JOURNAL OF INDUSTRIAL AND
MANAGEMENT OPTIMIZATION

doi:10.3934/jimo.2018190

A PRIMAL-DUAL INTERIOR-POINT METHOD CAPABLE OF RAPIDLY DETECTING INFEASIBILITY FOR NONLINEAR PROGRAMS

Yu-Hong Dai

LSEC, ICMSEC, Academy of Mathematics and Systems Science
Chinese Academy of Sciences, Beijing 100190, China
He is affiliated to School of Mathematical Sciences, University of Chinese Academy of Sciences

Xin-Wei Liu*

Institute of Mathematics, Hebei University of Technology
Tianjin 300401, China

Jie Sun

School of Science, Curtin University, Perth, Australia
and School of Business, National University of Singapore, Singapore

(Communicated by Wenxun Xing)

Abstract. With the help of a logarithmic barrier augmented Lagrangian function, we can obtain closed-form solutions of slack variables of logarithmic-barrier problems of nonlinear programs. As a result, a two-parameter primal-dual nonlinear system is proposed, which corresponds to the Karush-Kuhn-Tucker point and the infeasible stationary point of nonlinear programs, respectively, as one of two parameters vanishes. Based on this distinctive system, we present a primal-dual interior-point method capable of rapidly detecting infeasibility of nonlinear programs. The method generates interior-point iterates without truncation of the step. It is proved that our method converges to a Karush-Kuhn-Tucker point of the original problem as the barrier parameter tends to zero. Otherwise, the scaling parameter tends to zero, and the method converges to either an infeasible stationary point or a singular stationary point of the original problem. Moreover, our method has the capability to rapidly detect the infeasibility of the problem. Under suitable conditions, the method can be superlinearly or quadratically convergent to the Karush-Kuhn-Tucker point if the original problem is feasible, and it can be superlinearly or quadratically convergent to the infeasible stationary point when the problem is infeasible. Preliminary numerical results show that the method is efficient in solving some simple but hard problems, where the superlinear convergence to an infeasible stationary point is demonstrated when we solve two infeasible problems in the literature.

2010 *Mathematics Subject Classification.* Primary: 90C30, 90C51; Secondary: 90C26.

Key words and phrases. Nonlinear programming, constrained optimization, infeasibility, interior-point method, global and local convergence.

The first draft of this paper was completed on December 2, 2014. The first author is supported by the Chinese NSF grants (nos. 11631013, 11331012 and 81173633) and the National Key Basic Research Program of China (no. 2015CB856000). The second author is supported by the Chinese NSF grants (nos. 11671116 and 11271107) and the Major Research Plan of the NSFC (no. 91630202). The third author is supported by Grant DP-160101819 of Australia Research Council.

* Corresponding author: xxxx.

1. **Introduction.** Developing effective methods for nonlinear programs has always been an active area in optimization research. There are many interesting works in this area in recent years, which focus on various aspects of nonlinear programs. It is well known that, without assuming any constraint qualification, a local solution of nonlinear programs can be either a Karush-Kuhn-Tucker (KKT) point or a singular stationary point that is a Fritz-John (FJ) point at which the Mangasarian-Fromovitz constraint qualification (MFCQ) does not hold. A method is said to have strong global convergence if it can find either a KKT point or a singular stationary point, or even an infeasible point with first-order stationarity (i.e., an infeasible stationary point) for minimizing some kind of measure of constraint violations.

There are already many methods for nonlinear programs in the literature which are proved to have strong global convergence (see, for example, [1, 5, 6, 13, 24, 25, 26, 27, 39]). Some of them are shown to be of locally superlinear/quadratic convergence to the KKT point. However, it has been an open problem whether these methods are capable of rapidly converging to an infeasible stationary point before Byrd, Curtis and Nocedal [8] creatively presented a set of conditions to guarantee the superlinear convergence of their SQP algorithm to an infeasible stationary point. More recently, Burke, Curtis and Wang [5] considered the general program with equality and inequality constraints, and proved that their SQP method has strong global convergence and have rapid convergence to the KKT point, and have superlinear/quadratic convergence to an infeasible stationary point.

The aim of this paper is to present a primal-dual interior-point method capable of converging to an infeasible stationary point when a nonlinear constrained optimization problem is infeasible. In addition, this method is of strong global convergence and locally rapid convergence to the KKT point when the problem is feasible. Consider the nonlinear program with general inequality constraints

$$\text{minimize} \quad (\min) \quad f(x) \tag{1}$$

$$\text{subject to} \quad (\text{s.t.}) \quad c_i(x) \le 0, \ i \in \mathcal{I}, \tag{2}$$

where $x \in \Re^n$, $\mathcal{I} = \{1, 2, \ldots, m\}$ is an index set, f and c_i $(i \in \mathcal{I})$ are twice continuously differentiable real-valued functions defined on $\Re^n$. By introducing slack variables to the inequality constraints, problem (1)–(2) is reformulated as the program with equality and nonnegative constraints as follows:

$$\min \quad f(x) \tag{3}$$

$$\text{s.t.} \quad c_i(x) + y_i = 0, \ i \in \mathcal{I}, \tag{4}$$

$$y_i \ge 0, \ i \in \mathcal{I}, \tag{5}$$

where y_i $(i \in \mathcal{I})$ are slack variables.

The interior-point approach has been shown to be robust and efficient in solving linear and nonlinear programs (for example, see [2, 3, 9, 10, 13–22, 25, 27, 28, 31–33, 35, 36]. Among all interior-point methods, the primal-dual interior-point methods have drawn considerable attention. It is noted that, other than some feasible interior-point methods which requires all iterates to be (strictly) feasible for constraints, most of efficient interior-point methods for nonlinear programs are presented by combining a distinctive penalty strategy. These methods can roughly be summarized into three kinds by the order of using the penalty technique. The first kind of methods reformulate the original program to a problem with only equality constraints by interior barrier technique and then prompt the global convergence of these methods by different penalty functions, such as [9, 10, 13], [18]–[21], [25, 27, 28, 31]. The second kind of methods first use the penalty strategy to obtain a new formulation of the original program with only inequality constraints and then use the interior-point methods to solve the formulation, such as [22]. The third kind of methods use both penalty strategy and interior-point technique to transform the original problem with inequality constraints into a new formulation with only equality constraints (see [14]). The co-existence of penalty and barrier parameters brings new challenge to this kind of methods. As a reward to the challenge, the last kind of methods can be expected to have some exclusive global and/or local convergence properties such as the rapid detection of infeasibility.

Although every interior-point method has its novelty, they share some common features, for example, the iterates are usually the approximate solutions of some parametric primal-dual nonlinear system which converges to the KKT conditions of the original problem as the barrier parameter tends to zero, and should be interior points for nonnegative constraints. The interior-point condition can result in truncation of the step, which may cause the failure of global convergence to the KKT point even for a well-posed problem (see [34] for a counterexample). Besides, it could make the local convergence analysis of the primal-dual interior-point methods much sophisticated (e.g., [2, 11, 18, 20, 21, 37, 38]). By introducing the null-space technique, some interior-point methods such as [13, 25, 27] have been proved not to suffer the failure of global convergence. They have strong global convergence and can converge to an infeasible stationary point when the problem is infeasible, but they cannot detect the infeasibility *rapidly*.

Similar to the first kind of interior-point methods mentioned above, we consider the logarithmic-barrier problem

$$\min \quad f(x) - \beta \sum_{i\in\mathcal{I}} \ln y_i \tag{6}$$

$$\text{s.t.} \quad c_i(x) + y_i = 0,\ i \in \mathcal{I}, \tag{7}$$

where $\beta > 0$ is the barrier parameter, $y_i > 0$ ($i \in \mathcal{I}$) (that is, (x, y) is an interior point). With the help of a logarithmic barrier augmented Lagrangian function, we can obtain closed-form solutions of all slack variables of logarithmic-barrier problems of nonlinear programs. As a result, a two-parameter primal-dual nonlinear system is proposed, which corresponds to the KKT point and the infeasible stationary point of nonlinear programs, respectively, as one of two parameters vanishes. Based on this distinctive system, we present a primal-dual interior-point method capable of rapidly detecting infeasibility of nonlinear programs. Our method generates interior-point iterates without truncation of the step and can detect the infeasibility of the problem rapidly. It should be noted that rapid detection of infeasibility is one of important features of newly developed penalty-interior-point algorithm (see [14]) and SQP methods (see [5, 8]), and is a very useful property in practice.

Our method has similarity to the existing interior-point methods for nonlinear programs. Similar to [9, 10, 25, 27], we consider the problem with slack variables (6)–(7) and use similar null-space technique and the technique for updating slack variables. But unlike those existing methods, our method is based on a distinctive primal-dual system and uses a different merit function dependent on both primal and dual variables, which is similar to [18, 21]. We note that [18, 21] also use augmented Lagrangian functions in developing their interior-point methods, but they are not based on the problem (6)–(7) and have a different flavor with our method.

A recent work on interior-point methods is [14] which solves a two-parameter subproblem (or correspondingly a two-parameter primal-dual nonlinear system), but that system can only be proved to be asymptotically approximate the KKT conditions of the original problem as the barrier parameter tends to zero. Curtis [14] and Nocedal, Öztoprak and Waltz [28] have shown by numerical experiments that their interior-point methods have the ability to detect the infeasibility, but no theoretical proof is provided to show that those methods can detect infeasibility at quadratic or superlinear rate as we shall do.

Without assuming any constraint qualification or requiring any feasibility of constraints, we prove that our method globally converges to a KKT point of the original problem as the barrier parameter tends to zero. Otherwise, the scaling parameter tends to zero, and the method globally converges to either an infeasible stationary point or a singular stationary point of the original problem. Under suitable local conditions, we prove that the method not only can be superlinearly or quadratically convergent to the KKT point if the original problem is feasible, but also can be superlinearly or quadratically convergent to the infeasible stationary point if the problem is infeasible. Preliminary numerical results show that the method is efficient in solving some small but hard problems in the literature. The superlinear convergence have also been observed when we solve the infeasible problems given by [8].

This paper is organized as follows. In Section 2, we first give a closed-form solution on slack variables of the KKT system of the logarithmic barrier problem (6)–(7). A corresponding two-parameter primal-dual nonlinear system is followed. Then we describe our algorithm in Section 3. The strong global convergence results on the algorithm are proved in Section 4. In Section 5, under suitable assumptions, we show that the algorithm can be of locally quadratic or superlinear convergence to the KKT point or the infeasible stationary point of the original problem. The algorithm is implemented in Section 6, where preliminary numerical results for some small but hard problems from literature are reported. We conclude our paper in Section 7.

Throughout the article, a letter with subscript k (or l) is related to the kth (or lth) iteration, the subscript i indicates the ith component of a vector or the ith column of a matrix, and the subscript ki (or li) is the ith component of a vector or the ith column of a matrix at the kth (or lth) iteration. All vectors are column vectors, and $z = (x, u)$ means $z = [x^T, u^T]^T$, where "T" stands for the transpose. The expression $\theta_k = O(\tau_k)$ means that there exists a constant M independent of k such that $|\theta_k| \le M|\tau_k|$ for all k large enough, and $\theta_k = o(\tau_k)$ indicates that $|\theta_k| \le \epsilon_k|\tau_k|$ for all k large enough with $\lim_{k\to 0} \epsilon_k = 0$. If it is not specified, I is an identity matrix whose order may be showed in the subscript or be clear in the context, $\|\cdot\|$ is the Euclidean norm, $|\mathcal{S}|$ is the cardinality of set $\mathcal{S}$. For simplicity, we also use simplified notations for functions, such as $f_k = f(x_k)$, $\nabla f_k = \nabla f(x_k)$, $c_{ki} = c_i(x_k)$, $\nabla c_{ki} = \nabla c_i(x_k)$ and so on.

2. **A two-parameter primal-dual system for nonlinear programs.**

With the help of a logarithmic barrier augmented Lagrangian function, we can derive closed-form solutions on slack variables of the logarithmic barrier problem (6)–(7). A primal-dual nonlinear system with barrier and scaling parameters is then followed. Its solution corresponds to the KKT point and the infeasible stationary point of program (1)–(2), respectively, as one of two parameters vanishes. Based on this system, we present our primal-dual interior-point algorithm for nonlinear progrmas (1)–(2).

We consider the augmented Lagrangian function for the logarithmic barrier problem (6)–(7)

$$P_{(\beta,\rho)}(x,y,u)=\rho\left[f(x)-\beta\sum_{i\in\mathcal{I}}\ln y_i+u^T(c(x)+y)\right]+\frac{1}{2}\|c(x)+y\|^2, \tag{8}$$

where $\beta>0$ is the barrier parameter, $\rho>0$ is a scaling parameter, $c(x)=(c_i(x),\ i\in\mathcal{I})\in\Re^m$, $y=(y_i,\ i\in\mathcal{I})\in\Re^m$, u is a vector in $\Re^m$. The stationary conditions on $P_{(\beta,\rho)}(x,y,u)$ suggest the following equations:

$$\left\{\begin{array}{l}\rho\nabla f(x)+\sum_{i\in\mathcal{I}}[\rho u_i+c_i(x)+y_i]\nabla c_i(x)=0,\\ -\rho\beta y_i^{-1}+\rho u_i+c_i(x)+y_i=0,\quad i\in\mathcal{I},\\ \rho(c(x)+y)=0.\end{array}\right. \tag{9}$$

By multiplying y_i on both sides of the second equation, one has the equation

$$y_i^2+(c_i(x)+\rho u_i)y_i-\rho\beta=0,\quad i\in\mathcal{I}. \tag{10}$$

Thus, we have closed-form solutions on slack variables

$$y_i=\frac{1}{2}\left[\sqrt{(c_i(x)+\rho u_i)^2+4\rho\beta}-(c_i(x)+\rho u_i)\right],\quad i\in\mathcal{I},$$

where the negative root is not taken since $y_i>0$. Therefore,

$$c_i(x)+y_i=\frac{1}{2}\left[\sqrt{(c_i(x)+\rho u_i)^2+4\rho\beta}+(c_i(x)-\rho u_i)\right],\quad i\in\mathcal{I}. \tag{11}$$

If we set $\lambda_i=\rho u_i+c_i(x)+y_i$ for $i\in\mathcal{I}$, one has

$$\lambda_i=\frac{1}{2}[\sqrt{(c_i(x)+\rho u_i)^2+4\rho\beta}+(c_i(x)+\rho u_i)],\quad i\in\mathcal{I}. \tag{12}$$

Using (11) and (12), equations in (9) can be reformulated as the following system of equations on unknowns (x,u):

$$\left\{\begin{array}{l}\rho\nabla f(x)+\sum_{i\in\mathcal{I}}\frac{1}{2}\left[\sqrt{(c_i(x)+\rho u_i)^2+4\rho\beta}+(c_i(x)+\rho u_i)\right]\nabla c_i(x)=0,\\ \frac{1}{2}\rho\left[\sqrt{(c_i(x)+\rho u_i)^2+4\rho\beta}+(c_i(x)-\rho u_i)\right]=0,\quad i\in\mathcal{I},\end{array}\right. \tag{13}$$

where $\beta>0$ and $\rho>0$ are two parameters.

