CORE
Metadata, citation and similar papers at core.ac.uk
Provided by ScholarBank@NUS

SIAM J. OPTIM.
Vol. 14, No. 4, pp. 1163–1186
© 2004 Society for Industrial and Applied Mathematics

A ROBUST PRIMAL-DUAL INTERIOR-POINT ALGORITHM FOR NONLINEAR PROGRAMS*

XINWEI LIU† AND JIE SUN†

Abstract. We present a primal-dual interior-point algorithm for solving optimization problems with nonlinear inequality constraints. The algorithm has some of the theoretical properties of trust region methods, but works entirely by line search. Global convergence properties are derived without assuming regularity conditions. The penalty parameter ρ in the merit function is updated adaptively and plays two roles in the algorithm. First, it guarantees that the search directions are descent directions of the updated merit function. Second, it helps to determine a suitable search direction in a decomposed SQP step. It is shown that if ρ is bounded for each barrier parameter μ, then every limit point of the sequence generated by the algorithm is a Karush–Kuhn–Tucker point, whereas if ρ is unbounded for some μ, then the sequence has a limit point which is either a Fritz–John point or a stationary point of a function measuring the violation of the constraints. Numerical results confirm that the algorithm produces the correct results for some hard problems, including the example provided by Wächter and Biegler, for which many of the existing line search–based interior-point methods have failed to find the right answers.

Key words. nonlinear optimization, interior-point method, global convergence, regularity conditions

AMS subject classifications. 49M30, 49M37, 65K10, 90C22, 90C26, 90C30, 90C51

DOI. 10.1137/S1052623402400641

1. Introduction. Applying an interior-point approach to nonlinear programming has been the subject of intensive studies in recent years; see [1, 4, 5, 11, 12, 15, 16, 18, 23, 24, 25, 27, 28, 29]. For simplicity of presentation, we concentrate in this paper on inequality constrained nonlinear programs

$$\text{minimize } f(x) \quad \text{subject to } c(x) \le 0, \tag{1.1}$$

where $c(x) = (c_1(x), \ldots, c_m(x))^\top$, $f : \Re^n \to \Re$, and $c : \Re^n \to \Re^m$. We do not assume any convexity on f and c. However, we suppose that f and c are twice continuously differentiable throughout this paper.

The interior-point approach solves, as $\mu \downarrow 0$, the barrier problems

$$\text{minimize } f(x) - \mu \sum_{i=1}^{m} \ln y_i \quad \text{subject to } c(x) + y = 0. \tag{1.2}$$

The direction-finding Newton equations then include

$$c(x) + y + \nabla c(x)^\top d_x + d_y = 0. \tag{1.3}$$

Note that (1.3) is always feasible even if the linearized inequality

$$c(x) + \nabla c(x)^\top d_x \le 0 \tag{1.4}$$

*Received by the editors January 5, 2002; accepted for publication (in revised form) December 1, 2003; published electronically July 20, 2004. This research was partially supported by the Singapore-MIT Alliance and grant R-314-000-042/057-112 of the National University of Singapore.
http://www.siam.org/journals/siopt/14-4/40064.html
†Department of Decision Sciences and Singapore-MIT Alliance, National University of Singapore, Singapore (smaliuxw@nus.edu.sg, jsun@nus.edu.sg).

may be inconsistent, which presents difficulties in convergence of interior-point–based methods. The examples discussed by Byrd, Marazzi, and Nocedal [7] and Wächter and Biegler [26] show that the interior-point methods using (1.3) may not find a feasible point of the original problem or a point with stationary properties. We also notice that the global convergence analysis of most existing interior-point methods requires rather strong assumptions on regularity at all iterates. Wächter and Biegler [26] indicate that these assumptions may not hold even though the local minima have very good regularity properties.

A remedy to these problems is to apply sequential quadratic programming (SQP) techniques to the barrier problems and to use a trust region strategy to ensure the robustness of the algorithm. Such algorithms have recently been proposed by Byrd, Gilbert, and Nocedal [4] and Tseng [24], for example. The numerical experiments in [5] show that the trust region–type algorithm is very promising.

We provide a different approach in this paper. Instead of introducing additional trust region constraints, we use refined line search rules to generate a new iterate in a decomposed SQP framework. The search direction is determined by either a Newton-type step or a Cauchy-type step with the choice being made with reference to a penalty parameter in the merit function. In addition, we adjust the penalty parameter of the merit function adaptively. As a result, we have been able to analyze convergence without regularity conditions and to avoid the convergence problems mentioned above. However, unlike the trust region methods, the algorithm does not have the flexibility to allow the direct use of indefinite second order derivatives.

The convergence properties of the algorithm can be summarized as follows. Let ρ_k be the value of the penalty parameter of the merit function at iterate k. If $\{\rho_k\}_{k=0}^{\infty}$ is bounded independent of the barrier parameter μ, then every convergent subsequence produced by the algorithm converges to a Karush–Kuhn–Tucker (KKT) point of the problem. If $\rho_k \to \infty$ for some μ, then the sequence has a limit point that is either feasible with linearly dependent gradients of the active constraints (i.e., a Fritz–John point) or infeasible but stationary with respect to the function $\|\max[0, c(x)]\|$, which is obviously a measure of the violation of the constraints (ℓ_2-infeasibility for short).

Besides, we show that, if the penalty parameters are bounded, then the algorithm generates the identical search directions with the original primal-dual methods such as LOQO (see Shanno and Vanderbei [23, 25]) after a finite number of iterations. Thus, superlinear convergence may be derived by existing works, such as [6, 29], under suitable conditions; while in the unbounded case, the algorithm may have linear convergence. For brevity, we mainly consider global convergence in this paper. By the same token, practical implementation techniques are not discussed. The interested reader is referred to the related literature, such as [6, 8, 11, 15, 16, 23, 25, 28, 29], for details.

Our numerical results show that the proposed algorithm can find solutions of the examples in [7, 26] and the least ℓ_2-infeasibility solution for an infeasible example in [3], among others.

The paper is organized as follows. In section 2, we present a two-step decomposition scheme of SQP and specify the requirement for an approximate solution to the resulting unconstrained penalty subproblems. In section 3, this scheme is applied to the barrier problem 1.2 and we present a modified primal-dual system of equations that is used in the algorithm for the barrier problem. The global convergence of the algorithm is analyzed in section 4. In section 5 we present the overall algorithm for problem (1.1) and its global convergence results. We provide some computational formulae for the approximate solutions of the unconstrained penalty subproblems and report our preliminary numerical results in section 6.

We use standard notation from the literature of interior-point methods and nonlinear programming. For example, a letter with superscript k is related to the kth iteration; the subscript i is the ith component for a vector or the ith column for a matrix. The norm $\|\cdot\|$ is the Euclidean norm. We also use simplified notation, such as $f_k = f(x^k)$, $g^k = \nabla f(x^k)$, $c^k = c(x^k)$, and $A_k = \nabla c(x^k)$. For vector y, $Y = \mathrm{diag}(y)$ is the diagonal matrix whose ith diagonal element is y_i. All vector inequalities are understood componentwise. For two symmetric matrices A and B, $A \succ (\succeq) B$ means that $A - B$ is positive definite (semidefinite).

2. A decomposition scheme of SQP.

2.1. The basic idea. The barrier problem

$$\text{minimize } f(x) - \mu \sum_{i=1}^{m} \ln y_i \quad \text{subject to } c(x) + y = 0$$

is simply expressed as

$$\text{minimize } \psi_\mu(z) \tag{2.1}$$

$$\text{subject to } h(z) = 0, \tag{2.2}$$

where $z = (x, y)$, $h(z) = c(x) + y$, and $\psi_\mu(z) = f(x) - \mu \sum_{i=1}^{m} \ln y_i$. It is obvious that $\psi_\mu(z)$ is a continuously differentiable function for $y > 0$. At the current iteration point z, the SQP approach for (2.1)–(2.2) generates the search direction d_z by solving the quadratic programming problems

$$\text{minimize } \nabla\psi(z)^\top d + \frac{1}{2} d^\top Q d \tag{2.3}$$

$$\text{subject to } h(z) + \nabla h(z)^\top d = 0, \tag{2.4}$$

where Q is a positive definite approximation to the Lagrangian Hessian at z. Then the new iteration point z^+ is derived by a line search procedure,

$$z^+ = z + \alpha d_z, \tag{2.5}$$

where $\alpha \in (0, 1]$ is the steplength along d_z. This general framework requires regularity assumptions on $h(z)$ at all iterates. Otherwise, some of the slack variables may tend to zero too quickly and the algorithm may fail to find the right solution [26].

Our idea is rooted in the work of Fletcher [13, 14], Liu [19], and Yuan and Liu [20], although in the original works [19, 20] the authors need to *exactly* solve all the subproblems, including a nonsmooth unconstrained optimization problem. For the barrier problem, we first approximately solve the penalty optimization problem

$$\text{minimize}_{d \in \Re^n} \ \frac{1}{2} d^\top Q d + \rho \|h(z) + \nabla h(z)^\top d\|, \tag{2.6}$$

where $\rho > 0$ is the penalty parameter in the merit function

$$\phi(z; \rho) = \psi_\mu(z) + \rho \|h(z)\|. \tag{2.7}$$

Let $\tilde{d}_z$ be an approximate solution to (2.6). Then we generate the search direction d_z by solving the subproblem

$$\text{minimize } \nabla\psi(z)^\top d + \frac{1}{2}d^\top Q d \tag{2.8}$$

$$\text{subject to } \nabla h(z)^\top d = \nabla h(z)^\top \tilde{d}_z. \tag{2.9}$$

We consider subproblem (2.8)–(2.9) since it can provide us with the estimates of the multipliers, which are needed in the primal-dual interior-point approach. It can be proved (see Proposition 3.1) that, for sufficiently large ρ, the solution d_z to (2.8)–(2.9) is a descent direction of the merit function.

The idea is similar to the trust region interior-point method, in which the auxiliary step $\tilde{d}_z$ is generated by minimizing $\|h(z) + \nabla h(z)^\top d\|$ on a trust region; see [4, 9, 10, 21, 22]. Here, by adding a quadratic term, we remove the trust region constraint in deriving the auxiliary step for the modified system of primal-dual equations.

2.2. The approximate solution to subproblem (2.6). In this subsection we describe how to generate the approximate solution to subproblem (2.6). Subproblem (2.6) can be simply written as

$$\text{minimize } \; q(d) = \frac{1}{2}d^\top Q d + \rho\|r + R^\top d\|, \tag{2.10}$$

where $\rho > 0$, Q is any positive definite matrix, r is a vector, and R is a matrix with full column rank. It is easy to note that the exact solution is $d = 0$ if $r = 0$. Thus, in the following discussion, we assume that $r \neq 0$.

We generate the approximate solution $\tilde{d}_z$ to problem (2.10) by the following procedure.

PROCEDURE 2.1.

(1) *Compute the Q-weighted Newton step for minimizing $\|r + R^\top d\|$:*

$$\tilde{d}_z^N = -Q^{-1}R(R^\top Q^{-1}R)^{-1}r. \tag{2.11}$$

If $q(\tilde{d}_z^N) \leq \nu q(0)$ ($\nu \in (0,1)$ is a fixed constant), then $\tilde{d}_z = \tilde{d}_z^N$; else go to (2)*.*

(2) *Calculate the Q-weighted steepest descent step (Cauchy step)*

$$\tilde{d}_z^C = -Q^{-1}Rr. \tag{2.12}$$

Find $\tilde{d}_z$ in the subspace spanned by $\tilde{d}_z^N$ and $\tilde{d}_z^C$ (see details in section 6.1*) such that*

$$q(\tilde{d}_z) \leq \max\{\nu q(0), q(\alpha^C \tilde{d}_z^C)\}, \tag{2.13}$$

where $\alpha^C = \operatorname{argmin}_{\alpha\in[0,1]} q(\alpha\tilde{d}_z^C)$.

Let us point out that, when our algorithm produces a sequence converging to a KKT point of the barrier problem, the Q-weighted Newton step will eventually be accepted under suitable conditions, so the direction-finding process (2.6)–(2.9) will generate an identical direction with the original primal-dual interior-point methods (see section 3). Intuitively, the Newton step can be rejected only if $q(\tilde{d}_z^N) > \nu q(0)$, namely,

$$\frac{1}{2}r^\top(R^\top Q^{-1}R)^{-1}r > \rho\nu\|r\|. \tag{2.14}$$

With a moderate value of ρ, if $R^\top Q^{-1}R$ is nonsingular, the above relationship indicates that $\|r\|$ is large, or at least is of the order of ρ. This cannot happen for an iterate close to a KKT point x^* since this iterate must be nearly feasible, i.e., $\|r\|$ must be small. Later, we will present more detailed analysis on this point (see Propositions 3.2 and 3.3).

