A null-space primal-dual interior-point algorithm for nonlinear optimization with nice convergence properties*

Xinwei Liu† and Yaxiang Yuan‡

Abstract. We present a null-space primal-dual interior-point algorithm for solving nonlinear optimization problems with general inequality and equality constraints. The algorithm approximately solves a sequence of equality constrained barrier subproblems by computing a range-space step and a null-space step in every iteration. The ℓ_2 penalty function is taken as the merit function. Under very mild conditions on range-space steps and approximate Hessians, without assuming any regularity, it is proved that either every limit point of the iterate sequence is a Karush-Kuhn-Tucker point of the barrier subproblem and the penalty parameter remains bounded, or there exists a limit point that is either an infeasible stationary point of minimizing the ℓ_2 norm of violations of constraints of the original problem, or a Fritz-John point of the original problem. In addition, we analyze the local convergence properties of the algorithm, and prove that by suitably controlling the exactness of range-space steps and selecting the barrier parameter and Hessian approximation, the algorithm generates a superlinearly or quadratically convergent step. The conditions on guaranteeing the positiveness of slack variable vector for a full step are presented.

Key words: Global and local convergences, null-space technique, primal-dual interior-point methods, nonlinear optimization with inequality and equality constraints

---

*This work was partially supported by Grants 10571039, 10231060 of National Natural Science Foundation of China, and a grant from Hebei University of Technology.

†Department of Applied Mathematics, Hebei University of Technology, China

‡State Key Laboratory Of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, The Academy of Mathematics and Systems Science, CAS

1. Introduction

We consider to solve the nonlinear program with general inequality and equality constraints

$$\min \ f(x) \tag{1.1}$$
$$\text{s.t.} \ g_i(x) \le 0, \ i \in \mathcal{I}, \tag{1.2}$$
$$h_j(x) = 0, \ j \in \mathcal{E}, \tag{1.3}$$

where $x \in \Re^n$, $\mathcal{I} = \{1, 2, \ldots, m_I\}$ and $\mathcal{E} = \{m_I + 1, m_I + 2, \ldots, m\}$ are two index sets, f, $g_i(i \in \mathcal{I})$ and $h_j(j \in \mathcal{E})$ are differentiable real functions defined on $\Re^n$. Let $m_E = m - m_I$. By introducing slack variables to the inequality constraints, problem (1.1)–(1.3) is reformulated as

$$\min \ f(x) \tag{1.4}$$
$$\text{s.t.} \ g_i(x) + y_i = 0, \ i \in \mathcal{I}, \tag{1.5}$$
$$y_i \ge 0, \ i \in \mathcal{I}, \tag{1.6}$$
$$h_j(x) = 0, \ j \in \mathcal{E}, \tag{1.7}$$

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

A point (x, y) with $y > 0$ is called an interior point of problem (1.4)–(1.7), which is not necessarily a feasible point of problem (1.1)–(1.3). Interior point methods for problem (1.1)–(1.3) start from an interior point and relax the nonnegative slack variable constraints (1.6) to the objective by introducing a parameter μ and using the logarithmic barrier terms. For given $\mu > 0$, interior point methods for problem (1.1)–(1.3) solve the logarithmic barrier problem (approximately)

$$\min \ f(x) - \mu \sum_{i \in \mathcal{I}} \ln y_i \tag{1.8}$$
$$\text{s.t.} \ g_i(x) + y_i = 0, \ i \in \mathcal{I}, \tag{1.9}$$
$$h_j(x) = 0, \ j \in \mathcal{E}. \tag{1.10}$$

For every $\mu > 0$, a Karush-Kuhn-Tucker (KKT) point of problem (1.8)–(1.10) should be an interior point (i.e., $y_i > 0, i \in \mathcal{I}$) and be feasible for constraints (1.9) and (1.10) (in this case it is strictly feasible for inequality constraints (1.2)), and satisfies the KKT conditions

$$\nabla f(x) + \sum_{i \in \mathcal{I}} u_i \nabla g_i(x) + \sum_{j \in \mathcal{E}} v_j \nabla h_j(x) = 0, \tag{1.11}$$
$$-\mu y_i^{-1} + u_i = 0, \ i \in \mathcal{I}, \tag{1.12}$$

where (1.12) can be written equivalently as

$$YUe = \mu e, \tag{1.13}$$

with $Y = \operatorname{diag}(y_1, \ldots, y_{m_I})$, $U = \operatorname{diag}(u_1, \ldots, u_{m_I})$, and $e = (1, 1, ..., 1)^T$. If ∇f, $\nabla g_i(i \in \mathcal{I})$ and $\nabla h_j(j \in \mathcal{E})$ are continuous on $\Re^n$, the limit point (if it exists) of the sequence of KKT points of problem (1.8)–(1.10) as $\mu \to 0$ is likely to be a KKT point of the original program (1.1)–(1.3). Thus, it is important that the presented interior-point methods based on using slack variables can solve the logarithmic barrier problem (1.8)–(1.10) or the system of primal-dual equations comprising of (1.11), (1.13), (1.9) and (1.10) efficiently.

Many interior point methods for nonlinear programs generate their sequence of iterates by solving the above mentioned barrier problem or system of primal-dual equations approximately (for example, see [8, 9, 13, 25, 29, 33, 35]). Some of them are based on SQP techniques, others apply Newton's methods to the primal-dual equations. Both approaches generate search directions satisfying the linearizations of constraints (1.9) and (1.10).

It has been noted by Wächter and Biegler [30] that, if the starting point is infeasible, the algorithms that use the linearized constraints of (1.9) directly may fail in finding the solution of a very simple problem, although its solution is very regular in sense that the regularity and second-order necessary conditions hold.

There are three kinds of interior point methods which are proved to be robust and have strong global convergence properties in that they will not suffer the failure and can always converge to points with some strong or weak stationarity conditions. One is based on the perturbed KKT conditions and inertia control of coefficient matrix of normal equations, such as Chen and Goldfarb [7], Forsgren and Gill [12], Gertz and Gill [14], Gould, Orban and Toint [16], Tits, Wächter, Bakhtiari, Urban and Lawrance [26], etc.. Another is to modify the right-hand-side term in the linearized constraint equations so that the derived (scaled) directions are always bounded, for example, see Byrd, Gilbert and Nocedal [3], Byrd, Hribar and Nocedal [4], Tseng [27], Liu and Sun [20]. The third kind is to combine the interior point approach with filter method proposed by Fletcher and Leyffer [11], such as Benson, Shanno and Vanderbei [1], Ulbrich, Ulbrich and Vicente [28], Wächter and Biegler [31], and so on.

Based on an SQP decomposition scheme presented by [22] and using a line search procedure, [20] presents a robust line search primal-dual interior-point method for nonlinear programming problems with general nonlinear inequality constraints. The method adopts a similar technique to the Byrd-Omojokun decomposition but using the approximate second-order information and the current value of the penalty parameter to derive the normal range direction. Without assuming any regularity of constraints, it is shown that the algorithm always converges to some points with strong or weak stationarity. Furthermore, every limit point of the sequence of iterates is a KKT point of the original problem if the sequence of penalty parameters is bounded. An apparent restriction, however, is that it has no flexibility for using exact Hessian (the second derivative information) or even any not positive definite matrices in the subproblems for generating search directions.

The robust interior-point method is further extended to solve the mathematical program with equilibrium constraints (MPEC) (see [21]), in which case some additional equality constraints are incorporated. The global convergence results need assuming the linear independence of gradients of equality constraints at all iterates.

Subspace techniques play an important role in developing efficient algorithms for nonlinearly constrained optimization (see [10, 24, 32, 36]). We present a more general framework and method which attempts to improve the primal-dual interior-point algorithm in [20] and extend to solving nonlinear programming problem with general inequality and equality constraints by using the null space technique. The algorithm approximately solves a sequence of equality constrained barrier subproblems by computing a range-space step and a null-space step in every iteration. The ℓ_2 penalty function is taken as the merit function. Under very mild conditions on range-space steps and approximate Hessians, without assuming any regularity, it is proved that every limit point of the iterate sequence is a Karush-Kuhn-Tucker point of the barrier subproblem and the penalty parameter remains bounded, or there exists a limit point that is either an infeasible stationary point of minimizing the ℓ_2 norm of violations of constraints of the original problem, or a Fritz-John point of the original problem.

In addition, we analyze the local convergence properties of the algorithm, and prove that by suitably controlling the exactness of range-space steps and selecting the barrier parameter and Hessian approximation, the algorithm generates a superlinearly or quadratically convergent step. The conditions on guaranteeing the positiveness of slack variable vector for a full step are presented.

Our algorithm solves different subproblems from those in [1, 7, 12, 14, 16, 26, 27, 28, 31], and generates new iterates only by line search procedure (which is different from those using trust region strategy in [3, 4, 27, 28]). Moreover, [3, 27] only consider the nonlinear programming with inequality constraints, [31] solves the nonlinear programming with general equality constraints and simple nonnegative constraints. Although [1, 7, 26, 31] also use line search procedure, [1, 31] select the step-sizes for primal variables by filter techniques, and [7, 26] for primal variables by different merit functions and for dual variables by different rules. It is because we use range-space and null-space strategy, we can obtain the rapid convergence of the algorithm without assuming linear independence of gradients of equality constraints.

This paper is organized as follows. In Section 2, we present some results on the range-space step and the null-space step. Some definitions on the logarithmic barrier subproblem are also given for simplicity of presentations. The conditions on the range-space step and the (inner) algorithm for problem (1.8)–(1.10) are presented in Section 3. The global convergence results on the inner algorithm are proved in Section 4. In Section 5, we describe the overall algorithm for the original problem (1.1)–(1.3), and give its global convergence results. Some local convergence properties are proved in Section 6. Lastly we conclude the paper in Section 7. The numerical results presented in [20, 21] have illustrated the robustness and effectiveness of the algorithm in special situations.

We use the following notations. The lower letters represent the vectors, the capital letters the matrices, and the capital letters in calligraphic type the set of some indices. If not specified, the capital letter corresponding to a lower letter is a diagonal matrix with all components of the vector as its diagonal entries, for example, $Y = \mathrm{diag}\,(y)$, $U = \mathrm{diag}\,(u)$. The superscript of a vector and the subscript of a matrix correspond to the current iterate. The subscript of a vector stands for the corresponding component. Throughout the paper, I is the identity matrix with appropriate size, $\mathcal{I}$ is the set of indices of inequality constraints and $\mathcal{E}$ the set of indices of equality constraints, the norm $\|\cdot\|$ is the Euclidean norm if not specified.

2. The null-space step

The algorithm in [20] generates the search direction by an auxiliary step and a modified primal-dual step. The auxiliary step plays a role in balancing the Newton decrement of constraint violations and the length of search direction, because the search direction is too "long" will result in the step-size being too small and at last close to zero, which is the reason for failure of some line search interior point methods. Such an auxiliary step can be thought as a *range-space step*. The step is then used to modify the primal-dual equation so that a *null-space step* is derived for generating a "better" new estimate of the solution.