It is noted that, if $\beta=0$ and $\rho>0$, equations in (13) are reduced to the equations

$$\rho\nabla f(x)+\sum_{i\in\mathcal{I}}\frac{1}{2}[|c_i(x)+\rho u_i|+(c_i(x)+\rho u_i)]\nabla c_i(x)=0, \tag{14}$$

$$\frac{1}{2}[|c_i(x)+\rho u_i|+(c_i(x)-\rho u_i)]=0,\quad i\in\mathcal{I}. \tag{15}$$

Define index sets $\mathcal{A}(x)=\{i\in\mathcal{I}|c_i(x)+\rho u_i\geq 0\}$ and $\mathcal{N}(x)=\{i\in\mathcal{I}|c_i(x)+\rho u_i<0\}$. Then, by (15), for any solution (x,u) of the system (13) (if there exists), one has $c_i(x)=0$ for $i\in\mathcal{A}(x)$ and $u_i=0,\ i\in\mathcal{N}(x)$. Thus, $c_i(x)<0$ for $i\in\mathcal{N}(x)$, and (14) implies

$$\nabla f(x)+\sum_{i\in\mathcal{A}(x)}u_i\nabla c_i(x)=0. \tag{16}$$

Consequently, (x,u) is a KKT pair of the original problem (1)–(2).

If $\rho = 0$, the first equation in (13) is reduced to the equation

$$\sum_{i\in\mathcal{I}} \frac{1}{2}[|c_i(x)| + c_i(x)]\nabla c_i(x) = 0, \tag{17}$$

which shows that, if (x, u) satisfies the system (13), and x is infeasible to the problem (1)–(2), then x is a stationary point for minimizing $\frac{1}{2}\|\max(0, c(x))\|^2$, i.e., a stationary point for minimizing the ℓ_2 measure of residuals of the constraints, which is also called as an infeasible stationary point of problem (1)–(2) (see Definition 4.1).

The preceding argument shows that the proposed system (13) not only can reduce to the KKT conditions of the original problem as parameter β vanishes, but also can reduce to the stationary condition of an infeasible stationary point of the original problem as parameter ρ is zero. This feature is distinguished from all primal-dual systems used by the existing interior-point methods. It turns out that is a favorable and important characterization, since we want to develop an interior-point method which not only can converge to a KKT point of the original problem as the problem is feasible, but also can converge to an infeasible stationary point of the original problem as it is infeasible.

In next section, we will develop our primal-dual interior-point method for nonlinear programs based on the two-parameter system (13). For convenience of statement, we denote, for $i \in \mathcal{I}$,

$$y_i(x, u; \beta, \rho) = \frac{1}{2}\left[\sqrt{(c_i(x) + \rho u_i)^2 + 4\rho\beta} - (c_i(x) + \rho u_i)\right], \tag{18}$$

$$\lambda_i(x, u; \beta, \rho) = \frac{1}{2}\left[\sqrt{(c_i(x) + \rho u_i)^2 + 4\rho\beta} + (c_i(x) + \rho u_i)\right]. \tag{19}$$

That is, λ_i and y_i ($i \in \mathcal{I}$) are functions on (x, u), and are dependent on parameters β and ρ. If it is not confused in the context, we may use $\lambda_i = \lambda_i(x, u; \beta, \rho)$ and $y_i = y_i(x, u; \beta, \rho)$ for simplicity. Thus, $\lambda_i y_i = \rho\beta$ for $i \in \mathcal{I}$. Using (18) and (19), the two-parameter system (13) can be written as the concise form

$$\begin{cases} \rho\nabla f(x) + \sum_{i\in\mathcal{I}} \lambda_i(x, u; \beta, \rho)\nabla c_i(x) = 0, \\ c_i(x) + y_i(x, u; \beta, \rho) = 0, \quad i \in \mathcal{I}. \end{cases} \tag{20}$$

We need the following preliminary results for our method and its global and local analysis.

Lemma 2.1. *Given $\beta > 0$ and $\rho > 0$. For $i \in \mathcal{I}$, let y_i and λ_i be defined by (18) and (19), respectively.*
(1) If $c_i(x)$ is differentiable, then y_i and λ_i are differentiable on (x, u), and

$$\nabla_x y_i = -\frac{y_i}{y_i + \lambda_i}\nabla c_i(x), \quad \nabla_x \lambda_i = \frac{\lambda_i}{y_i + \lambda_i}\nabla c_i(x), \tag{21}$$

$$\frac{\partial y_i}{\partial u_{i'}} = \begin{cases} -\rho\frac{y_i}{y_i+\lambda_i}, & \textit{if } i' = i; \\ 0, & \textit{otherwise}, \end{cases} \quad \frac{\partial \lambda_i}{\partial u_{i'}} = \begin{cases} \rho\frac{\lambda_i}{y_i+\lambda_i}, & \textit{if } i' = i; \\ 0, & \textit{otherwise}. \end{cases} \tag{22}$$

(2) y_i is a monotonically decreasing function on u_i, and λ_i is a monotonically increasing function on u_i.
(3) y_i is smaller as $\beta > 0$ becomes smaller, and it will be also smaller as $\rho > 0$ becomes smaller provided $c_i(x) + y_i > 0$.

Proof. (1) Since $y_i\lambda_i = \rho\beta$, one has

$$y_i\nabla_x\lambda_i + \lambda_i\nabla_x y_i = 0. \tag{23}$$

By (19), $\lambda_i = \rho u_i + c_i(x) + y_i$. Thus,

$$\nabla_x \lambda_i = \nabla c_i(x) + \nabla_x y_i. \tag{24}$$

Substituting (24) into (23),

$$\nabla_x y_i = -\frac{y_i}{y_i + \lambda_i} \nabla c_i(x).$$

Again by (24), $\nabla_x \lambda_i = \frac{\lambda_i}{y_i+\lambda_i} \nabla c_i(x)$.

Similar to (23) and (24), one has

$$y_i \frac{\partial \lambda_i}{\partial u_i} + \lambda_i \frac{\partial y_i}{\partial u_i} = 0, \quad \frac{\partial \lambda_i}{\partial u_i} = \rho + \frac{\partial y_i}{\partial u_i}.$$

Thus,

$$\frac{\partial y_i}{\partial u_i} = -\rho \frac{y_i}{y_i + \lambda_i}, \quad \frac{\partial \lambda_i}{\partial u_i} = \rho \frac{\lambda_i}{y_i + \lambda_i}.$$

For $i \neq i'$, $\frac{\partial y_i}{\partial u_{i'}} = \frac{\partial \lambda_i}{\partial u_{i'}} = 0$ since y_i and λ_i do not depend on $u_{i'}$.

(2) The result follows immediately since $\frac{\partial y_i}{\partial u_i} < 0$ and $\frac{\partial \lambda_i}{\partial u_i} > 0$.

(3) It is obvious from (18) that y_i is smaller as β is smaller. If $c_i(x) + y_i > 0$, then $y_i(c_i(x) + y_i) > 0$, thus $u_i y_i = \frac{1}{\rho}(\lambda_i y_i - y_i(c_i(x) + y_i)) < \beta$ which implies

$$\frac{\partial y_i}{\partial \rho} = \frac{\beta - u_i y_i}{y_i + \lambda_i} > 0.$$

Hence, y_i is a nondecreasing function on ρ. □

3. **Our algorithm.** Our algorithm consists of the inner algorithm and the outer algorithm, where the inner algorithm tries to find an approximate solution of the system (13) for given parameters β and ρ, while the outer algorithm updates the parameters by the information derived from the inner algorithm.

3.1. **A well-behaved quadratic programming subproblem.** A quadratic programming subproblem is presented for deriving our search direction in this subsection. The subproblem is well-behaved since it is always feasible. Suppose that (x_k, u_k) is the current iterate. For given $\beta > 0$ and $\rho > 0$, let

$$B_k = H_k + \sum_{i \in \mathcal{I}} \frac{\lambda_{ki}}{y_{ki} + \lambda_{ki}} \nabla c_{ki} \nabla c_{ki}^T, \tag{25}$$

where H_k is the Hessian of the Fritz-John function $L_\rho(x, \lambda) = \rho f(x) + \lambda^T c(x)$ at (x_k, λ_k). In order to avoid the computation of second-order derivatives, we may take H_k to be an approximation to the Hessian in our algorithm. Using (21)–(22), the Newton's equations for (20) have the form

$$\left\{ \begin{array}{l} B_k d_x + \sum_{i \in \mathcal{I}} \rho \frac{\lambda_{ki}}{y_{ki}+\lambda_{ki}} d_{ui} \nabla c_{ki} = -(\rho \nabla f_k + \sum_{i \in \mathcal{I}} \lambda_{ki} \nabla c_{ki}), \\ \rho \frac{\lambda_{ki}}{y_{ki}+\lambda_{ki}} \nabla c_{ki}^T d_x - \rho^2 \frac{y_{ki}}{y_{ki}+\lambda_{ki}} d_{ui} = -\rho(c_{ki} + y_{ki}), \quad i \in \mathcal{I}. \end{array} \right. \tag{26}$$

For simplicity of statement, let

$$R_k = \begin{pmatrix} \frac{\lambda_{k1}}{y_{k1}+\lambda_{k1}} \nabla c_{k1} & \cdots & \frac{\lambda_{km}}{y_{km}+\lambda_{km}} \nabla c_{km} \\ -\rho \frac{y_{k1}}{y_{k1}+\lambda_{k1}} & \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdots & -\rho \frac{y_{km}}{y_{km}+\lambda_{km}} \end{pmatrix},$$

and $r_k = c_k + y_k$. The following result shows that one can obtain the solution of the system (26) by solving the feasible quadratic programming (QP) subproblem (27)–(28).

Lemma 3.1. *(1) For given $\beta > 0$ and $\rho > 0$, the solution to the QP problem*

$$\min \quad q_k(d) := (\nabla_x L_\rho(x_k, \lambda_k))^T d_x + \frac{1}{2} d^T Q_k d \tag{27}$$

$$s.t. \quad R_k^T d = -(c_k + y_k) \tag{28}$$

satisfies the system (26), where $d = (d_x, d_u) \in \Re^{n+m}$, $\nabla_x L_\rho(x_k, \lambda_k) = \rho \nabla f_k + \sum_{i\in\mathcal{I}} \lambda_{ki} \nabla c_{ki}$,

$$Q_k = \begin{pmatrix} H_k + \sum_{i\in\mathcal{I}} \frac{\rho\beta}{(y_{ki}+\lambda_{ki})^2} \nabla c_{ki} \nabla c_{ki}^T & \frac{\rho^2\beta}{(y_{k1}+\lambda_{k1})^2} \nabla c_{k1} & \cdots & \frac{\rho^2\beta}{(y_{km}+\lambda_{km})^2} \nabla c_{km} \\ \frac{\rho^2\beta}{(y_{k1}+\lambda_{k1})^2} \nabla c_{k1}^T & \frac{\rho^3\beta}{(y_{k1}+\lambda_{k1})^2} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\rho^2\beta}{(y_{km}+\lambda_{km})^2} \nabla c_{km}^T & 0 & \cdots & \frac{\rho^3\beta}{(y_{km}+\lambda_{km})^2} \end{pmatrix}.$$

(2) If

$$d_x^T H_k d_x + \sum_{i\in\mathcal{I}} \frac{\lambda_{ki}}{y_{ki} + \lambda_{ki}} \|\nabla c_{ki}^T d_x\|^2 > 0, \quad \forall d_x \in \Re^n, \tag{29}$$

the above QP has a unique solution, which implies that the system (26) is consistent.

Proof. (1) In addition to (28), the KKT conditions of the above QP contain the following equations:

$$\nabla_x L_\rho(x_k, \lambda_k) + (H_k + \sum_{i\in\mathcal{I}} \frac{\rho\beta}{(y_{ki} + \lambda_{ki})^2} \nabla c_{ki} \nabla c_{ki}^T) d_x + \sum_{i\in\mathcal{I}} \frac{\rho^2\beta}{(y_{ki} + \lambda_{ki})^2} d_{ui} \nabla c_{ki} + \sum_{i\in\mathcal{I}} \frac{\lambda_{ki}}{y_{ki} + \lambda_{ki}} \hat{\lambda}_{ki} \nabla c_{ki} = 0, \tag{30}$$

$$\frac{\rho^2\beta}{(y_{ki} + \lambda_{ki})^2} \nabla c_{ki}^T d_x + \frac{\rho^3\beta}{(y_{ki} + \lambda_{ki})^2} d_{ui} - \rho \frac{y_{ki}}{y_{ki} + \lambda_{ki}} \hat{\lambda}_{ki} = 0, \quad i \in \mathcal{I}, \tag{31}$$

where $\hat{\lambda}_{ki}$ $(i \in \mathcal{I})$ are the associated multipliers with (28). One can first have $\hat{\lambda}_k$ from (31), and then substitute it into (30) to derive the first equation of the system (26).

