Global convergence analysis of line search interior point methods for nonlinear programming without regularity assumptions[*]

Xinwei Liu[†]and Jie Sun[‡]

Abstract. It has been noticed by Wächter and Biegler that a number of interior point methods for nonlinear programming based on line search strategy may generate a sequence converging to an infeasible point. We show that by adopting a suitable merit function, a modified primal-dual equation, and a proper line search procedure, a class of interior point methods of line search type will generate a sequence such that either all limit points of the sequence are KKT points, or one of the limit points is a Fritz-John point, or one of the limit points is an infeasible point that is a stationary point of minimizing a function that measures the extent of violation to the constraint system. The analysis does not depend on regularity assumptions on the problem. Instead, it uses a set of satisfiable conditions on the algorithm's implementation to derive the desired convergence property.

Key words: nonlinear programming, interior point methods, convergence

---

[*] Research is partially supported by Grant R-314-000-026/028/042-112 of National University of Singapore and Singapore-MIT Alliance.

[†] Department of Decision Sciences and Singapore-MIT Alliance, National University of Singapore, Republic of Singapore. Email: smaliuxw@nus.edu.sg

[‡] Department of Decision Sciences and Singapore-MIT Alliance, National University of Singapore, Republic of Singapore. Fax (65)6779-2621. Email: jsun@nus.edu.sg

1. Introduction

Consider the nonlinear program

$$\min \; f(x) \tag{1.1}$$

$$\text{s.t. } \; c(x) \leq 0, \tag{1.2}$$

where $c(x) = (c_1(x), \ldots, c_m(x))^\top \in \Re^m$, $f(x)$ and $c_i(x), i = 1, \ldots, m$ are twice continuously differentiable functions. Interior point methods (IPMs) for problem (1.1)-(1.2) solve the log-barrier problem

$$\min \; f(x) - \mu \sum_{i=1}^{m} \ln y_i \tag{1.3}$$

$$\text{s.t. } \; c(x) + y = 0, \tag{1.4}$$

approximately while reducing μ to 0 (see, e.g., [6]). Depending on whether a line search strategy or a trust region strategy is used, they could be further categorized as line search IPMs or trust region IPMs. A model line search IPM for solving problem (1.3)-(1.4) is as follows.

Algorithm 1.1 *(A model algorithm for problem (1.3)-(1.4))*

Step 1 *Given* $z^0 := (x^0, y^0) \in \Re^n \times \Re^m_{++}$. *Let* $k := 0$;

Step 2 *Calculate a search direction* (d_x^k, d_y^k) *which satisfies the linearized constraints*

$$c(x^k) + y^k + \nabla c(x^k)^\top d_x^k + d_y^k = 0; \tag{1.5}$$

Step 3 *Select a step-size* $\alpha_k \in (0, 1]$ *such that*

$$y^k + \alpha_k d_y^k \geq \xi y^k, \tag{1.6}$$

where $\xi \in (0, 1)$ *is a constant. Let*

$$z^{k+1} := (x^{k+1}, y^{k+1}) = (x^k, y^k) + \alpha_k (d_x^k, d_y^k). \tag{1.7}$$

Step 4 *Let* $k := k + 1$ *and go to Step 2.*

Consider the problem (1.3)-(1.4) with constraint functions defined by

$$-x^2 + y_1 - a = 0, \tag{1.8}$$

$$-x + y_2 + b = 0. \tag{1.9}$$

Wächter and Biegler [15] showed that the sequence generated by Algorithm 1.1 can converge to an infeasible point if the algorithm is started from any point in a certain region. Thus, Algorithm 1.1 may not find a stationary point in general even if the solution is well defined.

Based on a trust region strategy, a new scheme

$$y^{k+1} = \max\{y^k + d_y^k, -c(x^{k+1})\} \tag{1.10}$$

for updating the slack vector y^k is introduced by Byrd, Gilbert, and Nocedal [1] to guarantee the convergence of their IPM to a point with certain stationary properties. A similar but more sophisticated technique is used in an infeasible IPM proposed by Tseng [13].

In our numerical experiment, we noted that by using a similar technique to (1.10) in Algorithm 1.1 (which is different from the algorithms in [1, 13] since it is based on a line search procedure) the algorithm could find the right solutions for some cases. However, without further modification it is not enough to guarantee global convergence for all tested problems; particularly, the algorithm could not find a point with stationary properties for the Wächter-Biegler example. Hence, we introduced a new IPM using a sequential quadratic programming strategy in [10]. It was shown in theory and by numerical experiments that the new algorithm behaves well in terms of global convergence. In particular, it can always find a point with certain stationary properties. In developing such an algorithm, we also realized that there is a set of implementation guidelines that can guarantee the good convergence behavior of the line search IPM, which should be identified and studied further.

In this paper we attempt to identify some essential factors for line search IPMs to have good convergence properties, which include an appropriate merit function, a suitable control procedure of the slack vector, and a proper steplength for the dual vector. We consider two algorithms. The first one, Algorithm 2.1, adopts the same merit function as that of [1, 13], but is based on the line search strategy. As a result, it generates a sequence that may have all limit points to be KKT points or may have at least one limit point to be a Fritz-John point. The negative side is that it may terminate at an infeasible point as Algorithm 1.1 did with the Wächter-Biegler example. The second one (Algorithm 4.1) is a refinement of Algorithm 2.1, which is more selective in computing the Newton step and in updating the dual variables. We show that Algorithm 4.1 will generate a sequence having the following property:

(a) Either all of its limit points are KKT points or

(b) one of its limit points is a Fritz-John point or

(c) one of its limit points is an infeasible point that is a stationary point of a function measuring the infeasibility.

We list some conditions that a specific implementation of Algorithm 4.1 should satisfy in order to have the aforementioned convergence property. We point out that those conditions are satisfied by the trust region IPM of [1], and the line search IPM of [10].

It should be noted that, while our analysis requires some general assumptions such as the boundedness of the generated sequence and the uniform positive-definiteness of the approximate Lagrangian Hessian matrices, it is entirely independent of the so-called regularity or nondegeneracy assumptions. Therefore, compared to other line search IPMs, the algorithms can be applied to a wider spectrum of problems with less a priori knowledge of the problems.

The paper is organized as follows. We present Algorithm 2.1 and a numerical example in Section 2. The example shows a possible behavior of the algorithm and is indicative to a relationship between the slack vector and the penalty parameter of the merit function. We show in Section 3 that the limit points of the sequence generated by Algorithm 2.1 are either all KKT points, or contain a Fritz-John point, or contain an infeasible point. In Section 4 we present Algorithm 4.1 and some key sufficient conditions that guarantee the generated sequence has either a Fritz-John point or an infeasible limit point with some stationary properties if not all of the limit points are KKT points. To demonstrate the satisfiability of those conditions, we prove in Section 5 that the conditions are satisfied by recent IPMs proposed by Byrd, Gilbert, Marazzi and Nocedal [1, 2] and by the authors [10].

2. The first Algorithm and a numerical example

We describe a special version of Algorithm 1.1 that incorporates a specific merit function as well as some techniques of IPMs used in [1, 4, 5, 7, 8, 9, 12, 14, 16].

Define the merit function as

$$\phi(z;\rho) = f(x) - \mu \sum_{i=1}^{m} \ln y_i + \rho \|c(x) + y\|, \tag{2.1}$$

where $z = (x, y)$, $x = (x_1, ..., x_n)^\top$, $y = (y_1, ..., y_m)^\top$, and $\rho > 0$ is a penalty parameter updated automatically in each iteration (see details below).

It has been proved [10, Proposition 3.1] that $\phi(z;\rho)$ is directional differentiable. Let $\phi^{'}(z^k, d^k; \rho)$ be its directional derivative at z^k along d^k and let

$$\pi_k(d^k;\rho) = {g^k}^\top d_x^k - \mu e^\top Y_k^{-1} d_y^k - \rho \left(\|c^k + y^k\| - \|c^k + y^k + A_k^\top d_x^k + d_y^k\| \right), \tag{2.2}$$

where $d^k = (d_x^k, d_y^k)$, $g^k = \nabla f(x^k)$, e is the vector of ones, $Y_k = \mathrm{diag}(y^k)$, $c^k = c(x^k)$, and $A_k = \nabla c(x^k)$. It has been shown [10, Proposition 3.1] that $\phi^{'}(z^k, d^k; \rho) \le \pi_k(d^k; \rho)$. Thus, the inequality $\pi_k(d^k;\rho) < 0$ implies that $\phi^{'}(z^k, d^k; \rho) < 0$, which means that d^k is a descent direction of the merit function.