We next provide a technical result on the decrement of the Cauchy step for later reference.

PROPOSITION 2.2. *There holds*

$$q(\alpha^C \tilde{d}_z^C) - q(0) \le \frac{1}{2}\left\{1 - \rho \min\left[\frac{1}{\|r\|}, \frac{\eta}{\|r\|}\right]\right\} r^\top (R^\top Q^{-1}R)r, \tag{2.15}$$

where $\eta = [r^\top(R^\top Q^{-1}R)r]/[r^\top(R^\top Q^{-1}R)^2 r]$.

Proof. Let $\chi(d) = \|r + R^\top d\|$. We have

$$\begin{aligned}\chi(0)^2 - \chi(\alpha\tilde{d}_z^C)^2 &= \|r\|^2 - \|(I - \alpha R^\top Q^{-1}R)r\|^2 \\ &= 2\alpha r^\top(R^\top Q^{-1}R)r - \alpha^2 r^\top (R^\top Q^{-1}R)^2 r.\end{aligned} \tag{2.16}$$

Suppose that $\tilde{\alpha} \in [0,1]$ minimizes $\chi(\alpha\tilde{d}_z^C)$. Then we have the following two cases:

(i) If $\eta \le 1$, then

$$\chi(0)^2 - \chi(\tilde{\alpha}\tilde{d}_z^C)^2 = \eta r^\top (R^\top Q^{-1}R)r, \tag{2.17}$$

which implies that

$$\chi(0) - \chi(\tilde{\alpha}\tilde{d}_z^C) \ge \frac{\eta}{2\|r\|} r^\top (R^\top Q^{-1}R)r. \tag{2.18}$$

(ii) If $\eta > 1$, then $\tilde{\alpha} = 1$ and $r^\top(R^\top Q^{-1}R)r > r^\top(R^\top Q^{-1}R)^2 r$; thus

$$\chi(0) - \chi(\tilde{\alpha}\tilde{d}_z^C) \ge \frac{1}{2\|r\|} r^\top (R^\top Q^{-1}R)r. \tag{2.19}$$

Then it follows from (2.18), (2.19), and $\tilde{\alpha} \le 1$ that

$$q(\tilde{\alpha}\tilde{d}_z^C) - q(0) \le \frac{1}{2}\left\{1 - \rho \min\left[\frac{1}{\|r\|}, \frac{\eta}{\|r\|}\right]\right\} r^\top (R^\top Q^{-1}R)r. \tag{2.20}$$

Since $q(\alpha^C \tilde{d}_z^C) \le q(\tilde{\alpha}\tilde{d}_z^C)$, we obtain (2.15). □

3. The algorithm for the barrier problem. We now specialize the formulae in the last section to the barrier problem (1.2) and present a modified primal-dual system of equations for generating the search directions. Later, based on this modification, we will propose our algorithm for the barrier problem.

By writing z as (x, y), $\psi_\mu(z)$ as $\psi_\mu(x, y)$, *and* $h(z)$ as $h(x, y)$, the barrier problem is

$$\text{minimize } \psi_\mu(x, y) = f(x) - \mu \sum_{i=1}^m \ln y_i \tag{3.1}$$

$$\text{subject to } h(x, y) = c(x) + y = 0, \tag{3.2}$$

where $y = (y_1, \ldots, y_m)^\top > 0$, and μ is a fixed positive scalar. The Lagrangian of problem (3.1)–(3.2) is

$$L(x, y, \lambda) = \psi_\mu(x, y) + \lambda^\top h(x, y), \tag{3.3}$$

and its Hessian is

$$\nabla^2 L(x,y,\lambda) = \begin{pmatrix} \nabla^2 \ell(x,\lambda) & \\ & \mu Y^{-2} \end{pmatrix}, \tag{3.4}$$

where $\lambda \in \Re^m$ is a multiplier vector associated with (3.2) and $\ell(x,\lambda) = f(x) + \lambda^\top c(x)$. The KKT conditions of program (3.1)–(3.2) can be written as

$$F_\mu(x,y,\lambda) = \begin{pmatrix} g(x) + A(x)\lambda \\ Y\Lambda e - \mu e \\ c(x) + y \end{pmatrix} = 0, \tag{3.5}$$

where $g(x) = \nabla f(x)$, $A(x) = \nabla c(x)$, $Y = \mathrm{diag}(y)$, $\Lambda = \mathrm{diag}(\lambda)$, and $e = (1,\ldots,1)^\top$.

Byrd, Marazzi, and Nocedal [7] showed that the algorithm using the norm of the residual function $\|F_\mu(x,y,\lambda)\|$ as the merit function may fail in converging to a stationary point of the problem. In this paper, as mentioned in (2.7), our merit function is

$$\phi_\mu(x,y;\rho) = \psi_\mu(x,y) + \rho\|h(x,y)\|, \tag{3.6}$$

where $\rho > 0$ is the penalty parameter and is updated automatically during the iterations. Then we have the following result.

PROPOSITION 3.1. *For any $\rho \ge 0$, $y > 0$, and $(d_x, d_y) \in \Re^{n+m}$, the directional derivative $\phi_\rho'((x,y);(d_x,d_y))$ of $\phi_\mu(x,y;\rho)$ along (d_x,d_y) exists, and*

$$\phi_\rho'((x,y);(d_x,d_y)) \le \pi_\rho((x,y);(d_x,d_y)), \tag{3.7}$$

where

$$\begin{aligned} &\pi_\rho((x,y);(d_x,d_y)) \\ &\quad = g(x)^\top d_x - \mu e^\top Y^{-1} d_y + \rho(\|c(x) + y + A(x)^\top d_x + d_y\| - \|c(x)+y\|). \end{aligned} \tag{3.8}$$

Proof. The first term on the right-hand side of (3.6), ψ_μ, is continuously differentiable. Its directional derivative is

$$\psi_\mu'((x,y);(d_x,d_y)) = g(x)^\top d_x - \mu e^\top Y^{-1} d_y. \tag{3.9}$$

Let $\theta(x,y) = \|h(x,y)\|$. Its directional differentiability follows from its convexity. Since

$$\begin{aligned} &\theta'((x,y);(d_x,d_y)) \\ &\quad = \lim_{\alpha\downarrow 0} \frac{[\theta(x+\alpha d_x, y+\alpha d_y) - \theta(x,y)]}{\alpha} \\ &\quad = \lim_{\alpha\downarrow 0} \frac{[\|c(x) + \alpha A(x)^\top d_x + y + \alpha d_y + o(\alpha)\| - \|c(x)+y\|]}{\alpha} \\ &\quad \le \lim_{\alpha\downarrow 0} \left[\frac{\|c(x) + y + \alpha(A(x)^\top d_x + d_y)\| - \|c(x)+y\|}{\alpha} + \frac{\|o(\alpha)\|}{\alpha} \right] \\ &\quad \le \|c(x) + y + A(x)^\top d_x + d_y\| - \|c(x)+y\| + \lim_{\alpha\downarrow 0} \frac{o(\alpha)}{\alpha}, \end{aligned}$$

where the last two inequalities follow from the triangle inequality and the convexity of the norm, the result follows immediately. $\square$

Suppose that (x^k, y^k) is the current iteration point and λ^k is the corresponding approximation of the multiplier vector. For problem (3.1)–(3.2), by substituting

$$Q = \begin{pmatrix} B_k & \\ & Y_k^{-1}\Lambda_k \end{pmatrix}, \quad R = \begin{pmatrix} A_k \\ I \end{pmatrix}, \quad d = \begin{pmatrix} d_x \\ d_y \end{pmatrix} \text{ and } r = (c^k + y^k) \tag{3.10}$$

into (2.10), our approach first approximately solves the problem

$$\text{minimize} \;\; q_k(d_x, d_y) = \frac{1}{2} d_x^\top B_k d_x + \frac{1}{2} d_y^\top S_k d_y + \rho_k \| c^k + y^k + A_k^\top d_x + d_y \|, \tag{3.11}$$

where $B_k \succ 0$ is an approximation to matrix $\nabla^2 \ell(x^k, \lambda^k)$, $S_k = Y_k^{-1}\Lambda_k$, $Y_k = \text{diag}(y^k)$, $\Lambda_k = \text{diag}(\lambda^k)$, $c^k = c(x^k)$, and $A_k = A(x^k)$, and ρ_k is the current value of the penalty parameter. The Q-weighted Newton step and the Q-weighted steepest descent step defined in Procedure 2.1 are, respectively,

$$(\tilde{d}_x^k)^N = -B_k^{-1} A_k (A_k^\top B_k^{-1} A_k + S_k^{-1})^{-1} (c^k + y^k), \tag{3.12}$$

$$(\tilde{d}_y^k)^N = -S_k^{-1} (A_k^\top B_k^{-1} A_k + S_k^{-1})^{-1} (c^k + y^k), \tag{3.13}$$

and

$$(\tilde{d}_x^k)^C = -B_k^{-1} A_k (c^k + y^k), \quad (\tilde{d}_y^k)^C = -S_k^{-1} (c^k + y^k). \tag{3.14}$$

Let $(\tilde{d}_x^k, \tilde{d}_y^k)$ be the approximate solution obtained through Procedure 2.1. We generate the search direction (d_x^k, d_y^k) for the new iterate by solving

$$\text{minimize} \;\; (g^k)^\top d_x - \mu e^\top Y_k^{-1} d_y + \frac{1}{2} d_x^\top B_k d_x + \frac{1}{2} d_y^\top S_k d_y \tag{3.15}$$

$$\text{subject to} \;\; A_k^\top d_x + d_y = A_k^\top \tilde{d}_x^k + \tilde{d}_y^k, \tag{3.16}$$

where $g^k = \nabla f(x^k)$. Since $(\tilde{d}_x^k, \tilde{d}_y^k)$ is a feasible solution to problem (3.15)–(3.16), by (3.8), we have the formula

$$\begin{aligned} &\pi_{\rho_k}((x^k, y^k); (d_x^k, d_y^k)) + \frac{1}{2} (d_x^k)^\top B_k d_x^k + \frac{1}{2} (d_y^k)^\top S_k d_y^k \\ &\quad \le \pi_{\rho_k}((x^k, y^k); (\tilde{d}_x^k, \tilde{d}_y^k)) + \frac{1}{2} (\tilde{d}_x^k)^\top B_k \tilde{d}_x^k + \frac{1}{2} (\tilde{d}_y^k)^\top S_k \tilde{d}_y^k, \end{aligned} \tag{3.17}$$

which plays an important role in our later global convergence analysis for the case $\rho_k \to \infty$.

The KKT conditions of problem (3.15)–(3.16) are

$$B_k d_x + A_k \tilde{\lambda} = -g^k, \tag{3.18}$$

$$S_k d_y + \tilde{\lambda} = \mu Y_k^{-1} e, \tag{3.19}$$

$$A_k^\top d_x + d_y = A_k^\top \tilde{d}_x^k + \tilde{d}_y^k, \tag{3.20}$$

which, by letting $d_\lambda = \tilde{\lambda} - \lambda^k$, can be equivalently written as the modified primal-dual system of equations

$$B_k d_x + A_k d_\lambda = -(g^k + A_k \lambda^k), \tag{3.21}$$

$$\Lambda_k d_y + Y_k d_\lambda = -(Y_k \Lambda_k e - \mu e), \tag{3.22}$$

$$A_k^\top d_x + d_y = A_k^\top \tilde{d}_x^k + \tilde{d}_y^k. \tag{3.23}$$

It is well known that the original primal-dual interior-point approach generates the search direction by solving the system of equations

$$B_k d_x + A_k d_\lambda = -(g^k + A_k \lambda^k), \tag{3.24}$$

$$\Lambda_k d_y + Y_k d_\lambda = -(Y_k \Lambda_k e - \mu e), \tag{3.25}$$

$$A_k^\top d_x + d_y = -(c^k + y^k), \tag{3.26}$$

which follows from the Newton method applied to (3.5); for example, see [11, 16, 23, 25, 28]. Then we have the following results.