(2) If (29) holds, then $\nabla^2 q_k$ is positive definite in the null space of R_k^T since

$$d^T Q_k d = d_x^T (H_k + \sum_{i\in\mathcal{I}} \frac{\lambda_{ki}}{y_{ki} + \lambda_{ki}} \nabla c_{ki} \nabla c_{ki}^T) d_x + \sum_{i\in\mathcal{I}} \frac{y_{ki}}{y_{ki} + \lambda_{ki}} \rho^2 d_{ui}^2 > 0,$$

for all $d \in \Re^{n+m}$ such that $R_k^T d = 0$. It follows from Lemma 16.1 of [29] that QP (27)–(28) has a unique solution. By (1), the unique solution also solves the system (26). □

The null-space technology in nonlinear optimization was initially presented by Byrd [7] for trust region methods. It has been proved to be very efficient in trust-region and line-search SQP and interior-point methods (for example, see [9, 25, 27]). In order to obtain the strong global convergence properties, we introduce this technique to the subproblem. Firstly, $d_k^c \in \Re^{n+m}$ is computed to satisfy some prescribed mild conditions presented in Assumption 4.3, and $d_k^c = 0$ as $r_k = 0$ (we refer the readers to [25, 27] for more details). Then we solve the following null-space quadratic programming subproblem

$$\begin{aligned} \min \quad \hat{q}_k(d) &:= \rho \nabla f_k^T d_x + \sum_{i \in \mathcal{I}} \frac{\rho\beta}{y_{ki} + \lambda_{ki}} (\nabla c_{ki}^T d_x + \rho d_{ui}) + \frac{1}{2} d_x^T H_k d_x \\ &+ \frac{1}{2} \sum_{i \in \mathcal{I}} \frac{\rho\beta}{(y_{ki} + \lambda_{ki})^2} (\nabla c_{ki}^T d_x + \rho d_{ui})^2 \qquad (32) \\ \text{s.t.} \quad R_k^T d &= R_k^T d_k^c, \qquad (33) \end{aligned}$$

where the right-hand-side term $-(c_k + y_k)$ of (28) is replaced by $R_k^T d_k^c$ and the scalar $(\lambda_k)^T R_k^T d_k^c$ of the objective in (27) is removed.

3.2. **The merit function.** In order to prompt global convergence of the algorithm, we introduce the merit function

$$\Phi_\xi(x, u; \beta, \rho) = \xi \rho f(x) - \xi \rho \beta \sum_{i \in \mathcal{I}} \ln y_i + \|c(x) + y\|,$$

where $\xi > 0$ is a penalty parameter which is updated in accordance with the directional derivative of $\Phi_\xi(x, u; \beta, \rho)$ along the search direction. The update of the scaling parameter ρ in the outer algorithm depends on the value of ξ. Although it has a similar form to those used in some existing primal-dual interior-point methods such as [13, 25, 27], it is essentially different in that y is a function on x and u.

The following result is helpful for us to select an appropriate penalty parameter ξ so that the search direction is a descent direction of the merit function.

Lemma 3.2. *For given $\beta > 0$ and $\rho > 0$, let $z_k = (x_k, u_k)$, and let $d_k = (d_{xk}, d_{uk})$ be the solution of subproblem (32)–(33), $\Phi_\xi^{'}(z_k; d_k)$ be the directional derivative of $\Phi_\xi(z; \beta, \rho)$ at z_k along the direction d_k.*
(1) There holds

$$\begin{aligned} \Phi_\xi^{'}(z_k; d_k) \quad &\le \quad \xi(\rho \nabla f_k^T d_{xk} + \sum_{i \in \mathcal{I}} \frac{\rho\beta}{y_{ki} + \lambda_{ki}} (\nabla c_{ki}^T d_{xk} + \rho d_{uki})) \\ &+ \|r_k + R_k^T d_k\| - \|r_k\|. \end{aligned}$$

(2) If $r_k = 0$, then $\Phi_\xi^{'}(z_k; d_k) \le -\frac{1}{2} \xi d_k^T Q_k d_k$.

Proof. (1) Let

$$\Theta(x, u) = \|c(x) + y\|. \qquad (34)$$

Then, by the proof of Proposition 3.1 of [25], $\Theta^{'}(z_k; d_k) \le \|r_k + R_k^T d_k\| - \|r_k\|$. Therefore,

$$\begin{aligned} &\Phi_\xi^{'}(z_k; d_k) \\ &\le \xi \rho \nabla f_k^T d_{xk} - \xi \rho \beta \sum_{i \in \mathcal{I}} y_{ki}^{-1} ((\nabla_x y_i)^T d_{xk} + (\nabla_u y_i)^T d_{uk}) + \|r_k + R_k^T d_k\| - \|r_k\| \\ &= \xi(\rho \nabla f_k^T d_{xk} + \sum_{i \in \mathcal{I}} \frac{\rho\beta}{y_{ki} + \lambda_{ki}} (\nabla c_{ki}^T d_{xk} + \rho d_{uki})) + \|r_k + R_k^T d_k\| - \|r_k\|, \end{aligned}$$

where the equality follows from Lemma 2.1(1).

(2) If $r_k = 0$, then, by (1), $\Phi'_\xi(z_k; d_k) \le \xi(\rho\nabla f_k^T d_{xk} + \sum_{i\in\mathcal{I}} \frac{\rho\beta}{y_{ki}+\lambda_{ki}}(\nabla c_{ki}^T d_{xk} + \rho d_{uki}))$. The result follows immediately since $d = 0$ is a feasible solution to the QP (32)–(33). □

Certain additional update techniques are used in primal-dual interior-point methods for nonlinear programs with strong global convergence (for example, see [9, 25, 27, 32]). A technique, which was introduced first in Byrd, Gilbert and Nocedal [9] and was examined to be efficient later, is to change $y_{k+1} = y_k + \alpha_k d_{y_k}$ to $y_{k+1} = \max\{y_k + \alpha_k d_{y_k}, -c(x_{k+1})\}$, so that $c(x_{k+1}) + y_{k+1} \ge 0$ at the $(k+1)$th iteration. However, this technique can not be applied to our method straightforward here since y_k depends on both x_k and u_k. The following result shows that $c(x_{k+1}) + y_{k+1} \ge 0$ can still hold provided u_{k+1} is appropriately updated, thus the strong global convergence is attained.

Lemma 3.3. *For given $\beta > 0$ and $\rho > 0$, if $c_i(x_{k+1}) \ge 0$, or $c_i(x_{k+1}) < 0$ but $u_{k+1,i} \le -\frac{\beta}{c_i(x_{k+1})}$ for any $i \in \mathcal{I}$, then $c_i(x_{k+1}) + y_{k+1,i} \ge 0$, where $y_{k+1,i} = y_i(x_{k+1}, u_{k+1}; \beta, \rho)$ is given by (18).*

Proof. If $c_i(x_{k+1}) \ge 0$, then $c_i(x_{k+1}) + y_{k+1,i} > 0$ since $y_{k+1,i} > 0$. In the remainder, we consider the case $c_i(x_{k+1}) < 0$.

If $c_i(x_{k+1}) - \rho u_{k+1,i} \ge 0$, by (11), one has $c_i(x_{k+1}) + y_{k+1,i} \ge 0$. In this case,

$$u_{k+1,i} < 0 < -\frac{\beta}{c_i(x_{k+1})}.$$

If $c_i(x_{k+1}) - \rho u_{k+1,i} < 0$, by (11), $c_i(x_{k+1}) + y_{k+1,i} \ge 0$ if and only if

$$\sqrt{(c_i(x_{k+1}) + \rho u_{k+1,i})^2 + 4\rho\beta} \ge -(c_i(x_{k+1}) - \rho u_{k+1,i}),$$

which is equivalent to $c_i(x_{k+1}) u_{k+1,i} \ge -\beta$.

Due to $c_i(x_{k+1}) < 0$, the result follows immediately. □

3.3. **The framework of our algorithm.** We denote by $\mathcal{F}$ the class of continuous functions $\theta : \Re_{++} \to \Re_{++}$ satisfying $\lim_{t\to 0} \theta(t) = 0$, and

$$\phi_{(\beta,\rho)}(x, u) = \begin{pmatrix} \rho\nabla f(x) + \sum_{i\in\mathcal{I}} \lambda_i(x, u; \beta, \rho)\nabla c_i(x) \\ \rho(c(x) + y(x, u; \beta, \rho)) \end{pmatrix}.$$

Now we are ready to describe our algorithmic framework for problem (1)–(2). The details on implementation of the algorithm will be provided in Section 6.

Algorithm 3.4. *(The algorithm for problem (1)–(2))*

Step 1 *Given $z_0 = (x_0, u_0) \in \Re^{n+m}$, $\beta_0 > 0$, $\rho_0 > 0$, $\delta \in (0, 1)$, $\sigma \in (0, \frac{1}{2})$, $\epsilon > 0$, and functions $\theta_1, \theta_2 \in \mathcal{F}$. Set $l := 0$.*

Step 2 *While $\beta_l > \epsilon$ and $\rho_l > \epsilon$, start the following inner algorithm.*

Step 2.0 Given $H_0 \in \Re^{n\times n}$, $\xi_0 = 1$, let $z_0 = (x_l, u_l)$. Evaluate y_0 and λ_0 by (18) and (19) with $\beta = \beta_l$ and $\rho = \rho_l$. Let $k := 0$.

Step 2.1 Obtain d_k^c, and solve the QP subproblem (32)–(33) to derive (d_{xk}, d_{uk}).

Step 2.2 Choose ξ_{k+1} with either $\xi_{k+1} = \xi_k$ or $\xi_{k+1} \le 0.5\xi_k$ such that

$$\pi_{\xi_{k+1}}(z_k; d_k) + (1 - \delta)(\|r_k\| - \|r_k + R_k^T d_k\|) \le -0.5\xi_{k+1} d_k^T Q_k d_k, \tag{35}$$

where $\pi_{\xi_{k+1}}(z_k; d_k) = \xi_{k+1}(\rho_l\nabla f_k^T d_{xk} + \sum_{i\in\mathcal{I}} \frac{\rho_l\beta_l}{y_{ki}+\lambda_{ki}}(\nabla c_{ki}^T d_{xk} + \rho_l d_{uki})) + \|r_k + R_k^T d_k\| - \|r_k\|$.

Step 2.3 Choose the step-size $\alpha_k \in (0,1]$ *to be the maximal in* $\{1, \delta, \delta^2, \dots\}$ *such that*

$$\Phi_{\xi_{k+1}}(x_k + \alpha_k d_{xk}, u_k + \alpha_k d_{uk}; \beta_l, \rho_l) - \Phi_{\xi_{k+1}}(x_k, u_k; \beta_l, \rho_l) \le \sigma \alpha_k \pi_{\xi_{k+1}}(z_k; d_k). \quad (36)$$

Step 2.4 Set $x_{k+1} = x_k + \alpha_k d_{xk}$ *and* $\hat{u}_{k+1} = u_k + \alpha_k d_{uk}$.

Step 2.5 Set

$$u_{k+1,i} = \begin{cases} \hat{u}_{k+1,i}, & \text{if} \quad c_i(x_{k+1}) \ge 0; \\ \min\{\hat{u}_{k+1,i}, -\frac{\beta_l}{c_i(x_{k+1})}\}, & \text{otherwise} \end{cases} \quad (37)$$

for every $i \in \mathcal{I}$. *Set* $z_{k+1} = (x_{k+1}, u_{k+1})$.

Step 3 If $\|\phi_{(\beta_l,\rho_l)}(x_{k+1}, u_{k+1})\|_\infty \le \rho_l \theta_1(\beta_l)$, *then update* β_l *to* $\beta_{l+1} \le 0.1\beta_l$, $\rho_{l+1} = \rho_l$; *else if* $\xi_{k+1} \le 0.1 \min(\rho_l^{0.5}, 1)$, *then update* ρ_l *to* $\rho_{l+1} \le \xi_{k+1}\rho_l$, $\beta_{l+1} = \beta_l$. *In these two cases, the inner algorithm is stopped. Let* $z_{l+1} = z_{k+1}$, $l := l + 1$ *and go to Step 2. Otherwise, evaluate* $y_{k+1} = y(x_{k+1}, u_{k+1}; \beta_l, \rho_l)$ *and* $\lambda_{k+1} = \lambda(x_{k+1}, u_{k+1}; \beta_l, \rho_l)$, *update* H_k *to* H_{k+1}, *let* $k := k + 1$ *and go to Step 2.1.*

Due to Step 2.2 of Algorithm 3.4, $\Phi'_{\xi_{k+1}}(z_k; d_k) < 0$. Thus, there is always a sufficiently small number $\alpha_k > 0$ such that (36) holds (for example, see Lemma 2.7 of [3]). That is, the inner algorithm of Algorithm 3.4 is well-defined.

Let $\hat{y}_{k+1,i} = y_i(x_{k+1}, \hat{u}_{k+1}; \beta_l, \rho_l)$ for $i \in \mathcal{I}$. It follows from (37) and the proof of Lemma 3.3 that $c_i(x_{k+1}) + \hat{y}_{k+1,i} \ge 0$ if and only if $u_{k+1,i} = \hat{u}_{k+1,i}$ (thus $y_{k+1,i} = \hat{y}_{k+1,i}$). If $c_i(x_{k+1}) + \hat{y}_{k+1,i} < 0$, then $u_{k+1,i} = -\frac{\beta_l}{c_i(x_{k+1})}$ and $c_i(x_{k+1}) + y_{k+1,i} = 0$ (in this case $y_{k+1,i} > \hat{y}_{k+1,i}$). Therefore,

$$c(x_{k+1}) + y_{k+1} \ge 0 \quad (38)$$

and $\|c(x_{k+1}) + y_{k+1}\| \le \|c(x_{k+1}) + \hat{y}_{k+1}\|$. Since the logarithmic function is monotonically nondecreasing, and, for any $i \in \mathcal{I}$, $y_{k+1,i} \ge \hat{y}_{k+1,i}$, one has $\ln y_{k+1,i} \ge \ln \hat{y}_{k+1,i}$ for every $i \in \mathcal{I}$. Note that the line search procedure guarantees $\Phi_{\xi_{k+1}}(x_{k+1}, \hat{u}_{k+1}; \beta_l, \rho_l) \le \Phi_{\xi_{k+1}}(x_k, u_k; \beta_l, \rho_l)$. Hence, for every $k \ge 0$,

$$\Phi_{\xi_{k+1}}(x_{k+1}, u_{k+1}; \beta_l, \rho_l) \le \Phi_{\xi_{k+1}}(x_k, u_k; \beta_l, \rho_l). \quad (39)$$

The well-definedness of the whole algorithm is based on the global convergence results of Algorithm 3.4. It will be proved, in the next section, that either the inner algorithm converges to a solution satisfying the system (20), in this situation the terminating condition $\|\phi_{(\beta_l,\rho_l)}(z_{k+1})\|_\infty \le \rho_l \theta_1(\beta_l)$ will hold in a finite number of iterations, or $\xi_{k+1} \to 0$ and the terminating condition $\xi_{k+1} \le 0.1 \min(\rho_l^{0.5}, 1)$ for the inner algorithm will be satisfied. Since the inner algorithm will always be terminated in a finite number of iterations, by Step 3 of Algorithm 3.4, either β_l or ρ_l will be reduced at least to a fixed fraction.