The algorithm under consideration is as follows:

Algorithm 2.1

Step 1 *Set* $z^0 := (x^0, y^0) \in \Re^n \times \Re^m_{++}$, $\lambda^0 \in \Re^m_{++}$, *and* $\rho_0 > 0$ *as well as positive parameters* $\xi < 1$, $\delta < 1$, $\sigma_0 < 1/2$. *Select* β_1 *and* β_2 *so that* $\beta_1 < 1 < \beta_2$ *and* $\beta_1 \mu e \le Y_0 \lambda^0 \le \beta_2 \mu e$. *Let* $k := 0$*;*

Step 2 *Calculate the primal search direction* (d_x^k, d_y^k) *and the dual direction* d_λ^k *by the following primal-dual system of equations:*

$$\begin{aligned} B_k d_x \qquad\qquad &+ A_k d_\lambda = -(g^k + A_k \lambda^k), && (2.3) \\ \Lambda_k d_y &+ Y_k d_\lambda = -(Y_k \Lambda_k e - \mu e), && (2.4) \\ A_k^\top d_x + \qquad d_y & \qquad\quad\;\, = -(c^k + y^k), && (2.5) \end{aligned}$$

where B_k is a positive definite approximation to the Lagrangian Hessian

$$\nabla^2 L(x^k, \lambda^k) = \nabla^2 f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla^2 c_i(x^k); \tag{2.6}$$

Step 3 *If*

$$\pi_k(d^k; \rho_k) \leq -\frac{1}{2} {d_x^k}^\top B_k d_x^k - \frac{1}{2} {d_y^k}^\top Y_k^{-1} \Lambda_k d_y^k, \tag{2.7}$$

let $\rho_{k+1} = \rho_k$; Otherwise, increase ρ_k to ρ_{k+1} such that (2.7) holds for $\pi_k(d^k; \rho_{k+1})$;

Step 4 *Compute the largest $\hat{\alpha}_k \in [0, 1]$ such that $y^k + \hat{\alpha}_k d_y^k \geq \xi y^k$, and select the least non-negative integer l such that*

$$\phi(z^k + \delta^l \hat{\alpha}_k d^k; \rho_{k+1}) - \phi(z^k; \rho_{k+1}) \leq \sigma_0 \delta^l \hat{\alpha}_k \pi_k(d^k; \rho_{k+1}); \tag{2.8}$$

Step 5 *If necessary, increase l until there exists a scalar $\gamma \in [0, 1]$ satisfying*

$$\beta_1 \mu e \leq (Y_k + \delta^l \hat{\alpha}_k D_y^k)(\lambda^k + \gamma d_\lambda^k) \leq \beta_2 \mu e, \tag{2.9}$$

where $D_y^k = diag(d_y^k)$. For this fixed l, let γ_k be the biggest $\gamma \in [0, 1]$ satisfying (2.9) and let $\alpha_k = \delta^l \hat{\alpha}_k$. Then update the primal-dual iterate by setting

$$z^{k+1} := (x^{k+1}, y^{k+1}) = (x^k, y^k) + \alpha_k (d_x^k, d_y^k) \text{ and} \tag{2.10}$$

$$\lambda^{k+1} := \lambda^k + \gamma_k d_\lambda^k; \tag{2.11}$$

Step 6 *If the stopping criterion holds, stop; else update the approximate Hessian B_k by B_{k+1}, let $k := k + 1$, and go to Step 2.*

Remark 1. Step 3 is well defined. Let us consider two cases. Note that since (d_x^k, d_y^k) satisfies (2.5), one has

$$\pi_k(d^k; \rho_k) = {g^k}^\top d_x^k - \mu e^\top Y_k^{-1} d_y^k - \rho_k \|c^k + y^k\|. \tag{2.12}$$

Case (i). If $\|c^k + y^k\| = 0$, then premultiply (2.3) by ${d_x^k}^\top$, premultiply (2.4) by ${d_y^k}^\top Y_k^{-1}$, add the two equations together and use that $A_k^\top d_x^k + d_y^k = 0$, we obtain

$${g^k}^\top d_x^k - \mu e^\top Y_k^{-1} d_y^k = -({d_x^k}^\top B_k d_x^k + {d_y^k}^\top Y_k^{-1} \Lambda_k d_y^k). \tag{2.13}$$

Thus, (2.7) follows from (2.12) and the fact that ${d_x^k}^\top B_k d_x^k + {d_y^k}^\top Y_k^{-1} \Lambda_k d_y^k \geq 0$.
Case (ii). If $\|c^k + y^k\| \neq 0$ and if (2.7) does not hold, then we simply select

$$\rho_{k+1} = \max\{2\rho_k,\ ({g^k}^\top d_x^k - \mu e^\top Y_k^{-1} d_y^k + \frac{1}{2} {d_x^k}^\top B_k d_x^k + \frac{1}{2} {d_y^k}^\top Y_k^{-1} \Lambda_k d_y^k) / \|c^k + y^k\|\} \tag{2.14}$$

to make (2.7) hold for $\pi_k(d^k; \rho_{k+1})$.

Remark 2. Step 4 is well defined. Note that $\phi'(z^k, d^k; \rho) \leq \pi_k(d^k; \rho)$. If $d^k \neq 0$, then $\pi_k(d^k; \rho) < 0$ because of (2.7). It follows that $\phi'(z^k, d^k; \rho) < 0$. Thus d^k is a descent direction of ϕ. If $d^k = 0$, then $\pi_k(d^k; \rho_k) = 0$. Thus $\alpha_k = 1$ and $\gamma_k = 1$ by Steps 4 and 5. It follows from system (2.3)-(2.5) that z^{k+1} is a KKT point of (1.3)-(1.4) and therefore the algorithm will stop at Step 6 (we assume that Step 6 includes a stopping criterion to check whether z^{k+1} is a KKT point).

Remark 3. It follows from (2.4) that the inequalities (2.9) can be satisfied for some $\gamma_k \geq 0$ if l is sufficiently large. Thus Step 5 is well-defined. It should be noted that one always has

$$\beta_1 \mu e \leq Y_{k+1} \Lambda_{k+1} e \leq \beta_2 \mu e \quad \forall\ k \tag{2.15}$$

due to (2.9)-(2.11). There have been many methods for the update of the multiplier vector λ^k in the literature. Essentially, all of them satisfy (2.15) for some explicit or implicit constants β_1 and β_2 by virtue of their respective assumptions, including regularity assumptions. Since we consider algorithms without regularity assumptions in this paper, we may not have so many selections on the update and therefore the selection criterion (2.9) is more stringent.

Consider the following example

$$\min\ x \ \text{ subject to } \ -x^2 - 1 \leq 0,\ -x + 5 \leq 0, \tag{2.16}$$

which is a special case of the Wächter-Biegler example. Its corresponding log-barrier problem with barrier parameter $\mu = 0.01$ is

$$\min\ x - 0.01(\ln y_1 + \ln y_2) \ \text{ s.t. } \ -x^2 + y_1 - 1 = 0,\ -x + y_2 + 5 = 0. \tag{2.17}$$

We apply Algorithm 2.1 to solve (2.17). The parameters are selected as $\xi = 0.005$, $\beta_1 = 0.01$, $\beta_2 = 10$, $\rho_0 = 1$, $\sigma_0 = 0.1$, $\delta = 0.8$, $\epsilon = 10^{-6}$. Matrix B_0 is the identity matrix and B_k is generated by Powell's dumped Newton update procedure (see, e.g., [11]). The initial point is $x^0 = -4$, $y^0 = \lambda^0 = (1, 1)^\top$. The numerical results of the first six iterations are displayed in Table 1, where RC_1 and RC_2 are the respective residues of the constraints. The results show that the sequence produced by Algorithm 2.1 may converge to an infeasible point and it appears that in this case one has $\rho_k \to \infty$. In the next section, we will prove that the behavior of ρ_k turns out to be a key indicator of the algorithm. That is, if $\rho_k \not\to \infty$, all of the limit points of $\{x^k\}$ are KKT points, while if $\rho_k \to \infty$, one of the limit points must be infeasible or irregular (explained later).

3. Convergence analysis of Algorithm 2.1

We need the following blanket assumptions.

Assumption 3.1

(1) f and $c_i (i = 1, \ldots, m)$ are twice continuously differentiable functions on $\Re^n$;

Table 1. Numerical results on problem (2.17)

IT	x	y_1	y_2	RC_1	RC_2	ρ
0	-4	1	1	-16	10	1
1	-3.7395	0.9250	0.0050	-14.0591	8.7445	2
2	-3.5820	0.0046	0.0021	-13.8260	8.5841	52.1213
3	-3.5806	0.0231e-03	0.2438e-03	-13.8208	8.5809	2.2158e+04
4	-3.5805	0.0116e-05	0.4261e-05	-13.8203	8.5805	2.2566e+06
5	-3.5805	0.0058e-07	0.2227e-07	-13.8202	8.5805	7.4596e+09
6	-3.5805	0.0029e-09	0.1114e-09	-13.8202	8.5805	2.7318e+14

(2) The sequence $\{x^k\}$ *is bounded. That is, there is a bounded set* $\Omega \subset \Re^n$ *independent of* k *such that* $x^k \in \Omega$ *for every nonnegative integer* k;

(3) There exist constants $\nu_1 \geq \nu_2 > 0$ *such that* $\nu_2 \|d\|^2 \leq d^\top B_k d \leq \nu_1 \|d\|^2$ *for any* $d \in \Re^n$ *and every* k.

Assumption 3.1 is commonly used in the analysis of SQP methods and IPMs for nonlinear programs although (3) could be restrictive. A variety of line search IPMs currently in use only assume that B_k is positive definite on the null space of the constraints — a much more reasonable assumption. However, Assumption 3.1 (3) can be readily implemented say, by using Powell's dumped Newton update procedure. Hence, it is not hard to do so in practice while in theory, such an assumption can simplify the analysis. Therefore, we still employ (3) in the sequel.

All results in the remainder of this paper are derived under Assumption 3.1. We will not declare it repeatedly. Since the analysis below does not require the so-called regularity (or nondegeneracy) assumptions, it may happen that Algorithm 2.1 generates a sequence converging to a point where the gradients of the active constraints are linearly dependent — a possible result in practice which was assumed not to happen by the usual regularity requirements.

The following result shows that the slack vector sequence is always bounded.