PROPOSITION 3.2. *The modified approach using* (3.21)–(3.23) *generates the same search directions as the original primal-dual interior-point methods using* (3.24)–(3.26) *if the weighted Newton step* (3.12)–(3.13) *is used.*

Proof. If $\tilde{d}_x^k = (\tilde{d}_x^k)^N$ and $\tilde{d}_y^k = (\tilde{d}_y^k)^N$, then $A_k^\top \tilde{d}_x^k + \tilde{d}_y^k = -(c^k + y^k)$. Thus the system (3.21)–(3.23) is the same as the system (3.24)–(3.26). $\square$

PROPOSITION 3.3. *Suppose that the two sets* $\{(x^k, y^k)\}_{k=0}^\infty$ *and* $\{(A_k^\top B_k^{-1} A_k + S_k^{-1})^{-1}\}_{k=0}^\infty$ *are bounded. Then there exists a positive constant* $\hat{\rho}$ *(which is not dependent on* k*) such that for* $\rho_k \ge \hat{\rho}$*, the Newton step* $((\tilde{d}_x^k)^N, (\tilde{d}_y^k)^N)$ *defined in* (3.12)–(3.13) *will be accepted by Procedure* 2.1.

Proof. We have

$$\begin{aligned} & q_k((\tilde{d}_x^k)^N, (\tilde{d}_y^k)^N) - \nu q_k(0,0) \\ &= \frac{1}{2}(c^k + y^k)^\top (A_k^\top B_k^{-1} A_k + S_k^{-1})^{-1}(c^k + y^k) - \nu \rho_k \|c^k + y^k\| \\ &\le \left[\frac{1}{2}\|(A_k^\top B_k^{-1} A_k + S_k^{-1})^{-1}(c^k + y^k)\| - \nu \rho_k\right] \|c^k + y^k\|. \end{aligned} \tag{3.27}$$

By the assumptions of the proposition, there exists a constant $\hat{\rho} > 0$ such that for all k we have

$$\|(A_k^\top B_k^{-1} A_k + S_k^{-1})^{-1}(c^k + y^k)\| \le 2\nu\hat{\rho}. \tag{3.28}$$

Thus, for every $\rho_k \ge \hat{\rho}$, $q_k((\tilde{d}_x^k)^N, (\tilde{d}_y^k)^N) \le \nu q_k(0,0)$. $\square$

In the following, we describe our algorithm for the barrier problem (3.1)–(3.2), which solves the problem (3.11) and the system of equations (3.21)–(3.23) at each iteration.

ALGORITHM 3.4 (the algorithm for problem (3.1)–(3.2)).

Step 1. *Given* $(x^0, y^0, \lambda^0) \in \Re^n \times \Re^m_{++} \times \Re^m_{++}$, $0 \prec B_0 \in \Re^{n\times n}$, $0 < \beta_1 < 1 < \beta_2$, $\rho_0 > 0$, $0 < \delta < 1$, $0 < \sigma_0 < \frac{1}{2}$, $\epsilon_1 > 0$, $\epsilon_2 > \epsilon_3 > 0$. *Let* $k := 0$.

Step 2. *Compute an approximate solution* $(\tilde{d}_x^k, \tilde{d}_y^k)$ *of problem* (3.11) *by Procedure* 2.1 *(see section* 6.1 *on its implementation).*

Step 3. *Calculate the search direction* $(d_x^k, d_y^k, d_\lambda^k)$ *by solving the system of equations* (3.21)–(3.23).

Step 4 *(update* ρ_k*). If*

$$\pi_{\rho_k}((x^k, y^k); (d_x^k, d_y^k)) \le -\frac{1}{2}(d_x^k)^\top B_k d_x^k - \frac{1}{2}(d_y^k)^\top S_k d_y^k, \tag{3.29}$$

then set $\rho_{k+1} = \rho_k$*; otherwise, we update* ρ_k *by*

$$\rho_{k+1} = \max\left\{\frac{\psi'_\mu((x^k, y^k); (d_x^k, d_y^k)) + \frac{1}{2}(d_x^k)^\top B_k d_x^k + \frac{1}{2}(d_y^k)^\top S_k d_y^k}{\Delta_k},\ 2\rho_k\right\}, \tag{3.30}$$

where

$$\pi_{\rho_k}((x^k, y^k); (d_x^k, d_y^k)) = (g^k)^\top d_x^k - \mu e^\top Y_k^{-1} d_y^k - \rho_k \Delta_k \tag{3.31}$$

and

$$\Delta_k = \|c^k + y^k\| - \|c^k + y^k + A_k^\top d_x^k + d_y^k\|. \tag{3.32}$$

Step 5 *(line search). Compute*

$$\hat{\alpha}_k = \frac{-0.995}{\min\{(y_i^k)^{-1}(d_y^k)_i, i = 1, \ldots, m; -0.995\}}. \tag{3.33}$$

Select the least nonnegative integer l *such that*

$$\begin{aligned} &\phi_\mu(x^k + \delta^l \hat{\alpha}_k d_x^k, y^k + \delta^l \hat{\alpha}_k d_y^k; \rho_{k+1}) - \phi_\mu(x^k, y^k; \rho_{k+1}) \\ &\quad \le \sigma_0 \delta^l \hat{\alpha}_k \pi_{\rho_k}((x^k, y^k); (d_x^k, d_y^k)). \end{aligned} \tag{3.34}$$

Let $\alpha_k = \delta^l \hat{\alpha}_k$*. The new primal iterate is generated as*

$$x^{k+1} = x^k + \alpha_k d_x^k, \tag{3.35}$$
$$y^{k+1} = \max\{y^k + \alpha_k d_y^k, -c^{k+1}\}. \tag{3.36}$$

Step 6 *(update dual iterate). If there exists* $\gamma \in [0, 1]$ *such that*

$$\beta_1 \mu e \le Y_{k+1}(\Lambda_k + \gamma D_\lambda^k) e \le \beta_2 \mu e, \tag{3.37}$$

where $D_\lambda^k = diag(d_\lambda^k)$*, then we select the maximum* $\gamma_k \in [0, 1]$ *satisfying* (3.37) *and then update* λ^k *by*

$$\lambda^{k+1} = \lambda^k + \gamma_k d_\lambda^k; \tag{3.38}$$

otherwise, we increase l *by* 1 *successively such that* (3.37) *holds, and then update the primal and dual iterates in the same way as in* (3.35), (3.36)*, and* (3.38)*.*

Step 7 *(check the stopping criteria). We terminate the algorithm if one of the following conditions is satisfied:*

(i) $\|F_\mu(x^{k+1}, y^{k+1}, \lambda^{k+1})\| < \epsilon_1$*;*

(ii) $\|c^{k+1} + y^{k+1}\| \ge \epsilon_2$ *and* $\|\begin{pmatrix} A_{k+1} \\ Y_{k+1} \end{pmatrix}(c^{k+1} + y^{k+1})\| < \epsilon_3$*;*

(iii) $\|c^{k+1} + y^{k+1}\| < \epsilon_3$ *and* $\det(A_{\mathcal{I}_{k+1}}^\top A_{\mathcal{I}_{k+1}}) < \epsilon_3$*,*

where $\mathcal{I}_{k+1} = \{i | c_i^{k+1} \ge -\epsilon_3\}$*, and* $A_{\mathcal{I}_{k+1}}$ *is a submatrix of* A_{k+1} *consisting of all columns indexed by* $\mathcal{I}_{k+1}$*. Else update the approximate Hessian* B_k *by* B_{k+1}*, let* $k := k + 1$*, and go to Step* 2*.*

We make the following remarks on the algorithm:

- The new primal and dual iterates are generated, respectively, by using different steplengths. Such a strategy has been used in [8, 28, 29]. We hope that $\gamma_k = 1$ can be accepted even if $\alpha_k < 1$.
- By (3.33), we have $y^k + \hat{\alpha}_k d_y^k \ge 0.005 y^k$. If $d_{yi}^k \ge 0$, we have $y_i^k + \alpha_k d_{yi}^k \ge y_i^k$; else $\alpha_k d_{yi}^k \ge \hat{\alpha}_k d_{yi}^k$ since $\alpha_k \le \hat{\alpha}_k$. Thus we always have $y^{k+1} \ge 0.005 y^k$ by (3.36).

- Formula (3.36) was first introduced in [4]; a similar, but more sophisticated, technique is also used in [24]. Since $y^{k+1} \geq y^k + \alpha_k d_y^k$ and $\|c^{k+1} + y^{k+1}\| \leq \|c^{k+1} + y^k + \alpha_k d_y^k\|$, we have

$$\begin{aligned} &\phi_\mu(x^{k+1}, y^{k+1}; \rho_{k+1}) - \phi_\mu(x^k, y^k; \rho_{k+1}) \\ &\quad \leq \phi_\mu(x^k + \alpha_k d_x^k, y^k + \alpha_k d_y^k; \rho_{k+1}) - \phi_\mu(x^k, y^k; \rho_{k+1}); \end{aligned} \tag{3.39}$$

thus $\phi_\mu(x^{k+1}, y^{k+1}; \rho_{k+1}) \leq \phi_\mu(x^k, y^k; \rho_{k+1})$ for all $k \geq 0$.
- A way to implement (3.37) will be introduced in section 6.2. The well-definedness of this step is shown in Lemma 4.4.
- Since we do not assume any regularity on the constraints, the stopping condition (i) may never hold, in which case the algorithm will terminate at condition (ii) or (iii) of Step 7 by the convergence results in the next section.

4. The analysis of global convergence. The global convergence of Algorithm 3.4 is analyzed in this section. Suppose that in the algorithm the tolerance ϵ_2 is small, tolerances ϵ_1 and ϵ_3 are very small, and an infinite sequence $\{(x^k, y^k, \lambda^k)\}$ is generated.

We need the following blanket assumption for all analysis in what follows.

ASSUMPTION 4.1.

(1) *Functions f and c are twice continuously differentiable functions on $\Re^n$.*

(2) *The set $\{x^k\}_{k=0}^\infty$ is bounded.*

(3) *There exist positive constants ν_1 and ν_2 such that $\nu_1 I \preceq B_k \preceq \nu_2 I$ for all k, where I stands for the identity matrix.*

Assumptions (1) and (2) are used in the convergence analysis of most algorithms for nonlinear programming. Assumption (3) guarantees the existence of the solution of system (3.21)–(3.23). Similar assumptions are also used by most line search–based interior-point methods for nonlinear programming. An exception is [8], in which the global convergence results are derived by assuming B_k to be uniformly positive definite and bounded on the null space of the linear equality constraints.

By Algorithm 3.4, for each integer $k \geq 0$, we have either $\rho_{k+1} = \rho_k$ or $\rho_{k+1} \geq 2\rho_k$. Thus, the sequence $\{\rho_k\}$ is a monotonically nondecreasing sequence.

LEMMA 4.2. *If there exist a positive integer $\hat{k}$ and a positive constant $\hat{\rho}$ such that $\rho_k = \hat{\rho}$ for all $k \geq \hat{k}$, then we have that*

(i) *both $\{y^k\}$ and $\{\lambda^k\}$ are bounded above and componentwise bounded away from zero. The same is true for the diagonal of S_k.*

(ii) *$\{(d_x^k, d_y^k, d_\lambda^k)\}$ is bounded.*

Proof. Without loss of generality, we suppose that $\rho_k = \hat{\rho}$ for all $k \geq 0$. By (3.34) and (3.39), $\phi_\mu(x^k, y^k; \hat{\rho})$ is monotonically decreasing; thus $\phi_\mu(x^k, y^k; \hat{\rho}) \leq \phi_\mu(x^0, y^0; \hat{\rho})$ for all k. Now we prove that y^k is bounded above by contradiction. Suppose that $\max_i\{y_i^k\} \to \infty$. We have also that

$$f_k - \mu \sum_{i=1}^m \ln y_i^k + \hat{\rho}\|c^k + y^k\| \leq \phi_\mu(x^0, y^0; \hat{\rho}). \tag{4.1}$$

Dividing both sides of (4.1) by $\max_i\{y_i^k\}$ and taking the limit when $k \to \infty$, we have that $\hat{\rho} \leq 0$ since each term approaches zero except $\lim_{k\to\infty} \|c^k + y^k\|/\max_i\{y_i^k\} \geq 1$. This is a contradiction.

By the fact that x^k and y^k are bounded and that

$$-\mu \sum_{i=1}^m \ln y_i^k \leq -f_k - \hat{\rho}\|c^k + y^k\| + \phi_\mu(x^0, y^0; \hat{\rho}), \tag{4.2}$$

y^k is componentwise bounded away from zero. It follows from (3.37) that λ^k is bounded above and componentwise bounded away from zero; so is the diagonal of S_k since $S_k = Y_k^{-1}\Lambda_k$.

