CORC TECHNICAL REPORT TR-2004-08

INTERIOR-POINT ℓ_2-PENALTY METHODS FOR NONLINEAR PROGRAMMING WITH STRONG GLOBAL CONVERGENCE PROPERTIES

LIFENG CHEN* AND DONALD GOLDFARB†

April 11, 2005

Abstract. We propose two line search primal-dual interior-point methods that approximately solve a sequence of equality constrained barrier subproblems. To solve each subproblem, our methods apply a modified Newton method and use an ℓ_2-exact penalty function to attain feasibility. Our methods have strong global convergence properties under standard assumptions. Specifically, if the penalty parameter remains bounded, any limit point of the iterate sequence is either a KKT point of the barrier subproblem, or a Fritz-John (FJ) point of the original problem that fails to satisfy the Mangasarian-Fromovitz constraint qualification (MFCQ); if the penalty parameter tends to infinity, there is a limit point that is either an infeasible FJ point of the inequality constrained feasibility problem (an infeasible stationary point of the infeasibility measure if slack variables are added) or a FJ point of the original problem at which the MFCQ fails to hold. Numerical results are given that illustrate these outcomes.

Key words. constrained optimization, nonlinear programming, primal-dual interior-point method, global convergence, penalty-barrier method, modified Newton method

AMS subject classifications. 49M37, 65F05, 65K05, 90C30

1. Introduction.

In this paper, we consider the general nonconvex nonlinear programming problem:

$$\text{(NP)}\qquad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & c_i(x) \geq 0,\ i \in \mathcal{I}, \ \text{ and } \ g_i(x) = 0,\ i \in \mathcal{E}, \end{array}$$

where x is a vector of dimension n, $\mathcal{I} = \{1, ..., m\}$, $\mathcal{E} = \{1, ..., p\}$, and the functions f, c and g are real valued and twice continuously differentiable on a domain $\Omega \subseteq \Re^n$. Ω will be specified in our standard assumptions.

To solve problem (NP), we propose two closely related barrier-sequential quadratic programming (SQP) methods[1], a *quasi-feasible method* and an *infeasible method*. Quasi-feasible methods require all iterates to be strictly feasible with respect to the inequality constraints and at each iteration, they "approximately" solve an equality constrained barrier subproblem for a fixed barrier parameter μ,

$$(\text{FP}_\mu)\qquad \begin{array}{ll} \text{minimize} & \varphi_\mu(x) = f(x) - \mu \sum_{i\in\mathcal{I}} \ln c_i(x) \\ \text{subject to} & c_i(x) > 0,\ i \in \mathcal{I}, \ \text{and} \ g_i(x) = 0,\ i \in \mathcal{E}. \end{array}$$

Later we will explicitly define what we mean by an approximate solution. However, we note that such a solution is not required to satisfy the equality constraints, $g_i(x) = 0$ ($i \in \mathcal{E}$), exactly. An optimal solution of problem (NP) is obtained by successively

*IEOR Department, Columbia University, New York, NY10027 (`lifeng.chen@columbia.edu`).
Research supported by the Presidential Fellowship of Columbia University.

†IEOR Department, Columbia University, New York, NY10027 (`gold@ieor.columbia.edu`).
Research supported in part by NSF Grant DMS 01-04282, DOE Grant DE-FG02-92EQ25126 and DNR Grant N00014-03-0514.

[1] This nomenclature is from [23].

solving problem (FP_μ) for decreasing values of the barrier parameter μ, $\mu \to 0$ (see, e.g., [16]). In many physical and engineering applications, the inequality constraints not only characterize the desired properties of the solution but also define a region in which the problem statement is meaningful (for example, $f(x)$ or some of the constraint functions $g_i(x)$ may not be defined outside the feasible region). Hence, quasi-feasible methods are well-suited to such problems.

Infeasible methods, on the other hand, first add slack variables to the inequality constraints of problem (NP) transforming it to

$$\text{(SP)} \qquad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & w_i \geq 0, \;\; c_i(x) - w_i = 0, \; i \in \mathcal{I}, \\ & g_i(x) = 0, \; i \in \mathcal{E}. \end{array}$$

These methods then solve a sequence of equality constrained barrier subproblems of the form,

$$(\mathrm{IP}_\mu) \qquad \begin{array}{ll} \text{minimize} & \varphi_\mu(x,w) = f(x) - \mu \sum_{i \in \mathcal{I}} \ln w_i \\ \text{subject to} & w_i > 0, \; c_i(x) - w_i = 0, \; i \in \mathcal{I}, \\ & g_i(x) = 0, \; i \in \mathcal{E}, \end{array}$$

obtaining a solution of problem (SP) (and hence, a solution of problem (NP)) as $\mu \to 0$. Infeasible methods have recently attracted more attention than quasi-feasible methods because: (i) they are closer in form to polynomial interior-point methods (IPMs) for linear and quadratic programming; (ii) they allow arbitrary starting points, since iterates are not required to satisfy $c_i(x) - w_i = 0$ ($i \in \mathcal{I}$) and hence $c_i(x)$ need not be positive; and (iii) they can easily compute the maximum line search step size.