2.1. The range-space step. For simplicity of statement, we surpass all subscripts and consider the problem

$$\min\ \psi(d) = (1/2)d^TQd + \rho\|R^Td + r\|, \tag{2.1}$$

where Q, R are matrices and r is a vector with appropriate sizes, $\rho > 0$ is a scalar. It is supposed that Q is positive definite in the null space of R^T, that is, there exists a small positive scalar β such that $d^TQd \geq \beta\|d\|^2$, $\forall d \in \{d : R^Td = 0\}$.

The global convergence results in [20] only need the approximate solution of problem (2.1) in which the objective function is reduced sufficiently. If Q is positive definite in whole space and R is of full column rank, then the Newton step and the Cauchy step in [20] can be derived by solving problems

$$\min\ (1/2)d^TQd \tag{2.2}$$

$$\text{s.t. } R^Td + r = 0, \tag{2.3}$$

and

$$\min\ (1/2)d^TQd + r^T(R^Td + r) \tag{2.4}$$

respectively.

In fact, it is easy to prove that problem (2.2)–(2.3) has a unique solution if only R is of full column rank and Q is positive definite in the null space of R^T (see Gould [15]). But the same conditions do not ensure problem (2.4) having any solution (an example is easy to find for illustrating it). Thus, in order to establish our global and local convergence results under milder and more desirable assumptions, we have to look for a novel type of approximate solutions of problem (2.1) and try to give a basic measure on the approximate solutions.

The following result plays an important role in our global convergence analysis.

Lemma 2.1 *Assume $P^2Rr \neq 0$, where P is a diagonal matrix with the same size as Q. Let $d^c = -P^2Rr$, $\alpha^c = argmin_{\alpha\in[0,1]}\|r + \alpha R^Td^c\|$. Then*

$$\|r\| - \|r + \alpha^cR^Td^c\| \geq (1/2)\min[1, \eta]\,\|PRr\|^2/\|r\|, \tag{2.5}$$

where $\eta = \|PRr\|^2/\|R^TP^2Rr\|^2$*. Moreover, if there exists a constant* $\nu > 0$ *such that* $\|P^TQP\| \leq \nu$*, then*

$$\psi(\alpha^c d^c) - \psi(0) \leq (1/2)\left\{\nu - \rho \min\left[1/\|r\|, \eta/\|r\|\right]\right\} \|PRr\|^2. \tag{2.6}$$

Proof. We firstly prove the inequality (2.5). Since

$$\|r\|^2 - \|r + \alpha R^T d^c\|^2 = 2\alpha r^T R^T P^2 R r - \alpha^2 r^T (R^T P^2 R)^2 r, \tag{2.7}$$

we have

$$\alpha^c = \min\{1, \eta\}, \tag{2.8}$$

where $\eta = \|PRr\|^2/\|R^TP^2Rr\|^2$. Moreover, by (2.7) and (2.8), if $\eta > 1$, then $\alpha^c = 1$, and

$$\|r\|^2 - \|r + \alpha^c R^T d^c\|^2 > \|PRr\|^2; \tag{2.9}$$

otherwise, $\alpha^c = \eta$, and

$$\|r\|^2 - \|r + \alpha^c R^T d^c\|^2 = \eta \|PRr\|^2. \tag{2.10}$$

Thus,

$$\|r\|^2 - \|r + \alpha^c R^T d^c\|^2 \geq \min[1, \eta] \|PRr\|^2. \tag{2.11}$$

Since $\|r\|^2 - \|r + \alpha^c R^T d^c\|^2 = (\|r\| + \|r + \alpha^c R^T d^c\|)(\|r\| - \|r + \alpha^c R^T d^c\|)$ and the inequality $\|r\| + \|r + \alpha^c R^T d^c\| \leq 2\|r\|$, by (2.11),

$$\begin{aligned} \|r\| - \|r + \alpha^c R^T d^c\| &\geq \min[1, \eta] \|PRr\|^2 / (\|r\| + \|r + \alpha^c R^T d^c\|) \\ &\geq [1/(2\|r\|)] \min[1, \eta] \|PRr\|^2, \end{aligned}$$

that is, (2.5) follows.

Now, (2.6) follows from (2.5) and the fact that

$$(1/2)(\alpha^c)^2 d^{cT} Q d^c \leq (1/2)\|P^TQP\|\|PRr\|^2. \tag{2.12}$$

This completes our proof. □

2.2. The null-space step. Let d_p be an approximate solution to (2.1). For any vector q with the same number of rows as matrices Q and R, consider the system of equations

$$\begin{aligned} Qd + R\lambda &= -q, & (2.13)\\ R^T d &= R^T d_p, & (2.14) \end{aligned}$$

which can be reformulated as

$$\begin{bmatrix} Q & R \\ R^T & 0 \end{bmatrix} \begin{bmatrix} d - d_p \\ \lambda \end{bmatrix} = \begin{bmatrix} -q - Qd_p \\ 0 \end{bmatrix}. \tag{2.15}$$

Correspondingly, by deleting the linearly dependent columns in R, we have the system

$$\begin{bmatrix} Q & R_{\mathcal{L}} \\ R_{\mathcal{L}}^T & 0 \end{bmatrix} \begin{bmatrix} d - d_p \\ \lambda_{\mathcal{L}} \end{bmatrix} = \begin{bmatrix} -q - Qd_p \\ 0 \end{bmatrix}, \tag{2.16}$$

where $R_{\mathcal{L}}$ consists of the largest number of linear independent columns of R, $\mathcal{L}$ is the set of subscripts of columns of $R_{\mathcal{L}}$.

If Q is positive definite in the null space of R^T, by applying the results in Gould [15], we know that equation (2.16) always has a unique solution. Again by Gould [15], $[d^{*T}\ \lambda^{*T}]^T$ solves equation (2.15) if and only if $[d^{*T}\ \lambda_{\mathcal{L}}^{*T}]^T$ is a solution of equation (2.16), where $\lambda_{\mathcal{L}}^* \in \Re^{|\mathcal{L}|}$ consists of components of λ^* with subscripts in $\mathcal{L}$, $|\mathcal{L}|$ is the cardinality of $\mathcal{L}$. Moreover, let $\hat{\lambda}_j^* = \lambda_i^*$ if $j \in \mathcal{L}$ is the ith index in $\mathcal{L}$, and otherwise $\hat{\lambda}_j^* = 0$, then $[d^{*T}\ \hat{\lambda^*}^T]^T$ is a solution of equation (2.15).

The next lemma is useful in proving our global convergence results.

Lemma 2.2 *Assume that Q is positive definite in the null space of R^T. Let $[d^{*T}\ \lambda^{*T}]^T$ be a solution to the problem (2.15). Then*

$$q^T d^* + (1/2){d^*}^T Q d^* \le q^T d_p + (1/2){d_p}^T Q d_p. \tag{2.17}$$

Proof. If Q is positive definite in the null space of R^T, equation (2.15) is equivalent to the quadratic programming problem

$$\min\ (1/2)d^T Q d + q^T d \tag{2.18}$$
$$\text{s.t. } R^T d = R^T d_p. \tag{2.19}$$

Since $[d^{*T}\ \lambda^{*T}]^T$ is a solution to the problem (2.15), d^* also solves problem (2.18)–(2.19). Because d_p is feasible, we have (2.17) immediately. □

The following lemma indicates, although λ^* is dependent on the conditioning of R, under suitable conditions on Q, $R\lambda^*$ is independent on R in some sense.

Lemma 2.3 *Assume that Q is positive definite in the null space of R^T, P is a nonsingular matrix with the same size as Q such that $\|P^TQP\|$ is bounded above and $d^T(P^TQP)d \ge \beta\|d\|^2$ for all $d \in \{d : (P^TR)^Td = 0\}$ with a constant $\beta > 0$. If both $\|P^{-1}d_p\|$ and $\|P^Tq\|$ are bounded, that is, there exits a constant ν independent of all variants such that $\|P^{-1}d_p\| \le \nu$ and $\|P^Tq\| \le \nu$, then $\|P^{-1}d^*\|$ and $\|P^TR\lambda^*\|$ are bounded, where $[d^{*T}\ \lambda^{*T}]^T$ is a solution of equation (2.15).*

Proof. Let $\hat{d} = P^{-1}(d - d_p)$. Problem (2.18)–(2.19) can be rewritten as

$$\min\ (1/2)(P\hat{d} + d_p)^T Q(P\hat{d} + d_p) + q^T(P\hat{d} + d_p) \tag{2.20}$$
$$\text{s.t. } (P^TR)^T\hat{d} = 0. \tag{2.21}$$

If $\hat{d}^* = P^{-1}(d^* - d_p)$, then $\hat{d}^*$ is in the null space of $(P^T R)^T$ and

$$(1/2)(P\hat{d}^* + d_p)^T Q(P\hat{d}^* + d_p) + q^T(P\hat{d}^* + d_p) \le (1/2)d_p^T Q d_p + q^T d_p. \tag{2.22}$$

That is,

$$(1/2)\hat{d^*}^T(P^T QP)\hat{d}^* + (P^{-1}d_p)^T(P^T QP)\hat{d}^* + (P^T q)^T\hat{d}^* \le 0. \tag{2.23}$$

If $\|P^{-1}d^*\|$ is large enough, since $\|P^{-1}d_p\|$ is bounded, $\|\hat{d}^*\|$ will be sufficiently large. By dividing $\|\hat{d}^*\|^2$ on the two-sides of (2.23), and letting $w^* = \hat{d}^*/\|\hat{d}^*\|$, we have that $\|w^*\| = 1$ and ${w^*}^T(P^T QP)w^*$ tends to be non-positive since $\|P^T QP\|$, $\|P^{-1}d_p\|$, and $\|P^T q\|$ are bounded, which contradicts that $P^T QP$ is positive definite in the null space of $(P^T R)^T$.

It follows from (2.15) that

$$Qd^* + R\lambda^* = -q. \tag{2.24}$$

Thus,

$$(P^T QP)(P^{-1}d^*) + P^T R\lambda^* = -P^T q, \tag{2.25}$$

that is,

$$P^T R\lambda^* = -P^T q - (P^T QP)(P^{-1}d^*). \tag{2.26}$$

Hence, $\|P^T R\lambda^*\|$ is bounded since $\|P^T q\|$, $\|P^{-1}d^*\|$ and $\|P^T QP\|$ are bounded above. □

Lemma 2.3 plays an important role in convergence analysis of our algorithm, and makes our algorithm obviously different from some interior point methods based on penalty and inertia control such as Chen and Goldfarb [7], Forsgren and Gill [12], Gertz and Gill [14], Gould, Orban, and Toint [16]. We note that the algorithm has many similarities with the trust region interior-point method proposed by Byrd, Gilbert and Nocedal [3] and Byrd, Hribar and Nocedal [4].

