A robust primal-dual interior point algorithm for nonlinear programs *

Xinwei Liu[†] and Jie Sun[‡]

Abstract. We present a primal-dual interior point algorithm of line-search type for nonlinear programs, which uses a new decomposition scheme of sequential quadratic programming. The algorithm can circumvent the convergence difficulties of some existing interior point methods. Global convergence properties are derived without assuming regularity conditions. The penalty parameter ρ in the merit function is updated automatically such that the search directions are descent directions for the merit function. It is shown that if ρ is bounded, then every limiting point of the sequence generated by the algorithm is a KKT point, whereas if ρ is unbounded, then the sequence has a limiting point which is either a Fritz-John point of the feasible set or an infeasible stationary point of minimizing the ℓ_2-norm of infeasibility. Numerical results confirm that the algorithm produces the correct results for some hard problems including the examples provided by Wächter and Biegler and Byrd et al. for which many of the existing line-search type interior point methods have failed to find the right answers.

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

---

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

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

[‡]Department of Decision Sciences and Singapore-MIT Alliance, National University of Singapore, Republic of Singapore. Fax (65)779-2621. Email: jsun@nus.edu.sg

1. Introduction

Applying interior point approaches to nonlinear programming has been the subject of intensive studies in recent years, see [12, 28, 18, 11, 30, 15, 16, 29, 26, 24, 5, 4, 25]. For simplicity of presentation, we concentrate on inequality constrained nonlinear programs in this paper although the presented algorithm can be readily extended to nonlinear programs with both equality and inequality constraints. The problem under consideration takes the following form

$$\min \ f(x) \quad \text{s.t. } c(x) \leq 0, \tag{1.1}$$

where $f(x) : \Re^n \to \Re$ and $c(x) = (c_1(x), \cdots, c_m(x))^\top : \Re^n \to \Re^m$. Both $f(x)$ and $c(x)$ are twice continuous differentiable and may be nonconvex. The interior point approach solves as $\mu \downarrow 0$ the log-barrier subproblems

$$\min \ f(x) - \mu \sum_{i=1}^{m} \log y_i \quad \text{s.t. } c(x) + y = 0. \tag{1.2}$$

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

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

may be inconsistent, which presents difficulties in convergence of interior point based methods. The examples given by Wächter and Biegler and Byrd et al.[27, 7] show that the interior point methods using (1.3) may not converge to any feasible point of the original problem or any point with stationary properties. We also notice that most global convergence analysis on existing interior point methods depend on assumptions on certain regularity conditions at all iterates. [27] indicates 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 trust region to ensure the robustness of the algorithm. Recently, such an algorithm has been published by Byrd et al.in [4] and the numerical experiments in [5] show that the corresponding algorithm is very promising.

We provide a different idea in this paper. Instead of introducing additional trust region constraints, we solve the SQP subproblems in a different way, adjust the penalty parameter of the merit function adaptively, and use refined line search rules. As a result, we have been able to analyze convergence without the regularity conditions and to avoid the convergence problems mentioned above.

Our algorithm needs calculate in each iteration at most one (scaled) Newton direction and one (scaled) steepest descent direction plus a line search procedure. If the penalty parameter keeps bounded, the algorithm generates the identical search directions with the original primal-dual method such as LOQO (Vanderbei and Shanno [26, 24]) after finite iterations. Particularly, the search direction will be the same if regularity conditions hold for all iterates and the initial penalty parameter is selected to be large enough.

Since our algorithm is based on a line search procedure, unlike the trust region case, the algorithm does not have the flexibility to use indefinite approximate Lagrangian Hessians. We rely on some well-known techniques [8, 26, 24] to keep these Hessians positive definite at all iterations. A possible gain by paying such a price is that the norm of search directions can be in the order of square root of the penalty parameter when this parameter tends to infinity (see Lemma 4.9), which help to stabilize the numerical behavior of the algorithm in this case.

The convergence properties of the algorithm can be summarized as follows. Let ρ_k be the penalty parameter used in the merit function at iterate k. If ρ_k is bounded, then every convergent subsequence produced by the algorithm converges to a KKT point of the problem. If $\rho_k \to \infty$, then the sequence has either a limiting point that is feasible with dependent gradients of the active constraints (a Fritz-John point) or that is infeasible but is stationary with respect to the minimization of the ℓ_2-norm of the infeasibility $\|\max[0, c(x)]\|$. Local convergence is not analyzed since in the former case it will be the same as the other interior point line-search methods and superlinear convergence may be derived by the existing works such as [30, 6] under suitable conditions; while in the latter, the algorithm will have at most linear convergence for a given barrier parameter since at last the auxiliary (scaled) Newton direction will not be accepted. The numerical results show that the proposed algorithm can find solutions of the contrary examples in [27, 7], and the least ℓ_2-infeasibility solution for an infeasible example in [3] among others.

For brevity, practical implementation techniques are not discussed in this paper. The interested reader is referred to the related literatures such as [11, 30, 6, 16, 15, 29, 8, 26, 24] for details.

The paper is organized as follows. In Section 2, we discuss a new decomposition scheme of SQP and specify the requirement for an approximate solution to the resulted piecewise quadratic subproblems. This scheme is applied to the logarithmic barrier problem in Section 3, where an interior point approach is presented. The global convergence of a proposed 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 the formulae for solving the piecewise quadratic subproblems and report our preliminary numerical results in Section 6.

We use standard notations in the literature of interior point methods and nonlinear programming. For example, a letter with superscript k is related to the k-th iteration, the subscript i is the i-th component for a vector or the i-th column for a matrix. $\|\cdot\|$ is the Euclidean norm. $c_k = c(x^k)$, $f_k = f(x^k)$, $A_k = A(x^k), Y = \mathrm{diag}(y)$, where y is an m-vector. 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

Let us temporarily ignore the concrete form of the problem in the first section and think of a general equality-constrained optimization problem

$$\min\ f(x) \tag{2.1}$$
$$\text{s.t. } h(x) = 0. \tag{2.2}$$

SQP for (2.1)-(2.2) generates the search directions by solving the quadratic programming problems

$$\min_d\ \nabla f(x)^\top d + \frac{1}{2} d^\top B d \tag{2.3}$$
$$\text{s.t. } h(x) + \nabla h(x)^\top d = 0, \tag{2.4}$$

where B is an approximate Lagrangian Hessian at iterate x and is generally supposed to be positive definite. The new iterate is derived by a line search procedure

$$x^+ = x + \alpha d, \tag{2.5}$$

where $\alpha \in (0, 1]$ is the steplength along d.

Without assuming regularity conditions on $h(x)$, our approach employs the following decomposition scheme to replace solving (2.3)-(2.4) directly. We first solve the unconstrained piecewise quadratic problem

$$\min_{d \in \Re^n}\ \frac{1}{2} d^\top B d + \rho \|h(x) + \nabla h(x)^\top d\|, \tag{2.6}$$

where $\rho > 0$ is a penalty parameter. This idea is closely related to the well-known Sℓ_1QP approach studied by Fletcher [13, 14] and the algorithms presented by Yuan and Liu [31, 19, 20] although ℓ_2-norm, rather than ℓ_1-norm, is used. Since we only need an approximate solution of (2.6) as described below, it does not add any essential computation compared to the ℓ_1-norm.

Let $\tilde{d}$ be an approximate solution to (2.6), the next step (usually called the range-space step) of the decomposition scheme generates the search direction d by solving the subproblem

$$\min \nabla f(x)^\top d + \frac{1}{2} d^\top B d \tag{2.7}$$

$$\text{s.t. } \nabla h(x)^\top d = \nabla h(x)^\top \tilde{d}. \tag{2.8}$$

We consider (2.7)-(2.8) instead of the subproblems in [20] since it can provide us with the estimate of the multipliers, which is needed in the primal-dual approach. It can be proved (see Proposition 3.1) that the solution d to (2.7)-(2.8) is a descent direction of the merit function

$$\phi(x;\rho) = f(x) + \rho\|h(x)\|, \tag{2.9}$$

where ρ is a suitably selected penalty parameter.

The novelty of our decomposition (2.6)-(2.8) stems from the quadratic term in (2.6). Traditionally, the direction $\tilde{d}$ is generated by minimizing $\|h(x) + \nabla h(x)^\top d\|$ on certain given region, see [4, 9, 10, 21, 23]. Note that problem (2.6) can be reformulated as a simple second-order cone program (by introducing new constraints such as $s \geq \rho\|h(x) + \nabla h(x)^\top d\|$). Thus, its exact solution is not difficult in theory. However, in the interior point framework we need not solve the problem exactly. It is only necessary to generate an approximate solution satisfying some loose conditions.