4. **Global convergence.** We present our global convergence results on Algorithm 3.4 in this section. Firstly, we consider the global convergence of the inner algorithm. For given $\beta_l > 0$ and $\rho_l > 0$, suppose that the inner algorithm does not terminate in a finite number of iterations. We prove that, if $\{\xi_k\}$ is bounded away from zero, then every limit point of sequence $\{(x_k, u_k)\}$ is a solution of the system (20); otherwise, $\xi_k \to 0$ as $k \to \infty$. It shows that the supposition will never happen. After that, the global convergence results of the whole algorithm are presented. The results show that the whole algorithm converges to a KKT point of the original problem provided $\beta_l \to 0$ but $\rho_l \not\to 0$, otherwise $\rho_l \to 0$ and there is one of the limit points of the sequence $\{x_l\}$ which is an infeasible stationary point or a singular stationary point of problem (1)–(2).

We need the following definitions.

Definition 4.1. $x^* \in \Re^n$ is called an infeasible stationary point of problem (1)–(2) if x^* is an infeasible point and

$$\sum_{i\in\mathcal{I}} a_i^* \nabla c_i(x^*) = 0, \tag{40}$$

where $a_i^* = \max\{c_i(x^*), 0\},\ i \in \mathcal{I}$.

Definition 4.2. $x^* \in \Re^n$ is called a singular stationary point of problem (1)–(2) if there is a nonzero vector $b^* \in \Re^m$ such that

$$\sum_{i\in\mathcal{I}} b_i^* \nabla c_i(x^*) = 0, \tag{41}$$

$$b_i^* \geq 0, \quad c_i(x^*) \leq 0, \quad b_i^* c_i(x^*) = 0, \quad i \in \mathcal{I}. \tag{42}$$

While Definition 4.1 shows that x^* is a stationary point for minimizing the constraint violations

$$\frac{1}{2}\sum_{i\in\mathcal{I}} |\max\{c_i(x), 0\}|^2, \tag{43}$$

Definition 4.2 implies that x^* is a Fritz-John point of problem (1)–(2) at which the Mangasarian-Fromovitz constraint qualification (MFCQ) does not hold.

It should be noticed that various definitions have been given for infeasible and singular stationary points, see [5, 6, 8, 13, 25, 26, 39]. These stationary points may either belong to a set of minimizers of the problem minimizing the measure of constraint violations like problem (43) or be the optimal solutions of some degenerate nonlinear programs, see Section 6.1 for the details. For example, [8] considered the infeasible stationary point to be a first-order optimal solution x^* of the problem

$$\begin{aligned} \min \quad & \sum_{i\in\{\mathcal{I}|c_i(x^*)>0\}} c_i(x) \\ \text{s.t.} \quad & c_i(x) = 0, \quad i \in \{i \in \mathcal{I} | c_i(x^*) = 0\}, \end{aligned}$$

whereas [25] identifies some singular stationary points at which the linear independence constraint qualification (LICQ) does not hold.

4.1. **Global convergence of the inner algorithm.** We consider the global convergence of the inner algorithm. Suppose that, for parameters $\beta_l > 0$ and $\rho_l > 0$, the inner algorithm of Algorithm 3.4 does not terminate in a finite number of iterations and $\{(x_k, u_k)\}$ is an infinite sequence generated by the algorithm. For the sake of global convergence analysis, we need the following blanket assumptions.

Assumption 4.3.
(1) The functions f and $c_i (i \in \mathcal{I})$ are twice continuously differentiable on $\Re^n$;
(2) The iterative sequence $\{x_k\}$ is in an open bounded set;
(3) The sequence $\{H_k\}$ is bounded, and for all $k \geq 0$ and $d \in \Re^n$, $d^T H_k d \geq \rho_l \gamma \|d\|^2$, where $\gamma > 0$ is a constant;
(4) For all $k \geq 0$, d_k^c satisfies the conditions:
(i) $\|d_k^c\| \leq \eta_1 \|R_k r_k\|$,
(ii) $\|r_k\| - \|r_k + R_k^T d_k^c\| \geq \eta_2 \|R_k r_k\|^2 / \|r_k\|$, where $\eta_1 > 0$ and $\eta_2 > 0$ are two constants.

The conditions in Assumption 4.3 (1)–(3) are the same as those commonly used in global convergence analysis of iterative methods for nonlinear optimization (for example, see [5, 9, 13, 22, 25, 27]). Assumption 4.3 (4) is for the strong global convergence of the algorithm, which is very mild and can be satisfied easily (see Section 2.2 of [25]).

The following results depend only on the merit function and can be proved in the same way as Lemma 5 of [9] and Lemma 4.2 of [27].

Lemma 4.4. *Suppose that Assumption 4.3 holds. Then $\{y_k\}$ is bounded, $\{\lambda_k\}$ is componentwise bounded away from zero and $\{u_k\}$ is lower bounded. Furthermore, if the penalty parameter ξ_k remains constant for all sufficiently large k, then $\{y_k\}$ is componentwise bounded away from zero, $\{\lambda_k\}$ and $\{u_k\}$ are bounded.*

Proof. The results on $\{y_k\}$ can be derived by [9, 27]. Due to $\lambda_{ki}y_{ki} = \rho_l\beta_l$, the results on $\{\lambda_k\}$ follow immediately.

For given $\beta_l > 0$ and $\rho_l > 0$, if $\{y_k\}$ is bounded, then, by (18), $u_{ki} > -\infty$ for all $k \geq 0$ and $i \in \mathcal{I}$. Otherwise, if $u_{ki} \to -\infty$ for some i, then $y_{ki} \to \infty$, which is a contradiction. If $\{y_k\}$ is componentwise bounded away from zero, then, by (18), $u_{ki} < \infty$ for all $k \geq 0$ and $i \in \mathcal{I}$. Thus, the results on $\{u_k\}$ are proved. □

The update rule on ξ_k is adaptive. It implies that the sequence $\{\xi_k\}$ is monotonically nonincreasing, which either remains a positive constant after a finite number of iterations or tends to zero as k tends to infinity. The next two results show that, if ξ_k is bounded away from zero, all step-sizes can be selected to be bounded away from zero.

Lemma 4.5. *Suppose that Assumption 4.3 holds. Let $d_k = (d_{xk}, d_{uk}) \in \Re^{n+m}$ be the solution of quadratic programming subproblem (32)–(33), and let $g_k \in \Re^m$ be the associated Lagrangian multiplier. If ξ_k remains a positive constant after a finite number of iterations, then $\{\|d_k\|\}$ and $\{\|R_k g_k\|\}$ are bounded.*

Proof. Since ∇f_k and d_k^c are bounded, H_k is bounded and uniformly positive definite, $\|d_{xk}\|$ and $|\sum_{i\in\mathcal{I}} \frac{\nabla c_{ki}^T d_{xk}+\rho_l d_{uki}}{y_{ki}+\lambda_{ki}}|$ are bounded due to $\hat{q}(d_k) \leq \hat{q}(d_k^c)$.

If ξ_k is bounded away from zero, in view of Lemma 4.4, both y_{ki} and λ_{ki} are bounded above and bounded away from zero. Thus, $\|d_k\|$ is bounded since $1/(y_{ki} + \lambda_{ki})$ for every $i \in \mathcal{I}$ is bounded away from zero.

In view of

$$\begin{aligned} &\rho_l \nabla f_k + H_k d_{xk} + \textstyle\sum_{i\in\mathcal{I}} \frac{\rho_l\beta_l}{y_{ki}+\lambda_{ki}}\left(1 + \frac{\nabla c_{ki}^T d_{xk}+\rho_l d_{uki}}{y_{ki}+\lambda_{ki}}\right)\nabla c_{ki} \\ &\qquad + \textstyle\sum_{i\in\mathcal{I}} \frac{\lambda_{ki}}{y_{ki}+\lambda_{ki}} g_{ki}\nabla c_{ki} = 0, \end{aligned} \tag{44}$$

$$\frac{\rho_l^2\beta_l}{y_{ki}+\lambda_{ki}}\left(1 + \frac{\nabla c_{ki}^T d_{xk}+\rho_l d_{uki}}{y_{ki}+\lambda_{ki}}\right) - \rho_l \frac{y_{ki}}{y_{ki}+\lambda_{ki}} g_{ki} = 0, \quad i \in \mathcal{I}, \tag{45}$$

and $\frac{\rho_l\beta_l}{y_{ki}+\lambda_{ki}} \leq \frac{\sqrt{\rho_l\beta_l}}{2}$ for $i \in \mathcal{I}$, and note that $\|d_k\|$ is bounded, one can deduce that $\|R_k g_k\|$ is bounded. □

Lemma 4.6. *Suppose that Assumption 4.3 holds. Let $\{\alpha_k\}$ be the sequence of step-sizes derived from (36) of Algorithm 3.4. If ξ_k remains a positive constant after a finite number of iterations, and*

$$\|R_k r_k\| \geq \hat{\eta}\|r_k\| \tag{46}$$

for some constant $\hat{\eta} > 0$ and for all $k \geq 0$, then $\{\alpha_k\}$ is bounded away from zero.

Proof. Due to Lemmas 4.4 and 4.5, for every $i \in \mathcal{I}$, one has

$$-\ln y_i(z_k + \alpha d_k; \beta_l, \rho_l) + \ln y_{ki} - \alpha \tfrac{1}{y_{ki} + \lambda_{ki}^+}(\nabla c_{ki}^T d_{xk} + \rho_l d_{uki}) = o(\alpha),$$
$$\Theta(x_k + \alpha d_{xk}, u_k + \alpha d_{uk}) = \|r_k + \alpha R_k^T d_k\| + o(\alpha)$$

for all $\alpha > 0$ sufficiently small, where $\Theta(x, u)$ is defined by (34). Therefore,

$$\begin{aligned} &\Phi_{\xi_{k+1}}(x_k + \alpha d_{xk}, u_k + \alpha d_{uk}; \beta_l, \rho_l) - \Phi_{\xi_{k+1}}(x_k, u_k; \beta_l, \rho_l) \\ &= \alpha \pi_{\xi_{k+1}}(z_k; d_k) + o(\alpha) \end{aligned} \tag{47}$$

for all $\alpha \in [0, \tilde{\alpha}]$, where $\tilde{\alpha} > 0$ is a sufficiently small scalar. Note that, due to (46),

$$(1-\sigma)\alpha\pi_{\xi_{k+1}}(z_k; d_k) \le \alpha(1-\sigma)(1-\delta)(\|r_k + R_k^T d_k\| - \|r_k\|) \le -\alpha\eta_3\|r_k\|, \tag{48}$$

where $\eta_3 = \eta_2\hat{\eta}^2(1-\sigma)(1-\delta)$. It follows from (47) and (48) that there exists a scalar $\hat{\alpha} \in (0, \tilde{\alpha}]$ such that

$$\Phi_{\xi_{k+1}}(x_k + \alpha d_{xk}, u_k + \alpha d_{uk}; \beta_l, \rho_l) - \Phi_{\xi_{k+1}}(x_k, u_k; \beta_l, \rho_l) \le \sigma\alpha\pi_{\xi_{k+1}}(z_k; d_k)$$

for all $\alpha \in (0, \hat{\alpha}]$ and all $k \ge 0$. Thus, by Step 2.3 of Algorithm 3.4, $\alpha_k \ge \hat{\alpha}$ for all $k \ge 0$. □

We prove that, if condition (46) holds, the penalty parameter ξ_k in the merit function will remain a positive constant after a finite number of iterations.

Lemma 4.7. *Suppose that Assumption 4.3 holds. If (46) holds for some scalar $\hat{\eta} > 0$ and for all $k \ge 0$, there is a constant $\hat{\xi} > 0$ such that $\xi_k = \hat{\xi}$ for all sufficiently large k.*

Proof. We achieve the result by proving that (35) holds with $\xi_k = \hat{\xi}$ as $\hat{\xi}$ is small enough.

Note that $\lambda_{ki} y_{ki} = \rho_l \beta_l$ and

$$\frac{1}{y_{ki} + \lambda_{ki}} = \frac{y_{ki}}{y_{ki}^2 + \lambda_{ki} y_{ki}} \le \frac{1}{\rho_l \beta_l} y_{ki}.$$

Hence, due to $\hat{q}_k(d_k) \le \hat{q}_k(d_k^c)$, Assumption 4.3 (4) (ii) and Lemma 4.4, one has

$$\begin{aligned} &\pi_{\xi_{k+1}}(z_k; d_k) + (1-\delta)(\|r_k\| - \|r_k + R_k^T d_k\|) + \frac{1}{2}\xi_{k+1} d_k^T Q_k d_k \\ &= \xi_{k+1}\hat{q}_k(d_k) + \delta(\|r_k + R_k^T d_k\| - \|r_k\|) \\ &\le \xi_{k+1}\hat{q}_k(d_k^c) + \delta(\|r_k + R_k^T d_k^c\| - \|r_k\|) \\ &\le \gamma_1 \xi_{k+1}\|d_k^c\| - \delta\eta_2\hat{\eta}^2\|r_k\|, \end{aligned}$$

where $\gamma_1 > 0$ is a scalar. Finally, it follows from Assumption 4.3 (4) (i) that (35) holds with $\xi_{k+1} = \hat{\xi}$ as $\hat{\xi} \le \delta\eta_2\hat{\eta}^2/(\gamma_1\eta_1)$. □

Now we prove that sequence $\{(x_k, u_k)\}$ generated by the inner algorithm of Algorithm 3.4 will converge to a solution of the system (20) provided (46) holds.