2.3. On the logarithmic barrier problem. For problem (1.8)–(1.10), we set

$$Q = \begin{bmatrix} B & \\ & Y^{-1}U \end{bmatrix},\ R = \begin{bmatrix} \nabla g & \nabla h \\ I & \end{bmatrix},\ P = \begin{bmatrix} I & \\ & Y \end{bmatrix},$$
$$r = \begin{bmatrix} g+y \\ h \end{bmatrix},\ q = \begin{bmatrix} \nabla f \\ -\mu Y^{-1}e \end{bmatrix},\ d = \begin{bmatrix} d_x \\ d_y \end{bmatrix},$$

where $B \in \Re^{n\times n}$, $Y = \mathrm{diag}\,(y)$, $U = \mathrm{diag}\,(u)$, $\nabla g = \nabla g(x)$, $\nabla h = \nabla h(x)$, $g = g(x)$, $h = h(x)$, $\nabla f = \nabla f(x)$, e is the all-one vector and $I = \mathrm{diag}\,(e)$. Then

$$R_{\mathcal{L}} = \begin{bmatrix} \nabla g & \nabla h_{\mathcal{J}} \\ I & \end{bmatrix},$$

where $\mathcal{J}$ is an index set which consists of subscripts of the largest number of linearly independent column vectors in ∇h, $\nabla h_{\mathcal{J}}$ and $h_{\mathcal{J}}$ respectively consist of $\nabla h_j(x)$ and $h_j(x)$ with $j \in \mathcal{J}$.

It is easy to note that $R_{\mathcal{L}}$ has full column rank. If $x \in \Re^n$ is in a bounded set $X \subset \Re^n$, and if g and h are continuous on the set, $\|R\|$, $\|R_{\mathcal{L}}\|$, $\|r\|$, $\|P^T q\|$ will be bounded on X. Furthermore, if $\|B\|$ is bounded, $y \in \Re^{m_I}$ is componentwise bounded away from zero and $u \in \Re^{m_I}$ is bounded, $\|Q\|$ will be bounded on X. Otherwise, if some $y_i (i \in \mathcal{I})$ is very close to zero, then $\|Q\|$, $\|q\|$ may tend to infinity. Moreover,

$$P^T Q P = \begin{bmatrix} B & \\ & YU \end{bmatrix}.$$

Thus, $\|P^T Q P\|$ is bounded if and only if $\|B\|$ and $\|YU\|$ are bounded.

Let Z_h be a matrix consisting of columns which are basis vectors of null space of ∇h^T, that is, $\nabla h^T Z_h = 0$. We have the following result.

Lemma 2.4 *Matrices Q and $P^T Q P$ are positive definite respectively in the null space of R^T and $(P^T R)^T$ if and only if*

$$Z_h^T (B + \nabla g Y^{-1} U \nabla g^T) Z_h \succ 0.$$

Proof. Since

$$\begin{bmatrix} \nabla g & \nabla h \\ I & \end{bmatrix}^T \begin{bmatrix} Z_h \\ -\nabla g^T Z_h \end{bmatrix} = 0,$$

$\bar{d} \in \{d : R^T d = 0\}$ and $\bar{d} \neq 0$ if and only if there exists a vector $\bar{s} \neq 0$ such that

$$\bar{d} = \begin{bmatrix} Z_h \\ -\nabla g^T Z_h \end{bmatrix} \bar{s}.$$

Similarly, since

$$\begin{bmatrix} \nabla g & \nabla h \\ Y & \end{bmatrix}^T \begin{bmatrix} Z_h \\ -Y^{-1} \nabla g^T Z_h \end{bmatrix} = 0,$$

$\tilde{d} \in \{d : (P^T R)^T d = 0\}$ and $\tilde{d} \neq 0$ imply that there is a nonzero $\tilde{s}$ such that

$$\tilde{d} = \begin{bmatrix} Z_h \\ -Y^{-1} \nabla g^T Z_h \end{bmatrix} \tilde{s}.$$

By the fact that

$$\bar{d}^T Q \bar{d} = \bar{s}^T Z_h^T (B + \nabla g Y^{-1} U \nabla g^T) Z_h \bar{s}, \tag{2.27}$$

$$\tilde{d}^T P^T Q P \tilde{d} = \tilde{s}^T Z_h^T (B + \nabla g Y^{-1} U \nabla g^T) Z_h \tilde{s}, \tag{2.28}$$

our lemma follows easily. □

3. The inner algorithm

The inner algorithm is presented for solving problem (1.8)–(1.10) for some given $\mu > 0$.

Define the merit function as

$$\phi(x,y;\rho) = f(x) - \mu \sum_{i\in\mathcal{I}} \ln y_i + \rho\|(g(x)+y, h(x))\|,$$

which is a logarithmic barrier penalty function, where $\rho > 0$ is a penalty parameter and $\mu > 0$ is a barrier parameter.

Suppose the current iteration point is (x^k, y^k, u^k, v^k) with $(y^k, u^k) > 0$. We compute a range-space step d_p^k which satisfies the following conditions:

$$(1)\ \ \|r^k\| - \|r^k + R_k^T d_p^k\| \geq \kappa_1 \|P_k R_k r^k\|^2/\|r^k\|; \tag{3.1}$$

$$(2)\ \ \|P_k^{-1} d_p^k\| \leq \kappa_2 \|R_k r^k\|, \tag{3.2}$$

where κ_1, κ_2 are positive constants, P_k, R_k, r^k are values of P, R, r at (x^k, y^k). The existence of d_p^k is guaranteed by Lemma 2.1.

The search direction d^k is obtained by solving equation (2.15). Let $z^k = (x^k, y^k)$. We then need to determine a step-size α_k such that

$$y^k + \alpha_k d_y^k \geq \xi y^k, \tag{3.3}$$

and $\phi(z^k + \alpha_k d^k; \rho_{k+1})$ is sufficiently descent comparing with $\phi(z^k; \rho_{k+1})$, that is,

$$\phi(z^k + \alpha_k d^k; \rho_{k+1}) - \phi(z^k; \rho_{k+1}) \leq \sigma \alpha_k \pi_k(d^k; \rho_{k+1}), \tag{3.4}$$

where $\xi \in (0,1)$, $\sigma \in (0, 1/2)$ are constants,

$$\pi_k(d^k; \rho_{k+1}) = {q^k}^T d^k + \rho_{k+1}(\|r^k + R_k^T d^k\| - \|r^k\|). \tag{3.5}$$

The selection on ρ_{k+1} is such that $\rho_{k+1} \geq \rho_k$ and

$$\pi_k(d^k; \rho_{k+1}) \leq (1-\tau)\rho_{k+1}(\|r^k + R_k^T d^k\| - \|r^k\|) - (1/2)(d^k - d_p^k)^T Q_k (d^k - d_p^k), \tag{3.6}$$

unless $P_k R_k r^k = 0$ and $r^k \neq 0$ (where $\tau \in (0,1)$). One advisable scheme is to select

$$\rho_{k+1} = \max\left\{(1/\tau)({q^k}^T d^k + (1/2)(d^k - d_p^k)^T Q_k (d^k - d_p^k))/(\|r^k\| - \|r^k + R_k^T d^k\|), 2\rho_k\right\}$$

if $P_k R_k r^k \neq 0$ and (3.6) does not hold, otherwise set $\rho_{k+1} = \rho_k$. Because Q_k may be symmetric indefinite, we have no freedom to use the same strategy as [20, 22] to update the penalty parameter. Due to this and without assuming the full rank of ∇h^k, our global analysis in next section is not trivial and is different from those of [20, 22].

We are now ready to present the algorithmic framework, which shares many similarities with those in [3, 4, 20, 22]. The main differences are in methods for deriving d_p^k and updating ρ_k and u^k, and terminating conditions.

Algorithm 3.1 *(The algorithmic framework for the barrier problem (1.8)-(1.10))*

Step 1 *Given* $(x^0, y^0, u^0, v^0) \in \Re^n \times \Re^{m_I}_{++} \times \Re^{m_I}_{++} \times \Re^{m_E}$, *constants* $\kappa_1, \kappa_2, \xi, \sigma, \tau$, *and* $\gamma_1, \gamma_2 (0 < \gamma_1 < 1 < \gamma_2)$, *initial penalty parameter* $\rho_0 > 0$. *Let* $k := 0$;

Step 2 *If* $P_k R_k r^k = 0$ *but* $r^k \neq 0$, *stop;*

Step 3 *Calculate* d_p^k *satisfying (3.1)–(3.2), and obtain* d^k *and* λ^k *by (2.15) (or (2.16));*

Step 4 *Update* ρ_k *to* ρ_{k+1} *such that (3.6) holds;*

Step 5 *Compute the maximal step-size* $\alpha_k \in (0, 1]$ *such that (3.3)–(3.4) hold. Set*

$$x^{k+1} = x^k + \alpha_k d_x^k, \quad y^{k+1} = \max\{y^k + \alpha_k d_y^k, -g(x^k + \alpha_k d_x^k)\}, \tag{3.7}$$

where d_x^k *and* d_y^k *consist of the first* n *and the remanent* m_I *elements of* d^k *respectively;*

Step 6 *Let*

$$d_u^k = \lambda_1^k - u^k, \ d_v^k = \lambda_2^k - v^k$$

with λ_1^k *and* λ_2^k *consisting of the first* m_I *and the remanent* $(m - m_I)$ *elements of* λ^k *respectively. For* $i = 1, \ldots, m_I$, *if* $y_i^{k+1} u_i^k \geq \gamma_2 \mu$ *and* $(d_u^k)_i \geq 0$, *set* $u_i^{k+1} = \gamma_2 \mu / y_i^{k+1}$; *else if* $y_i^{k+1} u_i^k \leq \gamma_1 \mu$ *and* $(d_u^k)_i \leq 0$, *set* $u_i^{k+1} = \gamma_1 \mu / y_i^{k+1}$; *otherwise, select* β_{ki} *(if possible, select* β_{ki} *being the maximum in* $(0, 1]$*) such that*

$$y_i^{k+1}(u_i^k + \beta_{ki}(d_u^k)_i) \in [\gamma_1 \mu, \gamma_2 \mu], \tag{3.8}$$

and set $u_i^{k+1} = u_i^k + \beta_{ki}(d_u^k)_i$. *Set* $v^{k+1} = v^k + d_v^k$;

Step 7 *Terminate the algorithm if* (x^{k+1}, y^{k+1}) *is a KKT point of problem (1.8)–(1.10); else update the related data, let* $k := k + 1$ *and go to Step 2.*

Fletcher and Johnson [10] have considered the numerical stability of solving the null space equation (2.16).