2.2. The approximate solution to subproblem (2.6)

Problem (2.6) can be simply written as

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

where $\rho > 0$, Q is any positive definite matrix, r is a vector, and R is a matrix, $\chi(d) = \|r + R^\top d\|$. 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 $\tilde{d}$ to problem (2.10) by the following procedure:

Procedure 2.1

(1) Compute the scaled Newton direction for minimizing $\chi(d)$:

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

If $q(\tilde{d}^N) \le \nu q(0)$ ($0 < \nu < 1$ is a fixed constant), then $\tilde{d} = \tilde{d}^N$; else go to the next step.

(2) Find a $\tilde{d}$ (see details in Section 6.1) such that the following conditions hold:

(i) $\tilde{d} \to 0$ implies that $Rr \to 0$;

(ii) $q(\tilde{d}) \le \max\{\nu q(0), q(\alpha^C \tilde{d}^C)\}$, where $\tilde{d}^C = -Q^{-1}Rr$ and

$$\alpha^C = \operatorname{argmin}_{\alpha\in(0,1]} q(\alpha \tilde{d}^C), \tag{2.12}$$

in which case we of course have $q(\alpha^C \tilde{d}^C) \le q(0)$.

Proposition 2.2 *There holds*

$$q(\alpha^C \tilde{d}^C) - q(0) \le \frac{1}{2}\left\{1 - \rho \min[\frac{1}{\|r\|}, \frac{\eta}{\|r\|}]\right\} r^\top (R^\top Q^{-1} R) r, \tag{2.13}$$

where

$$\eta = \frac{r^\top (R^\top Q^{-1}R) r}{r^\top (R^\top Q^{-1} R)^2 r}.$$

Proof. By the definition of $\chi(x)$, we have

$$\begin{aligned} \chi(0)^2 - \chi(\alpha\tilde{d}^C)^2 &= \|r\|^2 - \|(I - \alpha R^\top Q^{-1} R) r\|^2 \\ &= 2\alpha r^\top (R^\top Q^{-1} R) r - \alpha^2 r^\top (R^\top Q^{-1} R)^2 r. \end{aligned} \tag{2.14}$$

Suppose that $\tilde{\alpha} \in (0, 1]$ minimizes $\chi(\alpha\tilde{d}^C)$, then we have two cases:
(i) If $\eta \le 1$, then

$$\chi(0)^2 - \chi(\tilde{\alpha}\tilde{d}^C)^2 = \eta r^\top (R^\top Q^{-1} R) r, \tag{2.15}$$

which implies that

$$\chi(0) - \chi(\tilde{\alpha}\tilde{d}^C) \ge \frac{\eta}{2\|r\|} r^\top (R^\top Q^{-1} R) r. \tag{2.16}$$

(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

$$\chi(0) - \chi(\tilde{\alpha}\tilde{d}^C) \ge \frac{1}{2\|r\|} r^\top (R^\top Q^{-1} R) r. \tag{2.17}$$

Then it follows from (2.16), (2.17) and $\tilde{\alpha} \le 1$ that

$$q(\tilde{\alpha}\tilde{d}^C) - q(0) \le \frac{1}{2}\left\{1 - \rho \min[\frac{1}{\|r\|}, \frac{\eta}{\|r\|}]\right\} r^\top (R^\top Q^{-1} R) r. \tag{2.18}$$

Since $q(\alpha^C \tilde{d}^C) \le q(\tilde{\alpha}\tilde{d}^C)$, we obtain (2.13). ∎

3. A modified interior point approach

Recall that for any given barrier parameter $\mu > 0$ the interior point approach solves the logarithmic barrier problems:

$$\min\ f(x) - \mu \sum_{i=1}^{m} \ln y_i \tag{3.1}$$

$$\text{s.t. } c(x) + y = 0, \tag{3.2}$$

where $y = (y_1, \cdots, y_m)^\top > 0$.

Let $z = (x, y)$, $h(z) = c(x) + y$, and

$$\psi_\mu(z) = f(x) - \mu \sum_{i=1}^{m} \ln y_i. \tag{3.3}$$

The Lagrangian of (3.1)-(3.2) is

$$L(z, \lambda) = \psi_\mu(z) + \lambda^\top h(z) \tag{3.4}$$

and its Hessian is

$$\nabla^2 L(z, \lambda) = \begin{pmatrix} \nabla^2 \ell(x, \lambda) & \\ & \mu Y^{-2} \end{pmatrix}, \tag{3.5}$$

where $\ell(x, \lambda) = f(x) + \lambda^\top c(x)$.

The Karush-Kuhn-Tucker (KKT) conditions of the program (3.1)-(3.2) can be written as

$$F_\mu(z, \lambda) := \begin{pmatrix} g(x) + A(x)\lambda \\ Y\Lambda e - \mu e \\ c(x) + y \end{pmatrix} = 0, \tag{3.6}$$

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

It is indicated in [7] that the algorithm using the norm of residual function $\|F_\mu(z, \lambda)\|$ as the merit function may fail in converging to a stationary point of the problem. In this paper, we define the merit function as

$$\phi_\mu(z; \rho) = \psi_\mu(x) + \rho \|c(x) + y\|, \tag{3.7}$$

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

Proposition 3.1 *Suppose that $y > 0$. For any $d = (d_x, d_y) \in \Re^{n+m}$, the directional derivative $\phi_d^{'}(z;\rho)$ of $\phi_\mu(z;\rho)$ along d exists, and*

$$\phi_d^{'}(z;\rho) \le \pi_d(z;\rho), \tag{3.8}$$

where

$$\pi_d(z;\rho) := g(x)^\top d_x - \mu e^\top Y^{-1} d_y + \rho(\|c(x) + y + A(x)^\top d_x + d_y\| - \|c(x) + y\|). \tag{3.9}$$

Proof. $\psi_\mu(z)$ is differentiable, so it is directionally differentiable and

$$\psi_d^{'}(z) = \nabla\psi_\mu(z)^\top d = g(x)^\top d_x - \mu e^\top Y^{-1} d_y. \tag{3.10}$$

Let $\theta(z) = \|c(x) + y\|$. The directional differentiability follows from its convexity. Since

$$\begin{aligned} \theta_d^{'}(z) &= \lim_{\alpha\downarrow 0} \left[\theta(z + \alpha d) - \theta(z)\right]/\alpha \\ &= \lim_{\alpha\downarrow 0} \left[\|c(x) + \alpha A(x)^\top d_x + y + \alpha d_y + o(\alpha)\| - \|c(x) + y\|\right]/\alpha \\ &\le \|c(x) + y + A(x)^\top d_x + d_y\| - \|c(x) + y\| + \lim_{\alpha\downarrow 0} o(\alpha)/\alpha, \end{aligned} \tag{3.11}$$

the result follows immediately. ∎

In order to solve problem (3.1)-(3.2), according to the descriptions in Section 2, our approach first approximately solves the problem

$$\min_{(d_x,d_y)} \quad \varphi(d_x, d_y) = \frac{1}{2} d_x^\top B d_x + \frac{1}{2} d_y^\top S d_y + \rho\|c(x) + y + A(x)^\top d_x + d_y\|, \tag{3.12}$$

where $B \succ 0$ is an approximation to $\nabla_{xx}^2 \ell(x,\lambda)$ and $S = Y^{-1}\Lambda$. Let $(\tilde{d}_x, \tilde{d}_y)$ be an approximate solution to the program (3.12), we generate the search direction for the new iterate by solving

$$\min\ g(x)^\top d_x - \mu e^\top Y^{-1} d_y + \frac{1}{2} d_x^\top B d_x + \frac{1}{2} d_y^\top S d_y \tag{3.13}$$

$$\text{s.t. } A(x)^\top d_x + d_y = A(x)^\top \tilde{d}_x + \tilde{d}_y. \tag{3.14}$$

Since $(\tilde{d}_x, \tilde{d}_y)$ is feasible to the program (3.13)-(3.14), we have the formula

$$\pi_{d_z}(z;\rho) + \frac{1}{2}(d_x)^\top B d_x + \frac{1}{2}(d_y)^\top S d_y \le \pi_{\tilde{d}_z}(z;\rho) + \frac{1}{2}(\tilde{d}_x)^\top B \tilde{d}_x + \frac{1}{2}(\tilde{d}_y)^\top S_k \tilde{d}_y, \tag{3.15}$$

which will play an important role in our analysis for the case $\rho \to \infty$.