Lemma 4.8. *Let $\{(x_k, u_k)\}$ be the infinite sequence generated by the inner algorithm of Algorithm 3.4. Suppose that Assumption 4.3 holds, and assume that (46) holds for some scalar $\hat{\eta} > 0$ and for all $k \ge 0$. Then any limit point of $\{(x_k, u_k)\}$ is a solution of the system (20).*

Proof. Firstly , we prove that

$$\lim_{k\to\infty} \|r_k\| = 0 \quad \text{and} \quad \lim_{k\to\infty} \|d_k\| = 0. \tag{49}$$

Without loss of generality, we suppose that $\xi_k = \hat{\xi}$ for all $k \geq 0$. Then, by (39), $\{\Phi_{\hat{\xi}}(z_k; \beta_l, \rho_l)\}$ is a monotonically nonincreasing sequence. Note that it is also a bounded sequence. Thus,

$$\lim_{k \to \infty} \pi_{\hat{\xi}}(z_k; d_k) = 0 \tag{50}$$

since α_k is bounded away from zero. Using the last inequality of (48), one has

$$\lim_{k \to \infty} \|r_k\| = 0,$$

which implies $\lim_{k \to \infty} \|d_k^c\| = 0$. In view of (35), $\lim_{k \to \infty} \|d_k\| = 0$.

It follows from Lemma 4.4 that $\{y_k\}$ and $\{\lambda_k\}$ are bounded above and componentwise bounded away from zero, $\{u_k\}$ is bounded. Without loss of generality, let $z^* = (x^*, u^*)$ be any limit point of $\{z_k\}$ and suppose that $\lim_{k \to \infty} x_k = x^*$ and $\lim_{k \to \infty} u_k = u^*$. Since $\lim_{k \to \infty} \|r_k\| = 0$ and note $c_k + y_k = \lambda_k - \rho_l u_k$, one has

$$\lim_{k \to \infty} \lambda_k = \rho_l u^* > 0, \quad \lim_{k \to \infty} y_k = -c^* > 0.$$

In view of $\lim_{k \to \infty} \|d_k\| = 0$, by taking the limit on $k \to \infty$ in both sides of (44) and (45), there holds $\lim_{k \to \infty} g_{ki} y_{ki} = \rho_l \beta_l$ for $i \in \mathcal{I}$ and $\lim_{k \to \infty}(\rho_l \nabla f_k + \sum_{i \in \mathcal{I}} g_{ki} \nabla c_{ki}) = 0$. Thus,

$$\lim_{k \to \infty} (g_k - \lambda_k) = 0 \quad \text{and} \quad \lim_{k \to \infty} (\rho_l \nabla f_k + \sum_{i \in \mathcal{I}} \lambda_{ki} \nabla c_{ki}) = 0.$$

That is, $\phi_{(\beta_l, \rho_l)}(x^*, u^*) = 0$. □

Now we are ready to present our global convergence results on the inner algorithm of Algorithm 3.4. It indicates that, for any given $\beta_l > 0$ and $\rho_l > 0$, the inner algorithm of Algorithm 3.4 will be terminated in a finite number of iterations.

Theorem 4.9. *Given $\beta_l > 0$ and $\rho_l > 0$ are two scalars. Let $\{(x_k, u_k)\}$ be the infinite sequence generated by the inner algorithm of Algorithm 3.4. Suppose that Assumption 4.3 holds. Then one of the following statements is true:*
(1) $\|R_k r_k\| \geq \hat{\eta} \|r_k\|$ for some scalar $\hat{\eta} > 0$ and for all $k \geq 0$, ξ_k remains a positive constant for all sufficiently large k, and any limit point of $\{(x_k, u_k)\}$ is a solution of the system (20);
(2) $\xi_k \to 0$ as $k \to \infty$, and there exists some infinite index subset $\mathcal{K}$ such that

$$\lim_{k \in \mathcal{K}, k \to \infty} \|R_k r_k\| / \|r_k\| = 0,$$

where $r_k \geq 0$ for all $k \geq 0$.

Proof. The result (1) follows from the preceding Lemma 4.7 and Lemma 4.8. The result (2) is straightforward and can be taken as a corollary of Lemma 4.7, where $r_k \geq 0$ since (38). □

4.2. **Global Convergence results of the whole algorithm.** Now we consider the global convergence of the whole algorithm. Without loss of generality, we let $\epsilon = 0$ and let $\{(x_l, u_l)\}$ be an infinite sequence generated by the outer algorithm of Algorithm 3.4. It is shown that, either we have $\beta_l \to 0$ and $\rho_l \geq \hat{\rho}$ for some positive scalar $\hat{\rho}$ and for all l, and every limit point (x^*, u^*) of sequence $\{(x_l, u_l)\}$ is a KKT pair of the original problem (1)–(2), or we have $\rho_l \to 0$ and there exists a limit point x^* of the sequence $\{x_l\}$ which is either an infeasible stationary point or a singular stationary point of the problem (1)–(2).

Theorem 4.10. *Suppose that Assumption 4.3 holds for every given parameters $\beta_l > 0$ and $\rho_l > 0$. Let $\epsilon = 0$, and let $\{(x_l, u_l)\}$ be an infinite sequence generated by the outer algorithm of Algorithm 3.4. Then one of the following two cases will happen.*
(1) $\rho_l \ge \hat\rho$ for some positive scalar $\hat\rho$ and for all l, $\beta_l \to 0$ as $l \to \infty$, every limit point (x^*, u^*) of sequence $\{(x_l, u_l)\}$ is a KKT pair of the original problem (1)–(2).
(2) $\rho_l \to 0$ as $l \to \infty$, and there exists a limit point x^* of the sequence $\{x_l\}$ which is either an infeasible stationary point or a singular stationary point of the problem (1)–(2).

Proof. Since, for every given parameters $\beta_l > 0$ and $\rho_l > 0$, the inner algorithm of Algorithm 3.4 is terminated in a finite number of iterations, we have either the case with $\|\phi_{(\beta_l,\rho_l)}(z_{l+1})\|_\infty \le \rho_l\theta_1(\beta_l)$ for all sufficiently large l or the case that there exists an infinite subsequence $\{\rho_{l_k}\}$ of sequence $\{\rho_l\}$ such that $\rho_{l_k} \le 0.1\rho_{l_k-1}^{1.5}$ for all k.

If l_0 is a positive integer such that $\|\phi_{(\beta_l,\rho_l)}(z_{l+1})\|_\infty \le \rho_l\theta_1(\beta_l)$ for all $l \ge l_0$, then, by Step 3 of Algorithm 3.4, $\rho_l \ge \rho_{l_0}$ for all l and $\beta_l \to 0$ as $l \to \infty$. Thus,

$$\lim_{l\to\infty} \|\phi_{(\beta_l,\rho_l)}(x_{l+1}, u_{l+1})\|_\infty = 0.$$

In view of the argument on the system (13) in section 2, the above equation implies that every limit point (x^*, u^*) of sequence $\{(x_l, u_l)\}$ is a KKT pair of the original problem (1)–(2).

In the following, we consider the latter case. If $\rho_{l_k} \le 0.1\rho_{l_k-1}^{1.5}$ for all k, then $\rho_{l_k} \le 0.1\rho_{l_{k-1}}^{1.5}$ for all k since $\{\rho_l\}$ is a nonincreasing sequence. Thus, $\rho_l \to 0$ as $l \to \infty$. It follows from the result (2) of Theorem 4.9 that

$$\lim_{k\to\infty} \|R_{l_k} r_{l_k}\|/\|r_{l_k}\| = 0,$$

which is equivalent to

$$\lim_{k\to\infty} \sum_{i\in\mathcal{I}} \frac{\lambda_{l_k i}}{y_{l_k i} + \lambda_{l_k i}} \frac{(c_{l_k i} + y_{l_k i})}{\|r_{l_k}\|} \nabla c_{l_k i} = 0, \tag{51}$$

$$\lim_{k\to\infty} \frac{y_{l_k i}}{y_{l_k i} + \lambda_{l_k i}} \frac{(c_{l_k i} + y_{l_k i})}{\|r_{l_k}\|} = 0, \quad i \in \mathcal{I}. \tag{52}$$

Since $\{x_l\}$ and $\{y_l\}$ are bounded sequences, there exist convergent subsequences, for which, without loss of generality, we suppose

$$\lim_{k\to\infty} x_{l_k} = x^* \quad \text{and} \quad \lim_{k\to\infty} y_{l_k} = y^*.$$

If $\lim_{k\to\infty} \|r_{l_k}\| = 0$, then x^* is a feasible point of the original problem (1)–(2). Without loss of generality, we suppose

$$\lim_{k\to\infty} \frac{\lambda_{l_k i}}{y_{l_k i} + \lambda_{l_k i}} = \nu_i^*,\ i \in \mathcal{I}, \quad \lim_{k\to\infty} \frac{c_{l_k} + y_{l_k}}{\|r_{l_k}\|} = b^*.$$

Then $b^* \ne 0$. Since $c_l + y_l \ge 0$ for all $l \ge 0$, one has $b^* \ge 0$. By (52), $(1-\nu_i^*)b_i^* = 0,\ i \in \mathcal{I}$. Thus, for $i \in \mathcal{I}$, $b_i^* = \nu_i^* b_i^*$, i.e., $b_i^* = 0$ as $\nu_i^* = 0$, $\nu_i^* = 1$ as $b_i^* \ne 0$. Note that $\nu_i^* = 1$ implies $y_i^* = 0$ and $c_i^* = 0$. Hence, $b_i^* c_i^* = 0, \forall i \in \mathcal{I}$. Finally, by (51), (41) holds. That is, x^* is a singular stationary of the problem (1)–(2).

Due to (52) we have

$$y_i^*(c_i^* + y_i^*) = 0, \quad i \in \mathcal{I}. \tag{53}$$

Since $c_l+y_l \geq 0$ for all $l \geq 0$, it follows from (53) that, for $i \in \mathcal{I}$, $c_i^*+y_i^* = \max\{c_i^*, 0\}$. If $\lim_{k\to\infty} \|r_{l_k}\| \neq 0$, then x^* is an infeasible point of the original problem (1)–(2). By (51) and (52), one has $\sum_{i\in\mathcal{I}} \max\{c_i^*, 0\}\nabla c_i^* = 0$. Therefore, (40) follows immediately. □

The preceding theorem shows that, for any given $\epsilon > 0$, Algorithm 3.4 will be terminated at either the case $\beta_l \leq \epsilon$ or the case $\rho_l \leq \epsilon$.

5. **Local convergence.**

In this section, we prove that, under suitable conditions, the step at x_l in Algorithm 3.4 can be a superlinearly or quadratically convergent step, no matter whether the sequence $\{x_l\}$ converges to a KKT point or an infeasible stationary point of the original problem. Thus, the whole algorithm is capable of rapidly converging to a KKT point when the problem is feasible, and, in particular, rapidly converging to an infeasible stationary point when a problem is infeasible.

Let $\rho_l \to \rho^*$ and $\beta_l \to \beta^*$ as $l \to \infty$, $\nu_l \in \Re^m$ be a vector with components $\nu_{li} = \lambda_{li}/(y_{li} + \lambda_{li})$, $i \in \mathcal{I}$. For simplicity, we suppose that $\|\rho_l u_l\| \leq M$ for some constant $M > 0$ and for all $l \geq 0$. This supposition is reasonable from the global convergence analysis in previous section, and it does not hinder $\|u_l\|$ tend to ∞. If $\{x_l\}$ converges to a KKT point, then u_l is bounded and the supposition holds obviously. If it is other than that case, since the inner algorithm is terminated finitely for every l, one can select ρ_l such that the supposition holds. With this supposition, $\|\lambda_l\|$ is bounded.

We need the following blanket assumptions for local convergence analysis, in which Assumption 5.1 (3) and (4) are weaker than that commonly used in nonlinear programs.

Assumption 5.1.
(1) $x_l \to x^*$ and $\nu_l \to \nu^*$ as $l \to \infty$. Correspondingly, $y_l \to y^*$ and $\lambda_l \to \lambda^*$ as $l \to \infty$;
(2) The functions f and c_i $(i \in \mathcal{I})$ are twice differentiable on $\Re^n$, and their second derivatives are Lipschitz continuous at some neighborhood of x^*;
(3) The gradients $\nabla c_i(x^*)$ $(i \in \mathcal{W}^* \cap \mathcal{I}^*)$ are linearly independent, where $\mathcal{W}^* = \{i \in \mathcal{I}|\nu_i^* \neq 0\}$, $\mathcal{I}^* = \{i \in \mathcal{I}|c_i(x^*) = 0\}$;
(4) $d^T H^* d > 0$ for all $d \neq 0$ such that $\nu_i^* \nabla c_i(x^*)^T d = 0$, $i \in \mathcal{I}^*$, where $H^* = \rho^*\nabla^2 f(x^*) + \sum_{i\in\mathcal{I}} \lambda_i^* \nabla^2 c_i(x^*)$.

5.1. **Rapid convergence to a KKT point.**

We focus on how the barrier parameter β_l is updated at (x_l, u_l) results in that (d_{xl}, d_{ul}) is a superlinearly or quadratically convergent step, so that our algorithm is capable of rapidly converging to the KKT point. In addition to Assumption 5.1, we also need the following general conditions.

Assumption 5.2.
(1) $\rho^* > 0$ and $\beta^* = 0$;
(2) $u_l \to u^*$ as $l \to \infty$. Thus, $z_l \to z^*$ as $l \to \infty$.