It follows from (3.4) and (3.7) that

$$\begin{aligned} \phi(z^{k+1}; \rho_{k+1}) - \phi(z^k; \rho_{k+1}) &\leq \phi(z^k + \alpha_k d^k; \rho_{k+1}) - \phi(z^k; \rho_{k+1}) \\ &\leq \sigma \alpha_k \pi_k(d^k; \rho_{k+1}). \end{aligned} \tag{3.9}$$

Since $y_i^k (d_u^k)_i + u_i^k (d_y^k)_i = \mu - y_i^k u_i^k$, we have

$$(d_u^k)_i = [\alpha_k \mu + (1 - \alpha_k) y_i^k u_i^k - y_i^{k+1} u_i^k] / (\alpha_k y_i^k).$$

It shows that, if $y_i^k u_i^k \in [\gamma_1 \mu, \gamma_2 \mu]$, then $(d_u^k)_i > 0$ if $y_i^{k+1} u_i^k < \gamma_1 \mu$ and $(d_u^k)_i < 0$ if $y_i^{k+1} u_i^k > \gamma_2 \mu$. Thus, in Step 6, there always is a $\beta_{ki} > 0$ (which may not be in $(0, 1]$) such that (3.8) holds. Accordingly, for every $i = 1, \ldots, m_I$,

$$y_i^{k+1} u_i^{k+1} \in [\gamma_1 \mu, \gamma_2 \mu]. \tag{3.10}$$

If the algorithm terminates at iterate k, then either (x^k, y^k) is a KKT point of problem (1.8)–(1.10) by Step 7 (in which case x^k is an approximate KKT point of the original problem

(1.1)–(1.3) as μ is small enough), or $P_kR_kr^k = 0$ but $r^k \neq 0$ by Step 2. In latter case, x^k is infeasible for the original problem, and by expressions in Section 2.3, we have

$$\begin{cases} \nabla g^k(g^k + y^k) + \nabla h^k h^k = 0, \\ Y_k(g^k + y^k) = 0. \end{cases}$$

Since $y_i^k > 0$ for every $i \in \mathcal{I}$, we have that $g_i^k = -y_i^k < 0$ and $\sum_{j\in\mathcal{E}} \nabla h_j^k h_j^k = 0$. Thus, x^k is a stationary point for minimizing violations of equality constraints under strict feasibility for inequality constraints.

It is easy to notice that the algorithm with the current stopping conditions may not terminate in a finite number of iterations. We study this situation in the next section.

4. The global convergence

We now suppose that Algorithm 3.1 does not terminate in a finite number of iterations. An infinite sequence $\{(x^k, y^k, u^k, v^k)\}$ is then generated. Comparing to that of [20], we use the following milder assumptions for global convergence analysis.

Assumption 4.1

(1) Functions f, g and h are twice continuously differentiable on $\Re^n$;

(2) $\{x^k\}$ is in an open bounded set;

(3) $\{B_k\}$ is bounded. Moreover, for every k,

$$d_x^T B_k d_x + \sum_{i\in\mathcal{I}_k^0} 1/(\gamma_2\mu)(u_i^k \nabla g_i^{k^T} d_x)^2 \geq \beta\|d_x\|^2, \quad \forall d_x \neq 0 : \nabla h_j^{k^T} d_x = 0, j \in \mathcal{E}, \quad (4.1)$$

where $\mathcal{I}_k^0 = \{i \in \mathcal{I} : g_i^k = 0\}$, u^k is the k-th multiplier estimate corresponding to (1.9), $\beta > 0$ is a constant;

(4) $\{d_p^k\}$ satisfies conditions (3.1)–(3.2).

Since for every $\bar{d} \neq 0$ in the null space of R_k^T there holds

$$\begin{aligned} \bar{d}^T Q_k \bar{d} &= d_x^T B_k d_x + \sum_{i\in\mathcal{I}} (y_i^k u_i^k)^{-1} (u_i^k \nabla g_i^{k^T} d_x)^2 \\ &\geq d_x^T B_k d_x + \sum_{i\in\mathcal{I}_k^0} 1/(\gamma_2\mu)(u_i^k \nabla g_i^{k^T} d_x)^2, \end{aligned} \quad (4.2)$$

where $d_x \neq 0$ is a vector in $\Re^n$ such that $\nabla h_j^{k^T} d_x = 0, \forall j \in \mathcal{E}$. By Lemma 2.4, (4.1) implies that Q_k and $P_k^T Q_k P_k$ are positive definite respectively in the null space of R_k^T and $(P_k^T R_k)^T$, and

$$\begin{cases} \bar{d}^T Q_k \bar{d} \geq \beta\|\bar{d}\|^2, & \forall \bar{d} \in \{d : R_k^T d = 0\}; \\ \tilde{d}^T (P_k^T Q_k P_k)\tilde{d} \geq \beta\|\tilde{d}\|^2, & \forall \tilde{d} \in \{d : (P_k^T R_k)^T d = 0\}. \end{cases} \quad (4.3)$$

Locally, if there holds

$$d_x^T B_k d_x \geq \beta \|d_x\|^2, \quad \forall d_x \neq 0 : \nabla h_j^{k^T} d_x = 0, j \in \mathcal{E}; u_i^k \nabla g_i^{k^T} d_x = 0, i \in \mathcal{I}_k^0,$$

which is analogous to the second-order sufficient conditions of problem (1.1)–(1.3), we see that (4.1) holds for all sufficiently small μ.

If $\{B_k\}$ is bounded, then, by (3.10), there is a constant $\nu_0 > 0$ such that $\|P_k^T Q_k P_k\| \leq \nu_0$ for all k.

Recently, Griva, Shanno and Vanderbei [17] proved the global convergence of their interior-point method under some standard assumptions. Their assumptions are attractive since they are only related to the problem itself and are independent of any iterations.

The following result can be obtained from Lemma 2.3.

Lemma 4.2 *Suppose that Assumption 4.1 holds. Sequences* $\{P_k^{-1} d^k\}$ *and* $\{P_k^T R_k \lambda^k\}$ *are bounded.*

Proof. If Assumption 4.1 holds, then the conditions in Lemma 2.3 are satisfied. Then, by Lemma 2.3, $\{P_k^{-1} d^k\}$ and $\{P_k^T R_k \lambda^k\}$ are bounded. □

The next results depend only on the merit function and items (1) and (2) of Assumption 4.1.

Lemma 4.3 *Suppose that Assumption 4.1 holds. Then* $\{y^k\}$ *is bounded,* $\{u^k\}$ *is componentwise bounded away from zero. Furthermore, if* ρ_k *remains constant for all sufficiently large* k*, then* $\{y^k\}$ *is componentwise bounded away from zero, and* $\{u^k\}$ *is bounded above.*

Proof. The boundedness of $\{y^k\}$ follows from proofs in [3, 20].

Let $\hat{y}^0 = y^0$, $\hat{y}^{k+1} = y^k + \alpha_k d_y^k$, and $\hat{z}^k = (x^k, \hat{y}^k)$. If there exists a constant $\hat{\rho}$ such that $\rho_k = \hat{\rho}$ for all sufficiently large k, then, by (3.4) and (3.9),

$$\phi(z^{k+1}; \hat{\rho}) \leq \phi(\hat{z}^{k+1}; \hat{\rho}) \leq \phi(z^k; \hat{\rho}). \tag{4.4}$$

Thus, $\{\phi(z^k; \hat{\rho})\}$ is monotonically decreasing. Therefore, by the boundedness of $\{y^k\}$, (1) and (2) of Assumption 4.1, we can deduce that $\{y^k\}$ is componentwise bounded away from zero.

By (3.10), $\{u^k\}$ is componentwise bounded away from zero if $\{y^k\}$ is bounded and is bounded above if $\{y^k\}$ is componentwise bounded away from zero. □

Since there are equality constraints, $\{\rho_k\}$ may not be bounded even if $\{y^k\}$ is componentwise bounded away from zero, which is different from the situation that there are only inequality constraints (see Lemma 4.7 (ii) of [20]) or there exist equality constraints with full rank Jacobian at all iterates (see Lemma 5 of [21]).

Lemma 4.4 *Suppose that Assumption 4.1 holds. If there is a constant* $\eta_0 > 0$ *such that*

$$\|P_k R_k r^k\| / \|r^k\| \geq \eta_0, \tag{4.5}$$

for all k*, then* $\{\rho_k\}$ *is bounded.*

Proof. To prove that $\{\rho_k\}$ is bounded, it is sufficient if we show that there exists a constant $\hat{\rho}$ such that (3.6) holds with $\rho_{k+1} = \hat{\rho}$ for every k.

By (3.1) and (4.5),

$$\|r^k\| - \|r^k + R_k^T d_p^k\| \geq \kappa_1 \eta_0^2 \|r^k\|. \tag{4.6}$$

It follows from equation (2.15) that

$$q^{k^T} d_p^k + d^{k^T} Q_k d_p^k = -\lambda^{k^T} R_k^T d_p^k. \tag{4.7}$$

Hence,

$$\begin{aligned}
&(1/2)(d^k - d_p^k)^T Q_k (d^k - d_p^k) \\
&= (1/2) d^{k^T} Q_k d^k - d^{k^T} Q_k d_p^k + (1/2) d_p^{k^T} Q_k d_p^k \\
&= (1/2) d^{k^T} Q_k d^k + q^{k^T} d_p^k + \lambda^{k^T} R_k^T d_p^k + (1/2) d_p^{k^T} Q_k d_p^k.
\end{aligned} \tag{4.8}$$

Since

$$\begin{aligned}
&\pi_k(d^k; \hat{\rho}) - (1-\tau)\hat{\rho}(\|r^k + R_k^T d^k\| - \|r^k\|) + (1/2)(d^k - d_p^k)^T Q_k (d^k - d_p^k) \\
&= q^{k^T} d^k + (1/2) d^{k^T} Q_k d^k + \tau\hat{\rho}(\|r^k + R_k^T d^k\| - \|r^k\|) \\
&\quad + q^{k^T} d_p^k + \lambda^{k^T} R_k^T d_p^k + (1/2) d_p^{k^T} Q_k d_p^k \\
&\leq q^{k^T} d_p^k + (1/2) d_p^{k^T} Q_k d_p^k + \tau\hat{\rho}(\|r^k + R_k^T d_p^k\| - \|r^k\|) \\
&\quad + q^{k^T} d_p^k + \lambda^{k^T} R_k^T d_p^k + (1/2) d_p^{k^T} Q_k d_p^k \\
&\leq 2(P_k^T q^k)^T (P_k^{-1} d_p^k) + (P_k^{-1} d_p^k)^T (P_k^T Q_k P_k)(P_k^{-1} d_p^k) + (P_k^T R_k \lambda^k)^T (P_k^{-1} d_p^k) \\
&\quad + \tau\hat{\rho}(\|r^k + R_k^T d_p^k\| - \|r^k\|) \\
&\leq \left\{\kappa_2(2\|P_k^T q^k\| + \|P_k^T R_k \lambda^k\| + \|P_k^T Q_k P_k\| \|P_k^{-1} d_p^k\|)\|R_k\| - \tau\hat{\rho}\kappa_1\eta_0^2\right\} \|r^k\|,
\end{aligned} \tag{4.9}$$

where the equality is obtained by (3.5) and (4.8), the first inequality is derived from (2.15) and (2.17), and the third from (3.2) and (4.6). By Assumption 4.1, Lemma 4.2 and (3.2), $\|P_k^T q_k\|$, $\|P_k^T R_k \lambda^k\|$, $\|P_k^{-1} d_p^k\|$, and $\|R_k\|$ are bounded. The boundedness of $\|P_k^T Q_k P_k\|$ follows from boundednesses of $\|B_k\|$ and (3.10). Thus, we can have the desired result immediately. $\square$

The above lemma implies two corollaries:

Corollary 4.5 *Suppose that Assumption 4.1 holds. If $\{y^k\}$ is componentwise bounded away from zero, $\{\nabla h_j^k, j \in \mathcal{E}\}$ are linearly independent for all k, then $\{\rho_k\}$ is bounded.*

Corollary 4.6 *Suppose that Assumption 4.1 holds. If ρ_k tends to infinity as $k \to \infty$, then there exists a subset $\mathcal{K} \in \mathcal{N} = \{1, 2, \dots\}$ such that*

$$\lim_{k \in \mathcal{K}, k \to \infty} \|P_k R_k r^k\| / \|r^k\| = 0. \tag{4.10}$$

The first corollary shows that, under suitable assumptions on slack variable vectors and regularity of equality constraints, ρ_k will remain constant for all sufficiently large k.