Our quasi-feasible IPM is based on a line search exact ℓ_2-penalty approach for obtaining an approximate solution to problem (FP_μ). Specifically, we consider the problem

$$(\ell_2\mathrm{FP}_\mu) \qquad \begin{array}{ll} \text{minimize} & \Phi_{\mu,r}(x) = \varphi_\mu(x) + r\|g(x)\| \\ \text{subject to} & c_i(x) > 0, \; i \in \mathcal{I}, \end{array}$$

where r is the penalty parameter and $\|\cdot\|$ denotes the Euclidean vector norm.

To approximately solve (IP_μ), our infeasible IPM considers the problem

$$(\ell_2\mathrm{IP}_\mu) \qquad \begin{array}{ll} \text{minimize} & \Phi_{\mu,r}(x,w) = \varphi_\mu(x,w) + r\psi(x,w) \\ \text{subject to} & w_i > 0, \; i \in \mathcal{I}, \end{array}$$

where

$$\psi(x,w) = \left\| \begin{bmatrix} c(x) - w \\ g(x) \end{bmatrix} \right\|,$$

and r is the penalty parameter. The merit functions $\Phi_{\mu,r}(x)$ and $\Phi_{\mu,r}(x,w)$ are not differentiable. They are exact in the sense that if r is greater than a certain threshold value, a KKT point of the barrier subproblem (FP_μ) or (IP_μ) is a stationary point of the merit function, i.e., the directional derivative of $\Phi_{\mu,r}$ in any direction is nonnegative if it exists. Hence, r does not have to be increased to infinity.

Our methods apply a modified Newton method to the necessary conditions for local optimal solutions for problems ($\ell_2\mathrm{FP}_\mu$) and ($\ell_2\mathrm{IP}_\mu$) subject to the additional constraints that x and (x,w) lie in the open sets $\{x \mid \|g(x)\| > 0\}$ and $\{(x,w) \mid \psi(x,w) >$

$0\}$, respectively. Alternatively, each of these inner algorithms can be viewed as a modified Newton method applied to perturbed KKT conditions for problems (FP$_\mu$) and (IP$_\mu$), respectively. We note that our infeasible IPM is essentially our quasi-feasible IPM applied to the function (SP) in which advantage is taken of the special structure associated with the slack variables.

Our algorithms achieve strong global convergence results by employing a novel feasibility control strategy that forces $\|g(x)\|$ (resp. $\psi(x,w)$) to zero when the steps generated by the algorithms $\triangle x$ (resp. $(\triangle x, \triangle w)$) tend to zero, while keeping the penalty parameter r bounded above.

1.1. Related work. Recently, there has been a growing interest in IPMs for solving nonconvex nonlinear programming problems; e.g., see the the survey paper by Forsgren, Gill and Wright [23]. Both line search and trust-region inner algorithms have been proposed for solving subproblems (FP$_\mu$) and (IP$_\mu$). These methods generally use SQP methods or Newton-like methods to compute a step direction at each iteration and employ either penalty-based merit functions or filter techniques to enforce convergence. Merit function based line search barrier methods include those proposed by Argaez and Tapia [1], Gay, Overton and Wright [24], Moguerza and Prieto [36], Liu and Sun [33, 34], Bakhtiari and Tits [2], Conn, Gould and Toint [13], El-Bakry, Tapia, Tsuchiya and Zhang [16], Yamashita [49], and the algorithm that is the basis for the software package LOQO [38, 43, 31]. Merit function based trust region barrier-SQP methods are largely based on the Byrd-Omojokun algorithm [6, 37] and include methods proposed by Byrd et al. [7, 8, 11], Waltz, Morales, Nocedal and Orban [48], Dennis, Heinkenschloss and Vicente [14], Conn, Gould, Orban and Toint [12], Tseng [41], and Yamashita, Yabe and Tanabe [51].

