvex nonlinear programming, *Computational Optimization and Applications*, **13** (1999), 231–252.

[36] 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** (2006), 25–57.

[37] S. J. Wright, *Primal-dual Interior-Point Methods*, SIAM, 1987.

[38] Z. B. Xu, X. Y. Chang, F. M. Xu and H. Zhang, $L_{1/2}$ Regularization: A thresholding representation theory and a fast solver, *Neural Networks and Learning Systems, IEEE Transactions on*, **23** (2012), 1013–1027.

[39] X. Q. Yang and Z. Q. Meng, Lagrange multipliers and calmness conditions of order p, *Mathematics of Operations Research*, **32** (2007), 95–101.

[40] 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** (2003), 351–364.

[41] Y. Y. Ye, *Interior Point Algorithms: Theory and Analysis*, Wiley-Interscience, 1997.