Lemma 4.7 *Suppose that Assumption 4.1 holds. If (4.5) is satisfied for some constant $\eta_0 > 0$, then $\{d^k\}$ is bounded. Thereby, $\{\alpha_k\}$ is bounded away from zero.*

Proof. By Lemma 4.4, (4.5) implies that ρ_k remains constant for all sufficiently large k. Then by Lemma 4.3 $\{y^k\}$ is componentwise bounded away from zero. Therefore, by Lemma 4.2, $\{d^k\}$ is bounded.

The boundedness of $\{d^k\}$ and that $\{y^k\}$ is componentwise bounded away from zero imply that there exists a constant $\tilde{\alpha} \in (0, 1]$ such that (3.3) holds for every $\alpha_k \in (0, \tilde{\alpha}]$.

Suppose that $\rho_k = \hat{\rho}$ for all sufficiently large k. Since $\{d^k\}$ is bounded and $\{y^k\}$ is componentwise bounded away from zero, similar to the proof of Lemma 4.3 of [20], we have

$$\begin{aligned}
&\phi(z^k + \alpha d^k; \hat{\rho}) - \phi(z^k; \hat{\rho}) - \pi_k(\alpha d^k; \hat{\rho}) \\
&\le c_1 \alpha^2 \|P_k^{-1} d^k\|^2 \\
&\le c_1 \alpha^2 \|P_k^{-1}(d^k - d_p^k) + P_k^{-1} d_p^k\|^2 \\
&\le c_1 \alpha^2 \|P_k^{-1}(d^k - d_p^k)\|^2 + c_2 \alpha^2 \|R_k r^k\|
\end{aligned} \tag{4.11}$$

for all sufficiently small $\alpha > 0$, where $c_2 > c_1 > 0$ are constants. The last inequality follows from (3.2) and the boundednesses of $\|P_k^{-1} d_p^k\|$ and $\|P_k^{-1} d^k\|$. On the other hand, if (4.5) holds, then

$$\begin{aligned}
&\pi_k(\alpha d^k; \hat{\rho}) - \sigma \alpha \pi_k(d^k; \hat{\rho}) \\
&\le (1 - \sigma) \alpha \pi_k(d^k; \hat{\rho}) \\
&\le (1 - \sigma) \alpha [(1 - \tau) \hat{\rho} (\|r^k + R_k^T d^k\| - \|r^k\|) - (1/2)(d^k - d_p^k)^T Q_k (d^k - d_p^k)] \\
&\le -c_3 \alpha \|R_k r^k\| - c_4 \alpha \|P_k^{-1}(d^k - d_p^k)\|^2,
\end{aligned} \tag{4.12}$$

where $c_3 > 0$ and $c_4 > 0$ are constants, the second inequality is obtained by the update rule of penalty parameter, and the third by (3.1), (4.5) and that $\|P_k R_k r^k\| \ge c_5 \|R_k r^k\|$ for some constant $c_5 > 0$ since y^k is componentwise bounded away from zero, $(d^k - d_p^k)^T Q_k (d^k - d_p^k) \ge \beta \|P_k^{-1}(d^k - d_p^k)\|^2$ by (4.3) (where β is coming from Assumption 4.1 (3)).

It follows from (4.11) and (4.12) that there exists a $\bar{\alpha} \in (0, \tilde{\alpha}]$ such that

$$\phi(z^k + \alpha d^k; \hat{\rho}) - \phi(z^k; \hat{\rho}) \le \sigma \alpha \pi_k(d^k; \hat{\rho})$$

for all $\alpha \in (0, \bar{\alpha}]$ and all k. Thus, by Step 5 of Algorithm 3.1, $\alpha_k \ge \bar{\alpha}$ for all k. $\square$

Lemma 4.8 *Suppose that Assumption 4.1 holds. If we have (4.5) for some constant $\eta_0 > 0$, then as $k \to \infty$*

$$R_k r^k \to 0, \tag{4.13}$$

and thereby

$$r^k \to 0,\ d^k \to 0,\ Y_k U_k e - \mu e \to 0,\ \nabla f^k + \nabla g^k u^k + \nabla h^k v^k \to 0. \tag{4.14}$$

Proof. By Lemma 4.4 and (4.5), we can assume $\rho_k = \hat{\rho}$ for all sufficiently large k. Since $\phi(z^k; \hat{\rho})$ is monotonically decreasing, (3.4) and Lemma 4.7 imply that

$$\lim_{k\to\infty} \pi_k(d^k, \hat{\rho}) = 0. \tag{4.15}$$

By (3.6) and (3.1),

$$\begin{aligned} \pi_k(d^k, \hat{\rho}) &\le -(1-\tau)\hat{\rho}(\|r^k\| - \|r^k + R_k^T d_p^k\|) \\ &\le -(1-\tau)\hat{\rho}\kappa_1 \|P_k R_k r^k\|^2 / \|r^k\|. \end{aligned} \tag{4.16}$$

It follows from (4.15), (4.16) and (4.5) that

$$\lim_{k\to\infty} \|P_k R_k r^k\| = 0, \tag{4.17}$$

which by Lemma 4.3 implies $R_k r^k \to 0$ as $k \to \infty$. Hence, $r^k \to 0$ by (4.5). In this case, $\|d_p^k\| \to 0$ as $k \to \infty$ by (3.2).

By equation (2.15) that

$$q^{k^T} d^k + d^{k^T} Q_k d^k = -\lambda^{k^T} R_k^T d_p^k. \tag{4.18}$$

Hence, by (2.17),

$$-(1/2) d^{k^T} Q_k d^k \le \lambda^{k^T} R_k^T d_p^k + q^{k^T} d_p^k + (1/2) d_p^{k^T} Q_k d_p^k. \tag{4.19}$$

Since

$$\pi_k(d^k, \hat{\rho}) \le q^{k^T} d^k \le q^{k^T} d_p^k + (1/2) d_p^{k^T} Q_k d_p^k - (1/2) d^{k^T} Q_k d^k,$$

it follows from (4.19) and (4.15) that

$$\lim_{k\to\infty} q^{k^T} d^k = 0\ . \tag{4.20}$$

Thus, (4.18) implies that $(-\lambda^{k^T} R_k^T d_p^k - d^{k^T} Q_k d^k) \to 0$ for $k \to \infty$. Therefore,

$$\lim_{k\to\infty} d^{k^T} Q_k d^k = 0. \tag{4.21}$$

Because $\|d_p^k\| \to 0$, and, by (4.3), $(d^k - d_p^k)^T Q_k (d^k - d_p^k) \ge \beta \|d^k - d_p^k\|^2$, we have that

$$\lim_{k\to\infty} \|d^k\| = 0\ . \tag{4.22}$$

By (2.15), we have

$$B_k d_x^k + \nabla g^k (u^k + d_u^k) + \nabla h^k (v^k + d_v^k) = -\nabla f^k, \tag{4.23}$$

$$U_k d_y^k + Y_k (u^k + d_u^k) = \mu e. \tag{4.24}$$

Hence, by (4.22), as $k \to \infty$,

$$\nabla f^k + \nabla g^k(u^k + d_u^k) + \nabla h^k(v^k + d_v^k) \to 0, \tag{4.25}$$

$$Y_k(u^k + d_u^k) - \mu e \to 0. \tag{4.26}$$

By (4.26) and $\|d_y^k\| \to 0$, we have $\beta_{ki} = 1$ for all $i \in \mathcal{I}$ and all sufficiently large k. Consequently, $u^{k+1} = u^k + d_u^k$ for all sufficiently large k. Then, by (4.22), (4.25)–(4.26), we have (4.14). □

We present our global convergence results on Algorithm 3.1 in the following theorem.

Theorem 4.9 *Suppose that Assumption 4.1 holds, for any given $\mu > 0$, $\{x^k\}$ is an infinite sequence generated by Algorithm 3.1. Then we have one of the following results:*

(1) The penalty parameter ρ_k remains constant for all sufficiently large k, every limit point of $\{x^k\}$ is a KKT point of problem (1.8)–(1.10);

(2) There exists a limit point x^* of sequence $\{x^k\}$ which is either an infeasible stationary point of minimizing the ℓ_2 norm of violations of constraints, or a Fritz-John point of problem (1.1)–(1.3).

Proof. For an infinite sequence $\{x^k\}$ generated by Algorithm 3.1, we have either that (i) (4.5) holds for some constant $\eta_0 > 0$ and for all k or (ii) there exists a subset $\mathcal{K} \in \mathcal{N}$ such that (4.10) holds.

In case (i), by Lemma 4.4, ρ_k remains constant for all sufficiently large k. Again by Lemma 4.8, we have (4.14), that is, for every limit point x^*,

$$\begin{cases} g^* + y^* &= 0, \\ h^* &= 0, \\ y_i^* u_i^* &= \mu, \ i \in \mathcal{I}, \\ \nabla f^* + \nabla g^* u^* + \nabla h^* v^* &= 0, \end{cases} \tag{4.27}$$

where $g^* = g(x^*)$, $h^* = h(x^*)$, $\nabla f^* = \nabla f(x^*)$, $\nabla g^* = \nabla g(x^*)$, $\nabla h^* = \nabla h(x^*)$. The system (4.27) shows that x^* is precisely the KKT point of problem (1.8)–(1.10). Thus, the result (1) follows.

Next, we consider the case (ii). By (4.10), there exists a subsequence $\{x^k : k \in \mathcal{K}\}$ such that

$$\lim_{k \in \mathcal{K}, k \to \infty} P_k R_k r^k / \|r^k\| = 0. \tag{4.28}$$

Since $\{r^k\}$ is bounded, there exists a subsequence $\{x^k : k \in \hat{\mathcal{K}} \subseteq \mathcal{K}\}$ such that $x^k \to x^*$ and $r^k \to r^*$ as $k \in \hat{\mathcal{K}}$ and $k \to \infty$. If x^* is infeasible for the original problem (1.1)–(1.3), then $\|r^*\| \neq 0$, in this case (4.28) implies that

$$\lim_{k \in \hat{\mathcal{K}}, k \to \infty} P_k R_k r^k = 0, \tag{4.29}$$

or equivalently,

$$\nabla g^*(g^* + y^*) + \nabla h^* h^* = 0, \tag{4.30}$$
$$Y^*(g^* + y^*) = 0. \tag{4.31}$$

Since x^* is infeasible for the original problem (1.1)–(1.3), then $g^* + y^* = \max\{g^*, 0\}$, thus x^* satisfies

$$\nabla g^* g_+^* + \nabla h^* h^* = 0 \tag{4.32}$$

where $g_+^* = \max\{g^*, 0\}$, which shows that x^* is a stationary point for minimizing the ℓ_2 norm of violations of equality and inequality constraints.