The optimization problem (3.13)-(3.14) can be equivalently written as the primal-dual system of equations

$$Bd_x + A(x)d_\lambda = -(g(x) + A(x)\lambda), \tag{3.16}$$

$$\Lambda d_y + Yd_\lambda = -(Y\Lambda e - \mu e), \tag{3.17}$$

$$A(x)^\top d_x + d_y = A(x)^\top \tilde{d}_x + \tilde{d}_y. \tag{3.18}$$

In particular, note that the scaled Newton direction of (3.12) is (cf.(2.11))

$$\tilde{d}_x^N = -B^{-1}A(A^\top B^{-1}A + S^{-1})^{-1}(c+y), \tag{3.19}$$

$$\tilde{d}_y^N = -S^{-1}(A^\top B^{-1}A + S^{-1})^{-1}(c+y). \tag{3.20}$$

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

$$Bd_x + A(x)d_\lambda = -(g(x) + A(x)\lambda), \tag{3.21}$$

$$\Lambda d_y + Yd_\lambda = -(Y\Lambda e - \mu e), \tag{3.22}$$

$$A(x)^\top d_x + d_y = -(c(x) + y), \tag{3.23}$$

which follow from the Newton method applied to (3.6), for example, see [11, 15, 16, 29, 26, 24].

Proposition 3.2 *Our method generates identical directions as the original primal-dual interior point methods if the scale Newton direction is used.*

Proof. The proposition follows from the fact that if $\tilde{d}_x = \tilde{d}_x^N$, $\tilde{d}_y = \tilde{d}_y^N$, then $A(x)^\top \tilde{d}_x + \tilde{d}_y = -(c(x) + y)$. ∎

Proposition 3.3 *Suppose that both (x, y) and $(A^\top B^{-1}A + S^{-1})^{-1}$ are bounded, where B and S are positive definite. Then for large ρ, the Newton direction $(\tilde{d}_x^N, \tilde{d}_y^N)$ defined in (3.19) and (3.20) will be accepted by Procedure 2.1.*

Proof. We have

$$\begin{aligned}\varphi(\tilde{d}_x^N, \tilde{d}_y^N) - \nu\varphi(0,0) &= \frac{1}{2}(c+y)^\top (A^\top B^{-1}A + S^{-1})^{-1}(c+y) - \nu\rho\|c+y\| \\ &\le \left[\frac{1}{2}\|(A^\top B^{-1}A + S^{-1})^{-1}(c+y)\| - \nu\rho\right]\|c+y\|. \end{aligned} \tag{3.24}$$

By the supposition, there exists a constant $\varrho > 0$ such that

$$\|(A^\top B^{-1}A + S^{-1})^{-1}(c+y)\| \le \varrho. \tag{3.25}$$

Thus, for any $\rho > \varrho/\nu$, $\varphi(\tilde{d}_x^N, \tilde{d}_y^N) \le \nu\varphi(0,0)$. ∎

If $(A^\top B^{-1}A+S^{-1})^{-1}$ is unbounded, the scaled Newton direction may not be accepted. In the following, we describe an algorithm that solves (3.12) and (3.16)-(3.18) at each iteration.

Algorithm 3.4 *(The algorithm for problem (3.1)-(3.2))*

Step 1 Given $(x^0, y^0, \lambda^0) \in \Re^n \times \Re^m_{++} \times \Re^m_{++}$, $0 < \beta_1 < 1 < \beta_2$, $\rho_0 > 0$, $\delta < 1$, $\sigma_0 < \frac{1}{2}$. $\mu > 0$. ϵ_1, $\epsilon_2 > 0$. Let $k := 0$;

Step 2 Compute an auxiliary direction $(\tilde{d}_x^k, \tilde{d}_y^k)$ by Procedure 2.1

Step 3 Calculate the search direction $(d_x^k, d_y^k, d_\lambda^k)$ by (3.16)-(3.18);

Step 4 (update ρ) If

$$\pi_{d_z^k}(z^k;\rho_k) \le -\frac{1}{2}(d_x^k)^\top B_k d_x^k - \frac{1}{2}(d_y^k)^\top S_k d_y^k, \tag{3.26}$$

let $\rho_{k+1} = \rho_k$; Otherwise, update ρ_k by

$$\rho_{k+1} = \max\left\{\frac{\psi'_{d_z^k}(z^k) + \frac{1}{2}(d_x^k)^\top B_k d_x^k + \frac{1}{2}(d_y^k)^\top S_k d_y^k}{\Delta_k},\ 2\rho_k\right\}, \tag{3.27}$$

where $\pi_{d_z^k}(z^k;\rho_k) = (g^k)^\top d_x^k - \mu e^\top Y_k^{-1} d_y^k - \rho_k \Delta_k$ and

$$\Delta_k = \|c_k + y^k\| - \|c_k + y^k + A_k^\top d_x^k + d_y^k\|; \tag{3.28}$$

Step 5 (Line search) Compute

$$\hat{\alpha}_k = \frac{-0.995}{\min\{(y_i^k)^{-1}(d_y^k)_i, i = 1, \cdots, m; -0.995\}}. \tag{3.29}$$

Select the least non-negative integer l such that

$$\phi_\mu(z^k + \delta^l \hat{\alpha}_k d_z^k; \rho_{k+1}) - \phi_\mu(z^k;\rho_{k+1}) \le \sigma_0 \delta^l \hat{\alpha}_k \pi_{d_z^k}(z^k;\rho_{k+1}). \tag{3.30}$$

Let $\alpha_k = \delta^l \hat{\alpha}_k$ and the new primal iterate is generated by

$$x^{k+1} = x^k + \alpha_k d_x^k, \tag{3.31}$$

$$y^{k+1} = \max\{y^k + \alpha_k d_y^k, -c_{k+1}\}; \tag{3.32}$$

Step 6 (Update dual iterates) If for any $i = 1, \cdots, m$,

$$y_i^{(k+1)}\lambda_i^k \geq \beta_2\mu, \ \ (d_\lambda^k)_i \geq 0, \ \ \text{or} \ \ y_i^{(k+1)}\lambda_i^k \leq \beta_1\mu, \ \ (d_\lambda^k)_i \leq 0, \tag{3.33}$$

let $\lambda^{k+1} = \lambda^k$; Otherwise, select the maximum $\gamma_k \in (0, 1]$ such that $\lambda^{k+1} = \lambda^k + \gamma_k d_\lambda^k$ satisfies

$$\max\{\beta_2\mu e, Y_{k+1}\Lambda_k e\} \geq Y_{k+1}\Lambda_{k+1}e \geq \min\{\beta_1\mu e, Y_{k+1}\Lambda_k e\}; \tag{3.34}$$

Step 7 (Check the stopping criterion) If $\|F_\mu(z^{k+1}, \lambda^{k+1})\| < \epsilon_1$, or

$$\min\{y_i^k : i = 1, \cdots, m\} < \epsilon_2, \tag{3.35}$$

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

We make some comments on the algorithm. We generate new primal and dual iterates respectively by using different steplengths. We hope that the dual steplength $\gamma_k = 1$ can be accepted without being influenced by α_k. The similar strategy is used by [30, 29, 8]. (3.29) is computed such that $y^{k+1} \geq 0.005y^k$. Moreover, (3.32) is introduced by [4] and a similarly more sophisticated technique is used by [25]. Since $y^{k+1} \geq y^k + \alpha_k d_y^k$ and $\|y^{k+1} + c_{k+1}\| \leq \|y^k + \alpha_k d_y^k + c_{k+1}\|$, we have

$$\phi_\mu(z^{k+1}; \rho_{k+1}) - \phi_\mu(z^k; \rho_{k+1}) \leq \phi_\mu(z^k + \alpha_k d_z^k; \rho_{k+1}) - \phi_\mu(z^k; \rho_{k+1}). \tag{3.36}$$

Since we do not assume any regularity on the constraints, the stopping condition $\|F_\mu(z^{k+1}, \lambda^{k+1})\| < \epsilon$ may never hold, in which case the algorithm will terminate at (3.35).

4. Convergence Analysis of Algorithm 3.4

The global convergence of Algorithm 3.4 is analyzed in this section. We suppose the tolerances $\epsilon_1 = 0$, $\epsilon_2 = 0$ in Algorithm 3.4, and an infinite sequence $\{(x^k, y^k, \lambda^k)\}$ is generated. Let $z^k = (x^k, y^k)$, $d_z^k = (d_x^k, d_y^k)$ and $\tilde{d}_z^k = (\tilde{d}_x^k, \tilde{d}_y^k)$.

We need the following general assumption:

Assumption 4.1 *(1) $\{x^k\}$ is bounded; (2) There exist 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 unit matrix.*

Assumption 4.1 (1) is used by most globally convergent algorithms for nonlinear programming, including many algorithms based on the SQP approach; (2) is needed for the existence of the direction. Similar assumptions are used by all interior point methods using line search for nonlinear programming.