(ii) By Assumption 4.1(3), matrix $\hat{B}_k = B_k + A_k Y_k^{-1}\Lambda_k A_k^\top$ is invertible. By simple computation, the system (3.21)–(3.23) can be written as

$$\begin{pmatrix} B_k & A_k \\ A_k^\top & -\Lambda_k^{-1}Y_k \end{pmatrix}\begin{pmatrix} d_x^k \\ d_\lambda^k \end{pmatrix} = \begin{pmatrix} -(g^k + A_k\lambda^k) \\ (Y_k - \mu\Lambda_k^{-1})e + (A_k^\top \tilde{d}_x^k + \tilde{d}_y^k) \end{pmatrix} \tag{4.3}$$

and

$$d_y^k = (\mu\Lambda_k^{-1} - Y_k)e - \Lambda_k^{-1}Y_k d_\lambda^k. \tag{4.4}$$

Since

$$\begin{pmatrix} B_k & A_k \\ A_k^\top & -\Lambda_k^{-1}Y_k \end{pmatrix}^{-1} = \begin{pmatrix} \hat{B}_k^{-1} & \hat{B}_k^{-1}A_k Y_k^{-1}\Lambda_k \\ \Lambda_k Y_k^{-1}A_k^\top \hat{B}_k^{-1} & P_k \end{pmatrix}, \tag{4.5}$$

where $P_k = -Y_k^{-1}\Lambda_k + Y_k^{-1}\Lambda_k A_k^\top \hat{B}_k^{-1} A_k Y_k^{-1}\Lambda_k$, the boundedness of (d_x^k, d_λ^k) follows from (4.3). By (4.4), d_y^k is bounded. $\square$

By Lemma 4.2, there exist constants $b_1 > 0$ and $b_2 > 0$ such that $y^k \geq b_1 e$ and $\|d_y^k\| \leq b_2$ for all k. If $\hat{\alpha}_1 = \min\{1, 0.995b_1/b_2\}$, then $y^k + \hat{\alpha}_1 d_y^k \geq 0.005y^k$. Thus, for all $\alpha \in [0, \hat{\alpha}_1]$,

$$y^k + \alpha d_y^k \geq 0.005y^k. \tag{4.6}$$

Lemma 4.3. *If $\{\rho_k\}$ is bounded, then there is a constant $\hat{\alpha}_2 \in (0, \hat{\alpha}_1]$ such that, for every $\alpha \in (0, \hat{\alpha}_2]$ and for all $k \geq 0$, there holds that*

$$\phi_\mu(x^k + \alpha d_x^k, y^k + \alpha d_y^k; \rho_{k+1}) - \phi_\mu(x^k, y^k; \rho_{k+1}) \leq \alpha\sigma_0\pi_{\rho_{k+1}}((x^k, y^k); (d_x^k, d_y^k)). \tag{4.7}$$

Proof. Without loss of generality, we suppose that $\rho_k = \hat{\rho}$ for all $k \geq 0$. Then (3.29) holds at all iterates. For $\alpha \in (0, \hat{\alpha}_1]$, by (4.6), we have

$$(Y_k + \alpha D_y^k)^{-1} \preceq 200Y_k^{-1}, \tag{4.8}$$

where $D_y^k = \mathrm{diag}(d_y^k)$. Thus, for $\alpha \in (0, \hat{\alpha}_1]$,

$$\begin{aligned} &-\sum_{i=1}^m \ln[y_i^k + \alpha(d_y^k)_i] + \sum_{i=1}^m \ln y_i^k + \alpha e^\top {Y_k}^{-1} d_y^k \\ &\quad = e^\top \int_0^\alpha [Y_k^{-1} - (Y_k + tD_y^k)^{-1}] d_y^k dt \\ &\quad = e^\top \int_0^\alpha Y_k^{-1}(Y_k + tD_y^k)^{-1}(tD_y^k) d_y^k dt \leq 100\alpha^2 \|{Y_k}^{-1} d_y^k\|^2. \end{aligned} \tag{4.9}$$

Since f and c are twice continuously differentiable, there are positive constants b_3 and b_4 such that

$$f(x^k + \alpha d_x^k) - f(x^k) - \alpha g(x^k)^\top d_x^k \leq \frac{1}{2}\alpha^2 b_3 \|d_x^k\|^2 \tag{4.10}$$

and

$$\begin{aligned}&\|c(x^k+\alpha d_x^k)+y^k+\alpha d_y^k\|-\|c(x^k)+y^k+\alpha A(x_k)^\top d_x^k+\alpha d_y^k\|\\ &\quad\le \|c(x^k+\alpha d_x^k)-c(x^k)-\alpha A(x^k)^\top d_x^k\|\le \frac{1}{2}\alpha^2 b_4\|d_x^k\|^2.\end{aligned}\tag{4.11}$$

The constants b_3 and b_4 are the first order Lipschitzian constants of f and c, respectively.

Let $b_5=\max\{100\mu,\frac{1}{2}(b_3+\hat{\rho}b_4)\}$. Since

$$\begin{aligned}&\pi_{\hat{\rho}}((x^k,y^k);(\alpha d_x^k,\alpha d_y^k))\\ &\quad=\alpha\psi_\mu'((x^k,y^k);(d_x^k,d_y^k))+\hat{\rho}(\|c^k+y^k+\alpha A_k^\top d_x^k+\alpha d_y^k\|-\|c^k+y^k\|)\end{aligned}\tag{4.12}$$

by (3.8), it follows from (4.9), (4.10), and (4.11) that

$$\begin{aligned}&\phi_\mu(x^k+\alpha d_x^k,y^k+\alpha d_y^k;\hat{\rho})-\phi_\mu(x^k,y^k;\hat{\rho})-\pi_{\hat{\rho}}((x^k,y^k);(\alpha d_x^k,\alpha d_y^k))\\ &\quad\le \alpha^2 b_5(\|d_x^k\|^2+\|Y_k^{-1}d_y^k\|^2).\end{aligned}\tag{4.13}$$

It is easy to note that $\pi_{\hat{\rho}}((x^k,y^k);(\alpha d_x^k,\alpha d_y^k))$ is a convex function on $\alpha\in[0,1]$. Thus, we have

$$\pi_{\hat{\rho}}((x^k,y^k);(\alpha d_x^k,\alpha d_y^k))-\alpha\pi_{\hat{\rho}}((x^k,y^k);(d_x^k,d_y^k))\le 0,\tag{4.14}$$

and as a result,

$$\begin{aligned}&\pi_{\hat{\rho}}((x^k,y^k);(\alpha d_x^k,\alpha d_y^k))-\alpha\sigma_0\pi_{\hat{\rho}}((x^k,y^k);(d_x^k,d_y^k))\\ &\quad\le \alpha(1-\sigma_0)\pi_{\hat{\rho}}((x^k,y^k);(d_x^k,d_y^k))\\ &\quad\le -\frac{1}{2}\alpha(1-\sigma_0)\hat{\delta}(\|d_x^k\|^2+\|Y_k^{-1}d_y^k\|^2),\end{aligned}\tag{4.15}$$

where $\hat{\delta}=\min\{\nu_1,\beta_1\mu\}$ and the last inequality follows from (3.29), Assumption 4.1(3), and (3.37).

Let $\hat{\alpha}_2=\min\{\hat{\alpha}_1,(1-\sigma_0)\hat{\delta}/(2b_5)\}$. Then, by (4.13) and (4.15), (4.7) holds for all $\alpha\in[0,\hat{\alpha}_2]$ and $k\ge 0$. ∎

LEMMA 4.4. *Under the assumption of* Lemma 4.2, *if* $\beta_1\mu e\le Y_k\Lambda_k e\le\beta_2\mu e$, *then there exists a constant* $\hat{\alpha}_3\in(0,1]$ *such that*

$$\beta_1\mu e\le(\Lambda_k+\alpha D_\lambda^k)\max\{y^k+\alpha d_y^k,-c(x^k+\alpha d_x^k)\}\le\beta_2\mu e\tag{4.16}$$

for all $\alpha\in[0,\hat{\alpha}_3]$ *and all* k.

Proof. At first, we prove that

$$\beta_1\mu e\le(Y_k+\alpha D_y^k)(\Lambda_k+\alpha D_\lambda^k)e\le\beta_2\mu e\tag{4.17}$$

for all $\alpha\in[0,\bar{\alpha}_3]$ and all k, where $\bar{\alpha}_3\in(0,1]$ is a constant.

By (3.22), we have $(Y_k+\alpha D_y^k)(\Lambda_k+\alpha D_\lambda^k)e=\alpha\mu e+(1-\alpha)Y_k\Lambda_k e+\alpha^2 D_y^k D_\lambda^k e$. Thus,

$$(Y_k+\alpha D_y^k)(\Lambda_k+\alpha D_\lambda^k)e\ge\beta_1\mu e+\alpha(1-\beta_1)\mu e+\alpha^2 D_y^k D_\lambda^k e,\tag{4.18}$$

$$(Y_k+\alpha D_y^k)(\Lambda_k+\alpha D_\lambda^k)e\le\beta_2\mu e-\alpha(\beta_2-1)\mu e+\alpha^2 D_y^k D_\lambda^k e.\tag{4.19}$$

Since (d_y^k, d_λ^k) is bounded and $0 < \beta_1 < 1 < \beta_2$, there exists a constant $\bar{\alpha}_3 \in (0, 1]$ such that (4.17) holds for all $\alpha \in [0, \bar{\alpha}_3]$ and all $k \geq 0$.

If $\max\{y^k + \alpha d_y^k, -c(x^k + \alpha d_x^k)\} = y^k + \alpha d_y^k$ for all $k \geq 0$ and all $\alpha \in [0, \bar{\alpha}_3]$, then the lemma follows from (4.17) directly. Now we suppose that, for some k and some constant $\bar{\alpha}_3' \in (0, \bar{\alpha}_3]$, we have $y_i^k + \alpha d_{yi}^k < -c_i(x^k + \alpha d_x^k)$ for all $\alpha \in (0, \bar{\alpha}_3']$. We prove that there exists a constant $\tilde{\alpha}_3 \in (0, 1]$ not dependent on k such that, for all $\alpha \in [0, \tilde{\alpha}_3]$,

$$-(\lambda_i^k + \alpha d_{\lambda i}^k) c_i(x^k + \alpha d_x^k) \leq \beta_2 \mu. \tag{4.20}$$

For convenience of statement, we define $p_i(\alpha) = -(\lambda_i^k + \alpha d_\lambda^k) c_i(x^k + \alpha d_x^k)$. Then $p_i(0) = -c_i^k \lambda_i^k$. We show that there exists a positive constant $\bar{\epsilon}$ such that we have either $p_i(0) \leq \beta_2 \mu - \bar{\epsilon}$ or $p_i'(0) \leq -\bar{\epsilon} < 0$. Then (4.20) follows from the continuity of function p_i and the boundedness of (d_x^k, d_λ^k).

By (3.36), we have $c^k + y^k \geq 0$ and $\lambda_k > 0$ for $k \geq 1$. Thus, $p_i(0) \leq y_i^k \lambda_i^k$. For any given small constant $\epsilon > 0$ satisfying $\beta_2 \mu - c\epsilon > \mu$ ($c > 1$ is a constant), if $c_i^k + y_i^k \geq \epsilon$, or $c_i^k + y_i^k < \epsilon$ and $y_i^k \lambda_i^k \leq \beta_2 \mu - \epsilon$, then $p_i(0) \leq \beta_2 \mu - \bar{\epsilon}$ for some constant $\bar{\epsilon} > 0$. Now suppose $c_i^k + y_i^k < \epsilon$ and $y_i^k \lambda_i^k > \beta_2 \mu - \epsilon$. Then, by Procedure 2.1 and Lemma 4.2, there exists a small positive constant ϵ' dependent on ϵ such that $A_{ki}^\top d_x^k + d_{yi}^k \geq -\epsilon'$. Thus, $p_i'(0) = -\lambda_i^k A_{ki}^\top d_x^k - c_i^k d_{\lambda i}^k \leq \lambda_i^k d_{yi}^k + y_i^k d_{\lambda i}^k + \epsilon''$ for some small positive constant ϵ''. By (3.22), we have $p_i'(0) \leq \mu - y_i^k \lambda_i^k + \epsilon'' < \epsilon + \epsilon'' - (\beta_2 - 1)\mu < 0$ since $\beta_2 > 1$. $\square$

Let $\hat{\alpha}_4 = \min\{\hat{\alpha}_2, \hat{\alpha}_3\}$, where $\hat{\alpha}_2$ and $\hat{\alpha}_3$ are defined as in Lemmas 4.3 and 4.4, respectively. Then $0 < \hat{\alpha}_4 \leq 1$. By Step 5 of Algorithm 3.4, $\alpha_k > \delta \hat{\alpha}_4$ for all k, which implies that our line search procedure is well defined.