If x^* is feasible for the original problem (1.1)–(1.3). Then $g^* + y^* = 0$. Let $w^k = r^k/\|r^k\|$. Since $\|w^k\| = 1$ and $g^k + y^k \geq 0, \forall k$, there exists a subsequence $\{w^k : k \in \tilde{\mathcal{K}} \subseteq \hat{\mathcal{K}}\}$ such that $w^k \to w^*$ with two parts $w_1^* \in \Re_+^{|\mathcal{I}|}$ and $w_2^* \in \Re^{|\mathcal{E}|}$ as $k \in \tilde{\mathcal{K}}$ and $k \to \infty$, and, by (4.28),

$$\lim_{k \in \tilde{\mathcal{K}}, k \to \infty} P_k R_k w^k = 0. \tag{4.33}$$

In other words, there exists $u^* = w_1^* \in \Re_+^{|\mathcal{I}|}$ and $v^* = w_2^* \in \Re^{|\mathcal{E}|}$ such that

$$\nabla g^* u^* + \nabla h^* v^* = 0, \quad \text{and} \quad g_i^* u_i^* = 0,\ u_i^* \geq 0,\ i \in \mathcal{I}.$$

Thus, x^* is a Fritz-John point of the original problem. Thereby, result (2) is proved. □

The global convergence results suggest us how to define the stopping conditions such that the inner algorithm can always terminate at some approximate points with strong or weak stationary properties in a finite number of iterations. For any given sufficiently small $\epsilon_\mu > 0$ (which may depend on μ), the algorithm can be terminated if we have either

$$\|(r_d, r_c^\mu, r_g, r_h)\|_\infty < \epsilon_\mu, \tag{4.34}$$

where

$$\begin{aligned} r_d &= \nabla f^k + \nabla g^k u^k + \nabla h^k v^k, \\ r_c^\mu &= r_c - \mu e \quad (r_c = Y_k u^k), \\ r_g &= g^k + y^k, \\ r_h &= h^k, \end{aligned}$$

or

$$\|P_k R_k r^k\|/\|r^k\| < \epsilon, \tag{4.35}$$

where

$$P_k = \begin{bmatrix} I & \\ & Y_k \end{bmatrix}, \quad R_k = \begin{bmatrix} \nabla g^k & \nabla h^k \\ I & \end{bmatrix}, \quad r^k = \begin{bmatrix} g^k + y^k \\ h^k \end{bmatrix},$$

$\epsilon < \epsilon_\mu$ is a sufficiently small positive scalar.

5. The overall algorithm and its global convergence

We present the outer algorithm for the original problem (1.1)-(1.3), which together with the inner algorithm (Algorithm 3.1) forms the overall algorithm for solving nonlinear optimization with general equality and inequality constraints.

Algorithm 5.1 *(The outer algorithm for problem (1.1)-(1.3))*

Given *initial point* (x^0, y^0, u^0, v^0) *with* $y^0 > 0$ *and* $u^0 > 0$, *initial barrier parameter* $\mu_0 > 0$, *initial penalty parameter* $\rho_0 > 0$, *constants* $\kappa_1, \kappa_2, \xi, \sigma, \tau, \gamma_1, \gamma_2$ *in Step 1 of Algorithm 3.1,* $\chi \in (0, 1)$, $\epsilon > 0$. *Let* $j := 0$;

While $\mu_j \geq \epsilon$;

Apply Algorithm 3.1 to solve the barrier problem (1.8)–(1.10) (using μ_j *instead of* μ*) and terminate the inner algorithm if either (4.34) for* $\epsilon_\mu = \theta(\mu_j)$ *or (4.35) is met, where* $\theta : \Re_{++} \to \Re_{++}$ *is a real function such that* $\theta(\mu)$ *is monotonic and* $\lim_{\mu \to +0} \theta(\mu) = 0$.

Set $(x^{j+1}, y^{j+1}, u^{j+1}, v^{j+1}) = (x^k, y^k, u^k, v^k)$, $\rho_0 = \rho_{k+1}$.

If (4.35) is met, break and stop.

Set $\mu_{j+1} = \chi \mu_j$, $j := j + 1$;

end

It can be seen that (4.35) holds if Algorithm 3.1 terminates at Step 2. In this case, Algorithm 5.1 will stop. By the remarks on the inner algorithm, the termination point is strictly feasible for inequality constraints of the original problem and stationary for minimizing the ℓ_2 norm of violations of equality constraints, that is an infeasible Fritz-John point of the inequality constrained feasibility problem (FNP) defined in Chen and Goldfarb [7]. However, if Algorithm 3.1 terminates at Step 7, then (4.34) is satisfied and the overall algorithm proceeds to a smaller μ_j.

If, for every μ_j, Algorithm 3.1 terminates in satisfying (4.34), Algorithm 5.1 will generate a sequence $\{(x^j, y^j, u^j, v^j)\}$ and will terminate in a finite number of iterations. Otherwise, the outer algorithm will terminate for some μ_j, in which case the inner algorithm will terminate in satisfying (4.35).

Theorem 5.2 *Suppose that Assumption 4.1 holds for every* μ_j. *Then we have the following results.*

(a) For every μ_j, *the inner algorithm (Algorithm 3.1) terminates in satisfying (4.34), the outer algorithm (Algorithm 5.1) terminates at an approximate KKT point of the original problem (1.1)–(1.3) after a finite number of iterations;*

(b) For some μ_j, *the inner algorithm (Algorithm 3.1) terminates in satisfying (4.35), the outer algorithm (Algorithm 5.1) terminates at a point strictly feasible for inequality constraints (1.2) of the original problem and which is a stationary point for minimizing the* ℓ_2 *norm of violations of equality constraints (1.3) (an infeasible Fritz-John point of the*

inequality constrained feasibility problem (FNP)), or an approximately infeasible stationary point of minimizing the ℓ_2 norm of violations of constraints, or an approximately Fritz-John point of problem (1.1)–(1.3).

Proof. (a) Suppose that Algorithm 5.1 terminates in j_0 iterations. Then $\mu_{j_0} < \epsilon/\tau$. It follows from (4.34) that

$$\|(r_d, r_c, r_g, r_h)\|_\infty \leq \theta(\mu_{j_0}) + \mu_{j_0} \leq \theta(\epsilon/\tau) + \epsilon/\tau. \tag{5.1}$$

Let $\epsilon^{'} = \theta(\epsilon/\tau) + \epsilon/\tau$. Then $\epsilon^{'}$ can be sufficiently small if we select a very small ϵ, in which case, x^{j_0} together with multipliers u^{j_0} and v^{j_0} satisfies approximately the KKT conditions of problem (1.1)–(1.3) with maximal tolerance $\epsilon^{'}$.

(b) The results can be derived by the remarks on Algorithm 3.1 and the item (2) of Theorem 4.9. □

An infeasible Fritz-John point of the inequality constrained feasibility problem (FNP) is also an infeasible stationary point of problem (1.1)–(1.3), but not vice versa.

Theorem 5.3 *Suppose that Algorithm 5.1 breaks and stops for some μ_j, Assumption 4.1 holds for μ_j. Then we have the alternative results:*

(a) A point strictly feasible for inequality constraints (1.2) of the original problem and which is a stationary point for minimizing the ℓ_2 norm of violations of equality constraints (1.3) is derived;

(b) Either an approximately infeasible stationary point of minimizing the ℓ_2 norm of violations of constraints, or an approximate Fritz-John point of problem (1.1)–(1.3) is derived.

Proof. The results are obtained directly from the remarks on Algorithm 3.1 and Theorem 4.9(2). □

6. Local convergence

Under suitable assumptions, we prove that, by controlling the exactness of range-space steps, the algorithm generates a superlinear step. Moreover, we present the conditions on guaranteeing the positiveness of slack variable vector for a full step. Thus, locally superlinear convergence can be obtained if $\alpha_k = 1$ can be accepted by the merit function. Our analysis in this section is similar to that in Byrd, Liu and Nocedal [5]. We also need to use some results presented by Yamashita and Yabe [34]. For simplicity, we rearrange the over-all sequence as $\{(x^k, y^k, u^k, v^k)\}$. Let $z^k = (x^k, y^k), w^k = (z^k, u^k, v^k), d_w = (d_x, d_y, d_u, d_v)$.

Assumption 6.1

(1) $f(x)$, $g(x)$ *and* $h(x)$ *are twice differentiable, and their second derivatives are Lipschitz continuous at* x^**;*

(2) $x^k \to x^*$ *as* $k \to \infty$*, where* x^* *is a KKT point of problem (1.1)–(1.3),* $u^* \in \Re_+^{m_I}$ *and* $v^* \in \Re^{m_E}$ *are respectively the Lagrangian multiplier vectors of constraints (1.2) and (1.3);*

(3) Let $\mathcal{I}^* = \{i \in \mathcal{I} : g_i(x^*) = 0\}$, $\mathcal{J}^* \subseteq \mathcal{E}$. $\nabla g_i(x^*), i \in \mathcal{I}^*$ *are linearly independent, and set* $\{\nabla g_i(x^*) : i \in \mathcal{I}^*\} \cup \{\nabla h_j(x^*) : j \in \mathcal{J}^*\}$ *consists of the largest number of linearly independent column vectors of* $\{\nabla g_i(x^*) : i \in \mathcal{I}^*\} \cup \{\nabla h_j(x^*) : j \in \mathcal{E}\}$*. Moreover,* $\mathcal{J}_k = \mathcal{J}^*$*, where* $\mathcal{J}_k \subseteq \mathcal{E}$ *is an index set consisting of the largest number of linearly independent column vectors of* $\{\nabla h_j(x^k) : j \in \mathcal{E}\}$*;*

(4) $v_j^* = 0, \forall j \notin \mathcal{J}^*$ *and* $j \in \mathcal{E}$*;* $y_i^* + u_i^* > 0, \forall i \in \mathcal{I}$*;*

(5) $d^T \nabla^2 L(x^*, u^*, v^*) d > 0$, $\forall d \in \{d \neq 0 : \nabla h_j(x^*)^T d = 0, j \in \mathcal{J}^*; \quad \nabla g_i(x^*)^T d = 0, i \in \mathcal{I}^*\}$*, where* $L(x, u, v) = f(x) + u^T g(x) + v^T h(x)$*;*

Since $g_i(x^*) + y_i^* = 0, \forall i \in \mathcal{I}$, we have $\mathcal{I}^* = \{i \in \mathcal{I} : y_i^* = 0\}$. Consider our algorithm in this paper, we do not assume that all vectors in set $\{\nabla g_i(x^*) : i \in \mathcal{I}^*\} \cup \{\nabla h_j(x^*) : j \in \mathcal{E}\}$ are linearly independent. Instead, we assume that $\{\nabla g_i(x^*) : i \in \mathcal{I}^*\} \cup \{\nabla h_j(x^*) : j \in \mathcal{J}^* \subseteq \mathcal{E}\}$ are linearly independent and $\mathcal{J}_k = \mathcal{J}^*$ for all sufficiently large k.