Lemma 3.2 $\{y^k\}$ *is bounded.*

Proof. By (2.7) and (2.8), $\phi(z^{k+1}; \rho_{k+1}) \leq \phi(z^k; \rho_{k+1})$ for every $k \geq 0$. The boundedness of $\{x^k\}$ implies that there exists a constant $\eta_1 > 0$ such that $|f_k| < \eta_1$. Thus,

$$\begin{aligned} \frac{1}{\rho_{k+1}}\phi(z^{k+1}; \rho_{k+1}) - \frac{1}{\rho_k}\phi(z^k; \rho_k) &\leq (\frac{1}{\rho_k} - \frac{1}{\rho_{k+1}})(-f_k + \mu \sum_{i=1}^m \ln y_i^k) \\ &\leq (\frac{1}{\rho_k} - \frac{1}{\rho_{k+1}})(\eta_1 + \mu m \ln \|y^k\|). \end{aligned} \tag{3.1}$$

It follows from (3.1) that

$$\frac{1}{\rho_{k+1}}\phi(z^{k+1}; \rho_{k+1}) \leq \frac{1}{\rho_0}\phi(z^0; \rho_0) + (\frac{1}{\rho_0} - \frac{1}{\rho_{k+1}})(\eta_1 + \mu m \max_{0 \leq j \leq k+1} \ln \|y_j\|). \tag{3.2}$$

On the other hand,

$$\frac{1}{\rho_{k+1}}\phi(z^{k+1};\rho_{k+1}) \geq -\frac{1}{\rho_{k+1}}(\eta_1 + \mu m \max_{0\leq j\leq k+1} \ln\|y_j\|) + \|y^{k+1}\| - \|c^{k+1}\|. \tag{3.3}$$

Thus, by (3.2) and (3.3), there is a constant $\eta_2 > 0$ such that

$$\eta_2 + \frac{\mu m}{\rho_0} \max_{0\leq j\leq k+1} \ln\|y_j\| \geq \|y^{k+1}\|, \;\; \forall k \geq 0. \tag{3.4}$$

If $\{y^k\}$ is unbounded, then there exists a subset $\mathcal{K}$ such that for all $k \in \mathcal{K}$,

$$\eta_2 + \frac{\mu m}{\rho_0} \ln\|y^{k+1}\| \geq \|y^{k+1}\| \tag{3.5}$$

and $\|y^{k+1}\| \to \infty$ as $k \to \infty$. Dividing two sides of the above inequality by $\|y^{k+1}\|$ and taking limit, we have that $0 \geq 1$. This contradiction implies that $\{y^k\}$ is bounded. ■

The next result shows a relation between the slack vector y and the penalty parameter ρ.

Lemma 3.3 *$\{y^k\}$ is componentwise bounded away from zero if and only if $\{\rho_k\}$ is bounded.*

Proof. If $\{\rho_k\}$ is bounded, then there exists a positive integer k_0 and a positive scalar ρ^* such that $\rho_k = \rho^*$ for every $k \geq k_0$. By Algorithm 2.1, for $k \geq k_0$, we have that $\{\phi(z^k;\rho^*)\}$ is a monotonically nonincreasing sequence. Thus,

$$\phi(z^k;\rho^*) \leq \max_{0\leq j\leq k_0} \phi(z_j;\rho_j) \;\forall\; k \geq k_0. \tag{3.6}$$

Since f_k, c^k, and y^k are bounded above, it follows from (2.1) that

$$-\mu\sum_{i=1}^{m} \ln y_i^k \leq -f_k - \rho^*\|c^k + y^k\| + \max_{0\leq j\leq k_0} \phi(z_j;\rho_j) \;\forall\; k \geq k_0. \tag{3.7}$$

Thus, $\{y^k\}$ is componentwise bounded away from zero due to the boundedness of the right-hand-side of (3.7).

Now we suppose that $\{y^k\}$ is componentwise bounded away from zero. Let

$$\tilde{d}_x^k = -A_k(A_k^\top A_k + Y_k^2)^{-1}(c^k + y^k), \tag{3.8}$$

$$\tilde{d}_y^k = -Y_k^2(A_k^\top A_k + Y_k^2)^{-1}(c^k + y^k). \tag{3.9}$$

Then $c^k + y^k + A_k^\top \tilde{d}_x^k + \tilde{d}_y^k = 0$ and there exists a constant $\eta > 0$ such that

$$\|\tilde{d}_x^k\| \leq \eta\|c^k + y^k\|, \;\; \|Y_k^{-1}\tilde{d}_y^k\| \leq \eta\|c^k + y^k\|. \tag{3.10}$$

Since (d_x^k, d_y^k) solves the system (2.3)-(2.5), (d_x^k, d_y^k) must be an optimal solution to the quadratic program

$$\min \; {g^k}^\top d_x - \mu e^\top Y_k^{-1} d_y + \frac{1}{2} d_x^\top B_k d_x + \frac{1}{2} d_y^\top Y_k^{-1}\Lambda_k d_y \tag{3.11}$$

$$\text{s.t.} \;\; A_k^\top d_x + d_y = A_k^\top \tilde{d}_x^k + \tilde{d}_y^k \tag{3.12}$$

(where the optimal multiplier is $\lambda^k + d_\lambda^k$).

Since $(\tilde{d}_x^k, \tilde{d}_y^k)$ is a feasible solution to this quadratic program, we obtain

$$\begin{aligned}
&\pi_k(d^k;\rho_k) + \frac{1}{2}{d_x^k}^\top B_k d_x^k + \frac{1}{2}{d_y^k}^\top Y_k^{-1}\Lambda_k d_y^k \\
&\le \pi_k(\tilde{d}^k;\rho_k) + \frac{1}{2}{\tilde{d}_x^k}{}^\top B_k \tilde{d}_x^k + \frac{1}{2}{\tilde{d}_y^k}{}^\top Y_k^{-1}\Lambda_k \tilde{d}_y^k \\
&\le (\tilde{\eta} - \rho_k)\|c^k + y^k\|,
\end{aligned} \tag{3.13}$$

where the last inequality is due to (3.10) and the boundedness of $\|c^k + y^k\|$, and $\tilde{\eta} > 0$ is a constant independent of k. Thus (2.7) holds whenever $\rho_k \ge \tilde{\eta}$, which implies that ρ_k would not increase after a finite number of iterations; i.e. $\rho_k \not\to \infty$. ∎

Lemma 3.4 *Consider the linear system of equations*

$$H_k u = v^k, \tag{3.14}$$

where

$$H_k = \begin{pmatrix} B_k & 0 & A_k \\ 0 & \Lambda_k & Y_k \\ A_k^\top & I & 0 \end{pmatrix}, v^k = \begin{pmatrix} v_1^k \\ v_2^k \\ v_3^k \end{pmatrix}, \text{ and } u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}.$$

If both $\{\rho_k\}$ and $\{v^k\}$ are bounded, then the solution of equation (3.14) is unique and bounded.

Proof. Since $\{\rho_k\}$ is bounded, by Lemma 3.2 and Lemma 3.3, $\{y^k\}$ is bounded and is componentwise bounded away from zero. Furthermore, by (2.15), $\{\lambda^k\}$ is also bounded and componentwise bounded away from zero.

By substituting the second equation into the rest equations of (3.14), we have that

$$\begin{pmatrix} B_k & A_k \\ A_k^\top & -\Lambda_k^{-1}Y_k \end{pmatrix}\begin{pmatrix} u_1 \\ u_3 \end{pmatrix} = \begin{pmatrix} v_1^k \\ v_3^k - \Lambda_k^{-1}v_2^k \end{pmatrix} \tag{3.15}$$

and $u_2 = \Lambda_k^{-1}v_2^k - \Lambda_k^{-1}Y_k u_3$. Let $\tilde{B}_k = B_k + A_k Y_k^{-1}\Lambda_k A_k^\top$. Then by Assumption 3.1 and (2.15), $\chi_1 \le \theta_{\min}[\tilde{B}_k] \le \theta_{\max}[\tilde{B}_k] \le \chi_2$ for some constants $\chi_2 \ge \chi_1 > 0$ independent of k, where $\theta_{\max}[H]$ and $\theta_{\min}[H]$ denote the biggest and the smallest eigenvalues of matrix H, respectively. This indicates that $\tilde{B}_k$ is invertible for every $k \ge 0$.

Set $M_k = -Y_k^{-1}\Lambda_k + Y_k^{-1}\Lambda_k A_k^\top \tilde{B}_k^{-1} A_k Y_k^{-1}\Lambda_k$. Since

$$\begin{pmatrix} B_k & A_k \\ A_k^\top & -\Lambda_k^{-1}Y_k \end{pmatrix}^{-1} = \begin{pmatrix} \tilde{B}_k^{-1} & \tilde{B}_k^{-1}A_k Y_k^{-1}\Lambda_k \\ \Lambda_k Y_k^{-1} A_k^\top \tilde{B}_k^{-1} & M_k \end{pmatrix}, \tag{3.16}$$

the uniqueness of the solution of (3.14) follows immediately. Moreover, since $\{v^k\}$, $\|\tilde{B}_k^{-1}\|$, $\|\tilde{B}_k^{-1}A_k Y_k^{-1}\Lambda_k\|$, and $\|M_k\|$ are all bounded, the boundedness of the solution follows. ∎

Corollary 3.5 *The boundedness of* $\{\rho_k\}$ *implies that both* $\{(d_x^k, d_y^k)\}$ *and* $\{d_\lambda^k\}$ *are bounded.*

Proof. Set $v_1^k = -(g^k + A_k\lambda^k)$, $v_2^k = -(Y_k\Lambda_k e - \mu e)$, and $v_3^k = -(c^k + y^k)$. Then by Lemma 3.2 and Lemma 3.3 $\{v^k\}$ is bounded if $\{\rho_k\}$ is bounded. Thus the result follows by Lemma 3.4. ■

The next two lemmas show that stepsizes α_k for the primal iterates are bounded away from zero if $\{\rho_k\}$ is bounded.