LEMMA 4.5. *If $\rho_k = \hat{\rho}$ for all $k \geq \hat{k}$ and if $\{(x^k, y^k, \lambda^k)\}$ is an infinite sequence generated by* Algorithm 3.4*, then we have*

$$\lim_{k \to \infty} d_x^k = 0, \ \lim_{k \to \infty} d_y^k = 0, \tag{4.21}$$

$$\lim_{k \to \infty} \|c^{k+1} + y^{k+1}\| = 0, \tag{4.22}$$

$$\lim_{k \to \infty} Y_{k+1} \Lambda_{k+1} e = \mu e, \tag{4.23}$$

$$\lim_{k \to \infty} \|g^{k+1} + A_{k+1} \lambda^{k+1}\| = 0. \tag{4.24}$$

Proof. It follows from Lemma 4.2 that the sequence $\{\phi_\mu(x^k, y^k; \hat{\rho})\}$ is bounded. Combined with its monotonicity, the limit of $\{\phi_\mu(x^k, y^k; \hat{\rho})\}$ exists as $k \to \infty$. Since $\alpha_k > \delta \hat{\alpha}_4 > 0$ and $\pi_{\hat{\rho}}((x^k, y^k); (d_x^k, d_y^k)) \leq 0$ for all k, by taking the limit on the two sides of (3.34), we have $\lim_{k \to \infty} \pi_{\hat{\rho}}((x^k, y^k); (d_x^k, d_y^k)) = 0$, which implies that $\lim_{k \to \infty} (d_x^k, d_y^k) = 0$ by (3.29) and Lemma 4.2.

By (4.21) and (3.23), we have $A_k^\top \tilde{d}_x^k + \tilde{d}_y^k \to 0$ as $k \to \infty$. If $(\tilde{d}_x^k, \tilde{d}_y^k)$ satisfies $q_k(\tilde{d}_x^k, \tilde{d}_y^k) \leq \nu q_k(0, 0)$, then

$$\|c^k + y^k + A_k^\top \tilde{d}_x^k + \tilde{d}_y^k\| - \nu \|c^k + y^k\| \leq 0, \tag{4.25}$$

which implies that (4.22) holds. Otherwise, since $q_k(\tilde{d}_x^k, \tilde{d}_y^k) \leq q_k(0, 0)$, for $k \to \infty$ we have

$$0 \geq -\frac{1}{2\rho_k} \left(\tilde{d}_x^{k\top} B_k \tilde{d}_x^k + \tilde{d}_y^{k\top} S_k \tilde{d}_y^k \right) \geq \|c^k + y^k + A_k^\top \tilde{d}_x^k + \tilde{d}_y^k\| - \|c^k + y^k\| \to 0. \tag{4.26}$$

It follows that $(\tilde{d}_x^k, \tilde{d}_y^k) \to 0$ as $k \to \infty$. Thus, by Procedure 2.1, formulae (3.12)–(3.14), Lemma 4.2, and Assumption 4.1, we have $\lim_{k\to\infty} \|c^k + y^k\| = 0$. This proves (4.22) by (4.21).

It follows from (3.22) that $Y_k(\lambda^k + d_\lambda^k) = \mu e - \Lambda_k d_y^k$. Thus, by (4.21) and Lemma 4.2, $\lim_{k\to\infty} Y_{k+1}(\lambda^k + d_\lambda^k) = \lim_{k\to\infty} Y_k(\lambda^k + d_\lambda^k) = \mu e$. Then, by Step 6 of Algorithm 3.4, we have $\lambda^{k+1} = \lambda^k + d_\lambda^k$ for sufficiently large k; thus (4.23) holds. Moreover, for sufficiently large k, by (3.21), we have

$$g^k + A_k \lambda^{k+1} = -B_k d_x^k. \tag{4.27}$$

Thus, (4.24) follows immediately from Assumption 4.1 and (4.21). □

It follows from Lemmas 4.2 and 4.5 that the weighted Newton step will be accepted at last if $\{\rho_k\}_{k=0}^\infty$ is bounded, since (3.28) is satisfied after a finite number of iterations.

Now we consider the case of $\rho_k \to \infty$. For simplicity of statement, we give the following definitions.

Definition 4.6.

(1) *$x^* \in \Re^n$ is called a singular stationary point of the problem* (1.1) *if $c(x^*) \le 0$ and $A_i(x^*)$, $i \in \mathcal{I}$, are linearly dependent, where $\mathcal{I} = \{i | c_i(x^*) = 0, i = 1, \ldots, m\}$;*

(2) *$x^* \in \Re^n$ is called an infeasible stationary point of the problem* (1.1), *if x^* is an infeasible point of the problem* (1.1) *and $A(x^*)c(x^*)_+ = 0$, where $c(x^*)_+ = \max\{c(x^*), 0\}$.*

It is easy to see that both the singular stationary point and the infeasible stationary point have some first order stationary properties. Similar definitions are also used in [2, 20, 30]. A singular stationary point is also a Fritz–John point, where the linearly independent constraint qualification does not hold. An infeasible stationary point is also a stationary point for minimizing $\|c(x)_+\|$ because $A(x^*)c(x^*)_+ = 0$. Moreover, if all constraint functions are convex, then the infeasible stationary point is the "least infeasible solution" in ℓ_2 sense.

Lemma 4.7. *If $\rho_k \to \infty$, then*

(i) *the sequence $\{y^k\}$ is bounded;*

(ii) *$\{y^k\}$ is not componentwise bounded away from zero.*

Proof. (i) By (3.34), we have $\phi_\mu(x^{k+1}, y^{k+1}; \rho_{k+1}) \le \phi_\mu(x^k, y^k; \rho_{k+1})$ for all $k \ge 0$. The boundedness of $\{x^k\}$ implies that there exists a constant $b_7 > 0$ such that $|f_k| < b_7$. Thus,

$$\begin{aligned} &\frac{1}{\rho_{k+1}}\phi_\mu(x^{k+1}, y^{k+1}; \rho_{k+1}) - \frac{1}{\rho_k}\phi_\mu(x^k, y^k; \rho_k) \\ &\quad \le \left(\frac{1}{\rho_k} - \frac{1}{\rho_{k+1}}\right)(-\psi_\mu(x^k, y^k)) \\ &\quad \le \left(\frac{1}{\rho_k} - \frac{1}{\rho_{k+1}}\right)(b_7 + \mu m \ln \|y^k\|). \end{aligned} \tag{4.28}$$

It follows from (4.28) that

$$\begin{aligned} &\frac{1}{\rho_{k+1}}\phi_\mu(x^{k+1}, y^{k+1}; \rho_{k+1}) \\ &\quad \le \frac{1}{\rho_0}\phi_\mu(x^0, y^0; \rho_0) + \left(\frac{1}{\rho_0} - \frac{1}{\rho_{k+1}}\right)\left(b_7 + \mu m \max_{0 \le j \le k+1} \ln \|y^j\|\right). \end{aligned} \tag{4.29}$$

On the other hand, we have

$$\frac{1}{\rho_{k+1}}\phi_\mu(x^{k+1}, y^{k+1}; \rho_{k+1})$$

$$\geq -\frac{1}{\rho_{k+1}}\left(b_7 + \mu m \max_{0\leq j\leq k+1} \ln \|y^j\|\right) + \|y^{k+1}\| - \|c^{k+1}\|. \tag{4.30}$$

Thus, by (4.29) and (4.30), there is a constant $b_8 > 0$ such that

$$b_8 + \frac{\mu m}{\rho_0} \max_{0\leq j\leq k+1} \ln \|y^j\| \geq \|y^{k+1}\| \quad \text{for all } k \geq 0, \tag{4.31}$$

which implies that $\{y^k\}$ is bounded.

(ii) If $\{y^k\}$ is componentwise bounded away from zero, then, by (i) and (3.37), the sequence $\{\lambda^k\}$ is also bounded above and componentwise bounded away from zero. Thus, matrix S_k is uniformly bounded. Let $\mathcal{K} = \{k|\rho_k < \rho_{k+1}\}$. Then $\mathcal{K}$ is an infinite index set. It follows from Assumption 4.1 and Proposition 3.3 that there exists a positive constant $\hat{\rho}$ such that the weighted Newton step defined by (3.12) and (3.13) is accepted at iterate $k \in \mathcal{K}$ if $\rho_k > \hat{\rho}$. Thus, $\Delta_k = \|c^k + y^k\|$ by Proposition 3.2 and (3.32). Moreover, there exists a constant $b_9 > 0$ such that, for sufficiently large $k \in \mathcal{K}$,

$$\|\tilde{d}_x^k\| \leq b_9\|c^k + y^k\|, \;\; \|\tilde{d}_y^k\| \leq b_9\|c^k + y^k\|, \;\text{ and } \|S_k\tilde{d}_y^k\| \leq b_9\|c^k + y^k\|. \tag{4.32}$$

Hence, by the boundedness of $\|c^k+y^k\|$ and Assumption 4.1(3), there exists a constant $b_{10} > 0$ such that, for all sufficiently large $k \in \mathcal{K}$,

$$\begin{aligned} &\pi_{\rho_k}((x^k, y^k); (\tilde{d}_x^k, \tilde{d}_y^k)) + \frac{1}{2}(\tilde{d}_x^k)^\top B_k \tilde{d}_x^k + \frac{1}{2}(\tilde{d}_y^k)^\top S_k \tilde{d}_y^k \\ &\leq b_{10}\|c^k + y^k\| - \rho_k\|c^k + y^k\|, \end{aligned} \tag{4.33}$$

which, by (3.17), implies that we have (3.29) for all iterates $k \in \mathcal{K}$ with $\rho_k \geq \max\{\hat{\rho}, b_{10}\}$. This contradicts the fact that $\mathcal{K}$ is an infinite set. ∎

By Lemma 4.7 and (3.37), λ^k is componentwise bounded away from zero. Thus, both Λ_k^{-1} and S_k^{-1} are bounded above.

LEMMA 4.8. *Let $\mathcal{K} = \{k \mid \rho_k < \rho_{k+1}\}$. If $\rho_k \to \infty$ and if $\tilde{\mathcal{K}}$ is any subset of $\mathcal{K}$ such that $(x^k, y^k) \to (x^*, y^*)$ as $k \in \tilde{\mathcal{K}}$ and $k \to \infty$, then*

$$\det[(A_{\mathcal{J}}^*)^\top A_{\mathcal{J}}^*] = 0, \tag{4.34}$$

where $\mathcal{J} = \{i|y_i^* = 0, i = 1, \ldots, m\}$.

Proof. We prove this lemma by contradiction. Suppose that there is a set $\tilde{\mathcal{K}} \subseteq \mathcal{K}$ such that, as $k \in \tilde{\mathcal{K}}$ and $k \to \infty$, $(x^k, y^k) \to (x^*, y^*)$ and $A_i(x^*)$, $i \in \mathcal{J}$, are linearly independent. Then, by Assumption 4.1 and (3.37), there exists a constant $b_{11} > 0$ such that $A(x^*)^\top (B^*)^{-1} A(x^*) + G^* \succeq b_{11} I$, where I is the identity matrix, and for simplicity we assume that $B_k \to B^*$ and $S_k^{-1} \to G^*$ as $k \in \tilde{\mathcal{K}}$ and $k \to \infty$. Thus, by the continuity of $A(x)$, there exists a constant $b_{12} > 0$ such that

$$\|(A_k^\top B_k^{-1} A_k + S_k^{-1})^{-1}\| \leq b_{12} \tag{4.35}$$

for all sufficiently large $k \in \tilde{\mathcal{K}}$. It follows from (3.27) that the weighted Newton step defined by (3.12) and (3.13) is accepted. Hence, we have the same results as (4.32) and (4.33), which result in a contradiction to the definition of $\mathcal{K}$. ∎

LEMMA 4.9. *If $\rho_k \to \infty$, then there must be a limit point which is either a singular stationary point or an infeasible stationary point.*

In order to prove Lemma 4.9, we need to prove three other lemmas first.