Barrier Newton-like and barrier-SQP methods that use a primal-dual framework exhibit excellent performance in the neighborhood of a solution. In particular, under the assumption of strict complementarity and a suitable constraint qualification, the inner iterations can be terminated in such a way that the combined sequence of inner iterates converges to an isolated solution at a Q-superlinear rate; see, e.g., [9, 28]. However, in spite of such desirable local behavior, the global convergence theory for these methods, whether based on a line search or trust region framework, is not completely satisfactory. Some well-posed test examples illustrating the failure of several line search IPMs are analyzed in Wächter and Biegler [44] and Byrd, Marazzi and Nocedal [10]. These failures occur at infeasible non-stationary points where the Jacobian of the active constraints is rank deficient. To establish acceptable global convergence for problems with general equality constraints, many barrier-SQP methods have to resort to restrictive assumptions. For example, some methods require that linear independence of the equality constraint gradients or a MFCQ-like condition holds on a closed infeasible region containing all iterates (e.g., [1, 16, 49, 8, 24, 41]). Some rely on the assumption that the penalty parameter is bounded (e.g., [36]), while others assume that the multiplier approximation sequence is uniformly bounded or the Newton (quasi-Newton) system matrix is uniformly nonsingular (e.g., [42, 15, 51]). Others focus on practical implementational aspects and provide heuristic remedies for possible rank deficiency (e.g., [38, 43, 11, 4]).

Instead of using a merit function to drive a method towards both optimality and feasibility, one can use a filter. The first such method was proposed by Fletcher and Leyffer [19] in the context of (non-interior-point) SQP methods. Other filter methods have been proposed by Fletcher et at. [20, 18], Gonzaga, Karas and Vanti [30], Benson, Shanno and Vanderbei [3], Wächter and Biegler [45, 46, 47], and M.

Ulbrich, S. Ulbrich and Vicente [42]. The numerical results obtained by the *filter*-SQP code [19] and IPOPT, a filter IPM code [47], show that filter methods enjoy excellent robustness compared with penalty-based methods, especially in achieving feasibility. Global convergence results have been established for variants of the filter mechanism [20, 18, 45, 42], but this is still an ongoing area of research.

An interesting issue for barrier-SQP methods is how to ensure global convergence when the problem to be solved is primal degenerate. In this case, a linear independence constraint qualification (LICQ) may fail to hold at some feasible points in spite of the presence of MFCQ. Most existing barrier-SQP methods cannot rule out convergence to singular feasible non-critical points (e.g., [7, 33, 34, 42, 45]), unless strong conditions can be guaranteed, such as boundedness of the multiplier approximation sequence. We refer readers to Sporre and Forsgren [39] for the relation between convergence failure and divergence of the multiplier approximation sequence.

In contrast to barrier-SQP methods, penalty-barrier methods[2] treat equality constraints via unconstrained or inequality constrained minimization of a composite function that usually includes a quadratic or ℓ_1 penalty for violating such constraints; see, e.g., [17, 29]. Penalty-barrier methods have been proposed by Forsgren and Gill [22], Gertz and Gill [25], Goldfarb, Polyak, Scheinberg and Yuzefovich [26], Yamashita and Yabe [50], Tits, Wächter, Bakhtiari, Urban and Lawrence [40], and Gould, Orban and Toint [29] and global convergence results have been proved for these methods.

The paper is organized as follows. We first present some notation, definitions and the motivation behind the development of our algorithms in the next section. In Section 3 we present a detailed description of our feasibility control strategy, after studying the continuity and directional differentiability of the ℓ_2-exact penalty function $\Phi_{\mu,r}$. We then propose a quasi-feasible inner algorithm for solving the barrier subproblem (FP$_\mu$) and show that it is well defined. Its global convergence and that of a quasi-feasible IPM based on it is also analyzed in this section. Section 4 is devoted to presenting an infeasible IPM with slack variables and analyzing its global convergence. In section 5, we present numerical results that illustrate the convergence behavior of our methods. In the last section, we give some concluding remarks.

2. Preliminaries. We will use the following notation. A superscript (subscript) k will be used to denote the kth iterate of a vector or matrix (scalars). The superscript $\top$ will denote transposition. For a vector y, $\mathrm{diag}(y)$ is the diagonal matrix whose ith diagonal element is y_i. The gradient of $f(x)$ and the $m \times n$ and $p \times n$ Jacobians of $c(x)$ and $g(x)$ will be denoted by $\nabla f(x)$, $\nabla c(x)^\top$ and $\nabla g(x)^\top$, respectively. For a symmetric matrix $\mathcal{A}$ its inertia will be denoted by $\mathrm{In}(\mathcal{A})$, where $\mathrm{In}(\mathcal{A}) = (i_+, i_-, i_0)$ and i_+, i_-, i_0 are respectively the numbers of positive, negative and zero eigenvalues of $\mathcal{A}$. Given two vectors x and y of the same dimension l, we say $x \geq (>) y$ if and only if $x_i \geq (>) y_i$, $\forall i = 1, ..., l$, and $\min(x, y)$ is a vector whose ith element is $\min(x_i, y_i)$. For a symmetric matrix $\mathcal{A}$, $\mathcal{A} \succ (\succeq) 0$ means $\mathcal{A}$ is positive definite (semi-definite). For a positive integer q, $e \in \Re^q$ is the vector of all ones and $I \in \Re^{q \times q}$ is the unit matrix. $\Re^q_+$ denotes the set of nonnegative vectors in $\Re^q$.