The following index sets are used throughout this subsection: $\mathcal{P}^* = \{i \in \mathcal{I}|c_i(x^*)+\rho^* u_i^* > 0\}$, $\mathcal{Z}^* = \{i \in \mathcal{I}|c_i(x^*) + \rho^* u_i^* = 0\}$, $\mathcal{N}^* = \{i \in \mathcal{I}|c_i(x^*) + \rho^* u_i^* < 0\}$. Assumption 5.2 shows that (x^*, u^*) is a KKT pair, and $c_i(x^*) = 0$ for $i \in \mathcal{P}^* \cup \mathcal{Z}^*$, $c_i(x^*) < 0$ for $i \in \mathcal{N}^*$. It follows from (18) and (19) that $y_i^* = 0$ and $\lambda_i^* > 0$ for $i \in \mathcal{P}^*$, and $y_i^* > 0$ and $\lambda_i^* = 0$ for $i \in \mathcal{N}^*$. They imply that $\nu_i^* = 1$ for $i \in \mathcal{P}^*$, $\nu_i^* = 0$ for $i \in \mathcal{N}^*$. Hence, $\mathcal{W}^* \subseteq \mathcal{I}^*$ and $\mathcal{E}^* = \mathcal{E}$.

Similar to Lemma 16.1 of [29], one can prove the following result. We omit the proof for brevity.

Lemma 5.3. *Suppose that Assumptions 5.1 and 5.2 hold. Then the matrix*

$$\Omega^* = \begin{pmatrix} B^* & [\nu_i^*\nabla c_i(x^*), i \in \mathcal{W}^*] & 0 \\ [\nu_i^*\nabla c_i(x^*), i \in \mathcal{W}^*]^T & -\mathrm{diag}(1-\nu_i^*, i \in \mathcal{W}^*) & 0 \\ 0 & 0 & -I_{|\mathcal{I}\backslash\mathcal{W}^*|} \end{pmatrix}$$

is nonsingular, where $B^* = H^* + \sum_{i\in\mathcal{I}} \nu_i^*\nabla c_i(x^*)\nabla c_i(x^*)^T$, $[\nu_i^*\nabla c_i(x^*), i \in \mathcal{W}^*]$ *is a matrix with* $\nu_i^*\nabla c_i(x^*)$ $(i \in \mathcal{W}^*)$ *as its column vectors,* $\mathrm{diag}(1-\nu_i^*,\ i \in \mathcal{W}^*)$ *is a diagonal matrix with* $(1-\nu_i^*)$ $(i \in \mathcal{W}^*)$ *as its diagonal entries,* $I_{|\mathcal{I}\backslash\mathcal{W}^*|}$ *is an identity matrix of order* $|\mathcal{I}\backslash\mathcal{W}^*|$.

For simplicity of notations, we suppose that $\rho_l = \rho^*$ for all $l \ge 0$ in this subsection. Let $J^* = \lim_{l\to\infty} J_l$, where $J_l = \nabla\phi_{(\beta_l,\rho_l)}(z_l)$. Then $J^* = D^*\Omega^*D^*$, where D^* is a diagonal matrix with n 1s and m ρ^*s. Due to Lemma 5.3, J^* is nonsingular. It follows from the Implicit Function Theorem (p.128 of [30]) that there exists a $\hat\beta > 0$ such that the equation $\phi_{(\beta_l,\rho_l)}(z) = 0$ has a unique solution $z^*(\beta_l)$ for all $\beta_l \le \hat\beta$, and there holds

$$\|z^*(\beta_l) - z^*\| \le M\beta_l < \epsilon, \tag{54}$$

where $\epsilon > 0$ is small enough and

$$M = \max_{\|z-z^*\|<\epsilon} \left\| [\nabla\phi_{(\beta_l,\rho_l)}(z)]^{-1} \frac{\partial\phi_{(\beta_l,\rho_l)}(z)}{\partial\beta} \right\|$$

is a constant independent of β_l.

The following two lemmas can be obtained in a way similar to Lemmas 2.1 and 2.3 in [11]. We will not give their proofs for brevity.

Lemma 5.4. *Suppose that Assumptions 5.1 and 5.2 hold. Then there are sufficiently small scalars* $\epsilon > 0$ *and* $\hat\beta > 0$, *and positive constants* M_0 *and* L_0, *such that, for all* $\beta_l \le \hat\beta$ *and* $z \in \{z \mid \|z - z^*(\beta_l)\| < \epsilon\}$, $\nabla\phi_{(\beta_l,\rho_l)}(z)$ *is invertible,*

$$\|[\nabla\phi_{(\beta_l,\rho_l)}(z)]^{-1}\| \le M_0, \tag{55}$$

and

$$\|\nabla\phi_{(\beta_l,\rho_l)}(z) - \nabla\phi_{(\beta_l,\rho_l)}(z^*(\beta_l))\| \le L_0\|z - z^*(\beta_l)\|. \tag{56}$$

Moreover, for $z \in \{z \mid \|z - z^*(\beta_l)\| < \epsilon\}$ *and* $\beta_l \le \hat\beta$, *one has*

$$\|\nabla\phi_{(\beta_l,\rho_l)}(z)^T(z - z^*(\beta_l)) - \phi_{(\beta_l,\rho_l)}(z)\| \le L_0\|z - z^*(\beta_l)\|^2. \tag{57}$$

Lemma 5.5. *Suppose that Assumptions 5.1 and 5.2 hold. Then there are sufficiently small scalars* $\epsilon > 0$ *and* $\hat\beta > 0$, *such that for* $z \in \{z \mid \|z - z^*(\beta_l)\| < \epsilon\}$ *and* $\beta_l \le \hat\beta$,

$$\|z - z^*(\beta_l)\| \le 2M_0\|\phi_{(\beta_l,\rho_l)}(z)\|, \quad \|\phi_{(\beta_l,\rho_l)}(z)\| \le 2M_1\|z - z^*(\beta_l)\|, \tag{58}$$

where $M_1 = \sup_{\|z-z^*(\beta_l)\|<\epsilon} \|\nabla\phi_{(\beta_l,\rho_l)}(z)\|$.

Using Lemmas 5.4 and 5.5, we can prove the following results.

Theorem 5.6. *Suppose that Assumptions 5.1 and 5.2 hold, and* $\beta_l = O(\|z_l - z^*\|^2)$. *If* d_l^c *is computed such that* $\|r_l + R_l^T d_l^c\| = O(\|r_l\|^2)$, *then*

$$\|z_l + d_l - z^*\| = O(\|z_l - z^*\|^2). \tag{59}$$

That is, d_l *is a quadratically convergent step. If, instead,* $\beta_l = o(\|z_l - z^*\|)$, $\|r_l + R_l^T d_l^c\| = o(\|r_l\|)$, *then* $\|z_l + d_l - z^*\| = o(\|z_l - z^*\|)$. *That is, the step is superlinear.*

Proof. In order to prove the result, we show

$$\limsup_{k\to\infty} \|z_l + d_l - z^*\|/\|z_l - z^*\|^2 \le \gamma, \tag{60}$$

where $\gamma > 0$ is a constant.

Let $\phi_l = \phi_{(\beta_l,\rho_l)}(z_l)$, $J_l = \nabla\phi_{(\beta_l,\rho_l)}(z_l)$, $\hat{\phi}_l$ be the vector which is different from ϕ_l in that the last m components are replaced by $-\rho_l R_l^T d_l^c$, $\hat{d}_l$ be the unique solution of the equation $J_l d = -\phi_l$. Then, due to $\|r_l\| \le \|\phi_l\|$, by (55) and (58),

$$\|d_l - \hat{d}_l\| = \|J_l^{-1} \begin{pmatrix} 0 \\ r_l + R_l^T d_l^c \end{pmatrix}\| = O(\|z_l - z^*(\beta_l)\|^2). \tag{61}$$

Furthermore, by (55) and (57),

$$\begin{aligned} \|z_l + \hat{d}_l - z^*(\beta_l)\| &\le \|[J_l]^{-1}\|\|J_l(z_l - z^*(\beta_l)) - \phi_l\| \\ &\le M_0 L_0 \|z_l - z^*(\beta_l)\|^2. \end{aligned} \tag{62}$$

Using (54), (61) and (62), one has

$$\begin{aligned} &\|z_l + d_l - z^*\| \\ &\le \|z_l + \hat{d}_l - z^*(\beta_l)\| + \|d_l - \hat{d}_l\| + \|z^*(\beta_l) - z^*\| \\ &\le M_0 L_0 \|z_l - z^*(\beta_l)\|^2 + O(\|z_l - z^*(\beta_l)\|^2) + M\beta_l. \end{aligned} \tag{63}$$

If $\beta_l = O(\|z_l - z^*\|^2)$, that is, $\beta_l \le M_2\|z_l - z^*\|^2$ for some constant $M_2 > 0$, then

$$\|z_l - z^*(\beta_l)\| \le \|z_l - z^*\| + \|z^* - z^*(\beta_l)\| \le (1 + MM_2\|z_l - z^*\|)\|z_l - z^*\|.$$

Thus, (60) follows immediately from (63).

The result for the case $\beta_l = o(\|z_l - z^*\|)$ can be proved similarly. □

5.2. **Rapid convergence to an infeasible stationary point.** In this subsection, we consider the rate of convergence to an infeasible stationary point. We prove that d_{xl} can be a superlinearly or quadratically convergent step provided the penalty parameter ρ_l is appropriately updated at z_l. The barrier parameter $\beta_l \in (0, \beta_0]$ can be any finite number.

Other than the general assumptions in Assumption 5.1, we also need some additional conditions for local analysis in this subsection.

Assumption 5.7.
(1) $\rho^* = 0$, x^* is an infeasible stationary point;
(2) $\rho_l u_l \to 0$ as $l \to \infty$.

The above assumption does not prevent $\|u_l\|$ from tending to ∞. Since the inner algorithm is terminated finitely for every l, one can update ρ_l appropriately such that Assumption 5.7 (2) holds. With this assumption, $\|\lambda_l\|$ is bounded.

For simplicity, we set $\hat{u}_l = \rho_l u_l$. Let $\mathcal{P}^* = \{i \in \mathcal{I} | c_i(x^*) > 0\}$, and $\mathcal{N}^* = \{i \in \mathcal{I} | c_i(x^*) < 0\}$. In virtue of (18) and (19), $\lambda_i^* > 0$ and $y_i^* = 0$ for $i \in \mathcal{P}^*$, $\lambda_i^* = 0$ and $y_i^* > 0$ for $i \in \mathcal{N}^*$. They imply $\nu_i^* = 1$ for $i \in \mathcal{P}^*$, and $\nu_i^* = 0$ for $i \in \mathcal{N}^*$. Thus, $\mathcal{P}^* \subseteq \mathcal{W}^* \subseteq \mathcal{P}^* \cup \mathcal{I}^*$.

Let us consider the system

$$F_{(\beta,\rho)}(x, \hat{u}) = 0, \tag{64}$$

where $F_{(\beta,\rho)}(x, \hat{u}) = \begin{pmatrix} \rho\nabla f + \sum_{i\in\hat{\mathcal{I}}^*} \hat{\lambda}_i \nabla c_i + \sum_{i\in\mathcal{P}^*} (c_i + \hat{u}_{li})\nabla c_i c_i + \hat{y}_i, \quad i \in \hat{\mathcal{I}}^* \end{pmatrix}$, $\hat{\mathcal{I}}^* = \mathcal{I}^* \cap \mathcal{W}^*$, $\hat{\lambda}_i = \frac{1}{2}[\sqrt{(c_i + \hat{u}_i)^2 + 4\rho\beta} + c_i + \hat{u}_i]$, $\hat{y}_i = \frac{1}{2}[\sqrt{(c_i + \hat{u}_i)^2 + 4\rho\beta} -$

$c_i - \hat{u}_i]$. Obviously, when $\rho = 0$ and x^* is an infeasible stationary point of problem (1)–(2), $(x^*, 0)$ is a solution of (64).

Although our algorithm is totally different from those in [5, 8], we can similarly establish the following local convergence results.

Lemma 5.8. *Suppose that Assumptions 5.1 and 5.7 hold. Let $\hat{u}_{\mathcal{I}\backslash\hat{\mathcal{I}}^*} = 0$. Then there exists a constant $\hat{\rho} > 0$ such that, for $\rho \le \hat{\rho}$, the system (64) has a unique solution $(x^*(\rho), \hat{u}^*(\rho))$ with $\hat{u}_i^*(\rho) = 0$ for $i \in \mathcal{I}\backslash\hat{\mathcal{I}}^*$, and*

$$\left\| \begin{pmatrix} x^*(\rho) - x^* \\ \hat{u}^*(\rho) \end{pmatrix} \right\| \le M\rho \tag{65}$$

for some positive constant M independent of ρ.

Proof. Let $\hat{F}_{(\beta,\rho)}(x, \hat{u}_{\hat{\mathcal{I}}^*}) = F_{(\beta,\rho)}(x, \hat{u})$ with $\hat{u}_{\mathcal{I}\backslash\hat{\mathcal{I}}^*} = 0$. Note that $\hat{F}_{(\beta^*,0)}(x^*, 0) = 0$ and $\hat{F}_{(\beta,\rho)}(x, \hat{u}_{\hat{\mathcal{I}}^*})$ is continuously differentiable on $(x, \hat{u}_{\hat{\mathcal{I}}^*})$. Furthermore,

$$\nabla \hat{F}_{(\beta,\rho)}(x, \hat{u}_{\hat{\mathcal{I}}^*}) = \begin{pmatrix} G(x, \hat{u}_{\hat{\mathcal{I}}^*}) & [\nu_i^* \nabla c_i(x^*), i \in \hat{\mathcal{I}}^*] \\ [\nu_i^* \nabla c_i(x^*), i \in \hat{\mathcal{I}}^*]^T & -\mathrm{diag}(1 - \nu_i^*, i \in \hat{\mathcal{I}}^*) \end{pmatrix}, \tag{66}$$

where

$$G(x, \hat{u}_{\hat{\mathcal{I}}^*}) = \rho \nabla^2 f + \sum_{i \in \hat{\mathcal{I}}^*} \hat{\lambda}_i \nabla^2 c_i + \sum_{i \in \mathcal{P}^*} c_i \nabla^2 c_i + \sum_{i \in \hat{\mathcal{I}}^*} \hat{\nu}_i \nabla c_i \nabla c_i^T + \sum_{i \in \mathcal{P}^*} \nabla c_i \nabla c_i^T.$$

Let $J_F^* = \lim_{\rho \to 0} \nabla \hat{F}_{(\beta,\rho)}(x, \hat{u}_{\hat{\mathcal{I}}^*})$. By items (3) and (4) of Assumption 5.1 and Assumption 5.7, J_F^* is nonsingular. Thus, the result follows immediately by applying the Implicit Function Theorem (p.128 of [30]). ☐

Corresponding to the mapping $F_{(\beta,\rho)}(x, \hat{u})$, we set

$$\hat{\phi}_{(\beta,\rho)}(x, u) = \begin{pmatrix} \rho \nabla f + \sum_{i \in \mathcal{I}} \lambda_i \nabla c_i \\ c_i + y_i, \quad i \in \hat{\mathcal{I}}^* \end{pmatrix}.$$

Lemma 5.9. *Suppose that Assumptions 5.1 and 5.7 hold. Then, for all sufficiently large l,*

$$\|\hat{\phi}_{(\beta_l,\rho_l)}(x_l, u_l) - F_{(\beta_l,\rho_l)}(x_l, \hat{u}_l)\| \le M\rho_l \tag{67}$$

for some positive constant M independent of ρ_l.