By Algorithm 3.4, $\{\rho_k\}$ is a monotonically nondecreasing sequence: either $\rho_{k+1} = \rho_k$ or $\rho_{k+1} \geq 2\rho_k$ for any integer $k \geq 0$. We have the following result:

Lemma 4.2 *Under Assumption 4.1, if $\rho_k = \hat{\rho}$ for all $k \geq \hat{k}$ ($\hat{k}$ is a positive integer and $\hat{\rho}$ is a positive constant), then we have: (1) $\{y^k\}$ and $\{\lambda^k\}$ are bounded above and are componentwise bounded away from zero for $k \geq 0$. The same is true for the diagonal of S_k. (2) $\{d^k\}$ is bounded, where $d^k = (d_z^k, d_\lambda^k)$.*

Proof. (1) Without loss of generality, we suppose that $\rho_k = \hat{\rho}$ for all $k \geq 0$. By Algorithm 3.4, $\phi_\mu(z^k; \hat{\rho})$ is monotonically decreasing, thus $\phi_\mu(z^k; \hat{\rho}) \leq \phi_\mu(z^0; \hat{\rho})$ for all k. We prove that y^k is bounded above by contradiction. If $\max_i\{y_i^k\} \to \infty$, since

$$f_k - \mu \sum_j \ln y_j^k + \hat{\rho}\|c_k + y^k\| \leq \phi_\mu(z^0; \hat{\rho}), \tag{4.1}$$

dividing $\max_i\{y_i^k\}$ on two sides of (4.1) and taking limit on $k \to \infty$, we have $\hat{\rho} \leq 0$ since every 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

$$-\mu \sum_{i=1}^m \ln y_i^k \leq -f_k - \hat{\rho}\|c_k + y^k\| + \phi_\mu(z_0; \hat{\rho}), \tag{4.2}$$

y^k is componentwise bounded away from zero. It follows from Step 6 of Algorithm 3.4 that λ^k is bounded above and is componentwise bounded away from zero, so is the diagonal of S_k due to $S_k = Y_k^{-1}\Lambda_k$.

(2) By Assumption 4.1 (2), $\hat{B}_k = B_k + A_k Y_k^{-1} \Lambda^k A_k^\top$ is invertible. By simple computation, (3.16)-(3.18) 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

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

is bounded, 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. d_y^k is bounded by (4.4). ∎

By Lemma 4.2, assume that $y^k \geq b_1 e$ and $\|d_y^k\| \leq b_2$ for all k, if $\hat{\alpha}_1 = \min\{1, 0.995 b_1/b_2\}$, then for all $0 \leq \alpha \leq \hat{\alpha}_1$,

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

The following result shows that the procedure for selection of steplength is well-defined.

Lemma 4.3 *Under the assumption of Lemma 4.2, there is a constant* $0 < \hat{\alpha}_2 \leq \hat{\alpha}_1$*, such that for any* $\alpha \in (0, \hat{\alpha}_2]$ *and for all* k*, there holds*

$$\phi_\mu(z^k + \alpha d_z^k; \hat{\rho}) - \phi_\mu(z^k; \hat{\rho}) \leq \alpha \sigma_0 \pi_{d_z^k}(z^k; \hat{\rho}). \tag{4.7}$$

Proof. Without loss of generality, suppose that $\rho_k = \hat{\rho}$ for all k. Then it follows from Step 4 of Algorithm 3.4 that (3.26) holds for all iterations. For $\alpha \in (0, \hat{\alpha}_1]$, by (4.6), we have

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

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

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

Since $f(x)$ and $c(x)$ are second-order continuously differentiable, there are constants $b_3 > 0$ and $b_4 > 0$ 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

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

where b_3 and b_4 are the first-order Lipschitzian constants of f and c respectively.

By the definition of $\pi_d(z; \rho)$, we have

$$\pi_{\alpha d_z^k}(z^k; \hat{\rho}) = \alpha \psi'_{d_z^k}(z^k) + \hat{\rho}(\|c(x^k) + y^k + \alpha A(x^k)^\top d_x^k + \alpha d_y^k\| - \|c(x^k) + y^k\|). \tag{4.12}$$

Then by (4.9), (4.10) and (4.11), let $b_5 = \max\{100\mu, \frac{1}{2}(b_3 + \hat{\rho} b_4)\}$, we have

$$\phi_\mu(z^k + \alpha d_z^k; \mu, \hat{\rho}) - \phi_\mu(z^k; \mu, \hat{\rho}) - \pi_{\alpha d_z^k}(z^k; \hat{\rho}) \le \alpha^2 b_5(\|d_x^k\|^2 + \|Y_k^{-1} d_y^k\|^2). \tag{4.13}$$

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

$$\pi_{\alpha d_z^k}(z^k; \hat{\rho}) - \alpha \pi_{d_z^k}(z^k; \hat{\rho}) \le 0, \tag{4.14}$$

which results in that

$$\begin{aligned} &\pi_{\alpha d_z^k}(z^k; \hat{\rho}) - \alpha\sigma_0 \pi_{d_z^k}(z^k; \hat{\rho}) \le \alpha(1 - \sigma_0)\pi_{d_z^k}(z^k; \hat{\rho}) \\ &\le -\frac{1}{2}\alpha(1 - \sigma_0)\hat{\delta}(\|d_x^k\|^2 + \|Y_k^{-1} d_y^k\|^2). \text{ (by (3.26))} \end{aligned} \tag{4.15}$$

where $\hat{\delta} = \min\{\lambda_{\min}(B_k), k = 0, 1, \cdots; b_6\}$ and $\lambda_{\min}(B_k)$ is the minimum eigenvalue of B_k (which is not smaller than ν by Assumption 4.1 (2)), $y_i^k \lambda_i^k \ge b_6$ for $i = 1, \cdots, m$ and all $k \ge 0$. Let

$$\hat{\alpha}_2 = \min\left\{\hat{\alpha}_1, \frac{(1 - \sigma_0)\hat{\delta}}{2b_5}\right\}, \tag{4.16}$$

then for all $0 \le \alpha \le \hat{\alpha}_2$, it follows from (4.13) and (4.15) that

$$\phi_\mu(z^k + \alpha d_z^k; \hat{\rho}) - \phi_\mu(z^k; \hat{\rho}) \le \alpha\sigma_0 \pi_{d_z^k}(z^k; \hat{\rho}). \tag{4.17}$$

∎

By Step 5 of Algorithm 3.4, for $\tilde{\alpha}_k = \alpha_k/\delta$, we have

$$\phi_\mu(z^k + \tilde{\alpha}_k d_z^k; \hat{\rho}) - \phi_\mu(z^k; \hat{\rho}) > \tilde{\alpha}_k \sigma_0 \pi_{d_z^k}(z^k; \hat{\rho}). \tag{4.18}$$

Thus, by (4.7), $\tilde{\alpha}_k > \hat{\alpha}_2$. Hence $\alpha_k > \delta\hat{\alpha}_2$ for all k, which implies that our line search procedure is well-defined.

Lemma 4.4 *Under Assumption 4.1, if $\rho_k = \hat{\rho}$ for all $k \ge \hat{k}$, $\{z^k\}$ and $\{\lambda^k\}$ are infinite sequences generated by Algorithm 3.4, then we have*

$$\lim_{k\to\infty} d_z^k = 0, \tag{4.19}$$

$$\lim_{k\to\infty} \|c_k + y^k\| = 0, \tag{4.20}$$

$$\lim_{k\to\infty} Y_k \Lambda_k e = \mu e, \tag{4.21}$$

$$\lim_{k\to\infty} \|g_k + A_k \lambda^k\| = 0. \tag{4.22}$$

Proof. It follows from Lemma 4.2 that sequence $\{\phi_\mu(z^k;\hat\rho)\}$ is bounded. Combining its monotonicity, the limit of $\{\phi_\mu(z^k;\hat\rho)\}$ exists as $k\to\infty$. Since $\alpha_k > \delta\hat\alpha_2 > 0$ and $\pi_{d_z^k}(z^k;\hat\rho)\le 0$ for all k, by taking limit on two-sides of (3.30), we have $\lim_{k\to\infty}\pi_{d_z^k}(z^k;\hat\rho)=0$, which implies that $\lim_{k\to\infty} d_z^k = 0$ by (3.26) and Lemma 4.2.

By (4.19) and (3.18), we have $A_k^\top \tilde d_x^k + \tilde d_y^k \to 0$. If $\tilde d_z^k$ satisfies $\varphi(\tilde d_z^k)\le \nu\varphi(0)$, then

$$\|c_k + y^k + A_k^\top \tilde d_x^k + \tilde d_y^k\| - \nu\|c_k + y^k\| \le 0, \tag{4.23}$$

which implies (4.20). Otherwise, since $\varphi(\tilde d_z^k)\le\varphi(0)$, we have

$$0 \ge -\frac{1}{2\rho_k}\left(\tilde d_x^{k\top} B_k \tilde d_x^k + \tilde d_y^{k\top} S_k \tilde d_y^k\right) \ge \|c_k + y^k + A_k^\top \tilde d_x^k + \tilde d_y^k\| - \|c_k + y^k\| \to 0. \tag{4.24}$$

It follows that $(\tilde d_x^k, \tilde d_y^k)\to 0$ by the fact that B_k and S_k are uniformly bounded and ρ_k is bounded. Thus, from Procedure 2.1 (2)(i), we have $Rr\to 0$. Since $R=\begin{pmatrix} A_k \\ I\end{pmatrix}$ is of full rank, we obtain that $r\to 0$, i.e., $\lim_{k\to\infty}\|c_k+y^k\|=0$. This proves (4.20).