Lemma 3.6 *If* $\{\rho_k\}$ *is bounded, then there exists a constant* $\tau \in (0, 1]$ *such that* $y^k + \alpha d_y^k \geq \xi y^k$ *and*

$$\phi(z^k + \alpha d^k; \rho_{k+1}) - \phi(z^k; \rho_{k+1}) \leq \sigma_0 \alpha \pi_k(d^k; \rho_{k+1}) \tag{3.17}$$

for any $\alpha \in [0, \tau]$, *where* $\sigma_0 < \frac{1}{2}$ *is a positive constant.*

Proof. Since $\{\rho_k\}$ is bounded, by Lemma 3.3 and Lemma 3.4, $\{y^k\}$ is bounded away from zero and $\{d_y^k\}$ is bounded. Suppose that $y^k \geq \zeta_1 e$ and $\|d_y^k\| \leq \zeta_2$ for all k (where $\zeta_1 > 0$ and $\zeta_2 > 0$ are constants). Let $\hat{\tau} = \min\{1, (1-\xi)\zeta_1/\zeta_2\}$. Then $y^k + \alpha d_y^k \geq \xi y^k$ for any $\alpha \in [0, \hat{\tau}]$.

Suppose that $\rho_k = \rho^*$ for every $k \geq k_0$. For $\alpha \in [0, \hat{\tau}]$, since $y^k + \alpha d_y^k \geq \xi y^k$,

$$\begin{aligned}
&-\sum_{i=1}^m \ln[y_i^k + \alpha d_{yi}^k] + \sum_{i=1}^m \ln y_i^k + \alpha e^\top {Y_k}^{-1} d_y^k \\
&= e^\top \int_0^\alpha [Y_k^{-1} - (Y_k + tD_y^k)^{-1}] d_y^k dt \quad (\text{where } D_y^k = \operatorname{diag}(d_y^k)) \\
&= e^\top \int_0^\alpha Y_k^{-1}(Y_k + tD_y^k)^{-1}(tD_y^k) d_y^k dt \leq \frac{1}{2\xi}\alpha^2 \|{Y_k}^{-1} d_y^k\|^2.
\end{aligned} \tag{3.18}$$

Since $f(x)$ and $c(x)$ are twice continuously differentiable,

$$\begin{aligned}
&f(x^k + \alpha d_x^k) - f(x^k) - \alpha g(x^k)^\top d_x^k \\
&= \int_0^\alpha [g(x^k + td_x^k) - g(x^k)]^\top d_x^k dt \leq \frac{1}{2}\alpha^2 \eta_1 \|d_x^k\|^2,
\end{aligned} \tag{3.19}$$

and

$$\begin{aligned}
&\|c(x^k + \alpha d_x^k) + y^k + \alpha d_y^k\| - \|c(x^k) + y^k + \alpha A(x^k)^\top d_x^k + \alpha d_y^k\| \\
&\leq \|c(x^k + \alpha d_x^k) - c(x^k) - \alpha A(x^k)^\top d_x^k\| \\
&= \| \int_0^\alpha [A(x^k + td_x^k) - A(x^k)]^\top d_x^k dt\| \leq \frac{1}{2}\alpha^2 \eta_2 \|d_x^k\|^2,
\end{aligned} \tag{3.20}$$

where η_1 and η_2 are the first-order Lipschitz constants of f and c, respectively. Let $\eta_3 = \max\{\frac{\mu}{2\xi}, \frac{1}{2}(\eta_1 + \rho^*\eta_2)\}$. Then by (3.18), (3.19) and (3.20),

$$\phi(z^k + \alpha d^k; \rho^*) - \phi(z^k; \rho^*) - \pi_k(\alpha d^k; \rho^*) \leq \alpha^2 \eta_3 (\|d_x^k\|^2 + \|Y_k^{-1} d_y^k\|^2). \tag{3.21}$$

Since $\pi_k(\alpha d^k;\rho^*)$ is convex on $\alpha \in [0,1]$, $\pi_k(\alpha d^k;\rho^*) - \alpha\pi_k(d^k;\rho^*) \le 0$. Thus, by Assumption 3.1 and (2.15),

$$\begin{aligned}&\pi_k(\alpha d^k;\rho^*) - \alpha\sigma_0\pi_k(d^k;\rho^*) \le \alpha(1-\sigma_0)\pi_k(d^k;\rho^*) \qquad (3.22)\\ &\le -\frac{1}{2}\alpha\theta(1-\sigma_0)(\|d_x^k\|^2 + \|Y_k^{-1}d_y^k\|^2), \quad \text{(by (2.7))}\end{aligned}$$

where $\theta = \min\{\nu_2, \beta_1\mu\}$. Let $\tau_0 = \min\{\hat{\tau}, (1-\sigma_0)\theta/(2\eta_3)\}$. Then for $0 \le \alpha \le \tau_0$, it follows from (3.21) and (3.22) that

$$\phi(z^k + \alpha d^k;\rho^*) - \phi(z^k;\rho^*) \le \alpha\sigma_0\pi_k(d^k;\rho^*) \tag{3.23}$$

for $k \ge k_0$. By taking $\tau = \min\{\tau_0, \alpha_0, \ldots, \alpha_{k_0-1}\}$, we have the desired result. ∎

Lemma 3.7 *Suppose that $\{\rho_k\}$ is bounded. Then there exists a constant $\tilde{\tau} \in (0,1]$ such that for any $\alpha \in [0,\tilde{\tau}]$ and for every $k \ge 0$,*

$$\min\{\beta_1\mu e, Y_k\Lambda_k e\} \le (Y_k + \alpha D_y^k)(\lambda^k + \alpha d_\lambda^k) \le \max\{\beta_2\mu e, Y_k\Lambda_k e\}, \tag{3.24}$$

where $0 < \beta_1 < 1 < \beta_2$.

Proof. By (2.4),

$$\begin{aligned}&(Y_k + \alpha D_y^k)(\lambda^k + \alpha d_\lambda^k) = \alpha\mu e + (1-\alpha)Y_k\Lambda_k e + \alpha^2 D_y^k d_\lambda^k\\ &= \alpha\beta_1\mu e + (1-\alpha)Y_k\Lambda_k e + \alpha((1-\beta_1)\mu e + \alpha D_y^k d_\lambda^k) \qquad (3.25)\\ &= \alpha\beta_2\mu e + (1-\alpha)Y_k\Lambda_k e + \alpha(-(\beta_2-1)\mu e + \alpha D_y^k d_\lambda^k). \qquad (3.26)\end{aligned}$$

We now consider (3.24) componentwise. Let $\mathcal{I}_0 = \{i : (d_y^k)_i(d_\lambda^k)_i = 0, i = 1,\ldots,m\}$, $\mathcal{I}_1 = \{i : (d_y^k)_i(d_\lambda^k)_i < 0, i = 1,\ldots,m\}$, and $\mathcal{I}_2 = (\{1,\ldots,m\}\backslash\mathcal{I}_0)\backslash\mathcal{I}_1$. For $i \in \mathcal{I}_0$, (3.24) holds trivially for all $\alpha \in [0,1]$. For $i \in \mathcal{I}_1$, by (3.25), if $\alpha \le \min(1, \min_i\{-(1-\beta_1)/((d_y^k)_i(d_\lambda^k)_i) : i \in \mathcal{I}_1\})$, then (3.24) holds. For $i \in \mathcal{I}_2$, by (3.26), if $\alpha \le \min(1, \min_i\{(\beta_2-1)\mu/((d_y^k)_i(d_\lambda^k)_i) : i \in \mathcal{I}_2\})$, then (3.24) holds.

Since $\{\rho_k\}$ is bounded, by Lemma 3.4 there exists a constant $\zeta > 0$ such that $\|D_y^k d_\lambda^k\| \le \zeta$. By letting $\tilde{\tau} = \min\{1, (1-\beta_1)\mu/\zeta, (\beta_2-1)\mu/\zeta\}$, we have the desired result. ∎

We are now ready to present the global convergence result of Algorithm 2.1.

Theorem 3.8 *Suppose that $\{z^k\}$ is a sequence generated by Algorithm 2.1. Then one of the following claims must be true.*

(1) The penalty parameter sequence $\{\rho_k\}$ is bounded, $\{y^k\}$ is bounded above and is componentwise bounded away from zero, and all limit points of $\{z^k\}$ are KKT points of the problem (1.3)-(1.4).

(2) The penalty parameter sequence $\{\rho_k\}$ is unbounded, $\{y^k\}$ is bounded above and is not componentwise bounded away from zero, and there exists one limit point of $\{z^k\}$, say z^*, which is either an infeasible point of problem (1.3)-(1.4), or a feasible point at which A_i^*, $i \in \mathcal{I}$ are linearly dependent, where $A_i^* = \nabla c_i(x^*)$ and $\mathcal{I} = \{i : c_i(x^*) = 0\}$.

Proof. If $\{y^k\}$ is componentwise bounded away from zero, then by Lemma 3.3 $\{\rho_k\}$ is bounded. By Lemma 3.6 and Lemma 3.7, there exists a constant $\hat{\alpha} \in (0, 1]$ such that $\alpha_k \geq \hat{\alpha}$ for every k. On the other hand, since $\{y^k\}$ is bounded above and is bounded away from zero and $\{\rho_k\}$ is bounded, by Assumption 3.1, $\{\phi(z^k; \rho_{k+1})\}$ is bounded. Without loss of generality, assume $\rho_k = \rho_0$ for every k. Then by (2.8), $\{\phi(z^k; \rho_0)\}$ is a monotonically nonincreasing sequence, thus it has the limit as $k \to \infty$. Again by (2.8), $\pi_k(d^k; \rho_0) \to 0$ as $k \to \infty$. Thus, by (2.7) and Assumption 3.1, we have that $(d_x^k, d_y^k) \to 0$ as $k \to \infty$. By (2.3)-(2.5) we have (1).

If $\{y^k\}$ is not componentwise bounded away from zero, then by Lemma 3.3 $\{\rho_k\}$ is unbounded. By Lemma 3.2 and Assumption 3.1, $\{z^k\}$ is bounded. Thus, we have at least one limit point. If a limit point is infeasible to the system (1.3)-(1.4), then we complete the proof. Otherwise, all limit points are feasible. We prove that there is a limit point at which $\{A_i^* : i \in \mathcal{I}\}$ are linearly dependent.