LEMMA 4.10. *If $\{(\tilde{d}_x^k, \tilde{d}_y^k)\}$ is a sequence such that $q_k(\tilde{d}_x^k, \tilde{d}_y^k) \leq \omega q_k(0, 0)$ for $0 < \omega \leq 1$, then $\|\tilde{d}_x^k\|/\sqrt{\rho_k}$ and $\|Y_k^{-1}\tilde{d}_y^k\|/\sqrt{\rho_k}$ are uniformly bounded above.*

Proof. Let $(\hat{d}_x^k, \hat{d}_y^k) = (\tilde{d}_x^k/\sqrt{\rho_k}, Y_k^{-1}\tilde{d}_y^k/\sqrt{\rho_k})$. Then by $q_k(\tilde{d}_x^k, \tilde{d}_y^k) \le \omega q_k(0,0)$, we have

$$\frac{1}{2}\hat{d}_x^{k\top} B_k \hat{d}_x^k + \frac{1}{2}\hat{d}_y^{k\top} Y_k \Lambda_k \hat{d}_y^k + \|c^k + y^k + \sqrt{\rho_k} A_k^\top \hat{d}_x^k + \sqrt{\rho_k} Y_k \hat{d}_y^k\| \le \omega \|c^k + y^k\|. \tag{4.36}$$

The boundedness of $(\hat{d}_x^k, \hat{d}_y^k)$ follows from the uniform lower boundedness of the quadratic terms by Assumption 4.1 and (3.37). □

LEMMA 4.11. *Suppose that* $(\tilde{d}_x^k, \tilde{d}_y^k)$ *is an approximate solution of program* (3.11) *such that* $q_k(\tilde{d}_x^k, \tilde{d}_y^k) \le q_k(\alpha_k^C(\tilde{d}_x^k)^C, \alpha_k^C(\tilde{d}_y^k)^C)$*, where* $((\tilde{d}_x^k)^C, (\tilde{d}_y^k)^C)$ *is the weighted steepest descent step* (*see Procedure* 2.1 *and* (3.10))*,* $\alpha_k^C \in [0,1]$ *minimizes the function* $q_k(\alpha(\tilde{d}_x^k)^C, \alpha(\tilde{d}_y^k)^C)$*. Then there exist positive constants* $\tilde{\rho}$ *and* $\tilde{\omega}$ *such that, for* $\rho_k \ge \tilde{\rho}$*, we have*

$$q_k(\tilde{d}_x^k, \tilde{d}_y^k) - q_k(0,0) \le -\tilde{\omega}\rho_k \left\| \begin{pmatrix} A_k \\ Y_k \end{pmatrix} (c^k + y^k) \right\|^2. \tag{4.37}$$

Proof. By (3.10), the value of η in Proposition 2.2 is

$$\eta_k = \|(A_k^\top B_k^{-1} A_k + S_k^{-1})^{1/2}(c^k + y^k)\|^2 / \|(A_k^\top B_k^{-1} A_k + S_k^{-1})(c^k + y^k)\|^2. \tag{4.38}$$

It follows from Assumption 4.1 and (3.37) that

$$\begin{aligned} &(c^k + y^k)^\top \begin{pmatrix} A_k \\ I \end{pmatrix}^\top \begin{pmatrix} B_k^{-1} & \\ & Y_k \Lambda_k^{-1} \end{pmatrix} \begin{pmatrix} A_k \\ I \end{pmatrix} (c^k + y^k) \\ &\ge \omega_1 \left\| \begin{pmatrix} A_k \\ Y_k \end{pmatrix} (c^k + y^k) \right\|^2, \end{aligned} \tag{4.39}$$

where $\omega_1 = \min\{\nu_2^{-1}, \beta_2^{-1}\mu^{-1}\}$. By Assumption 4.1 and Lemma 4.7(i), there is a constant $\omega_2 > 0$ such that $\|c^k + y^k\| \le \omega_2$. Let $\tilde{\rho}_1 = 2\omega_2$. Then, for $\rho_k \ge \tilde{\rho}_1$, we have $1 - (\rho_k/\|c^k + y^k\|) \le -\rho_k/(2\omega_2)$. If $\eta_k \ge 1$, by Proposition 2.2, we have (4.37) if $\tilde{\omega} \le \omega_1/(4\omega_2)$.

Now we suppose that $\eta_k < 1$. By Assumption 4.1, Lemma 4.7, and (3.37), there is a constant $\omega_3 > 0$ such that $\|(A_k^\top B_k^{-1} A_k + S_k^{-1})^{1/2}\|^2 \|c^k + y^k\| \le \omega_3$ for all k. Since $\eta_k \ge 1/\|(A_k^\top B_k^{-1} A_k + S_k^{-1})^{1/2}\|^2$ by (4.38), if we select $\tilde{\rho}_2 = 2\omega_3$, then, for $\rho_k \ge \tilde{\rho}_2$, we have $1 - (\rho_k \eta_k/\|c^k + y^k\|) \le -\rho_k/(2\omega_3)$. Thus, for $\rho_k \ge \tilde{\rho}_2$, it follows from Proposition 2.2 and (4.39) that (4.37) holds if $\tilde{\omega} \le \omega_1/(4\omega_3)$.

Let $\tilde{\omega} = \min\{\omega_1/(4\omega_2), \omega_1/(4\omega_3)\}$. Then the result follows by taking the constant $\tilde{\rho} = \max\{\tilde{\rho}_1, \tilde{\rho}_2\}$. □

LEMMA 4.12. *Let* $\mathcal{K} = \{k \mid \rho_k < \rho_{k+1}\}$*. If* $\rho_k \to \infty$*, then*

$$\left\| \begin{pmatrix} A_k \\ Y_k \end{pmatrix} (c^k + y^k) \right\| \to 0 \tag{4.40}$$

as $k \in \mathcal{K}$ *and* $k \to \infty$*.*

Proof. Suppose that (4.40) does not hold. Then there exist an infinite subset $\tilde{\mathcal{K}} \subseteq \mathcal{K}$, positive constants τ_1 and τ_2 such that

$$\left\| \begin{pmatrix} A_k \\ Y_k \end{pmatrix} (c^k + y^k) \right\| \ge \tau_1, \tag{4.41}$$

and $\|c^k + y^k\| \ge \tau_2$ for all $k \in \tilde{\mathcal{K}}$.

The approximate solution $(\tilde{d}_x^k, \tilde{d}_y^k)$ is generated such that either $q_k(\tilde{d}_x^k, \tilde{d}_y^k) \le \nu q_k(0,0)$ or $q_k(\tilde{d}_x^k, \tilde{d}_y^k) \le q_k(\alpha_k^C(\tilde{d}_x^k)^C, \alpha_k^C(\tilde{d}_y^k)^C)$ (which implies that $q_k(\tilde{d}_x^k, \tilde{d}_y^k) \le q_k(0,0)$). Then, by Lemma 4.10, there is a constant $\tau_3 > 0$ such that $\|\tilde{d}_x^k\| \le \tau_3\sqrt{\rho_k}$, $\|Y_k^{-1}\tilde{d}_y^k\| \le \tau_3\sqrt{\rho_k}$.

If $q_k(\tilde{d}_x^k, \tilde{d}_y^k) \le \nu q_k(0,0)$ for all $k \in \tilde{\mathcal{K}}$, then there exists a constant $\tau_4 > 0$ such that

$$\begin{aligned} &\pi_{\rho_k}\ ((x^k, y^k); (\tilde{d}_x^k, \tilde{d}_y^k)) + \frac{1}{2}(\tilde{d}_x^k)^\top B_k \tilde{d}_x^k + \frac{1}{2}(\tilde{d}_y^k)^\top S_k \tilde{d}_y^k \\ &\le (g^k)^\top \tilde{d}_x^k - \mu e^\top Y_k^{-1}\tilde{d}_y^k - (1-\nu)\rho_k\|c^k + y^k\| \\ &\le \tau_4\sqrt{\rho_k} - (1-\nu)\tau_2\rho_k. \end{aligned} \tag{4.42}$$

Thus, by (3.17), we can select a positive constant $\hat{\rho}$ such that (3.29) holds for all $\rho_k \ge \hat{\rho}$. This contradicts the definition of $\mathcal{K}$. Hence, there must exist an infinite subset $\hat{\mathcal{K}}$ of $\tilde{\mathcal{K}}$ such that $q_k(\tilde{d}_x^k, \tilde{d}_y^k) \le q_k(\alpha_k^C(\tilde{d}_x^k)^C, \alpha_k^C(\tilde{d}_y^k)^C)$ for all $k \in \hat{\mathcal{K}}$. It follows from Lemma 4.11 that (4.37) holds for all $k \in \hat{\mathcal{K}}$. Then, by (4.41), there is a positive constant b_{13} such that, for all $k \in \hat{\mathcal{K}}$,

$$q_k(\tilde{d}_x^k, \tilde{d}_y^k) - q_k(0,0) \le -b_{13}\tau_1^2\rho_k. \tag{4.43}$$

Thus, we have

$$\begin{aligned} &\pi_{\rho_k}\ ((x^k, y^k); (\tilde{d}_x^k, \tilde{d}_y^k)) + \frac{1}{2}(\tilde{d}_x^k)^\top B_k \tilde{d}_x^k + \frac{1}{2}(\tilde{d}_y^k)^\top S_k \tilde{d}_y^k \\ &\le (g^k)^\top \tilde{d}_x^k - \mu e^\top Y_k^{-1}\tilde{d}_y^k - b_{13}\tau_1^2\rho_k \\ &\le \tau_4\sqrt{\rho_k} - b_{13}\tau_1^2\rho_k \end{aligned} \tag{4.44}$$

for all sufficiently large $k \in \bar{\mathcal{K}}$, which implies a contradiction to the definition of $\mathcal{K}$. $\square$

Proof of Lemma 4.9. Since (x^k, y^k) is bounded, without loss of generality, we suppose that $(A_k, c^k, x^k, y^k, Y_k) \to (A^*, c^*, x^*, y^*, Y^*)$ as $k \in \mathcal{K}$ and $k \to \infty$, where $\mathcal{K}$ is defined as in Lemma 4.12, $A^* = A(x^*)$, and $c^* = c(x^*)$. If the limit point (x^*, y^*) is such that $c^* + y^* = 0$, i.e., $c_i^* = 0$ if and only if $y_i^* = 0$, then this limit point is a singular stationary point by Lemma 4.8 since $\mathcal{I} = \mathcal{J}$, where $\mathcal{I}$ and $\mathcal{J}$ are defined as in Definition 4.6 and Lemma 4.8, respectively. Now we consider the case of $\|c^* + y^*\| \ne 0$. By Lemma 4.12,

$$\begin{pmatrix} A^* \\ Y^* \end{pmatrix}(c^* + y^*) = 0, \tag{4.45}$$

and so for any i,

$$y_i^* > 0 \ \Rightarrow\ c_i^* + y_i^* = 0 \ \Rightarrow\ c_i^* < 0. \tag{4.46}$$

Since $c^k + y^k \ge 0$ and $y^k \ge 0$ for all $k \ge 1$ by the algorithm, for each i such that $c_i^* + y_i^* \ne 0$, one has $y_i^* = 0$ by (4.45), and hence $c_i^* > 0$, implying that x^* is infeasible. Then $c^* + y^* = c_+^* = \max\{c^*, 0\}$. It follows from (4.45) that $A^* c_+^* = 0$. Therefore, x^* is an infeasible stationary point. The proof is finished. $\square$

Now we can state our global convergence theorem on Algorithm 3.4.

THEOREM 4.13. *Suppose that* $\{(x^k, y^k, \lambda^k)\}$ *is an infinite sequence generated by applying Algorithm* 3.4 *to the barrier problem* (3.1)–(3.2)*, and suppose that Assumption* 4.1 *holds. The penalty parameter sequence* $\{\rho_k\}$ *is automatically updated and monotonically nondecreasing.*

(i) *If* $\{\rho_k\}$ *is bounded, then any cluster point of* $\{(x^k, y^k, \lambda^k)\}$ *is a KKT point of the barrier problem* (3.1)–(3.2). *In this case,* $\{y^k\}$ *is componentwise bounded away from zero,* $\{x^k\}$ *is asymptotically strictly feasible for the constraints* (1.1), *and* $g^k + A_k\lambda^k \to 0$.

(ii) *If* $\rho_k \to \infty$, *then* $\{y^k\}$ *is not componentwise bounded away from zero, and there is at least one cluster point of* $\{(x^k, y^k, \lambda^k)\}$ *which is either a singular stationary point or an infeasible stationary point. In the latter case, if* (x^k, y^k) *is asymptotically feasible for constraints* (3.2), *then* $\{x^k\}$ *is asymptotically feasible for and close to the boundary of constraints* (1.1). *At the limit the gradients of active constraints of* (1.1) *are linearly dependent. If* (x^k, y^k) *is not asymptotically feasible for constraints* (3.2), *then at the limit point* x^* *we have* $A^*c_+^* = 0$.

Proof. Part (i) follows from Lemma 4.5. Part (ii) can be derived directly by Lemma 4.9. □

5. The overall interior-point algorithm and its convergence. We denote by $\mathcal{F}$ the class of continuous functions $\theta : \Re_{++} \to \Re_{++}$ satisfying $\lim_{\mu\to 0}\theta(\mu) = 0$. Now we present our algorithm for nonlinearly constrained optimization (1.1).

ALGORITHM 5.1 (the line search–based interior-point algorithm for (1.1)).