It follows from (3.17) that $Y_k(\Lambda_k + D_\lambda^k)e = \mu e - \Lambda_k d_y^k$. Thus by (4.19) and Lemma 4.2, $\lim_{k\to\infty} Y_{k+1}(\Lambda_k + D_\lambda^k)e = \lim_{k\to\infty} Y_k(\Lambda_k + D_\lambda^k)e = \mu e$, by Step 6 of Algorithm 3.4, which implies that $\lambda^{k+1}=\lambda^k + d_\lambda^k$ for sufficiently large k and (4.21) holds. Moreover, for sufficiently large k, since (3.16), we have

$$g_k + A_k\lambda^{k+1} = -B_k d_x^k. \tag{4.25}$$

Thus, (4.22) follows immediately from Assumption 4.1 and $\lim_{k\to\infty} d_x^k = 0$. ■

(4.20) implies that the scaled Newton direction will be accepted at last if ρ is bounded since (3.25) is satisfied definitely after finite iterations.

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

Definition 4.5 *(1) $x^*\in\Re^n$ is called as a feasible stationary point of the problem (1.1), if $c(x^*)\le 0$ and $A_i(x^*)$, $i\in I$ are linearly dependent, where $I=\{i: c_i(x^*)=0, i=1,\cdots,m\}$; (2) $x^*\in\Re^n$ is called an infeasible stationary point of the problem (1.1), if x^* is an infeasible point to the problem (1.1) and $A(x^*)c(x^*)_+=0$, where $c(x^*)_+=\max\{c(x^*),0\}$.*

It is easy to see that both feasible and infeasible stationary points satisfy some first-order stationary properties. Similar definitions are also used by [2, 31, 20]. A feasible stationary point is also a Fritz-John point. An infeasible stationary point is also a stationary point for minimizing $\|c(x)_+\|$, which is the ℓ_2 norm of the violation of constraints. Of course, if the constraint functions are convex, then it is the "least-infeasible solution" in ℓ_2 sense. We have the following result:

Lemma 4.6 *Under Assumption 4.1, if $\rho_k \to \infty$, then:*
(1) the sequence $\{y^k\}$ is also bounded;
(2) $\{y^k\}$ is not componentwise bounded away from zero.

Proof. (1) By (3.30), we have $\phi_\mu(z^{k+1};\rho_{k+1}) \le \phi_\mu(z^k;\rho_{k+1})$ 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,

$$\begin{aligned} \frac{1}{\rho_{k+1}}\phi_\mu(z^{k+1};\rho_{k+1}) - \frac{1}{\rho_k}\phi_\mu(z^k;\rho_k) &\le \left(\frac{1}{\rho_k} - \frac{1}{\rho_{k+1}}\right)(-\psi_\mu(z^k)) \\ &\le \left(\frac{1}{\rho_k} - \frac{1}{\rho_{k+1}}\right)(b_7 + \mu m \ln \|y^k\|). \end{aligned} \tag{4.26}$$

It follows from (4.26) that

$$\frac{1}{\rho_{k+1}}\phi_\mu(z^{k+1};\rho_{k+1}) \le \frac{1}{\rho_0}\phi_\mu(z^0;\rho_0) + \left(\frac{1}{\rho_0} - \frac{1}{\rho_{k+1}}\right)\left(b_7 + \mu m \max_{0\le j\le k+1} \ln \|y^j\|\right). \tag{4.27}$$

On the other hand, we have

$$\frac{1}{\rho_{k+1}}\phi_\mu(z^{k+1};\rho_{k+1}) \ge -\frac{1}{\rho_{k+1}}\left(b_7 + \mu m \max_{0\le j\le k+1} \ln \|y^j\|\right) + \|y^{k+1}\| - \|c_{k+1}\|. \tag{4.28}$$

Thus, by (4.27) and (4.28), there is a constant $b_8 > 0$ such that

$$b_8 + \frac{\mu m}{\rho_0} \max_{0\le j\le k+1} \ln \|y^j\| \ge \|y^{k+1}\|, \ \ \forall k \ge 0, \tag{4.29}$$

which implies the desired result.

(2) If $\{y^k\}$ is bounded away from zero, then by (1) and Step 6 of Algorithm 3.4, λ^k is also bounded above and componentwise bounded away from zero. Thus S_k is uniformly bounded. It follows from Proposition 3.3 that the scaled Newton directions defined by (3.19) and (3.20) are accepted for all sufficiently large k and large ρ_k. Hence, there exists a constant $b_9 > 0$,

$$\|\tilde{d}_x^k\| \le b_9\|c_k + y^k\|, \ \ \|\tilde{d}_y^k\| \le b_9\|c_k + y^k\| \tag{4.30}$$

and

$$\|S_k \tilde{d}_y^k\| \le b_9\|c_k + y^k\|. \tag{4.31}$$

Moreover, by the proof of Proposition 3.2, we have $\Delta_k = \|c_k + y^k\|$ for sufficiently large k. Hence, by the equivalence of quadratic programming (3.13)-(3.14) and the system of equations (3.16)-(3.18), we have a constant $b_{10} > 0$ such that

$$\pi_{\tilde{d}_z^k}(z^k;\rho_k) + \frac{1}{2}(\tilde{d}_x^k)^\top B_k \tilde{d}_x^k + \frac{1}{2}(\tilde{d}_y^k)^\top S_k \tilde{d}_y^k \le b_{10}\|c_k + y^k\| - \rho_k\|c_k + y^k\|, \tag{4.32}$$

which by (3.15) implies that we have (3.26) for $\rho_k \ge \hat{\rho}$, where $\hat{\rho} > 0$ is a constant. This contradicts that $\rho_k \to \infty$. ∎

Lemma 4.6 implies that the sequence $\{\|c_k + y^k\|\}$ is bounded. A result similar to Lemma 4.6 (1) is also proved by [4]. By Lemma 4.6 (1) and Step 6 of Algorithm 3.4, λ^k is componentwise bounded away from zero, thus $(\Lambda^k)^{-1}$ and S_k^{-1} are bounded above. We have the following result:

Lemma 4.7 *Let $\mathcal{K} = \{k \mid \rho_k < \rho_{k+1}\}$. Under Assumption 4.1, if $\rho_k \to \infty$, then at any limiting point of $\{x^k, y^k\}_{k\in\mathcal{K}}$, the columns $A_i(x^*)$, $i \in J$ are linearly dependent, where $J = \{i : y_i^* = 0, i = 1, \cdots, m\}$.*

Proof. We prove it by contradiction. Suppose that there is a limiting point (x^*, y^*) of the sequence, $(x_k, y_k) \to (x^*, y^*)$ for $k \in \mathcal{K}$ and $k \to \infty$, $A_i(x^*)$, $i \in J$ are linearly independent. Then $A(x^*)^\top B^{*-1} A(x^*) + G^*$ is positive definite, where for simplicity we assume that $B_k \to B^*$ and $S_k^{-1} \to G^*$ for $k \in \mathcal{K}$ and $k \to \infty$. Thus, by the continuity of $A(x)$, there exists a constant $b_{11} > 0$ such that

$$\|A_k^\top B_k^{-1} A_k + S_k^{-1}\| \ge b_{11} \tag{4.33}$$

for large $k \in \mathcal{K}$. It follows from (3.24) that the scaled Newton directions defined by (3.19) and (3.20) are accepted. Hence

$$\|\tilde{d}_x^k\| = O(\|c_k + y^k\|), \;\; \|\tilde{d}_y^k\| = O(\|c_k + y^k\|) \tag{4.34}$$

and

$$\|S_k \tilde{d}_y^k\| = O(\|c_k + y^k\|). \tag{4.35}$$

Thus we have

$$\pi_{\tilde{d}_z^k}(z^k;\rho_k) + \frac{1}{2}(\tilde{d}_x^k)^\top B_k \tilde{d}_x^k + \frac{1}{2}(\tilde{d}_y^k)^\top S_k \tilde{d}_y^k \le O(\|c_k + y^k\|) - \rho_k\|c_k + y^k\|, \tag{4.36}$$

by (3.15), which contradicts the definition of $\mathcal{K}$. ∎

Lemma 4.8 *Under Assumption 4.1, if $\rho_k \to \infty$, then there must be a limiting point that is either feasible stationary or infeasible stationary.*

In order to prove Lemma 4.8, we need prove two lemmas at first.