Let $r(z) = \begin{bmatrix} g(x) + y \\ h(x) \end{bmatrix}$, $r_{\mathcal{J}}(z) = \begin{bmatrix} g(x) + y \\ h_{\mathcal{J}}(x) \end{bmatrix}$, $L_{\mathcal{J}}(x, u, v) = f(x) + u^T g(x) + v_{\mathcal{J}}^T h_{\mathcal{J}}(x)$,

$$F_\mu(w) = \begin{bmatrix} \nabla L_{\mathcal{J}}(x, u, v) \\ YUe - \mu e \\ r_{\mathcal{J}}(z) \end{bmatrix},$$

and $F(w) = F_0(w)$. Then

$$F_\mu'(w^k) = F'(w^k) = \begin{bmatrix} \nabla^2 L_{\mathcal{J}_k}(x^k, u^k, v^k) & 0 & \nabla g^k & \nabla h_{\mathcal{J}_k}^k \\ 0 & U_k & Y_k & 0 \\ \nabla g^{kT} & I & 0 & 0 \\ \nabla h_{\mathcal{J}_k}^{k\ T} & 0 & 0 & 0 \end{bmatrix}.$$

Define

$$G_k = \begin{bmatrix} B_k & 0 & \nabla g^k & \nabla h_{\mathcal{J}_k}^k \\ 0 & U_k & Y_k & 0 \\ \nabla g^{kT} & I & 0 & 0 \\ \nabla h_{\mathcal{J}_k}^{k\ T} & 0 & 0 & 0 \end{bmatrix}.$$

Then d_w^k is computed by

$$G_k d_w = -F_{\mu_k}(w^k) + \begin{bmatrix} 0 \\ 0 \\ r_{\mathcal{J}_k}^k + R_{\mathcal{J}_k}^T d_p^k \end{bmatrix}. \tag{6.1}$$

The following results are presented in Lemma 2.1 of [5] and Lemmas 1 and 7 of [34].

Lemma 6.2 *Suppose that Assumption 6.1 holds. There is a scalar $\epsilon > 0$ such that for all $w : \|w - w^*\| \leq \epsilon$,*

$$\|r(z) - r(z^*)\| \leq L_0\|z - z^*\|, \quad \|F^{'}(w) - F^{'}(w^*)\| \leq M_0\|w - w^*\|,$$

$F^{'}(w)$ is invertible and

$$\|F^{'}(w)^{-1}\| \leq M \tag{6.2}$$

for some constant $M > 0$. Moreover, for all $w, w^{'} : \|w - w^*\| \leq \epsilon, \|w^{'} - w^*\| \leq \epsilon$, we have that

$$\|F^{'}(w)(w - w^{'}) - F(w) + F(w^{'})\| \leq L\|w - w^{'}\|^2, \tag{6.3}$$

for some constant $L > 0$. Furthermore, there exists a $\delta > 0$ such that, if $\|w^k - w^*\| \geq \epsilon$, $\|B_k - \nabla^2 L(x^*, u^*, v^*)\| \leq \delta$, then

$$\|G_k - F^{'}(w^*)\| \leq M_1, \quad \|G_k^{-1}\| \leq M_2 \tag{6.4}$$

for some constants $M_1, M_2 > 0$.

Assumption 6.1 implies that, for all sufficiently small μ, there exists a neighborhood of w^* such that the barrier problem (1.8)–(1.10) has a unique solution which we denote by $w^*(\mu)$ if it has a solution. The next results follow from Lemma 2.2 and Lemma 2.3 of [5].

Lemma 6.3 *Suppose that Assumption 6.1 holds. Then there is $\bar{\mu} > 0$ such that for all $\mu \leq \bar{\mu}$, the system $F_\mu(w) = 0$ has a solution satisfying*

$$\|w^*(\mu) - w^*\| \leq \bar{M}\mu < \epsilon, \tag{6.5}$$

where $\bar{M}$ is a constant independent of μ. For all w sufficiently close to $w^*(\mu)$ and $\mu < \bar{\mu}$,

$$\|w - w^*(\mu)\| \leq 2M\|F_\mu(w)\|, \quad \|F_\mu(w)\| \leq 2\hat{M}\|w - w^*(\mu)\|, \tag{6.6}$$

where $\hat{M} = \sup_{\|w - w^*\| < \epsilon} \|F^{'}(w)\|$.

Using Lemma 6.2 and Lemma 6.3, we can prove the following results:

Theorem 6.4 *Suppose that Assumption 6.1 holds, $\|r^k\| - \|r^k + R_k^T d_p^k\| \geq \eta_k\|r^k\| (0 < \eta_k \leq 1)$ for all sufficiently large k, and w^k is sufficiently close to w^*.*

(1) If $\mu_k = o(\|F(w^k)\|)$, $\|(B_k - \nabla^2 L(x^*, u^*, v^*))d_x^k\| = o(\|d_x^k\|)$, and $1 - \eta_k = o(1)$, then

$$\lim_{k \to \infty} \frac{\|w^k + d_w^k - w^*\|}{\|w^k - w^*\|} = 0. \tag{6.7}$$

(2) If $\mu_k = O(\|F(w^k)\|^2)$, $B_k = \nabla^2 L(x^k, u^k, v^k)$, and $1 - \eta_k = O(\|r^k\|)$, then

$$\|w^k + d_w^k - w^*\| = O(\|w^k - w^*\|^2). \tag{6.8}$$

Proof. (1) By Lemma 8 of [34], $\|(B_k - \nabla^2 L(x^*, u^*, v^*))d_x^k\| = o(\|d_x^k\|)$ implies that

$$\|(G_k - F^{'}(w^*))d_w^k\| = o(\|d_w^k\|).$$

Let $\bar{d}_w^k$ be the solution of the system

$$F^{'}(w^k)d_w = -F_{\mu_k}(w^k).$$

Then

$$\begin{array}{rcl} d_w^k - \bar{d}_w^k & = & F^{'}(w^k)^{-1}(F^{'}(w^k)d_w^k + F_{\mu_k}(w^k)) \\ & = & F^{'}(w^k)^{-1}(F^{'}(w^k) - G_k)d_w^k + F^{'}(w^k)^{-1}(0, 0, 0, r_{\mathcal{J}_k}^k + R_{\mathcal{J}_k}^T d_p^k) \\ & \le & MM_0\|w^k - w^*\|\|d_w^k\| + o(\|d_w^k\|) + M(\|r_{\mathcal{J}_k}^k + R_{\mathcal{J}_k}^T d_p^k\|), \end{array} \tag{6.9}$$

where the second equality follows from (6.1) and the last inequality from Lemma 6.2.

Since, by Lemma 6.2,

$$\begin{array}{rcl} \|w^k + \bar{d}_w^k - w^*(\mu_k)\| & \le & \|F^{'}(w^k)^{-1}\|\|F^{'}(w^k)(w^k - w^*(\mu_k)) - F_{\mu_k}(w^k)\| \\ & \le & \|F^{'}(w^k)^{-1}\|\|F^{'}(w^k)(w^k - w^*(\mu_k)) - F(w^k) + F(w^*(\mu_k))\| \\ & \le & ML\|w^k - w^*(\mu_k)\|^2, \end{array} \tag{6.10}$$

and by Lemma 6.3

$$\begin{array}{rcl} d_w^k & = & -G_k^{-1}F_{\mu_k}(w^k) + G_k^{-1}(0, 0, 0, r_{\mathcal{J}_k}^k + R_{\mathcal{J}_k}^T d_p^k) \\ & \le & 2M_2\hat{M}(\|w^k - w^*(\mu_k)\|) + M_2(\|r_{\mathcal{J}_k}^k + R_{\mathcal{J}_k}^T d_p^k\|), \end{array} \tag{6.11}$$

we have that

$$\begin{array}{l} \|w^k + d_w^k - w^*\| \\ \le \|w^k + \bar{d}_w^k - w^*(\mu_k)\| + \|d_w^k - \bar{d}_w^k\| + \|w^*(\mu_k) - w^*\| \\ \le ML\|w^k - w^*(\mu_k)\|^2 + MM_0\|w^k - w^*\|(2M_2\hat{M}\|w^k - w^*(\mu_k)\| \\ \quad +M_2\|r_{\mathcal{J}_k}^k + R_{\mathcal{J}_k}^T d_p^k\|) + o(\|w^k - w^*(\mu_k)\|) + M(\|r_{\mathcal{J}_k}^k + R_{\mathcal{J}_k}^T d_p^k\|) + \bar{M}\mu_k \\ \le (2ML + 2MM_2M_0\hat{M})\|w^k - w^*\|^2 + (2MM_2M_0\hat{M}\|w^k - w^*\| + 2ML\bar{M}\bar{\mu} + 1)\bar{M}\mu_k \\ \quad +M(M_0M_2\|w^k - w^*\| + 1)(1 - \eta_k)\|r^k\| + o(\|w^k - w^*\|). \end{array} \tag{6.12}$$

By the fact that $\|F(w^k)\| = O(\|w^k - w^*\|)$ and $\|r^k\| = O(\|z^k - z^*\|) \le O(\|w^k - w^*\|)$,

$$\frac{\|w^k + d_w^k - w^*\|}{\|w^k - w^*\|} \le (2ML + 2MM_2M_0\hat{M})\|w^k - w^*\| + o(1). \tag{6.13}$$

Hence, (6.7) follows immediately.

(2) If $B_k = \nabla^2 L(x^k, u^k, v^k)$, then $G_k = F^{'}(w^k)$. Thus,

$$\|d_w^k - \bar{d}_w^k\| \le M\|r_{\mathcal{J}_k}^k + R_{\mathcal{J}_k}^T d_p^k\| \le M\|r^k + R_k^T d_p^k\|. \tag{6.14}$$

Therefore,

$$\begin{aligned}\|w^k + d_w^k - w^*\| &\leq \|w^k + \bar{d}_w^k - w^*(\mu_k)\| + \|d_w^k - \bar{d}_w^k\| + \|w^*(\mu_k) - w^*\| \\ &\leq ML\|w^k - w^*(\mu_k)\|^2 + M\|r^k + R_k^T d_p^k\| + \bar{M}\mu_k \\ &\leq 2ML\|w^k - w^*\|^2 + M(1-\eta_k)\|r^k\| + (2ML\bar{M}\bar{\mu} + 1)\bar{M}\mu_k. \quad (6.15)\end{aligned}$$

If $\mu_k = O(\|F(w^k)\|^2)$, $1 - \eta_k = O(\|r^k\|)$, we can obtain the result by $\|F(w^k)\| = O(\|w^k - w^*\|)$ and $\|r^k\| = O(\|w^k - w^*\|)$. □

In the following theorem, we consider the iterate w^k at which the terminating condition (4.34) is satisfied. Thus, a new barrier parameter μ_{k+1} is taken for generating d_w^k. We prove that the full step d_w^k will be accepted if w^k sufficiently close to w^*, μ_k is sufficiently small and parameters $\mu_{k+1}, \epsilon_{\mu_k}, \eta_k$, and ξ are selected elaborately.