Proof. For sufficiently large l, one has $c_{li} < 0,\ i \in \mathcal{N}^*$ and $c_{li} > 0,\ i \in \mathcal{P}^*$. Thus, for $i \in \mathcal{N}^*$,

$$\begin{aligned} \lambda_{li} &= \frac{1}{2}\left(\sqrt{(c_{li} + \rho_l u_{li})^2 + 4\rho_l \beta_l} + c_{li} + \rho_l u_{li}\right) \\ &= \frac{2\beta_l \rho_l}{\sqrt{(c_{li} + \rho_l u_{li})^2 + 4\rho_l \beta_l} - c_{li} - \rho_l u_{li}} \\ &\le \frac{\sqrt{M}}{m+p}\,\rho_l, \end{aligned} \tag{68}$$

and for $i \in \mathcal{P}^*$,

$$\begin{aligned} (\lambda_{li} - c_{li} - \hat{u}_{li}) &= \frac{1}{2}\left(\sqrt{(c_{li} + \rho_l u_{li})^2 + 4\rho_l \beta_l} - c_{li} - \rho_l u_{li}\right) \\ &= \frac{2\beta_l \rho_l}{\sqrt{(c_{li} + \rho_l u_{li})^2 + 4\rho_l \beta_l} + c_{li} + \rho_l u_{li}} \\ &\le \frac{\sqrt{M}}{m+p}\,\rho_l \end{aligned} \tag{69}$$

for some positive constant M independent of ρ_l. Therefore, for sufficiently large l, there holds

$$\begin{aligned}\|\hat{\phi}_{(\beta_l,\rho_l)}(x_l,u_l) - F_{(\beta_l,\rho_l)}(x_l,\hat{u}_l)\| &\leq \|\sum_{i\in\mathcal{P}^*}(\lambda_{li}-c_{li}-\hat{u}_{li})\nabla c_{li} + \sum_{i\in\mathcal{N}^*}\lambda_{li}\nabla c_{li}\| \\ &\leq M\rho_l \end{aligned} \tag{70}$$

provided $\|\nabla c_{li}\| \leq \sqrt{M}$ for $i \in \mathcal{P}^* \cup \mathcal{N}^*$. Then the result follows immediately from items (1) and (3) of Assumption 5.1. ☐

For simplicity, we denote $\hat{z}^*(\rho) = (x^*(\rho), \hat{u}^*_{\hat{\mathcal{I}}^*}(\rho))$, $\hat{z} = (x, \hat{u}_{\hat{\mathcal{I}}^*})$, $\hat{w}^* = (x^*, 0_{\hat{\mathcal{I}}^*})$. The following two lemmas can be obtained in a way similar to Lemmas 5.4 and 5.5 and hence their proofs are neglected here for brevity.

Lemma 5.10. *Suppose that Assumptions 5.1 and 5.7 hold. Let $\hat{F}_l(\hat{z}) = \hat{F}_{(\beta_l,\rho_l)}(\hat{z})$. Then there are sufficiently small scalars $\epsilon > 0$ and $\hat{\rho} > 0$, and positive constants M_0 and L_0, such that, for all $\rho_l \leq \hat{\rho}$ and $\hat{z} \in \{\hat{z} | \|\hat{z} - \hat{z}^*(\rho_l)\| < \epsilon\}$, $\nabla \hat{F}_l(\hat{z})$ is invertible,*

$$\|[\nabla \hat{F}_l(\hat{z})]^{-1}\| \leq M_0, \tag{71}$$

and

$$\|\nabla \hat{F}_l(\hat{z})^T(\hat{z} - \hat{z}^*(\rho_l)) - \hat{F}_l(\hat{z})\| \leq L_0\|\hat{z} - \hat{z}^*(\rho_l)\|^2. \tag{72}$$

Lemma 5.11. *Suppose that Assumptions 5.1 and 5.7 hold. Then there are sufficiently small scalars $\epsilon > 0$ and $\hat{\rho} > 0$, such that, for all $\rho_l \leq \hat{\rho}$ and $\hat{z} \in \{\hat{z} | \|\hat{z} - \hat{z}^*(\rho_l)\| < \epsilon\}$,*

$$\|\hat{F}_l(\hat{z})\| \leq 2M_1\|\hat{z} - \hat{z}^*(\rho_l)\|,$$

where $M_1 = \sup_{\|\hat{z}-\hat{z}^*(\rho_l)\|<\epsilon} \|\nabla \hat{F}_l(\hat{z})\|$.

Denote $(r_l)_{\hat{\mathcal{I}}^*} = (c_{li} + y_{li}, i \in \hat{\mathcal{I}}^*)$, $(R_l^T)_{\hat{\mathcal{I}}^*} = (R_{li}^T, i \in \hat{\mathcal{I}}^*)$, $\hat{d}_l = (d_{xl}, \rho_l(d_{ul})_{\hat{\mathcal{I}}^*})$, where (d_{xl}, d_{ul}) is the solution of QP (32)–(33). Let $\tilde{d}_l$ be the unique solution of the equation $\nabla \hat{F}_l(\hat{z}_l)^T d = -\hat{F}_l(\hat{z}_l)$. Now we are ready to provide the following local convergence result when the whole algorithm converges to an infeasible stationary point.

Theorem 5.12. *Suppose that Assumptions 5.1 and 5.7 hold. If $\rho_l = O(\|x_l - x^*\|)^2$, and d_l^c is computed such that $\|(r_l)_{\hat{\mathcal{I}}^*} + (R_l^T)_{\hat{\mathcal{I}}^*} d_l^c\| = O(\|(r_l)_{\hat{\mathcal{I}}^*}\|^2)$, then*

$$\|x_l + d_{xl} - x^*\| = O(\|x_l - x^*\|^2). \tag{73}$$

If, instead, $\rho_l = o(\|x_l - x^*\|)$, and $\|(r_l)_{\hat{\mathcal{I}}^*} + (R_l^T)_{\hat{\mathcal{I}}^*} d_l^c\| = o(\|(r_l)_{\hat{\mathcal{I}}^*}\|)$, then the convergence is superlinear.

Proof. Assume that $\rho_l = O(\|x_l - x^*\|)^2$. In order to prove the result, we first show that

$$\limsup_{l\to\infty} \|\hat{z}_l + \hat{d}_l - \hat{z}^*\|/\|\hat{z}_l - \hat{z}^*\|^2 \leq \gamma, \tag{74}$$

where $\gamma > 0$ is a constant.

Due to $\|\hat{z}_l + \tilde{d}_l - \hat{z}^*(\rho_l)\| \leq \|[\nabla \hat{F}_l^T]^{-1}\|\|\nabla \hat{F}_l^T(\hat{z}_l - \hat{z}^*(\rho_l)) - \hat{F}_l\|$, by (71) and (72), one has

$$\|\hat{z}_l + \tilde{d}_l - \hat{z}^*(\rho_l)\| = O(\|\hat{z}_l - z^*(\rho_l)\|^2). \tag{75}$$

Note that

$$\begin{aligned}\|\hat{d}_l - \tilde{d}_l\| &\leq \|[\nabla \hat{F}_l^T]^{-1}(\hat{\phi}_l - F_l)\| + M_0\|(r_l)_{\hat{\mathcal{I}}^*} + (R_l^T)_{\hat{\mathcal{I}}^*} d_l^c\| \\ &= O(\rho_l) + O(\|\hat{z}_l - \hat{z}^*(\rho_l)\|^2)\end{aligned}$$

(by (71), Lemmas 5.9 and 5.11) and

$$\|\hat{z}_l + \hat{d}_l - \hat{z}^*\| \leq \|\hat{z}_l + \tilde{d}_l - \hat{z}^*(\rho_l)\| + \|\hat{d}_l - \tilde{d}_l\| + \|\hat{z}^*(\rho_l) - \hat{z}^*\|,$$

it follows from (65) and (75) that

$$\|\hat{z}_l + \hat{d}_l - \hat{z}^*\| = O(\|\hat{z}_l - \hat{z}^*\|^2). \tag{76}$$

Therefore, (74) is obtained.

Since $\|\hat{z}_l + \hat{d}_l - \hat{z}^*\|^2 = \|x_l + d_{xl} - x^*\|^2 + \rho_l^2(\|(u_l)_{\hat{\mathcal{I}}^*} + (d_{ul})_{\hat{\mathcal{I}}^*}\|^2)$ and $\|\hat{z}_l - \hat{z}^*\|^2 = \|x_l - x^*\|^2 + \rho_l^2(\|(u_l)_{\hat{\mathcal{I}}^*}\|^2)$, if $\rho_l = O(\|x_l - x^*\|)^2$, then

$$\|x_l + d_{xl} - x^*\| = O(\|x_l - x^*\|^2). \tag{77}$$

One can similarly prove the result for the case of $\rho_l = o(\|x_l - x^*\|)$. □

6. **Numerical experiments.** Our main motivation for the numerical experiments is to observe the performance of our algorithm when applied to solving some infeasible nonlinear programs in the literature. We implemented our algorithm in MATLAB (version R2008a). The numerical tests were conducted on a Lenovo laptop with the LINUX operating system (Fedora 11).

The initial parameters were chosen as follows: $\beta_0 = 0.1$, $\delta = 0.5$, $\sigma = 10^{-4}$, and $\epsilon = 10^{-8}$. The initial penalty was $\rho_0 = \min\{100, \max(1, \|\max(0, c(x_0))\|/|f(x_0)|)\}$, which depended on the initial point. Simply, we took $H_0 = \rho I$ (where $I \in \Re^{n\times n}$ is the identity matrix), H_k was updated similarly by the well-known Powell's damped BFGS update formula (for example, see [3, 29]).

The vector d_k^c was derived by Algorithm 6.1 of [25]. For solving the QP subproblem (32)–(33), we first computed the null-space matrix W_k of R_k^T by the MATLAB null-space routine, then computed the solution of the QP by forming the reduced Hessian explicitly and using the MATLAB routine of bi-conjugate gradients method with preconditioner generated by the sparse incomplete Cholesky-Infinity factorization, which was presented by Zhang [40] for avoiding numerically zero pivots in the sparse incomplete Cholesky factorization.

In the inner algorithm, ξ_{k+1} is further updated such that $\|\xi_{k+1}g_k\|_\infty \leq 0.1$ (where g_k is the multiplier of the QP (see Lemma 4.5)), and $\xi_{k+1}(\max(\max(\rho_l u_k, 0)))^{1.1} \leq 1$ so that $\xi_{k+1}\rho_l u_k \to 0$ as $\xi_{k+1} \to 0$. Functions θ_1 and θ_2 are defined as $\theta_1(\beta) = 10\beta$ and $\theta_2(\rho) = \rho$, respectively. In order to obtain rapid convergence, we update β_l to $\beta_{l+1} = \min(0.1\beta_l, \|\phi_{(\beta_l,\rho_l)}(z_{k+1})\|_\infty^{1.5})$ when we need to reduce β_l. If ρ_l needs to be updated, ρ_l is reduced to

$$\rho_{l+1} = \min\{\xi_{k+1}\rho_l, \|\psi_{(\beta_l,\xi_{k+1}\rho_l)}(z_{k+1})\|_\infty^2, (\|\lambda(z_{k+1};\beta_l,\rho_l)\|_\infty/\rho_l)^{-2}\}$$

provided $\|r_k\| - \|r_k + R_k^T d_k\| < 0.01\|r_k\|$, otherwise $\rho_{l+1} = \xi_{k+1}\rho_l$, where

$$\psi_{(\beta_l,\xi_{k+1}\rho_l)}(z_{k+1}) = \xi_{k+1}\rho_l \nabla f_{k+1} + \sum_{i\in\mathcal{I}} \lambda_i(z_{k+1};\beta_l,\xi_{k+1}\rho_l)\nabla c_{k+1,i}.$$

We use $\|\psi_{(\beta_l,\xi_{k+1}\rho_l)}(z_{k+1})\|_\infty$ to measure the convergence to the infeasible stationary point, which is the same as [8]. It is easy to note that $\|\psi_{(\beta_l,\xi_{k+1}\rho_l)}(z_{k+1})\|_\infty \to 0$ as $\rho_l \to 0$ due to Theorem 4.10. The whole algorithm was terminated as either

TABLE 1. Output for test problem (TP1)

l	f_l	v_l	$\|\phi_l\|_\infty$	$\|\psi_l\|_\infty$	β_l	ρ_l	k
0	5	16.6132	129.6234	129.6234	0.1000	3.3226	-
1	0.1606	2.0205	4.8082	0.7313	0.1000	0.0972	3
2	-0.0149	2.0002	0.0989	0.0445	0.1000	0.0020	4
3	-0.0036	2.0000	0.0029	0.0018	0.1000	3.1595e-06	3
4	-0.0029	2.0000	3.1674e-06	2.8185e-06	0.1000	1.0000e-09	1
5	0.0018	2.0000	1.0011e-09	6.7212e-10	-	-	-

$\beta_l < \epsilon$ or $\rho_l < \epsilon$, or the total number of iterations (that is, the number of solving QP (32)–(33)) was greater than 1000.

Two infeasible test problems are taken from Byrd, Curtis and Nocedal [8]. The results for them are reported respectively in Tables 1–2, in which the numbers in column l are the order numbers of outer iterations, $f_l = f(x_l)$, $v_l = \|\max\{0, c(x_l)\}\|$, $\|\phi_l\|_\infty = \|\phi_{(\beta_l,\rho_l)}(z_{k+1})\|_\infty$, $\|\psi_l\|_\infty = \|\psi_{(\beta_l,\xi_{k+1}\rho_l)}(z_{k+1})\|_\infty$, k is the number of inner iterations needed for changing parameters.