Lemma 4.9 *If $(\tilde{d}_x, \tilde{d}_y)$ is a point such that $\varphi(\tilde{d}_x, \tilde{d}_y) \le \nu\varphi(0,0)$ for $0 < \nu \le 1$, then $\|\tilde{d}_x\|/\sqrt{\rho}$ and $\|Y^{-1}\tilde{d}_y\|/\sqrt{\rho}$ are uniformly bounded above.*

Proof. To obtain the result, we only need to prove that $\tilde{d}_x/\sqrt{\rho}$ and $Y^{-1}\tilde{d}_y/\sqrt{\rho}$ are bounded. Let $(d_x^{'}, d_y^{'}) = (\tilde{d}_x/\sqrt{\rho}, Y^{-1}\tilde{d}_y/\sqrt{\rho})$, then $\varphi(\tilde{d}_x, \tilde{d}_y) \le \nu\varphi(0,0)$ implies that

$$\frac{1}{2}d_x^{'\top} B d_x^{'} + \frac{1}{2}d_y^{'\top} Y\Lambda d_y^{'} + \|c + y + \sqrt{\rho}A^\top d_x^{'} + \sqrt{\rho}Y d_y^{'}\| \le \nu\|c + y\|. \tag{4.37}$$

Then the boundedness of $(d_x^{'}, d_y^{'})$ follows from the uniform lower boundedness of the quadratic terms (Assumption 4.1 and (3.34)). ∎

Lemma 4.10 *Let $\mathcal{K} = \{k \mid \rho_k < \rho_{k+1}\}$. Under Assumption 4.1, if $\rho_k \to \infty$, $\|c_k + y^k\| \ne 0$, then there exists a subsequence in $\mathcal{K}$, such that*

$$\left\|\begin{pmatrix} A_k \\ Y_k \end{pmatrix}(c_k + y^k)\right\| \to 0 \tag{4.38}$$

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

Proof. Suppose that (4.38) does not hold. Then for all sufficiently large k, there exists a constant $\tau_1 > 0$ such that

$$\left\|\begin{pmatrix} A_k \\ Y_k \end{pmatrix}(c_k + y^k)\right\| \ge \tau_1 \tag{4.39}$$

if $\|c_k + y^k\| \ne 0$. Moreover, set $\|c_k + y^k\| \ge \tau_2$ for a positive constant τ_2.

By Procedure 2.1, $\tilde{d}_z^k$ is generated such that either $\varphi_k(\tilde{d}_z^k) \le \nu\varphi_k(0)$ or $\varphi_k(\tilde{d}_z^k) \le \varphi_k(\alpha_k^C(\tilde{d}_z^k)^C)$, where $(\tilde{d}_z^k)^C$ is a scaled steepest direction, $\alpha_k^C \in (0,1]$ is selected such that $\varphi_k(\alpha_k^C(\tilde{d}_z^k)^C) \le \varphi_k(0)$. Then by Lemma 4.9, 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 for all $k \in \mathcal{K}$ such that $\varphi_k(\tilde{d}_z^k) \le \nu\varphi_k(0)$, then

$$\pi_{\tilde{d}_z^k}(z^k; \rho_k) + \frac{1}{2}(\tilde{d}_x^k)^\top B_k \tilde{d}_x^k + \frac{1}{2}(\tilde{d}_y^k)^\top S_k \tilde{d}_y^k \tag{4.40}$$

$$\le (g^k)^\top \tilde{d}_x^k - \mu e^\top Y_k^{-1}\tilde{d}_y^k - (1-\nu)\rho_k\|c_k + y^k\| \tag{4.41}$$

$$\le \tau_4\sqrt{\rho_k} - (1-\nu)\tau_2\rho_k$$

for some constant $\tau_4 > 0$, by (3.15), there exists a large $\hat{\rho} > 0$ such that (3.26) holds for all $\rho_k \geq \hat{\rho}$, which implies that there must exist some subsequence such that $\varphi_k(\tilde{d}_z^k) \leq \varphi_k(\alpha_k^C(\tilde{d}_z^k)^C)$, in this case, by Proposition 2.2, since $\eta_k \to \infty$ as y_k is not componentwise bounded away from zero, we have

$$\varphi_k(\tilde{d}_z^k) - \varphi_k(0) \leq \frac{1}{2} b_{12} \left\{ 1 - \frac{\rho_k}{\|c_k + y^k\|} \right\} \left\| \begin{pmatrix} A_k \\ Y_k \end{pmatrix} (c_k + y^k) \right\|^2, \tag{4.42}$$

where $b_{12} \leq \min\{\nu_2^{-1}, \beta_2^{-1}\mu^{-1}\}$ is a positive constant, ν_2, β_2 are defined in Assumption 4.1 and (3.34) respectively. Thus by the uniform boundedness of B_k and $Y_k \Lambda_k$ and (4.39), there are positive constants b_{13} and b_{14} such that

$$\varphi_k(\tilde{d}_z^k) - \varphi_k(0) \leq b_{13} - b_{14} \tau_1^2 \rho_k \tag{4.43}$$

for $\rho_k \geq \tau_5 \geq \|c_k + y^k\|$, which results in that

$$\pi_{\tilde{d}_z^k}(z^k; \rho_k) + \frac{1}{2} (\tilde{d}_x^k)^\top B_k \tilde{d}_x^k + \frac{1}{2} (\tilde{d}_y^k)^\top S_k \tilde{d}_y^k \tag{4.44}$$

$$\leq (g^k)^\top \tilde{d}_x^k - \mu e^\top Y_k^{-1} \tilde{d}_y^k + \varphi_k(\alpha_k^C(\tilde{d}_z^k)^C) - \varphi_k(0) \tag{4.45}$$

$$\leq \tau_4 \sqrt{\rho_k} + b_{13} - b_{14} \tau_1^2 \rho_k.$$

(4.41) and (4.45) indicate there exists a large $\hat{\rho} > 0$ such that (3.26) holds for all $\rho_k \geq \hat{\rho}$, which is a contradiction to the definition of $\mathcal{K}$. ∎

Proof of Lemma 4.8. Since x^k and y^k are bounded, without loss of generality, suppose that $(A_k, x^k, c_k, y^k, Y_k) \to (A^*, x^*, c^*, y^*, Y^*)$ as $k \in \mathcal{K}$ and $k \to \infty$, where $\mathcal{K}$ is defined in Lemma 4.10. If the limiting point (x^*, y^*) is such that $c(x^*) + y^* = 0$, then this limiting point is a feasible stationary point since $I = J$, where I and J are defined in Definition 4.5 and Lemma 4.7 respectively. Now we consider the case that $c(x^*) + y^* \neq 0$. By Lemma 4.10

$$\begin{pmatrix} A^* \\ Y^* \end{pmatrix} (c^* + y^*) = 0 \tag{4.46}$$

and so

$$y_i^* > 0 \ \Rightarrow \ c_i^* + y_i^* = 0 \ \Rightarrow \ c_i^* < 0. \tag{4.47}$$

Since $c^k + y^k \geq 0$ and $y^k \geq 0$ for all $k > 1$ by the algorithm, for each i such that $c_i^* + y_i^* \neq 0$, one has $y_i^* = 0$ by (4.46) and hence $c_i^* > 0$, implying that x^* is infeasible. Then (4.47) implies $c^* + y^* = c_+^* = \max\{c^*, 0\}$. It follows from (4.46) that $A^* c_+^* = 0$. Therefore x^* is an infeasible stationary point. The proof is finished. ∎

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

Theorem 4.11 *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 Assumption 4.1 holds.* $\{\rho_k\}$ *is the automatically updated and monotonically nondecreasing penalty parameter sequence. Then:*
(1) 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* $g(x^k) + A(x^k)\lambda^k \to 0$*;*
(2) 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 feasible stationary point or an infeasible stationary point, in which case, if* (x^k, y^k) *is asymptotically feasible for constraints (3.2), then* $\{x^k\}$ *is asymptotically feasible for constraints (1.1) and close to the boundary of constraints (1.1), at the limit the gradients of active constraints of (1.1) are linearly dependent; or if* (x^k, y^k) *is not asymptotically feasible for constraints (3.2), then at the limiting point* x^* *we have* $A^* c_+^* = 0$.

Proof. The part (1) follows from Lemma 4.4, part (2) and (3) can be derived directly by Lemma 4.7 and Lemma 4.8. ∎

5. The 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 interior point algorithm using line search 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 $\mu_0 > 0$, $\tau \in (0, 1)$ and some other parameters in Step 1 of Algorithm 3.4. Tolerance $\epsilon > 0$ and function $\theta \in \mathcal{F}$. Let $j := 0$.