Theorem 6.5 *Suppose that Assumption 6.1 holds, $\|r^k\| - \|r^k + R_k^T d_p^k\| \geq \eta_k\|r^k\|(0 < \eta_k \leq 1)$. If w^k is sufficiently close to w^*, μ_k is sufficiently small, $\|(r_d^k, r_c^k - \mu_k e, r_g^k, r_h^k)\|_\infty \leq \epsilon_{\mu_k}$, $\|(B_k - \nabla^2 L(x^*, u^*, v^*))d_x^k\| = o(\|d_x^k\|)$,*

$$[(\mu_k + \epsilon_{\mu_k})^2 + (1-\eta_k)\epsilon_{\mu_k}]/\mu_{k+1} \to 0, \quad (6.16)$$

and

$$\limsup \quad (\mu_k + \epsilon_{\mu_k})/\mu_{k+1} < 1/\xi, \quad (6.17)$$

then $y^k + d_y^k \geq \xi y^k$ for all sufficiently large k, where $r_d^k, r_c^k, r_g^k, r_h^k$ are defined in (4.34), ξ is defined in Algorithm 3.1.

Proof. The inequality $\|(r_d^k, r_c^k - \mu_k e, r_g^k, r_h^k)\|_\infty \leq \epsilon_{\mu_k}$ implies that $\|F_{\mu_k}(w^k)\| \leq \sqrt{n + 2m_I + m_E}\epsilon_{\mu_k}$ and $\|r^k\| \leq \sqrt{m_I + m_E}\epsilon_{\mu_k}$.

For every i, if $u_i^* > 0$, then by Assumption 6.1 $y_i^* = 0$. Thus,

$$y_i^k = y_i^k - y_i^* \leq \|w^k - w^*\| \leq \|w^k - w^*(\mu_k)\| + \|w^*(\mu_k) - w^*\| \leq M_3\epsilon_{\mu_k} + \bar{M}\mu_k, \quad (6.18)$$

where $M_3 = 2M\sqrt{n + 2m_I + m_E}$. Since

$$u_i^k(y_i^k + (d_y^k)_i) = \mu_{k+1} - y_i^k(d_u^k)_i, \quad (6.19)$$

and, by (6.11),

$$\begin{aligned}\|d_w^k\| &\leq 2M_2\hat{M}(\|w^k - w^*(\mu_k)\|) + M_2(\|r_{\mathcal{J}_k}^k + R_{\mathcal{J}_k}^T d_p^k\|) \\ &\leq (2M_2\hat{M}M_3 + M_2\sqrt{m_I + m_E}(1-\eta_k))\epsilon_{\mu_k}, \quad (6.20)\end{aligned}$$

we have

$$\begin{aligned}[y_i^k + (d_y^k)_i]/y_i^k &= [\mu_{k+1} - y_i^k(d_u^k)_i]/(y_i^k u_i^k) \\ &\geq [\mu_{k+1} - (M_3\epsilon_{\mu_k} + \bar{M}\mu_k)\|d_w^k\|]]/(y_i^k u_i^k) \quad (6.21) \\ &\geq [\mu_{k+1} - M_4((\epsilon_{\mu_k} + \mu_k)^2 + (1-\eta_k)\epsilon_{\mu_k})]/(y_i^k u_i^k) \\ &\geq (\mu_{k+1}/(y_i^k u_i^k))[1 - M_4((\epsilon_{\mu_k} + \mu_k)^2 + (1-\eta_k)\epsilon_{\mu_k})/\mu_{k+1}],\end{aligned}$$

where $M_4 = \max\{2M_2\hat{M}M_3, M_2\sqrt{m_I + m_E}\}\max(M_3, \bar{M})$. The last inequality follows from the fact that ϵ_{μ_k} and μ_k are sufficiently small.

Because $\|r_c^k - \mu_k e\|_\infty \le \epsilon_{\mu_k}$, we have $y_i^k u_i^k \le \mu_k + \epsilon_{\mu_k}$. Then it follows from (6.21), (6.16) and (6.17) that $y_i^k + (d_y^k)_i \ge \xi y_i^k$ for all sufficiently large k.

If, for some index i, $u_i^* = 0$, then by Assumption 6.1 $y_i^* > 0$. Since

$$\begin{aligned} y_i^k + (d_y^k)_i &\ge y_i^* - |y_i^k + (d_y^k)_i - y_i^*| \\ &\ge y_i^* - \|w^* - w^*(\mu_k)\| - \|w^k + d_w^k - w^*(\mu_k)\|, \end{aligned} \tag{6.22}$$

and

$$\begin{aligned} &\|w^k + d_w^k - w^*(\mu_k)\| \\ &\le \|w^k + \bar{d}_w^k - w^*(\mu_k)\| + \|d_w^k - \bar{d}_w^k\| \\ &\le ML\|w^k - w^*(\mu_k)\|^2 + MM_0(\|w^k - w^*(\mu_k)\| + \bar{M}\mu_k)(2M_2\hat{M}\|w^k - w^*(\mu_k)\| \\ &\quad + M_2(1-\eta_k)\|r^k\|) + o(\|w^k - w^*(\mu_k)\| + (1-\eta_k)\|r^k\|) + M(1-\eta_k)\|r^k\|, \end{aligned} \tag{6.23}$$

for sufficiently small μ_k and ϵ_{μ_k} and for w^k sufficiently close to w^*, we can have

$$\|w^* - w^*(\mu_k)\| + \|w^k + d_w^k - w^*(\mu_k)\| \le (1-\xi)y_i^* \tag{6.24}$$

for sufficiently large k. Thus, for every $i = 1, \ldots, m_I$, we have $y_i^k + (d_y^k)_i \ge \xi y_i^k$ for sufficiently large k. □

7. Conclusion

Using subspace techniques to improve the robustness of line search interior-point methods of nonlinearly constrained optimization is not a new methodology. It has firstly made success in trust region interior-point methods proposed by Byrd, Gilbert and Nocedal [3]. A kind of robust line search interior-point methods are then presented in Liu and Sun [20, 21]. Numerical results in Byrd, Hribar and Nocedal [4], Liu and Sun [20, 21] have illustrated that these methods are robust and efficient in solving some hard problems, including some simple but eccentric problems presented and discussed by Burke and Han [2], Byrd, Marazzi and Nocedal [6], Hock and Schittkowski [18], Wächter and Biegler [30], and problems from the collection MacMPEC of mathematical programs with equilibrium constraints [19, 23].

There are some restrictions on line search interior-point methods in [20, 21], however, which are undesirable and may circumscribe the applications of these methods. The method in [20] generates a weighted Newton step and a weighted Cauchy step in each iteration. The existence of these steps closely depends on the positive definiteness of the approximate Hessian. Although [21] does not clearly show how to calculate the range-space steps, its global convergence analysis is based on assumptions on the exact solution of problem (2.1), the positive definiteness of the approximate Hessian and that gradients of equality constraints are of full rank in all iterations. We get rid of all these limitations in this paper, and prove that the algorithm has strong global convergence properties under some much general conditions on range-space steps and approximate Hessian. Recently, using a different approach, Chen and Goldfarb [7] have shown that their method has similar strong global convergence properties.

Additionally, following the approaches proposed by Byrd, Liu and Nocedal [5] and using some results of Yamashita and Yabe [34], it is proved that by suitably controlling the exactness of range-space steps, selecting the barrier parameter and Hessian approximation, our algorithm generates a superlinearly or quadratically convergent step. The conditions on guaranteeing the positiveness of slack variable vector for a full step are also presented.

Acknowledgements. We thank the associate editor and two referees for their valuable comments and suggestions on an earlier version of the paper.

References

[1] H.Y. Benson, D.F. Shanno and R.J. Vanderbei, *Interior-point methods for nonconvex nonlinear programming: Filter methods and merit functions,* Comput. Optim. Appl., 23(2002), 257–272.

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

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

[4] 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.

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

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

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

[8] J.E. Dennis, M. Heinkenschloss, and L.N. Vicente, Trust-region interior-point SQP algorithms for a class of nonlinear programming problems, SIAM J. Control Optim., 36(1998), 1750–1794.

[9] 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.

[10] R. Fletcher and T. Johnson, *On the stability of null-space methods for KKT systems,* Numerical Analysis Report NA/167, Department of Mathematics and Computer Science, University of Dundee, Dundee DD1 4HN, Scotland, UK, Dec., 1995.

[11] R. Fletcher, and S. Leyffer, *Nonlinear Programming without a penalty function,* Math. Program., 91(2002), 239–269.

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

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

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

[15] N.I.M. Gould, *On practical conditions for the existence and uniqueness of solutions to the general equality quadratic programming problem,* Math. Program. 32(1985), 90–99.

[16] N.I.M. Gould, D. Orban and PH.L. Toint, *An interior-point l_1-penalty method for nonlinear optimization,* RAL-TR-2003-022, Rutherford Appleton Lab., Chilton, Oxfordshire, England, 2003.

[17] I. Griva, D.F. Shanno and R.J. Vanderbei, *Convergence analysis of a primal-dual interior-point method for nonlinear programming,* 2004.

[18] W. Hock and K. Schittkowski, *Test examples for nonlinear programming codes,* Lecture Notes in Eco. Math. Systems 187, Springer, New York, 1981.

[19] S. Leyffer, *MacMPEC – www-unix.mcs.anl.gov/leyffer/MacMPEC/.* 2002.

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

[21] X.-W. Liu and J. Sun, *Generalized stationary points and an interior point method for mathematical programs with equilibrium constraints,* Math. Program., 101(2004), 231–261.

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

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

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

[25] 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.

[26] A.L. Tits, A. Wächter, S. Bakhtiari, T.J. Urban, and C.T. Lawrence, *A primal-dual interior-point method for nonlinear programming with strong global and local convergence properties,* SIAM J. Optim., 14(2003), 173–199.

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

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

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

[30] 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.

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

[32] Z.H. Wang and Y.X. Yuan, *A subspace implementation of quasi-Newton trust region methods for unconstrained optimization*, Numer. Math., 104(2006), 241–269.

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

[34] 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.

[35] H. Yamashita, H. Yabe and T. Tanabe, *A globally and superlinearly convergent primal-dual interior point trust region method for large scale constrained optimization,* Math. Program., 102(2004), 111–151.

[36] Y.X. Yuan, *Subspace techniques in optimization,* Talk in 2006' China Mathematical Programming Meeting, Nanning, April 2006.