Step 1. *Given initial point* $(x^0, y^0, \lambda^0) \in \Re^n \times \Re^m_{++} \times \Re^m_{++}$, *initial barrier parameter* $\mu_0 > 0$, $\tau \in (0, 1)$, *tolerance* $\epsilon > 0$, *and function* $\theta \in \mathcal{F}$. *Let* $j := 0$.

Step 2. *For the given barrier parameter* μ_j, *we apply Algorithm* 3.4 *to the barrier problem* (3.1)–(3.2). *If the iterate* $(x^{k_j}, y^{k_j}, \lambda^{k_j})$ *satisfies*

$$\|F_{\mu_j}(x^{k_j}, y^{k_j}, \lambda^{k_j})\| < \theta(\mu_j), \tag{5.1}$$

then let

$$(x^{j+1}, y^{j+1}, \lambda^{j+1}) = (x^{k_j}, y^{k_j}, \lambda^{k_j}) \tag{5.2}$$

and $\rho_{j+1} = \rho_{k_j}$, *and go to Step* 3; *if one of conditions* (ii) *and* (iii) *of Algorithm* 3.4 *holds, stop.*

Step 3. *If* $\mu_j < \epsilon$ *stop; otherwise, let* $\mu_{j+1} = \tau\mu_j$, $j := j + 1$ *and go to Step* 2.

Now we consider the convergence of Algorithm 5.1. The result closely depends on how Algorithm 3.4 behaves for each μ_j. For $\theta(\mu_j) > 0$, if condition (5.1) is satisfied, then Algorithm 5.1 will proceed with μ_{j+1}. The global convergence results of the algorithm are as follows.

THEOREM 5.2. *Suppose that* $\theta \in \mathcal{F}$ *and* $\{(x^j, y^j, \lambda^j)\}$ *is a sequence generated by Algorithm* 5.1. *If Assumption* 4.1 *holds for each barrier problem, and if* $\{(x^k, y^k, \lambda^k)\}$ *is a sequence generated by Algorithm* 3.4, *then, for sufficiently small* ϵ, *Algorithm* 5.1 *may terminate in finitely many steps in one of the following two cases:*

(i) *For some* μ_j, *Algorithm* 5.1 *terminates at Step* 2. *If the termination point is an approximately feasible point, then it is an approximately singular stationary point. Otherwise, it is an approximately infeasible stationary point.*

(ii) *For each* μ_j, *Algorithm* 3.4 *terminates at* (5.1). *Then Algorithm* 5.1 *terminates at Step* 3, *in which case an approximate KKT point of the original problem* (1.1) *is obtained.*

Proof. The results follow immediately from Theorem 4.13 and Algorithm 5.1. □

6. Numerical experiment.

6.1. Formulae used in Procedure 2.1. We present an implementation of Procedure 2.1 in this subsection.

Suppose that the full Q-weighted Newton step is not accepted. Then we compute the weighted Cauchy step $\tilde{d}_z^C$ and try to get an approximate solution $\tilde{d}_z$ to (2.10) along the Q-weighted Newton step, or the so-called dog-leg step, so that (2.13) holds and $q(\tilde{d}_z)$ has as much reduction as possible. If this is impossible, then we do a line search along the Q-weighted steepest descent step and take the approximate solution $\tilde{d}_z$ to be either the truncated Q-weighted Newton step or the truncated Q-weighted steepest descent step so that $q(\tilde{d}_z)$ has more reduction. Thus, (2.13) holds. The details are as follows.

We first compute the optimal steplength along the Q-weighted Newton step to derive as much reduction as possible in this direction. Thus, we solve the single-variable minimizing problem

$$\text{minimize}_{\alpha\in[0,1]}\ \hat{q}(\alpha) = \frac{1}{2}\alpha^2 \tilde{d}_z^{N\top} Q \tilde{d}_z^N + \rho\|r + \alpha R^\top \tilde{d}_z^N\|. \tag{6.1}$$

By direct computation, we have the solution

$$\tilde{\alpha}_1 = \min\left\{\frac{\rho\|r\|}{r^\top(R^\top Q^{-1}R)^{-1}r}, 1\right\}. \tag{6.2}$$

Set $d_z^1 = \tilde{\alpha}_1 \tilde{d}_z^N$. Then we have $\hat{q}(\tilde{\alpha}_1) \le \hat{q}(0)$. It is more convenient in the implementation to compute a dog-leg step in the line segment spanned by the Q-weighted Newton step $\tilde{d}_z^N$ and the following scaled Cauchy step (where η is defined as in Proposition 2.2):

$$\tilde{d}_z^C = -\min\{\eta, 1\} Q^{-1} R r. \tag{6.3}$$

It is apparent that this scaling on $\tilde{d}_z^C$ will not result in any change in our theoretical results. If $\eta \le 1$, then $\tilde{d}_z^C$ is the so-called Cauchy point in minimizing $\|r + R^\top d\|^2$ with starting point $d = 0$. Let $d_z(\alpha) = \alpha \tilde{d}_z^N + (1-\alpha)\tilde{d}_z^C$. Then we calculate $\tilde{\alpha}_2$ by

$$\text{minimize}_{\alpha\in[0,1]} \tilde{q}(\alpha) = \frac{1}{2} d_z(\alpha)^\top Q d_z(\alpha) + \rho\|r + R^\top d_z(\alpha)\|. \tag{6.4}$$

By setting $\tilde{q}^{'}(\alpha) = 0$, we have

$$\alpha_2^* = \frac{\rho\|r + R^\top \tilde{d}_z^C\| - (\tilde{d}_z^N - \tilde{d}_z^C)^\top Q \tilde{d}_z^C}{(\tilde{d}_z^N - \tilde{d}_z^C)^\top Q (\tilde{d}_z^N - \tilde{d}_z^C)}. \tag{6.5}$$

If $\alpha_2^* \le 0$, then $\tilde{\alpha}_2 = 0$; else if $\alpha_2^* \ge 1$, then $\tilde{\alpha}_2 = 1$; else we have $\tilde{\alpha}_2 = \alpha_2^*$. If $\min\{\hat{q}(\tilde{\alpha}_1), \tilde{q}(\tilde{\alpha}_2)\} \le \nu q(0)$ (where ν is defined as in Procedure 2.1), we define $d_z^2 = d_z(\tilde{\alpha}_2)$, else we set $d_z^2 = \tilde{\alpha}_3 \tilde{d}_z^C$, where $\tilde{\alpha}_3 \in (0,1]$ minimizes the function

$$\bar{q}(\alpha) = \frac{1}{2}\alpha^2 (\tilde{d}_z^C)^\top Q \tilde{d}_z^C + \rho\|r + \alpha R^\top \tilde{d}_z^C\|. \tag{6.6}$$

We select the approximate solution $\tilde{d}_z$ from d_z^1 or d_z^2, whichever gives a lower value of $q(\tilde{d}_z)$.

The process for solving (2.10) approximately is summarized into the following algorithm.

ALGORITHM 6.1 (the algorithm for solving problem (2.10) approximately).

Step 1. *Compute the Newton step* $\tilde{d}_z^N$ *by* (2.11). *If* $q(\tilde{d}_z^N) \le \nu q(0)$, *then* $\tilde{d}_z = \tilde{d}_z^N$. *Stop.*

Step 2. *Compute the steepest descent step* $\tilde{d}_z^C$ *by* (2.13).

Step 3. *Calculate* $d_z^1 = \tilde{\alpha}_1 \tilde{d}_z^N$ *by* (6.2) *and* $d_z^2 = \tilde{\alpha}_2 \tilde{d}_z^N + (1-\tilde{\alpha}_2)\tilde{d}_z^C$ *by* (6.4). *If* $\min\{\hat{q}(\tilde{\alpha}_1), \tilde{q}(\tilde{\alpha}_2)\} \le \nu q(0)$, *then go to Step* 5.

Step 4. *Calculate* $d_z^2 = \tilde{\alpha}_3 \tilde{d}_z^C$ *by* (6.6). *If* $\hat{q}(\tilde{\alpha}_1) \le \bar{q}(\tilde{\alpha}_3)$, *we have the approximate solution* $\tilde{d}_z = d_z^1$; *else we select* $\tilde{d}_z = d_z^2$. *Stop.*

Step 5. *If* $\hat{q}(\tilde{\alpha}_1) \le \tilde{q}(\tilde{\alpha}_2)$, *then* $\tilde{d}_z = d_z^1$; *else we have* $\tilde{d}_z = d_z^2$. *Stop.*

6.2. Numerical results. The algorithm is programmed in MATLAB 6.1 and is run on a personal computer under Windows 98. In order to obtain rapid convergence, it is also necessary to carefully control the rate at which the barrier parameter μ and the tolerance $\theta(\mu)$ are decreased. This question has been studied in [6, 11, 29].

It is restrictive to require that (3.37) holds for given β_1 and β_2 for all iterates of Algorithm 3.4 in practice. In our implementation, we update the dual iterate flexibly by selecting the maximal $\gamma_k \in [0,1]$ such that

$$\min\{Y_{k+1}\Lambda_k e, \bar{\beta}_1 \mu e\} \le Y_{k+1}\Lambda_{k+1} e \le \max\{Y_{k+1}\Lambda_k e, \bar{\beta}_2 \mu e\}, \tag{6.7}$$

where $0 < \bar{\beta}_1 < 1 < \bar{\beta}_2$, $\Lambda_{k+1} = \operatorname{diag}(\lambda^{k+1})$, and $\lambda^{k+1} = \lambda^k + \gamma_k d_\lambda^k$. If $\{\rho_k\}_{k=0}^{\infty}$ is bounded, then, by Lemma 4.2 and (6.7), there exist β_1 and β_2 such that (3.37) holds for all iterates. In the case of $\rho_k \to \infty$, suppose that Algorithm 3.4 is terminated within a given number of iterations (for example, 300 iterations). Then, by the fact that $y^{k+1} \ge 0.005 y^k$ and (6.7), $Y_{k+1}\Lambda_{k+1}e \ge \min\{0.005 Y_k \Lambda_k e, \bar{\beta}_1 \mu e\}$. Thus, $Y_k \Lambda_k e \ge \beta_1 \mu e$ if we select $\beta_1 = 0.005^{300} \min\{\mu^{-1} Y_0 \Lambda_0 e, 200\bar{\beta}_1 e\}$. If $y_i^k \lambda_i^k \to \infty$ as $k \to \infty$ for some i, then, by (6.7), $\lambda_i^k \le \lambda_i^{k-1}$ and $\lambda_i^k \to \infty$ as $k \to \infty$ since $\{y^k\}$ is bounded. This is a contradiction. Thus, there exist a constant $\beta_2 > 0$ and an infinite index set $\mathcal{K}$ such that $Y_k \Lambda_k e \le \beta_2 \mu e$ for $k \in \mathcal{K}$. Hence, we have (3.37) for all $k \in \mathcal{K}$.

We select the initial parameters $\mu_0 = 0.01$, $\bar{\beta}_1 = 0.01$, $\bar{\beta}_2 = 10$, $\sigma_0 = 0.1$, $\delta = 0.8$, and the initial matrix B_0 to be the $n \times n$ identity matrix. The scalar in Algorithm 6.1 is $\nu = 0.98$. The choice of the initial penalty parameter ρ_0 is scale dependent and $\rho_0 = 1$ is chosen for our experiment. Simply, we select $\theta(\mu) = \mu$, $\tau = 0.01$, $\epsilon = 10^{-6}$. For conditions (ii) and (iii) of Step 7 of Algorithm 3.4, we select $\epsilon_2 = \epsilon$ and $\epsilon_3 = \epsilon^2$.

The approximate Lagrangian Hessian B_{k+1} is computed by the damped BFGS update formula

$$B_{k+1} = B_k - \frac{B_k s^k (s^k)^\top B_k}{(s^k)^\top B_k s^k} + \frac{w^k (w^k)^\top}{(s^k)^\top w^k}, \tag{6.8}$$

where

$$w^k = \begin{cases} \hat{w}^k & \text{if } (\hat{w}^k)^\top s^k \ge 0.2 (s^k)^\top B_k s^k, \\ \theta_k \hat{w}^k + (1-\theta_k) B_k s^k & \text{otherwise,} \end{cases} \tag{6.9}$$

and $\hat{w}^k = g^{k+1} - g^k + (A_{k+1} - A_k)\lambda^{k+1}$, $s^k = x^{k+1} - x^k$, $\theta_k = 0.8 (s^k)^\top B_k s^k / ((s^k)^\top B_k s^k - (s^k)^\top \hat{w}^k)$. For all test problems, we select the initial slack and dual variables as

$$y^0 = e, \quad \lambda^0 = e \tag{6.10}$$

if not specified.