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

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

$$\|Y_{k_j} \Lambda_{k_j} e - \mu_j e\| < \theta(\mu_j), \tag{5.2}$$

$$\|g(x^{k_j}) + A(x^{k_j})\lambda^{k_j}\| < \theta(\mu_j), \tag{5.3}$$

then let

$$(x^{j+1}, y^{j+1}, \lambda^{j+1}) = (x^{k_j}, y^{k_j}, \lambda^{k_j}) \tag{5.4}$$

and $\rho_{j+1} = \rho_{k_j}$, and go to Step 3; else if

$$\min\{y_i^k : i = 1, \cdots, m\} < \min\{10^{-3}, \theta(\mu_j)\}\epsilon, \tag{5.5}$$

stop; (Note: Algorithm 3.4 will not terminate until (5.1)-(5.3) or (5.5) is satisfied.)

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. It closely depends on how Algorithm 3.4 behaves for each μ_j. For any $\theta(\mu_j) > 0$, if it is in case (1) of Theorem 4.11, then Algorithm 3.4 will terminate finitely, and Algorithm 5.1 will proceed to a less μ_{j+1}.

Now we can give the theorem of the global convergence of the algorithm.

Theorem 5.2 *Suppose that f and c are twice continuously differentiable functions, $\theta \in \mathcal{F}$, $\{(x^j, y^j, \lambda^j)\}$ is a sequence generated by Algorithm 5.1. If for each barrier problem, Assumption 4.1 holds, $\{(x^k, y^k, \lambda^k)\}$ is a sequence generated by Algorithm 3.4, then for sufficiently small ϵ, Algorithm 5.1 may finitely terminate at the following two cases:*
(1) For some μ_j, Algorithm 3.4 terminates at (5.5). If the termination point is an approximately feasible point, then it is an approximately feasible stationary point. Otherwise, it is an approximately infeasible stationary point.
(2) For each μ_j, Algorithm 3.4 terminates at (5.1)-(5.3). Algorithm 5.1 terminates at Step 3, in which case the approximate KKT point of the original problem (1.1) is derived.

Proof. The results follow immediately from Theorem 4.11 and Algorithm 5.1. ∎

6. Numerical Experiment

6.1. Formulae used in Procedure 2.1 We give an implementable method of Procedure 2.1 in this part of the section.

If the scaled Newton direction is not accepted, then we compute the scaled steepest descent direction $(\tilde{d}_x^C, \tilde{d}_y^C)$, and try to get an approximate solution $(\tilde{d}_x, \tilde{d}_y)$ to (3.12) such that $\varphi(0,0) - \varphi(d_x, d_y)$ has as much more reduction as possible than $\varphi(0,0) - \varphi(\tilde{d}_x^C, \tilde{d}_y^C)$ with using the scaled Newton direction. Our method is also similar to [22].

We compute the scaled steepest descent direction by minimizing $\chi(d)^2$ with starting point $d = 0$, which results in the Cauchy point:

$$\tilde{d}^C = -\eta Q^{-1} R r, \tag{6.1}$$

where

$$\eta = \frac{r^\top (R^\top Q^{-1} R) r}{r^\top (R^\top Q^{-1} R)^2 r}. \tag{6.2}$$

Then we solve the single-variable minimizing problem:

$$\min_{\alpha \in (0,1]} \; q_1(\alpha) = \frac{1}{2}\alpha^2 \tilde{d}^{N\top} Q \tilde{d}^N + \rho \| r + \alpha R^\top \tilde{d}^N \|. \tag{6.3}$$

By direct computation, we have the solution

$$\tilde{\alpha}_1 = \min \left\{ \frac{\rho \|r\|}{r^\top (R^\top Q^{-1} R)^{-1} r}, 1 \right\}. \tag{6.4}$$

Set $d^1 = \tilde{\alpha}_1 \tilde{d}^N$, then we have $q_1(\tilde{\alpha}_1) \le \rho \|r\|$.

On the other hand, we let $d(\alpha) = \alpha \tilde{d}^N + (1-\alpha)\tilde{d}^C$, calculate $\tilde{\alpha}_2$ by

$$\min_{\alpha \in [0,1]} q_2(\alpha) = \frac{1}{2} d(\alpha)^\top Q d(\alpha) + \rho \| r + R^\top d(\alpha) \|. \tag{6.5}$$

By $q_2^{'}(\alpha) = 0$, we have

$$\alpha_2^* = \frac{\rho \| r + R^\top \tilde{d}^C \| - (\tilde{d}^N - \tilde{d}^C)^\top Q \tilde{d}^C}{(\tilde{d}^N - \tilde{d}^C)^\top Q (\tilde{d}^N - \tilde{d}^C)}. \tag{6.6}$$

If $\alpha_2^* \le 0$, $\tilde{\alpha}_2 = 0$; else if $\alpha_2^* \ge 1$, $\tilde{\alpha}_2 = 1$; else $\tilde{\alpha}_2 = \alpha_2^*$. If $q_2(\tilde{\alpha}_2) \le \rho\|r\|$, we define $d^2 = d(\tilde{\alpha}_2)$, else $d^2 = \tilde{\alpha}_3 \tilde{d}^C$, where $\tilde{\alpha}_3 \in (0,1]$ minimizes the function

$$q_3(\alpha) = \frac{1}{2}\alpha^2 (\tilde{d}^C)^\top Q \tilde{d}^C + \rho \| r + \alpha R^\top \tilde{d}^C \|. \tag{6.7}$$

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

We call it a full-Newton step if (2.11) is accepted. Similarly, we call d^1, $d(\tilde{\alpha}_2)$, and $\tilde{\alpha}_3 \tilde{d}^C$ the truncated-Newton step, the dog-leg step, and the Cauchy step respectively.

Now we specialize the above procedure to (3.12). The scaled steepest descent direction $(\tilde{d}_x^C, \tilde{d}_y^C)$ given by (6.1) are

$$\tilde{d}_x^C = -\eta B^{-1} A (c + y), \tag{6.8}$$

$$\tilde{d}_y^C = -\eta S^{-1} (c + y). \tag{6.9}$$

Thus, the process for solving (3.12) approximately is summarized into the following algorithm:

Algorithm 6.1 *(The algorithm for solving problem (3.12) approximately)*

Step 1 Compute the Newton direction $(\tilde{d}_x^N, \tilde{d}_y^N)$ by (3.19) and (3.20). If $\varphi(\tilde{d}_x^N, \tilde{d}_y^N) \leq \nu\varphi(0, 0)$, then $(\tilde{d}_x, \tilde{d}_y) = (\tilde{d}_x^N, \tilde{d}_y^N)$, stop.

Step 2 Compute the Cauchy direction $(\tilde{d}_x^C, \tilde{d}_y^C)$ by (6.8) and (6.9).

Step 3 Calculate $d^1 = \tilde{\alpha}_1(\tilde{d}_x^N, \tilde{d}_y^N)$ by (6.4), $d^2 = (d_x(\tilde{\alpha}_2), d_y(\tilde{\alpha}_2))$ by (6.5). If $q_2(\tilde{\alpha}_2) \leq \rho\|c + y\|$, go to Step 5.

Step 4 Calculate $d^2 = \tilde{\alpha}_3(\tilde{d}_x^C, \tilde{d}_y^C)$ by (6.7). If $q_1(\tilde{\alpha}_1) \leq q_3(\tilde{\alpha}_3)$, we have the approximate solution $(\tilde{d}_x, \tilde{d}_y) = d^1$; else $(\tilde{d}_x, \tilde{d}_y) = d^2$. Stop.

Step 5 If $q_1(\tilde{\alpha}_1) \leq q_2(\tilde{\alpha}_2)$, $(\tilde{d}_x, \tilde{d}_y) = d^1$; else $(\tilde{d}_x, \tilde{d}_y) = d^2$. Stop.

If $(\tilde{d}_x, \tilde{d}_y)$ is a multiple of $(\tilde{d}_x^C, \tilde{d}_y^C)$, then the condition (2) (i) of Procedure 2.1 holds obviously since B and $Y^{-1}\Lambda^{-1}$ is uniformly positive definite. Otherwise if $(\tilde{d}_x, \tilde{d}_y)$ is a multiple of $(\tilde{d}_x^N, \tilde{d}_y^N)$, since $(Rr)^\top Q^{-1}R(R^\top Q^{-1}R)^{-1}r = \|r\|^2$, $(\tilde{d}_x^N, \tilde{d}_y^N) \to 0$ implies that $(c + y) \to 0$. Thus the condition (2) (i) of Procedure 2.1 holds for the solution generated by Algorithm 6.1.

By Proposition 3.2 and Proposition 2.2, either the scaled Newton direction is accepted or the condition (1) and (2) hold.

6.2. Numerical results The algorithm is programmed in MATLAB code and run under version 6.1. Our goal is to observe how the modified approach behaves in comparing with the original interior point approach (3.21)-(3.23). 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 by [11, 30, 6].