The first test problem is the so-called *isolated* problem:

$$\text{(TP1)} \qquad \begin{array}{ll} \min & x_1 + x_2 \\ \text{s.t.} & x_1^2 - x_2 + 1 \le 0, \\ & x_1^2 + x_2 + 1 \le 0, \\ & -x_1 + x_2^2 + 1 \le 0, \\ & x_1 + x_2^2 + 1 \le 0. \end{array}$$

The standard initial point is $x_0 = (3, 2)$, its solution $x^* = (0, 0)$ is a strict minimizer of the infeasibility measure (43). The algorithm presented in [8] found this point. Our algorithm terminates at an approximate point to it. Table 1 shows that, when $\rho_3 = 3.1595e-06$ is reduced to $\rho_4 = 1.0000e-09$, rapid convergence emerged since $\|\psi_3\|_\infty$ is reduced superlinearly.

The second example is the *nactive* problem in [8]:

$$\text{(TP2)} \qquad \begin{array}{ll} \min & x_1 \\ \text{s.t.} & \frac{1}{2}(x_1 + x_2^2 + 1) \le 0, \\ & -x_1 + x_2^2 \le 0, \\ & x_1 - x_2^2 \le 0. \end{array}$$

The given initial point is $x_0 = (-20, 10)$. The point $x^* = (0, 0)$ derived by [8] was an infeasible stationary point with $\|\max(0, c^*)\| = 0.5$. Algorithm 3.4 terminates at a point approximating an infeasible stationary point $x^* = (-0.2000, 0.0000)$. Similar to that for (TP1), Table 2 indicates that, when $\rho_6 = 2.6880e-06$ is reduced to $\rho_7 = 1.0000e-09$, rapid convergence emerged since $\|\psi_6\|_\infty$ is reduced superlinearly.

In order to observe the strong global convergence of our algorithm, we also solve the counterexample presented by Wächter and Biegler [34] (also see Byrd, Marazzi and Nocedal [12]) and a standard test problem for which the minimizer is a singular stationary point of the nonlinear program by our algorithm.

TABLE 2. Output for test problem (TP2)

l	f_l	v_l	$\|\phi_l\|_\infty$	$\|\psi_l\|_\infty$	β_l	ρ_l	k
0	-20	126.6501	2.8052e+03	2.8052e+03	0.1000	6.3325	-
1	-172.5829	172.7978	1.0948e+03	6.2866	0.1000	0.8719	6
2	0.2155	0.7149	1.4269	0.7894	0.1000	0.3895	1
3	-0.1364	0.5550	0.3865	0.3865	0.0100	0.3895	3
4	-0.1416	0.5223	0.2864	0.2648	0.0100	0.1512	1
5	-0.1472	0.5140	0.1446	0.1446	0.0100	0.0209	4
6	-0.1997	0.4472	0.0084	0.0016	0.0100	2.6880e-06	3
7	-0.1999	0.4472	2.4923e-06	2.4923e-06	0.0100	1.0000e-09	1
8	-0.1999	0.4472	9.2732e-10	9.2732e-10	-	-	-

TABLE 3. Output for test problem (TP4)

l	f_l	v_l	$\|\phi_l\|_\infty$	$\|\psi_l\|_\infty$	β_l	ρ_l	k
0	20	2.8284	9.9557	9.9557	0.1000	1	-
1	0.2305	0.4167	0.8900	0.7008	0.0100	1	4
2	0.1652	0.1687	0.1631	0.0771	0.0100	0.3268	4
3	0.1690	0.1630	0.0503	0.0022	0.0100	4.7328e-06	1
4	0.8561	2.9531e-04	3.1379e-06	3.1379e-06	0.0100	1.0000e-09	14
5	0.9028	1.2372e-04	9.3463e-08	9.3463e-08	-	-	-

The counterexample has the following formulation:

$$\text{(TP3)}\qquad \begin{array}{ll} \min & x_1 \\ \text{s.t.} & x_1^2 - x_2 - 1 = 0, \\ & x_1 - x_3 - 2 = 0, \\ & x_2 \geq 0, \quad x_3 \geq 0. \end{array}$$

The initial point is $x_0 = (-4, 1, 1)$. This problem has a unique global minimizer $(2, 3, 0)$, at which gradients of the active constraints are linearly independent, and MFCQ holds. However, [34] showed that many line search interior-point methods could not find the minimizer, even failed to find a feasible solution. Our algorithm terminates at the approximate solution $x^* = (2.0000, 3.0000, 0.0000)$ in 16 iterations (including all numbers of inner iterations), where, for $l = 9$, $f_l = 2.0000$, $v_l = 6.0156e-06$, $\|\phi_l\|_\infty = 3.6515e-10$.

The standard test problem is the one taken from [23, Problem 13]:

$$\text{(TP4)}\qquad \begin{array}{ll} \min & (x_1 - 2)^2 + x_2^2 \\ \text{s.t.} & (1 - x_1)^3 - x_2 \geq 0, \\ & x_1 \geq 0, \quad x_2 \geq 0. \end{array}$$

The standard initial point $x_0 = (-2, -2)$ is an infeasible point. This problem was not solved in [31] and some other references, but the algorithms in [10, 32] got its approximate solution. Its optimal solution $x^* = (1, 0)$ is not a KKT point but is a singular stationary point, at which the gradients of active constraints are linearly dependent. Numerical results show that Algorithm 3.4 terminates at an approximate point to the solution, but it does not suggest rapid convergence for

TABLE 4. The last 4 inner iterations corresponding to $l = 4$ for test problem (TP4)

k	f_k	v_k	$\|\phi_k\|_\infty$	$\|\psi_k\|_\infty$	x_{k1}	x_{k2}
11	0.8500	5.7136e-04	5.6703e-04	5.6703e-04	1.0780	0.0001
12	0.8548	3.0434e-04	1.2222e-05	1.2222e-05	1.0754	-0.0002
13	0.8556	2.9845e-04	6.2125e-06	6.2125e-06	1.0750	-0.0002
14	0.8561	2.9531e-04	3.1379e-06	3.1379e-06	1.0747	-0.0002

either inner or outer iterations in Tables 3 and 4 where the last n columns are the components of iterates. In fact, we still do not have any theoretical result on rapid convergence to a singular stationary point of nonlinear programs in the literature.

In summary, the preceding numerical results not only demonstrate our global convergence results on Algorithm 3.4 for infeasible, hard and degenerate nonlinear programs, but also demonstrate our locally rapid convergence results on Algorithm 3.4 with convergence to an infeasible stationary point of a nonlinear program which is infeasible.

7. **Conclusion.**

Upon great success in solving large-scale linear programming problems, the interior-point approach has effectively been extended to solving general convex programming (such as semidefinite and cone programming) and nonconvex programming problems. The research on interior-point methods for nonlinear programs has been one of focuses of optimization area in recent years. Based on a distinctive two-parameter primal-dual nonlinear system, which corresponds to the KKT point and the infeasible stationary point of nonlinear programs, respectively, as one of two parameters vanishes, we have presented a new interior-point method for nonlinear programs in this paper. Our method always produces interior-point iterates without truncation of the step. The method not only can be globally convergent and locally quick convergent to KKT points when the problem is feasible, but also can globally converge to an infeasible stationary point and rapidly detect the infeasibility of the solved problem when the problem is infeasible. A possible future topic of the subsequent research is to consider similar methods in solving linear programming or semidefinite programming problems.

REFERENCES

[1] R. Andreani, E. G. Birgin, J. M. Martinez and M. L. Schuverdt, Augmented Lagrangian methods under the constant positive linear dependence constraint qualification, *Math. Program.*, **111** (2008), 5–32.

[2] P. Armand and J. Benoist, A local convergence property of primal-dual methods for nonlinear programming, *Math. Program.*, **115** (2008), 199–222.

[3] P. Armand, J. C. Gilbert and S. Jan-Jégou, A feasible BFGS interior point algorithm for solving convex minimization problems, *SIAM J. Optim.*, **11** (2000), 199–222.

[4] I. Bongartz, A. R. Conn, N. I. M. Gould and P. L. Toint, CUTE: Constrained and Unconstrained Testing Environment, *ACM Tran. Math. Software*, **21** (1995), 123–160.

[5] J. V. Burke, F. E. Curtis and H. Wang, A sequential quadratic optimization algorithm with rapid infeasibility detection, *SIAM J. Optim.*, **24** (2014), 839–872.

[6] J. V. Burke and S. P. Han, A robust sequential quadratic programming method, *Math. Program.*, **43** (1989), 277–303.

[7] R. H. Byrd, *Robust Trust-Region Method for Constrained Optimization*, Paper presented at the SIAM Conference on Optimization, Houston, TX, 1987.

[8] R. H. Byrd, F. E. Curtis and J. Nocedal, Infeasibility detection and SQP methods for nonlinear optimization, *SIAM J. Optim.*, **20** (2010), 2281–2299.

[9] R. H. Byrd, J. C. Gilbert and J. Nocedal, A trust region method based on interior point techniques for nonlinear programming, *Math. Program.*, **89** (2000), 149–185.

[10] R. H. Byrd, M. E. Hribar and J. Nocedal, An interior point algorithm for large-scale nonlinear programming, *SIAM J. Optim.*, **9** (1999), 877–900.

[11] R. H. Byrd, G. Liu and J. Nocedal, On the local behaviour of an interior point method for nonlinear programming, In *Numerical Analysis 1997 (eds. D. F. Griffiths and D. J. Higham)*, Addison-Wesley Longman, Reading, MA, **380** (1998), 37–56.

[12] R. H. Byrd, M. Marazzi and J. Nocedal, On the convergence of Newton iterations to non-stationary points, *Math. Program.*, **99** (2004), 127–148.

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

[14] F. E. Curtis, A penalty-interior-point algorithm for nonlinear constrained optimization, *Math. Program. Comput.*, **4** (2012), 181–209.

[15] 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, *J. Optim. Theory Appl.*, **89** (1996), 507–541.

[16] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley and Sons, New York, 1968; republished as Classics in Appl. Math. 4, SIAM, Philadelphia, 1990.

[17] R. Fletcher, *Practical Methods for Optimization. Vol. 2: Constrained Optimization*, John Wiley and Sons, Chichester, 1980.

[18] A. Forsgren and P. E. Gill, Primal-dual interior methods for nonconvex nonlinear programming, *SIAM J. Optim.*, **8** (1998), 1132–1152.

[19] A. Forsgren, Ph. E. Gill and M. H. Wright, Interior methods for nonlinear optimization, *SIAM Review*, **44** (2002), 525–597.

[20] D. M. Gay, M. L. Overton and M. H. Wright, A primal-dual interior method for nonconvex nonlinear programming, in *Advances in Nonlinear Programming,* (ed. Y.-X. Yuan), Kluwer Academic Publishers, Dordrecht, **14** (1998), 31–56.

[21] E. M. Gertz and Ph. E. Gill, A primal-dual trust region algorithm for nonlinear optimization, *Math. Program.*, **100** (2004), 49–94.

[22] N. I. M. Gould, D. Orban and Ph. L. Toint, An interior-point ℓ_1-penalty method for nonlinear optimization, in *Recent Developments in Numerical Analysis and Optimization*, Proceedings of NAOIII 2014, Springer, Verlag, **134** (2015), 117–150.

[23] W. Hock and K. Schittkowski, *Test Examples for Nonlinear Programming Codes*, Lecture Notes in Eco. and Math. Systems 187, Springer-Verlag, Berlin, New York, 1981.

[24] X. -W. Liu, G. Perakis and J. Sun, A robust SQP method for mathematical programs with linear complementarity constraints, *Comput. Optim. Appl.*, **34** (2006), 5–33.

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

[26] X. -W. Liu and Y. -X. Yuan, A robust algorithm for optimization with general equality and inequality constraints, *SIAM J. Sci. Comput.*, **22** (2000), 517–534.

[27] X. -W. Liu and Y. -X. Yuan, A null-space primal-dual interior-point algorithm for nonlinear optimization with nice convergence properties, *Math. Program.*, **125** (2010), 163–193.

[28] J. Nocedal, F. Öztoprak and R. A. Waltz, An interior point method for nonlinear programming with infeasibility detection capabilities, *Optim. Methods Softw.*, **29** (2014), 837–854.

[29] J. Nocedal and S. Wright, *Numerical Optimization*, Springer-Verlag New York, Inc., 1999.

[30] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York and London, 1970.

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

[32] P. Tseng, Convergent infeasible interior-point trust-region methods for constrained minimization, *SIAM J. Optim.*, **13** (2002), 432–469.

[33] M. Ulbrich, S. Ulbrich and L. N. Vicente, A globally convergent primal-dual interior-point filter method for nonlinear programming, *Math. Program.*, **100** (2004), 379–410.

[34] A. Wächter and L. T. Biegler, Failure of global convergence for a class of interior point methods for nonlinear programming, *Math. Program.*, **88** (2000), 565–574.

[35] A. Wächter and L. T. Biegler, Line search filter methods for nonlinear programming: Motivation and global convergence, *SIAM J. Optim.*, **16** (2005), 1–31.

[36] A. Wächter and L. T. Biegler, On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming, *Math. Program.*, **106** (2006), 25–57.

[37] M. H. Wright, Why a pure primal Newton barrier step may be infeasible? *SIAM J. Optim.*, **5** (1995), 1–12.

[38] S. J. Wright, On the convergence of the Newton/Log-barrier method, *Math. Program.*, **90** (2001), 71–100.

[39] Y. -X. Yuan, On the convergence of a new trust region algorithm, *Numer. Math.*, **70** (1995), 515–539.

[40] Y. Zhang, Solving large-scale linear programs by interior-point methods under the MATLAB environment, *Optim. Methods Softw.,* **10** (1998), 1–31.

Received July 2018; revised August 2018.

E-mail address: dyh@lsec.cc.ac.cn
E-mail address: mathlxw@hebut.edu.cn
E-mail address: jie.sun@curtin.edu.au