Assuming the contrary, we first show that there exists $\epsilon > 0$ such that $\theta_{\min}\left[A_k^\top A_k + Y_k^2\right] \geq \epsilon$ for sufficiently large k. Let $z^* = (x^*, y^*)$ be an arbitrary limit point of $\{z^k\}$. Then z^* is feasible and $\mathcal{I} = \{i : c_i(x^*) = 0\} = \{i : y_i^* = 0\}$. Since $\{A_i^* : i \in \mathcal{I}\}$ are linearly independent, there exists $\hat{\epsilon}_{\mathcal{I}} > 0$ such that $\theta_{\min}\left[(A^*)_{\mathcal{I}}^\top (A^*)_{\mathcal{I}}\right] \geq 2\hat{\epsilon}_{\mathcal{I}}$, where $(A^*)_{\mathcal{I}}$ consists of A_i^* indexed by $\mathcal{I}$. It then follows that for all sufficiently large $k \in \mathcal{K}$, $\theta_{\min}\left[(A_k)_{\mathcal{I}}^\top (A_k)_{\mathcal{I}}\right] \geq \hat{\epsilon}_{\mathcal{I}}$, where $\mathcal{K}$ is the index subset such that $\{z^k\}_{k \in \mathcal{K}} \to z^*$. Obviously, there exists $\bar{\epsilon}_{\mathcal{I}}$ such that $\theta_{\min}\left[(Y_k)_{\bar{\mathcal{I}}}^\top (Y_k)_{\bar{\mathcal{I}}}\right] \geq \bar{\epsilon}_{\mathcal{I}} > 0$ for sufficiently large $k \in \mathcal{K}$. Then $\theta_{\min}\left[A_k^\top A_k + Y_k^2\right] \geq \min\{\bar{\epsilon}_{\mathcal{I}}, \hat{\epsilon}_{\mathcal{I}}\}$. Now we can conclude that there exists $\epsilon > 0$ such that $\theta_{\min}\left[A_k^\top A_k + Y_k^2\right] \geq \epsilon$ for all sufficiently large k because the number of different $\mathcal{I}$s is finite and any index k belongs to a certain $\mathcal{K}$ associated with a certain $\mathcal{I}$.

Now we derive a contradiction. Due to the uniform positive lower bound of $\theta_{\min}\left[A_k^\top A_k + Y_k^2\right]$, consider $(\tilde{d}_x^k, \tilde{d}_y^k)$ defined by (3.8)-(3.9), the relationship (3.10) holds for some constant $\eta > 0$. Hence, by (3.13), $\{\rho_k\}$ must be bounded, which is a contradiction and (2) is therefore proved. ∎

A feasible point with linearly dependent gradients of the active constraints is a so-called Fritz-John point. A implication of Theorem 3.8 is that if $\{z^k\}$ converges, then it converges to either a KKT point, or a Fritz-John point, or an infeasible point.

4. A refined version of Algorithm 2.1 with improved convergence properties

Continuing our discussion of last section, we will show in this section that with some refinements of Algorithm 2.1 one of the infeasible limit points (if exists at all) also has some stationary property, namely it is a stationary point of minimizing the function $\|c(x)_+\|^2$, which is a measure of the infeasibility of x.

For brevity, we do not list all steps of the new algorithm, the identical steps with Algorithm 2.1 are omitted, only the modified steps are given.

Algorithm 4.1 *(A refined version of Algorithm 2.1)*

Step $2'$ *Calculate* $(d_x^k, d_y^k, d_\lambda^k)$ *by the modified primal-dual system of equations:*

$$\begin{aligned} B_k d_x \qquad\qquad & + A_k d_\lambda = -(g^k + A_k \lambda^k), && (4.1)\\ \Lambda_k d_y & + Y_k d_\lambda = -(Y_k \Lambda_k e - \mu e), && (4.2)\\ A_k^\top d_x + \quad d_y \qquad & = A_k^\top \hat{d}_x^k + \hat{d}_y^k, && (4.3) \end{aligned}$$

where $(\hat{d}_x^k, \hat{d}_y^k) \in \Re^{n+m}$ *is derived from some given problem.*

Step $5'$ *Increase* l *until there exists a scalar* $\gamma \in [0,1]$ *satisfying*

$$\min\{\beta_1 \mu e, Y_k \Lambda_k e\} \le (Y_k + \delta^l \hat{\alpha}_k D_y^k)(\Lambda_k + \gamma D_\lambda^k) e \le \max\{\beta_2 \mu e, Y_k \Lambda_k e\},$$

where $D_y^k = \operatorname{diag}(d_y^k)$, $D_\lambda^k = \operatorname{diag}(d_\lambda^k)$. *If* $-(\Lambda_k + \gamma D_\lambda^k) c(x^k + \delta^l \hat{\alpha}_k d_x^k) \le \beta_2 \mu e$ *for some* $\gamma \in [0,1]$, *then select* γ_k *to be the biggest* $\gamma \in [0,1]$ *satisfying*

$$\begin{aligned} \min\{\beta_1 \mu e, Y_k \Lambda_k e\} & \le (\Lambda_k + \gamma_k D_\lambda^k) \max\{y^k + \delta^l \hat{\alpha}_k d_y^k, -c(x^k + \delta^l \hat{\alpha}_k d_x^k)\} \\ & \le \max\{\beta_2 \mu e, Y_k \Lambda_k e\}, \qquad (4.4) \end{aligned}$$

and set $\alpha_k = \delta^l \hat{\alpha}_k$; *otherwise, increase* l *until (4.4) holds for some* $\gamma_k \in [0,1]$. *The new primal iterate is*

$$x^{k+1} := x^k + \alpha_k d_x^k, \;\; y^{k+1} := \max\{y^k + \alpha_k d_y^k, \; -c^{k+1}\}, \qquad (4.5)$$

and the new dual iterate is $\lambda^{k+1} := \lambda^k + \gamma_k d_\lambda^k$.

It is noted that Step $2'$ of Algorithm 4.1 is different from Step 2 of Algorithm 2.1 in the right-hand-side of equation (4.3). In the following we will show that under suitable conditions the stepsizes for the primal iterates are bounded away from zero. Then, by (4.4) and (4.5), there exist constants $\bar{\beta}_1 > 0$ and $\bar{\beta}_2 > 0$ such that

$$\bar{\beta}_1 \mu e \le Y_{k+1} \Lambda_{k+1} e \le \bar{\beta}_2 \mu e. \qquad (4.6)$$

Lemma 4.2 *The following results hold for Algorithm 4.1:*

(i) $\{y^k\}$ *is bounded above. If* $\{\rho_k\}$ *is bounded, then* $\{y^k\}$ *is componentwise bounded away from zero;*

(ii) $\{\lambda^k\}$ *is componentwise bounded away from zero. If* $\{\rho_k\}$ *is bounded, then* $\{\lambda^k\}$ *is bounded above;*

(iii) if $\{\rho_k\}$ *and* $\{(\hat{d}_x^k, \hat{d}_y^k)\}$ *are bounded, then* $\{(d_x^k, d_y^k, d_\lambda^k)\}$ *is bounded.*

Proof. Result (i) is true due to Lemma 3.2 and the first half of Lemma 3.3 since the related proofs do not depend on the modified steps.

Result (ii) follows from (i) and (4.6) directly.

Since Lemma 3.4 is still valid for Algorithm 4.1, Result (iii) follows from Result (i) and Lemma 3.4. ∎

Remark. Lemma 4.2 (iii) implies that Lemma 3.6 holds if $\{\rho_k\}$ and $\{(\hat{d}_x^k, \hat{d}_y^k)\}$ are bounded.

To simplify the statements, we define two sets: $\mathcal{S}_1^k = \{d \in \Re^n : d = P_k(c^k + y^k)\}$ and $\mathcal{S}_2^k = \{d \in \Re^m : d = Q_k(c^k + y^k)\}$, where P_k and Q_k are certain bounded matrices under Assumption 3.1 and the condition that y^k is bounded above and is componentwise bounded away from zero. For example, $-A_k(c^k + y^k) \in \mathcal{S}_1^k$ and $-Y_k^2(c^k + y^k) \in \mathcal{S}_2^k$. Some other examples include the weighted Newton step and the weighted Cauchy step defined in [10].

The next lemma says that (4.4) is well defined if $\{\rho_k\}$ is bounded.

Lemma 4.3 *If $\{\rho_k\}$ is bounded, $(\hat{d}_x^k, \hat{d}_y^k) \in \mathcal{S}_1^k \times \mathcal{S}_2^k$, then there exists a constant $\tilde{\alpha} \in (0, 1]$ such that for any $\alpha \in [0, \tilde{\alpha}]$ and for all k*

$$\min\{\beta_1 \mu e, Y_k \Lambda_k e\} \le (\Lambda_k + \alpha D_\lambda^k) \max\{y^k + \alpha d_y^k, -c(x^k + \alpha d_x^k)\} \le \max\{\beta_2 \mu e, Y_k \Lambda_k e\}. \tag{4.7}$$

Proof. If $\max\{y^k + \alpha d_y^k, -c(x^k + \alpha d_x^k)\} = y^k + \alpha d_y^k$ for all k, then the result can be shown in the same way as Lemma 3.7.