We select the initial parameters $\mu_0 = 0.01$, $\beta_1 = 0.01$, $\beta_2 = 10$, $\sigma_0 = 0.1$, $\delta = 0.8$, B_0 is the $n \times n$ unitary matrix. The scalar in Algorithm 6.1 is set $\nu = 0.98$. The choice of the initial penalty parameter ρ_0 is scale dependent and $\rho_0 = 1$ is chosen for our experiment. Simply, $\theta(\mu) = \mu$, $\tau = 0.01$. $\epsilon = 10^{-6}$.

We derive the approximate Lagrangian Hessian B_{k+1} by the damped BFGS update formula:

$$B_{k+1} = B_k - \frac{B_k s^k (s^k)^\top B_k}{(s^k)^\top B_k s^k} + \frac{w^k (w^k)^\top}{(s^k)^\top w^k}, \tag{6.10}$$

where

$$w^k = \begin{cases} \hat{w}^k, & \text{if } (\hat{w}^k)^\top s^k \geq 0.2(s^k)^\top B_k s^k, \\ \theta_k \hat{w}^k + (1 - \theta_k) B_k s^k, & \text{otherwise,} \end{cases} \tag{6.11}$$

and $\hat{w}^k = g_{k+1} - g_k + (A_{k+1} - A_k)\lambda^{k+1}$, $s^k = x^{k+1} - x^k$, $\theta_k = 0.8(s^k)^\top B_k s^k/((s^k)^\top B_k s^k - (s^k)^\top \hat{w}^k)$.

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	-0.0000e-09	-0.9324e-09	8	full-Newton
10	2.0017	3.0067	0.0017	0	0	8	

Firstly, we test three simple examples. One is the example presented by [27, 7]:

$$\text{(TP1)} \quad \begin{aligned} \min \quad & x_1 && (6.12) \\ \text{s.t.} \ & x_1^2 - x_2 - 1 = 0, && (6.13) \\ & x_1 - x_3 - 2 = 0, && (6.14) \\ & x_2 \geq 0, \ \ x_3 \geq 0. && (6.15) \end{aligned}$$

It is easy to note that the initial point $(x_1^0, x_2^0, x_3^0) = (-4, 1, 1)$ satisfies the conditions of Theorem 1 of [27]. There is only one stationary point for this problem, which is the global minimizer. Moreover, this problem is well-posed, since at the solution the sufficient second order optimality conditions, strict complementarity and nondegeneracy hold. However, it is proved by [27] that the original interior point methods using line search 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 $\|g_k + A_k\lambda^k\| = 6.3283e{-}14$, $\|Y_k\Lambda_k e - \mu_k e\| = 2.0000e - 08$, $\|c_k + y^k\| = 0.8232e - 17$. $\hat{\rho} = 8$. 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 (6.13) and (6.14) respectively, the last column in table shows the performance of Algorithm 6.1. We also solve this example by an original approach using (3.21)-(3.23) with y^{k+1} generated by (3.32) and $y^{k+1} = y^k + \alpha_k d_y^k$ respectively, the results are presented in Tables 2 and 3.

Table 2. Numerical results by the original approach with y^{k+1} generated by (3.32) 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

Table 3. Numerical results by the original 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

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 us that the original interior point approach terminates at the infeasible points as $\mu = 0.01$. The infeasibility can not be improved by decreasing μ since x_2 and x_3 are close to the boundary of the feasible region.

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

$$\text{(TP2)} \quad x^2 + 1 \leq 0, \;\; x \leq 0. \tag{6.16}$$

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 $x^* = -1.3740e-05$, and correspondingly the slack variables $y_1^* = 3.2974e-10$ and $y_2^* = 1.3740e-05$ after 29 iterations. It is easy to note that x^* is close to a point by which the norm $\|c(x)_+\|$ is minimized. Algorithm 6.1 takes 4 full-Newton steps at first and then uses truncated-Newton steps in later 25 iterations. $\hat{\rho} = 2.4669e+07$.

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

$$\begin{aligned} & & \min\ (x_1 - 2)^2 + x_2^2 & \qquad (6.17)\\ \text{(TP3)} & & \text{s.t. } (1 - x_1)^3 - x_2 \geq 0, & \qquad (6.18)\\ & & x_1 \geq 0, \;\; x_2 \geq 0. & \qquad (6.19) \end{aligned}$$

The standard initial point $(-2, -2)$ is an infeasible point. The solution $(1, 0)$ is a feasible stationary point, but not a KKT point, at which the gradients of active constraints are dependent. This problem has not been solved by [29, 26, 24], but has been solved by [5].

Let $\sigma = \max\{1, -0.5 \min(c_{0i} : i = 1, \cdots, m)\}$, the initial slack variables and the dual variables are given by

$$y^0 = \sigma e, \;\; \lambda^0 = (\mu_0/\sigma)e \tag{6.20}$$

Algorithm 5.1 applied to problem (TP3) terminates at the singular stationary point in 34 iterations and $\mu = 0.01$. $y^* = (0, 1, 0)$, $\lambda^* = (2.9474e+07, 0.0, 2.9474e+07)$. The residuals respectively are $\|g_k + A_k\lambda^k\| = 1.2510$, $\|Y_k\Lambda_k e - \mu_k e\| = 0.0301$, $\|c_k + y^k\| = 6.8744e-10$. $\hat{\rho} = 2.3682e+07$.

We also apply our algorithm to some other test problems taken from [17], which are numbered in the same way as [17]. For example, "TP022" is problem 22 in the book. The initial points are the same as [17]. The numerical results are reported in Table 4, where $Iter$ represents the number of iterations, $RD = \|g_k + A_k\lambda^k\|$, $RP = \|c_k + y^k\|$ and $RG = \|Y_k\Lambda_k e - \mu_k e\|$, $\hat{\rho}$ is the value of penalty parameter when the algorithm terminates.

Table 4. Numerical results by Algorithm 5.1

Problem	Iter	RD	RP	RG	$\hat{\rho}$
TP001	24	5.4435e-09	0	1.0000e-08	1
TP002	23	3.8488e-12	8.9937e-17	1.0000e-08	2
TP003	16	1.9997e-09	0	1.0000e-08	1
TP004	11	1.7656e-13	8.6821e-17	2.0001e-08	4
TP010	19	2.2597e-14	3.0675e-14	1.0000e-08	1
TP011	17	1.2561e-15	4.7613e-16	1.0000e-08	4
TP012	16	3.1355e-14	5.4408e-15	1.0000e-08	1
TP020	41	9.0994e-14	0.5203e-17	5.0000e-08	675.3022
TP021	26	9.8223e-09	3.6169e-15	5.0000e-08	1
TP022	12	1.0993e-12	1.6100e-16	2.0000e-08	1
TP023	19	7.1686e-12	7.1370e-15	9.0000e-08	2
TP024	20	2.5116e-12	2.1923e-16	5.0000e-08	1
TP038	41	1.4114e-10	0	8.0000e-08	1
TP043	24	2.7478e-10	7.1491e-13	3.0000e-08	2
TP044	16	1.7296e-13	1.3698e-15	1.0000e-08	1
TP076	20	2.4892e-09	6.8720e-16	7.0000e-08	1

Summary

We presented a primal-dual interior point algorithm in this paper, which is more robust than previous primal-dual interior point methods of line-search type in the sense that its convergence requires no assumptions on the regularity of constraints and it can converge to certain stationary points when the regularity assumptions are indeed not met. We achieved the goal by using a new SQP technique. In doing this, we developed an approximation scheme for solving the piecewise quadratic subproblems. The numerical results confirm the effectiveness of the algorithm.

Acknowledgment. The authors are grateful to Professor Philippe L. Toint for reading the draft of this paper and giving insightful comments on it.

References

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

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

[3] J.V. Burke and S.P. Han, *A robust sequential quadratic programming method,* Math. Program., 43(1989), 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), 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), 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., Addison-Wesley Longman, Reading, MA, 1997.

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

[8] 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. Dipillo and F. Giannessi, eds., Kluwer Academic Publishers, 1999.

[9] J.E. Dennis, 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), 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), 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*, JOTA, 89(1996), 507-541.

[12] 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.

[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), 1132-1152.

[16] 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.X. Yuan, eds., Kluwer Academic Publishers, 1998.

[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), 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 eds., Science Press, Beijing, New York, 1999, 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), 517-534.

[21] E.O. Omojokun, *Trust region algorithms for optimization with nonlinear equality and inequality constraints,* Ph.D Dissertation, University of Colorado, 1991.

[22] M.J.D. Powell, *A hybrid method for nonlinear equations,* in Numerical Methods for Nonlinear Algebra Equations, P. Rabinowitz, ed., Gordon & Breach, London, 1970, 87-114.

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

[24] 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.

[25] P. Tseng, *A convergent infeasible interior-point trust-region method for constrained minimization,* Research report, Dept. of Math., Univ. of Washington, 1999, submitted to SIAM J. Optim.

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

[27] 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.

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

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

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

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