TABLE 1. *Numerical results by Algorithm* 3.4 *when* $\mu = 0.01$.

IT	x_1	x_2	x_3	RC_1	RC_2	ρ	$\tilde{d}_x$
0	-4	1	1	14	-7	1	full-Newton
1	-3.6590	12.3880	0.0050	0	-5.6640	2	dog-leg
2	-2.2786	4.1919	0.0040	0	-4.2826	4	full-Newton
3	-1.3633	0.8586	0.0030	0	-3.3663	4	full-Newton
4	-1.0500	0.1025	0.0026	0	-3.0525	8	dog-leg
5	-0.8756	0.0005	0.0019	-0.2339	-2.8775	8	dog-leg
6	-0.4536	0.0015	0.0000	-0.7957	-2.4537	8	dog-leg
7	0.4972	0.0430e-03	0.5770e-03	-0.7528	-1.5033	8	dog-leg
8	1.4035	0.9697	0.0009	0	-0.5975	8	full-Newton
9	2.0008	3.0031	0.0008	0	-0.9324e-09	8	full-Newton
10	2.0017	3.0067	0.0017	0	0	8	

TABLE 2. *Numerical results by the ordinary approach with* y^{k+1} *generated by* (3.36) *when* $\mu = 0.01$.

IT	x_1	x_2	x_3	RC_1	RC_2	ρ
0	-4	1	1	14	-7	1
1	-3.6590	12.3880	0.0050	0	-5.6640	2
2	-1.9746	2.8990	0.0028	0	-3.9774	5.2958
3	-1.2442	0.5480	0.0018	0	-3.2460	11.9755
4	-1.0251	0.0508	0.0007	0	-3.0258	101.7079
5	-1.0004	0.8606e-03	0.1721e-03	0	-3.0006	4.4576e+03
6	-1.0000	0.0449e-04	0.1219e-04	0	-3.0000	1.1483e+06
7	-1.0000	0.0224e-06	0.1183e-06	0	-3.0000	7.7089e+08
8	-1.0000	0.1122e-09	0.5969e-09	0	-3.0000	9.1419e+12
9	-1.0000	0.0561e-11	0.2984e-11	0	-3.0000	3.0875e+17

First, we apply our algorithm to three simple examples. The first one is the example presented by Wächter and Biegler and further discussed by Byrd, Marazzi, and Nocedal [7, 26]:

$$\text{(TP1)}\qquad \begin{aligned} &\text{Minimize} && x_1 && (6.11)\\ &\text{subject to} && x_1^2 - x_2 - 1 = 0, && (6.12)\\ &&& x_1 - x_3 - 2 = 0, && (6.13)\\ &&& x_2 \geq 0, \ \ x_3 \geq 0. && (6.14) \end{aligned}$$

Note that the initial point $(x_1^0, x_2^0, x_3^0) = (-4, 1, 1)$ satisfies the conditions of Theorem 1 of [26]. There is a unique stationary point for this problem, which is the global minimizer. Moreover, this problem is well-posed, since at the solution the second order sufficient optimality condition, strict complementarity, and nondegeneracy hold. However, it is proved by [26] that many existing interior-point methods using line search (let us call them the "ordinary" interior-point methods for convenience) fail to converge to the stationary point.

Algorithm 5.1 terminates at the approximate KKT point $(2, 3, 0)$ with the Lagrangian multiplier $(0, 1)$ in 16 iterations. The residuals, respectively, are $\|g^k + A_k\lambda^k\| = 6.3283e\text{-}14$, $\|Y_k\Lambda_k e - \mu_k e\| = 2.0000e\text{-}08$, and $\|c^k + y^k\| = 0.8232e\text{-}17$. The value of the penalty parameter is $\hat{\rho} = 8$. In order to see the performance clearly, we give the numerical results of Algorithm 3.4 when $\mu = 0.01$, which is listed in Table 1, where RC_1 and RC_2 are residual values of constraints, (6.12) and (6.13), respectively. The last column in Table 1 shows the performance of Algorithm 6.1, where

TABLE 3. *Numerical results by the ordinary approach with* $y^{k+1} = y^k + \alpha_k d_y^k$ *when* $\mu = 0.01$.

IT	x_1	x_2	x_3	RC_1	RC_2	ρ
0	-4	1	1	14	-7	1
1	-3.6590	0.9438	0.0050	11.4442	-5.6640	2
2	-3.4809	0.0047	0.0029	11.1118	-5.4838	11.9086
3	-3.4789	0.0236e-03	0.3727e-03	11.1028	-5.4793	5.4425e+03
4	-3.4788	0.0118e-05	0.8007e-05	11.1017	-5.4788	3.8388e+05
5	-3.4787	0.0059e-07	0.4240e-07	11.1017	-5.4787	8.9516e+08
6	-3.4787	0.0029e-09	0.2121e-09	11.1017	-5.4787	3.3359e+13

"full-Newton" means that the approximate solution to (3.11) is the full weighted Newton step, and "dog-leg" represents the dog-leg step. In order to observe how the ordinary interior-point approach using (3.24)–(3.26) behaves, we also solve this example by solving (3.24)–(3.26) with y^{k+1} generated by (3.36) and $y^{k+1} = y^k + \alpha_k d_y^k$, respectively; the results are presented in Tables 2 and 3.

It is easy to note from Table 1 that Algorithm 3.4 terminates at the approximate feasible point when $\mu = 0.01$. The approximate feasibility will be further improved when μ is decreased in Algorithm 5.1. However, the results in Tables 2 and 3 show that the ordinary interior-point approach using (3.24)–(3.26) terminates at the infeasible points as $\mu = 0.01$. The infeasibility cannot be improved by decreasing μ since x_2 and x_3 are close to the boundary of the feasible region.

The last column of Table 1 shows that the weighted Newton steps are accepted as the iterates are nearly feasible, which is important for the algorithm to have rapid convergence.

Our second test example is taken from [3], which minimizes any objective function on an obviously infeasible set defined by the constraints:

$$\text{(TP2)} \quad x^2 + 1 \le 0, \ \ x \le 0. \tag{6.15}$$

We select to minimize x as the objective. The initial point is $x^0 = 4$. For $\mu = 0.01$, Algorithm 3.4 terminates at the point $x^* =$ -6.0363e-07, and correspondingly the slack variables $y_1^* =$ 6.3712e-13 and $y_2^* =$ 6.0363e-07 after 38 iterations. It is easy to see that x^* is close to a point by which the norm $\|c(x)_+\|$ is minimized. Algorithm 6.1 takes four full weighted Newton steps at first and then uses the truncated weighted Newton steps in 34 later iterations. The value of the penalty parameter is $\hat{\rho} = 1.2767e{+}10$.

The third simple test problem is a standard one taken from [17, Problem 13]:

$$\text{Minimize} \ \ (x_1 - 2)^2 + x_2^2 \tag{6.16}$$

$$\text{(TP3)} \quad \text{subject to} \ (1 - x_1)^3 - x_2 \ge 0, \tag{6.17}$$

$$x_1 \ge 0, \ \ x_2 \ge 0. \tag{6.18}$$

The standard initial point $(-2, -2)$ is an infeasible point. The optimal solution $(1, 0)$ is not a KKT point but is a singular stationary point, at which the gradients of active constraints are linearly dependent. This problem has not been solved in [23, 25, 28], but has been solved in [5, 24].

Algorithm 5.1 applied to problem (TP3) terminates at the singular stationary point in 44 iterations and $\mu = 0.01$, $y^* = (0, 1, 0)$, $\lambda^* = (3.4923e{+}10, 0.0, 3.4923e{+}10)$. The residuals, respectively, are $\|g^k + A_k\lambda^k\| = 1.2716$, $\|Y_k\Lambda_k e - \mu_k e\| = 0.0292$, and $\|c^k + y^k\| = 0.0$. The value of the penalty parameter is $\hat{\rho} = 2.6370e{+}10$.

TABLE 4. *Numerical results by Algorithm* 5.1.

Problem	Iter	RD	RP	RG	$\hat{\rho}$
TP001	25	2.5953e-11	0	1.0000e-08	1
TP002	22	3.8447e-12	0	1.0000e-08	2
TP003	16	1.9997e-09	0	1.0000e-08	1
TP004	10	1.7656e-13	8.8947e-17	2.0001e-08	4.9402
TP010	18	2.4976e-14	3.0564e-14	1.0000e-08	1
TP011	15	1.1383e-14	9.2021e-16	1.0000e-08	4
TP012	15	2.5011e-14	1.8881e-15	1.0000e-08	1
TP020	38	9.0994e-14	0.5983e-17	5.0000e-08	512
TP021	18	1.3468e-09	0	5.0000e-08	1
TP022	11	1.0991e-12	1.4037e-16	2.0000e-08	1
TP023	14	7.1677e-12	7.1056e-15	9.0000e-08	1
TP024	14	2.5103e-12	4.3581e-16	5.0000e-08	1
TP038	95	7.6785e-09	0	8.0000e-08	1
TP043	22	2.7486e-10	7.2071e-13	3.0000e-08	2
TP044	15	1.3328e-13	7.8580e-16	1.0000e-07	2
TP076	17	2.6222e-09	1.1974e-15	7.0000e-08	1

We also apply our algorithm to some other test problems taken from [17], which are numbered in the same way as that in [17]. For example, "TP022" is Problem 22 in the book. We use these test problems (but not all test problems) since they have only inequality constraints, and thus are suitable for testing the algorithm. The initial points are the same as in [17]. The numerical results are reported in Table 4, where "Iter" represents the number of iterations, $RD = \|g^k + A_k\lambda^k\|$, $RP = \|c^k + y^k\|$, $RG = \|Y_k\Lambda_k e - \mu_k e\|$, and $\hat{\rho}$ is the value of the penalty parameter when the algorithm terminates.

Acknowledgments. The authors are grateful to Professor Philippe L. Toint for reading the draft of this paper and giving insightful comments on it. We also appreciate the associate editor Professor Jorge Nocedal and the anonymous referees for their valuable suggestions and comments.

REFERENCES

[1] 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), pp. 199–222.

[2] J. V. Burke, *A sequential quadratic programming method for potentially infeasible mathematical programs,* J. Math. Anal. Appl., 139 (1989), pp. 319–351.

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

[4] 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), pp. 149–185.

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

[6] R. H. Byrd, G. Liu, and J. Nocedal, *On the local behavior of an interior point algorithm for nonlinear programming,* in Numerical Analysis 1997, D. F. Griffiths and D. J. Higham, eds., Longman, Harlow, UK, 1998, pp. 37–56.

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

[8] A. R. Conn, N. Gould, and Ph. L. Toint, *A primal-dual algorithm for minimizing a non-convex function subject to bound and linear equality constraints,* in Nonlinear Optimization and Related Topics, G. Dipillo and F. Giannessi, eds., Kluwer Academic Publishers, Dordrecht, 2000, pp. 15–49.

[9] J. E. Dennis, Jr., M. El-Alem, and M. C. Maciel, *A global convergence theory for general trust-region-based algorithms for equality constrained optimization,* SIAM J. Optim., 7 (1997), pp. 177–207.

[10] J. E. Dennis and L. N. Vicente, *On the convergence theory of trust-region-based algorithms for equality-constrained optimization,* SIAM J. Optim., 7 (1997), pp. 927–950.

[11] 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), pp. 507–541.

[12] 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.

[13] R. Fletcher, *Practical Methods for Optimization. Vol.* 2*: Constrained Optimization,* John Wiley and Sons, Chichester, 1981.

[14] R. Fletcher, *A model algorithm for composite nondifferentiable optimization problems,* Math. Prog. Stud., 17 (1982), pp. 67–76.

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

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

[17] W. Hock and K. Schittkowski, *Test Examples for Nonlinear Programming Codes,* Lecture Notes in Econom. and Math. Systems 187, Springer, New York, 1981.

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

[19] X.-W. Liu, *A globally convergent, locally superlinearly convergent algorithm for equality constrained optimization,* in Numerical Linear Algebra and Optimization, Y. X. Yuan, ed., Science Press, Beijing, New York, 1999, pp. 131–144.

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

[21] E. O. Omojokun, *Trust Region Algorithms for Optimization with Nonlinear Equality and Inequality Constraints,* Ph. D. Dissertation, University of Colorado, Boulder, CO, 1991.

[22] M. J. D. Powell and Y. Yuan, *A trust region algorithm for equality constrained optimization,* Math. Program., 49 (1991), pp. 189–211.

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

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

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

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

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

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

[29] H. Yamashita and H. Yabe, *Superlinear and quadratic convergence of some primal-dual interior point methods for constrained optimization,* Math. Program., 75 (1996), pp. 377–397.

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

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.