Otherwise, for some i, there exists a subset $\mathcal{K}$ such that for any given $\epsilon > 0$, we have $y_i^k < -c_i^k + \epsilon$ for $k \in \mathcal{K}$. By (4.5), $y_i^k + c_i^k \ge 0$. Since $\{\rho_k\}$ is bounded, by Lemma 4.2, $\{y^k\}$ is componentwise bounded away from zero. Thus, by the fact that $(\hat{d}_x^k, \hat{d}_y^k) \in \mathcal{S}_1^k \times \mathcal{S}_2^k$, there exists $\epsilon' > 0$ (which is dependent on ϵ) such that $(A_k)_i^\top d_x^k + (d_y^k)_i = (A_k)_i^\top \hat{d}_x^k + (\hat{d}_y^k)_i \ge -\epsilon'$ $\forall$ $k \in \mathcal{K}$. It follows from Assumption 3.1 and Lemma 4.2 that

$$\begin{aligned} -[\lambda_i^k + \alpha (d_\lambda^k)_i] c_i(x^k + \alpha d_x^k) &= -[\lambda_i^k + \alpha (d_\lambda^k)_i](c_i^k + \alpha (A_k)_i^\top d_x^k + O(\alpha^2)) \\ &= -c_i^k \lambda_i^k - \alpha [\lambda_i^k (A_k)_i^\top d_x^k + c_i^k (d_\lambda^k)_i] + O(\alpha^2). \end{aligned} \tag{4.8}$$

Since $-\lambda_i^k (A_k)_i^\top d_x^k \le \lambda_i^k \epsilon' + \lambda_i^k (d_y^k)_i$ and $-c_i^k (d_\lambda^k)_i \le y_i^k (d_\lambda^k)_i + \max\{0, -\epsilon (d_\lambda^k)_i\}$ for $k \in \mathcal{K}$, by (4.2), there is a constant $\epsilon'' > 0$ such that for all $k \in \mathcal{K}$

$$\begin{aligned} -[\lambda_i^k + \alpha (d_\lambda^k)_i] c_i(x^k + \alpha d_x^k) &\le y_i^k \lambda_i^k + \alpha [\lambda_i^k (d_y^k)_i + y_i^k (d_\lambda^k)_i + \epsilon'' + O(\alpha)] \\ &\le y_i^k \lambda_i^k + \alpha (\mu - y_i^k \lambda_i^k + \epsilon'' + O(\alpha)) \quad \text{(by (4.2))} \\ &\le \max\{\beta_2 \mu, y_i^k \lambda_i^k\} + \alpha [-(\beta_2 - 1)\mu + \epsilon'' + O(\alpha)]. \end{aligned} \tag{4.9}$$

The remaining proof is similar to the proof of Lemma 3.7. ■

Let

$$\psi_k(d; \rho_k) = \pi_k(d; \rho_k) + \frac{1}{2} d_x^\top B_k d_x + \frac{1}{2} d_y^\top Y_k^{-1} \Lambda_k d_y. \tag{4.10}$$

We require that $(\hat{d}_x^k, \hat{d}_y^k)$ satisfies the following conditions.

(1) $(\hat{d}_x^k, \hat{d}_y^k) \in \mathcal{S}_1^k \times \mathcal{S}_2^k$;

(2) If for any constant $\eta > 0$ there exists a constant $\hat{\eta} > 0$ and an infinite index set $\mathcal{K}$ such that for every $k \in \mathcal{K}$

$$\min_{\|(d_x, Y_k^{-1} d_y)\| \le \eta, (d_x, d_y) \in \mathcal{S}_1^k \times \mathcal{S}_2^k} \|c^k + y^k + A_k^\top d_x + d_y\| - \|c^k + y^k\| \le -\hat{\eta}, \tag{4.11}$$

then there is a $\hat{\rho} > 0$ such that for every $k \in \mathcal{K}$

$$\psi_k(\hat{d}^k; \hat{\rho}) \le 0; \tag{4.12}$$

(3) If for any constant $\eta > 0$ there exists an infinite index set $\mathcal{K}$ such that for every $k \in \mathcal{K}$

$$\theta_{\min}([A_k^\top \;\; Y_k]) \ge \eta, \tag{4.13}$$

then there exist constants $\bar{\eta} > 0$ and $1 > \tilde{\eta} > 0$ such that for every $k \in \mathcal{K}$,

$$\begin{cases} \|\hat{d}_x^k\| \le \bar{\eta}\|c^k + y^k\|, \\ \|Y_k^{-1}\hat{d}_y^k\| \le \bar{\eta}\|c^k + y^k\|, \\ \|c^k + y^k + A_k^\top \hat{d}_x^k + \hat{d}_y^k\| - \|c^k + y^k\| \le -\tilde{\eta}\|c^k + y^k\|. \end{cases} \tag{4.14}$$

Theorem 4.4 *Conditions (1)-(3) guarantee that if* $\{\rho_k\}$ *is bounded, then* $\lim_{k\to\infty}\|c^k + y^k\| = 0$ *and any limit point of* $\{z^k\}$ *is a KKT point of (1.3)-(1.4).*

Proof. By Lemma 4.2, $\{y^k\}$ is componentwise bounded away from zero. Thus, (4.13) holds for some constant $\eta > 0$ since $[A_k^\top \;\; Y_k]$ is of full raw rank, which by Condition (3) implies (4.14) for every k and some constants $\bar{\eta}$ and $\tilde{\eta}$.

Since $\{y^k\}$ and $\{(\hat{d}_x^k, \hat{d}_y^k)\}$ are bounded, it follows from Lemma 4.2 (3) that $\{(d_x^k, d_y^k)\}$ is bounded. By (4.5), $y^{k+1} \ge y^k + \alpha_k d_y^k$ and $\|c^{k+1} + y^{k+1}\| \le \|c^{k+1} + y^k + \alpha_k d_y^k\|$. Thus,

$$\phi(z^{k+1}; \rho_{k+1}) \le \phi(z^k + \alpha_k d^k; \rho_{k+1}). \tag{4.15}$$

Then by Lemma 3.6 and Lemma 4.3, there is a constant $\hat{\alpha} \in (0, 1]$ such that $\alpha_k \ge \hat{\alpha}$ for every k. The fact that $\{\rho_k\}$ is bounded implies that ρ_k is a constant for sufficiently large k, thus $\{\phi(z^k; \rho_k)\}$ is monotonically nonincreasing. Together with its lower boundedness, the sequence $\{\phi(z^k; \rho_k)\}$ has a limit. Hence, $\pi_k(d^k; \rho_k) \to 0$, which by (2.7), Assumption 3.1 and (4.6) implies that $(d_x^k, d_y^k) \to 0$ as $k \to \infty$. Then by (4.3) $A_k^\top \hat{d}_x^k + \hat{d}_y^k \to 0$ as $k \to \infty$. As a result of (4.14), we have that $\lim_{k\to\infty}\|c^k + y^k\| = 0$. The KKT property of the limit points follows from (4.1) and (4.2). ∎

Lemma 4.5 *Let* $\{z^k\}_{k\in\mathcal{K}} \to z^*$ *with* $\|c^* + y^*\| \ne 0$*. If for some constant* $\eta > 0$

$$\lim_{k\to\infty, k\in\mathcal{K}} \left[\min_{\|(d_x, Y_k^{-1}d_y)\| \le \eta, (d_x, d_y) \in \mathcal{S}_1^k \times \mathcal{S}_2^k} \|c^k + y^k + A_k^\top d_x + d_y\| - \|c^k + y^k\| \right] = 0, \tag{4.16}$$

then

$$\begin{pmatrix} A^* \\ Y^* \end{pmatrix} (c^* + y^*) = 0. \tag{4.17}$$

Moreover, if $\|c^* + y^*\| \ne 0$*, then* $A^* c_+^* = 0$*, where* $c^* = c(x^*)$*,* $A^* = A(x^*)$ *and* $c_+^* = \max\{c(x^*), 0\}$.

Proof. If (4.17) does not hold, then for sufficiently large $k \in \mathcal{K}$, there exists a constant $\xi > 0$ such that

$$\|\hat{g}^k\| := \| \begin{pmatrix} A_k \\ Y_k \end{pmatrix} (c^k + y^k)\| \ge \xi. \tag{4.18}$$

Let $\begin{pmatrix} \hat{d}_x^k \\ \hat{d}_y^k \end{pmatrix} = - \begin{pmatrix} A_k \\ Y_k^2 \end{pmatrix} (c^k + y^k)$ and $\hat{\alpha}_k$ be the optimal solution to the problem

$$\min_{\alpha\in(0,1]} \ \|c^k + y^k + \alpha A_k^\top \hat{d}_x^k + \alpha \hat{d}_y^k\| \tag{4.19}$$

$$\text{s.t. } \|(\alpha \hat{d}_x^k, \alpha Y_k^{-1} \hat{d}_y^k)\| \le \eta. \tag{4.20}$$

Then $\hat{\alpha}_k \le \min\{1, \eta/\|\hat{g}^k\|\}$. Since $\alpha_k = \|\hat{g}^k\|^2/\|(A_k^\top A_k + Y_k^2)(c^k + y^k)\|^2$ minimizes the objective function (4.19) over $\Re$, if $\alpha_k \le \min\{1, \eta/\|\hat{g}^k\|\}$, then $\hat{\alpha}_k = \alpha_k$, and

$$\|c^k + y^k\|^2 - \|c^k + y^k + A_k^\top(\hat{\alpha}_k \hat{d}_x^k) + \hat{\alpha}_k \hat{d}_y^k\|^2 = \|\hat{g}^k\|^4/\|(A_k^\top A_k + Y_k^2)(c^k + y^k)\|^2; \tag{4.21}$$

otherwise, $\alpha_k > \min\{1, \eta/\|\hat{g}^k\|\}$, and $\hat{\alpha}_k = \min\{1, \eta/\|\hat{g}^k\|\}$, then

$$\|c^k + y^k\|^2 - \|c^k + y^k + A_k^\top(\hat{\alpha}_k \hat{d}_x^k) + \hat{\alpha}_k \hat{d}_y^k\|^2 \ge \min\{\|\hat{g}^k\|^2, \ \eta\|\hat{g}^k\|\}. \tag{4.22}$$

Thus,

$$\begin{aligned} &\|c^k + y^k + A_k^\top(\hat{\alpha}_k \hat{d}_x^k) + \hat{\alpha}_k \hat{d}_y^k\| - \|c^k + y^k\| \\ &\le -\frac{1}{2\|c^k + y^k\|} \min\left\{ \frac{\|\hat{g}^k\|^4}{\|(A_k^\top A_k + Y_k^2)(c^k + y^k)\|^2}, \ \|\hat{g}^k\|^2, \ \eta\|\hat{g}^k\| \right\}. \end{aligned} \tag{4.23}$$

Since $\min_{\|(d_x, Y_k^{-1} d_y)\| \le \eta, (d_x, d_y) \in \mathcal{S}_1^k \times \mathcal{S}_2^k} \|c^k + y^k + A_k^\top d_x + d_y\| \le \|c^k + y^k + A_k^\top(\hat{\alpha}_k \hat{d}_x^k) + \hat{\alpha}_k \hat{d}_y^k\|$, it follows from (4.16) that $\|\hat{g}^k\| \to 0$, which contradicts (4.18). Thus (4.17) is valid. If $y_i^* \ne 0$, then $c_i^* + y_i^* = 0$, which in turn implies $c_i^* < 0$ and $(c_+^*)_i = 0$. Since $c^* + y^* \ge 0$ by (4.5), $y_i^* = 0$ implies $c_i^* \ge 0$. Thus, it follows from (4.17) that $A^* c_+^* = 0$. $\blacksquare$

The following result covers the case that $\rho_k \to \infty$.

Theorem 4.6 *Under Conditions (1)-(3), if $\rho_k \to \infty$, then there exists a limit point of $\{z^k\}$, say z^*, such that either $\|c^* + y^*\| = 0$ and A_i^*, $i \in \mathcal{I}$ are linearly dependent, or $\|c^* + y^*\| \ne 0$ and $A^* c_+^* = 0$, where $\mathcal{I} = \{i : c_i(x^*) = 0\}$, $A^* = A(x^*)$ and $c_+^* = \max\{c(x^*), 0\}$.*

Proof. Suppose that the result is not true. For any convergent subsequence $\{z^k : k \in \mathcal{K}\}$, let z^* be its limit point. Then if $\|c^* + y^*\| = 0$, we have that A_i^*, $i \in \mathcal{I}$ are linearly independent, in which case, there exists a sufficiently large $k_0 \in \mathcal{K}$ such that for every $k \ge k_0$ and $k \in \mathcal{K}$, $(A_k)_i, i \in \mathcal{I}$ are linearly independent. By Condition (3), (4.14) holds for every $k \ge k_0$ and $k \in \mathcal{K}$. Hence,

$$\psi_k(d^k; \rho_k) \le \psi_k(\hat{d}^k; \rho_k) \le (\eta_0 - \rho_k \tilde{\eta})\|c^k + y^k\| \tag{4.24}$$

for every $k \ge k_0$ and $k \in \mathcal{K}$ since $\|c^k + y^k\|$ is bounded above, where η_0 is a positive constant.

On the other hand, if $\|c^* + y^*\| \neq 0$, then $A^* c_+^* \neq 0$ which, by Lemma 4.5, results in that (4.11) holds for every sufficiently large $k \in \mathcal{K}$, thus (4.12) holds.

In both cases, it follows from (4.24) and (4.12) that there exists a $\tilde{\rho} > 0$ such that (2.7) holds for $\rho_k \geq \tilde{\rho}$ which contradicts the fact that $\rho_k \to \infty$. The result follows from the contradiction. ■

Back to the original problem (1.1)-(1.2), since $2A^* c_+^*$ is the gradient of the "infeasibility function" $\|c_+(x)\|^2$ at x^*, Theorem 4.6 says that at least one limit point of $\{x^k\}$ is either a Fritz-John point or an infeasible stationary point of the original problem. In the case where $\rho \not\to \infty$, Theorem 4.4 says that any limit point (x^*, y^*, λ^*) of the generated sequence will satisfy the following KKT conditions of (1.3)-(1.4):

$$\nabla f(x^*) + \nabla c(x^*)\lambda^* = 0 \tag{4.25}$$

$$c(x^*) + y^* = 0, y^* \geq 0 \tag{4.26}$$

$$Y^* \lambda^* = \mu e. \tag{4.27}$$

In addition, obviously we have $y^k > 0, \ \forall \ k$, which together with (2.15) (or (4.6)) implies $\lambda^k > 0, \ \forall \ k$. Now the above KKT conditions together with $\lambda^* \geq 0$ imply

$$\nabla f(x^*) + \nabla c(x^*)\lambda^* = 0 \tag{4.28}$$

$$c(x^*) \leq 0, \lambda^* \geq 0 \tag{4.29}$$

$$c(x^*)^\top \lambda^* = -m\mu. \tag{4.30}$$

That is, the point (x^*, λ^*) is an approximate KKT pair of the original problem with duality gap no more than $m\mu$.

5. Validity of the conditions in two different types of interior point methods

In this section, we prove that both the Byrd-Omojokun approach (see [1, 2]) and the quadratic programming approach (see [10]) satisfy Conditions (1)-(3).

Theorem 5.1 *(I) (The approach used in Byrd et al. [1, 2]) If* $(\hat{d}_x^k, \hat{d}_y^k)$ *is computed through the following problem*

$$\min \ \|c^k + y^k + A_k^\top d_x + d_y\| \tag{5.1}$$

$$s.t. \ \|(d_x, Y_k^{-1} d_y)\| \leq \eta, \tag{5.2}$$

$$(d_x, d_y) \in \mathcal{S}_1^k \times \mathcal{S}_2^k, \tag{5.3}$$

where $\eta > 0$ *is a constant, then Conditions (1)-(3) hold.*

(II) (The approach used in [10]) If $(\hat{d}_x^k, \hat{d}_y^k)$ *is generated by the quadratic program*

$$\min_{(d_x, d_y) \in \mathcal{S}_1^k \times \mathcal{S}_2^k} \frac{1}{2} d_x^\top B_k d_x + \frac{1}{2} d_y^\top Y_k^{-1} \Lambda_k d_y + \rho_k \|c^k + y^k + A_k^\top d_x + d_y\|, \tag{5.4}$$

then

(i) Conditions (1)-(2) hold;

(ii) Condition (3) holds for all k with $\rho_k \geq \hat{\rho}$, where $\hat{\rho} > 0$ is a constant.

Before proving the theorem, we need the following lemma.

Lemma 5.2 *Let $(\hat{d}_x^k, \hat{d}_y^k)$ be the solution of (5.4). Then there exists a positive constant ζ independent of k such that*

$$\|\hat{d}_x^k\|/\sqrt{\rho_k} \leq \zeta, \ \ \|Y_k^{-1}\hat{d}_y^k\|/\sqrt{\rho_k} \leq \zeta. \tag{5.5}$$

Proof. If $(\hat{d}_x^k, \hat{d}_y^k) \in \mathcal{S}_1^k \times \mathcal{S}_2^k$, then $(\bar{d}_x^k, \bar{d}_y^k) = (\hat{d}_x^k/\sqrt{\rho_k}, Y_k^{-1}\hat{d}_y^k/\sqrt{\rho_k}) \in \mathcal{S}_1^k \times \mathcal{S}_2^k$. Let $(\bar{d}_x, \bar{d}_y) = (d_x/\sqrt{\rho_k}, Y_k^{-1}d_y/\sqrt{\rho_k})$. Then problem (5.4) is equivalent to the following problem:

$$\min_{(\bar{d}_x,\bar{d}_y)\in\mathcal{S}_1^k\times\mathcal{S}_2^k} \quad \bar{\varphi}_k(\bar{d}) := \frac{1}{2}\bar{d}_x^\top B_k \bar{d}_x + \frac{1}{2}\bar{d}_y^\top Y_k\Lambda_k \bar{d}_y + \|c^k + y^k + \sqrt{\rho_k}A_k^\top \bar{d}_x + \sqrt{\rho_k}Y_k\bar{d}_y\|. \tag{5.6}$$

If $(\bar{d}_x^k, \bar{d}_y^k)$ is unbounded, then by Assumption 3.1 and (4.6), $\bar{\varphi}_k(\bar{d}^k)$ is unbounded, which contradicts the fact that $\bar{\varphi}_k(\bar{d}^k) \leq \bar{\varphi}_k(0) = \|c^k + y^k\|$. ■

Proof of Theorem 5.1. Condition (1) obviously holds for (I) and (II).

(I) Since $(\hat{d}_x^k, \hat{d}_y^k)$ solves problem (5.1)-(5.3), if (4.11) holds, then

$$\psi_k(\hat{d}^k; \rho_k) \leq \eta_1\eta - \rho_k\hat{\eta}, \tag{5.7}$$

where η_1 is a constant such that $\|g^k\| + \mu\sqrt{m} + \frac{1}{2}\|B_k\hat{d}_x^k\| + \frac{1}{2}\|(Y_k\Lambda_k)Y_k^{-1}\hat{d}_y^k\| \leq \eta_1$. Thus Condition (2) follows.

Now we prove that Condition (3) holds. Suppose that (4.13) holds. Let

$$\bar{d}_x^k = -A_k(A_k^\top A_k + Y_k^2)^{-1}(c^k + y^k), \ \ \bar{d}_y^k = -Y_k^2(A_k^\top A_k + Y_k^2)^{-1}(c^k + y^k), \tag{5.8}$$

then

$$c^k + y^k + A_k^\top \bar{d}_x^k + \bar{d}_y^k = 0, \tag{5.9}$$

and there exists a constant $\hat{\eta}$ such that

$$\left\|\begin{pmatrix} \bar{d}_x^k \\ Y_k^{-1}\bar{d}_y^k \end{pmatrix}\right\| \leq \hat{\eta}\|c^k + y^k\|. \tag{5.10}$$

If $\|c^k + y^k\| \leq \eta/\hat{\eta}$, then $(\bar{d}_x^k, \bar{d}_y^k)$ is the solution of (5.1)-(5.3). Let $(\hat{d}_x^k, \hat{d}_y^k) = (\bar{d}_x^k, \bar{d}_y^k)$, then the result follows from (5.9). Otherwise, by Lemma 3.2 and the boundedness of $\{x^k\}$, there is a positive constant κ such that $\|c^k + y^k\| \leq \kappa$. by selecting $0 < \tilde{\eta} < \min\{1, \eta/(\hat{\eta}\kappa)\}$ and letting

$$\tilde{d}_x^k = \tilde{\eta}\bar{d}_x^k, \ \ \tilde{d}_y^k = \tilde{\eta}\bar{d}_y^k, \tag{5.11}$$

it follows that (5.2) and (5.3) hold for $(\tilde{d}_x^k, \tilde{d}_y^k)$. Since $(\hat{d}_x^k, \hat{d}_y^k)$ solves (5.1)-(5.3), we have

$$\begin{aligned} &\|c^k + y^k + A_k^\top \hat{d}_x^k + \hat{d}_y^k\| - \|c^k + y^k\| \le \|c^k + y^k + A_k^\top \tilde{d}_x^k + \tilde{d}_y^k\| - \|c^k + y^k\| \\ &\le \tilde{\eta}(\|c^k + y^k + A_k^\top \bar{d}_x^k + \bar{d}_y^k\| - \|c^k + y^k\|) = -\tilde{\eta}\|c^k + y^k\|. \end{aligned} \tag{5.12}$$

(II) Let $(\tilde{d}_x^k, \tilde{d}_y^k)$ be a solution of (5.1)-(5.3) and

$$\varphi_k(d_x, d_y) = \frac{1}{2} d_x^\top B_k d_x + \frac{1}{2} d_y^\top Y_k^{-1} \Lambda_k d_y + \rho_k \|c^k + y^k + A_k^\top d_x + d_y\|. \tag{5.13}$$

Then

$$\varphi_k(\hat{d}_x^k, \hat{d}_y^k) \le \varphi_k(\tilde{d}_x^k, \tilde{d}_y^k). \tag{5.14}$$

Since B_k and $Y_k \Lambda_k$ are uniformly bounded, if (4.11) holds, then there exists a constant ζ such that

$$\varphi_k(\tilde{d}_x^k, \tilde{d}_y^k) - \rho_k \|c^k + y^k\| \le \zeta - \rho_k \hat{\eta}. \tag{5.15}$$

Moreover, by Lemma 5.2,

$$\|\hat{d}_x^k\| / \sqrt{\rho_k} \le \eta_2, \quad \|Y_k^{-1} \hat{d}_y^k\| / \sqrt{\rho_k} \le \eta_2 \tag{5.16}$$

for some constant $\eta_2 > 0$. Thus, (4.12) follows from the definition of ψ, (5.16), (5.14) and (5.15).

If (4.13) holds, then by (5.8)-(5.10)

$$\varphi_k(\bar{d}_x^k, \bar{d}_y^k) \le \eta_3 \|c^k + y^k\|^2 \tag{5.17}$$

for some constant $\eta_3 > 0$. On the other hand, there exists a constant $\eta_4 > 0$ such that

$$\varphi_k(\hat{d}_x^k, \hat{d}_y^k) \ge \frac{1}{2} \hat{d}_x^{k\,\top} B_k \hat{d}_x^k \ge \eta_4 \|\hat{d}_x^k\|^2, \tag{5.18}$$

$$\varphi_k(\hat{d}_x^k, \hat{d}_y^k) \ge \frac{1}{2} \hat{d}_y^{k\,\top} Y_k^{-1} \Lambda_k \hat{d}_y^k \ge \eta_4 \|Y_k^{-1} \hat{d}_y^k\|^2. \tag{5.19}$$

Since $\varphi_k(\bar{d}_x^k, \bar{d}_y^k) \ge \varphi_k(\hat{d}_x^k, \hat{d}_y^k)$, we obtain

$$\|\hat{d}_x^k\| \le \sqrt{\eta_3/\eta_4} \left\|c^k + y^k\right\|, \quad \|Y_k^{-1} \hat{d}_y^k\| \le \sqrt{\eta_3/\eta_4} \left\|c^k + y^k\right\|. \tag{5.20}$$

In the next, we prove that the last inequality of (4.14) holds. Since

$$\varphi_k(\hat{d}_x^k, \hat{d}_y^k) \ge \rho_0 \|c^k + y^k + A_k^\top \hat{d}_x^k + \hat{d}_y^k\|, \tag{5.21}$$

together with (5.17) and $\varphi_k(\bar{d}_x^k, \bar{d}_y^k) \ge \varphi_k(\hat{d}_x^k, \hat{d}_y^k)$, we have

$$\begin{aligned} \|c^k + y^k + A_k^\top \hat{d}_x^k + \hat{d}_y^k\| - \|c^k + y^k\| &\le (\eta_3/\rho_0)\|c^k + y^k\|^2 - \|c^k + y^k\| \\ &\le ((\eta_3/\rho_0)\|c^k + y^k\| - 1)\|c^k + y^k\|. \end{aligned} \tag{5.22}$$

Since the boundedness of $\{(x^k, y^k)\}$, there exists a constant $\eta_5 > 0$ such that $\|c^k + y^k\| \le \eta_5$. Thus, if $\eta_3\eta_5 < \rho_0$, then by letting $\tilde{\eta} = 1 - \eta_3\eta_5/\rho_0$, we have the desired result; otherwise, if $\eta_3\eta_5 \ge \rho_0$, since

$$\begin{aligned} &\|c^k + y^k + A_k^\top \hat{d}_x^k + \hat{d}_y^k\| - \|c^k + y^k\| \le \frac{1}{\rho_k}(\varphi_k(\hat{d}_x^k, \hat{d}_y^k) - \rho_k\|c^k + y^k\|) \qquad (5.23)\\ &\le \frac{1}{\rho_k}\left(\varphi_k(\bar{d}_x^k, \bar{d}_y^k) - \rho_k\|c^k + y^k\|\right) \le \left(\frac{1}{\rho_k}\eta_3\eta_5 - 1\right)\|c^k + y^k\|, \end{aligned}$$

for $\hat{\rho} \ge \eta_3\eta_5 + 1$, by selecting $\tilde{\eta} = 1 - \eta_3\eta_5/\hat{\rho}$, Condition (3) holds for all k with $\rho_k \ge \hat{\rho}$. ■

It should be noted that in a practical implementation of the algorithm we need not solve problems (5.1)-(5.3) and (5.4) exactly in order to satisfy the convergence conditions, see [1, 2, 10]. For example, the algorithm in [1] only requires the approximate solution to satisfy the so-called Cauchy Decrease Condition. Our results provide a possible theoretical explanation for the nice convergence properties of these methods.

In summary, we have provided a theoretical framework for the analysis of a line search IPM and its refinement in which a new merit function and some delicate direction-finding/line-search procedure are employed. All limit points of the sequence generated by these algorithms are shown to be KKT points if $\rho_k \not\to \infty$, whereas if $\rho_k \to \infty$, one of the limit point is either a Fritz-John point or a stationary point of the infeasibility function. To the authors' knowledge, the algorithms in [10] and this paper appear to be so far the only IPMs that have this nice property while entirely based on a line search strategy without regularity assumptions.

Acknowledgment. We thank Professor Khoo Boo Cheong, the co-chair of the high performance computation program of the Singapore-MIT Alliance, for his support in this research.

References

[1] 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.

[2] 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.

[3] R.H. Byrd, M. Marazzi and J. Nocedal, *On the convergence of Newton iterations to non-stationary points,* Report OTC 2001/01, Optimization Technology Center, Northwestern University, Evanston, IL 60208, USA

[4] A.R. Conn, N. Gould and Ph.L. Toint, *A primal-dual algorithm for minimizing a nonconvex function subject to bound and linear equality constraints,* Nonlinear Optimization and Applications 2, G. Di Pillo and F. Giannessi, eds., Kluwer Academic Publishers, 1999.

[5] 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,* JOTA, 89(1996), 507-541.

[6] A.V. Fiacco and G.P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques,* John Wiley and Sons, New York, 1968, republished by SIAM, Philadelphia, 1990.

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

[8] D.M. Gay, M.L. Overton and M.H. Wright, *A primal-dual interior method for nonconvex nonlinear programming*, Advances in nonlinear programming: Proceedings of the 96 International conference on nonlinear programming, Y. Yuan, eds., Kluwer Academic Publishers, 1998.

[9] L.S. Lasdon, J. Plummer and G. Yu, *Primal-dual and primal interior point algorithms for general nonlinear programs,* ORSA J. Comput., 7(1995), 321-332.

[10] X.W. Liu and J. Sun, *A robust primal-dual interior point algorithm for nonlinear programs,* Research report, Dept. of Decision Sciences, National Univ. of Singapore, 2001. http://www.bschool.nus.edu.sg/depart/ds/sunjiehomepage/

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

[12] 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.

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

[14] R.J. Vanderbei and D.F. Shanno, *An interior-point algorithm for nonconvex nonlinear programming,* Comput. Optim. Appl. 13(1999), 231-252.

[15] 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.

[16] H. Yamashita, *A global convergent primal-dual interior point method for constrained optimization,* Optim. Methods Softw., 10(1998), 448-469.