2.1. Definitions. Let $\mathcal{X}$ be the feasible set of problem (NP). Let $\mathcal{F}$ and $\mathcal{F}^o$ be the quasi-feasible set and strictly quasi-feasible set of problem (NP), respectively, i.e.,

$$\mathcal{F} = \{x \in \Re^n | c_i(x) \geq 0, i \in \mathcal{I}\}, \quad \mathcal{F}^o = \{x \in \Re^n | c_i(x) > 0, i \in \mathcal{I}\};$$

[2] This nomenclature is from [23].

and let

$$\bar{\mathcal{F}}^o = \{x \in \Re^n | c_i(x) > 0, i \in \mathcal{I}; \|g(x)\| > 0\}.$$

The KKT conditions for problem (NP) give rise to the following system of nonlinear equations in $x \in \mathcal{X}, \lambda \in \Re^m_+$ and $y \in \Re^p$,

$$\mathcal{R}(x, \lambda, y) = \begin{bmatrix} \nabla_x \mathcal{L}(x, \lambda, y) \\ C(x)\lambda \\ g(x) \end{bmatrix} = 0, \tag{2.1}$$

where $\mathcal{L}(x, \lambda, y)$ is the Lagrangian function for problem (NP)

$$\mathcal{L}(x, \lambda, y) = f(x) - \sum_{i \in \mathcal{I}} \lambda_i c_i(x) + \sum_{i \in \mathcal{E}} y_i g_i(x),$$

and λ and y are vectors of Lagrange multipliers associated with the inequality and equality constraints, respectively. For a quasi-feasible point $x \in \mathcal{F}$, denote by $\mathcal{I}^o(x)$ the active set with respect to inequality constraints, i.e., $\mathcal{I}^o(x) = \{i \in \mathcal{I} | c_i(x) = 0\}$. For a feasible point $x \in \mathcal{X}$, the active set is $\mathcal{I}^o(x) \cup \mathcal{E}$.

DEFINITION 2.1. *For a feasible point $x \in \mathcal{X}$, we say that the LICQ holds at x if the active constraint gradients are linearly independent.*

DEFINITION 2.2. *For a feasible point $x \in \mathcal{X}$, we say that the MFCQ holds at x if the gradients $\nabla g_i(x)$ for $i \in \mathcal{E}$ are linearly independent, and there exists a direction $d \in \Re^n$ such that $\nabla c_i(x)^\top d > 0, \ \forall i \in \mathcal{I}^o(x)$ and $\nabla g_i(x)^\top d = 0, \ \forall i \in \mathcal{E}$.*

Denote by $m^o(x)$ the number of indices in $\mathcal{I}^o(x)$. By Farkas lemma it follows that Definition 2.2 is equivalent to the following Definition 2.3.

DEFINITION 2.3. *For a feasible point $x \in \mathcal{X}$, we say that the MFCQ is not satisfied at x if there exist $\lambda \in \Re^{m^o(x)}_+$ and $y \in \Re^p$ with $(\lambda, y) \neq 0$ such that*

$$\sum_{i \in \mathcal{I}^o(x)} \lambda_i \nabla c_i(x) + \sum_{i \in \mathcal{E}} y_i \nabla g_i(x) = 0. \tag{2.2}$$

DEFINITION 2.4. *A feasible point $x \in \mathcal{X}$ is called a Fritz-John (FJ) point of problem (NP) if there exist $z \geq 0$, $\lambda \in \Re^{m^o(x)}_+$ and $y \in \Re^p$ with $(z, \lambda, y) \neq 0$ such that*

$$z \nabla f(x) = \sum_{i \in \mathcal{I}^o(x)} \lambda_i \nabla c_i(x) + \sum_{i \in \mathcal{E}} y_i \nabla g_i(x). \tag{2.3}$$

A FJ point x is a KKT point of problem (NP) if the MFCQ holds at x. Since quasi-feasible methods maintain strict feasibility of all iterates with respect to inequality constraints (i.e., $c_i(x) > 0$ $(i \in \mathcal{I})$), the following feasibility problem arises naturally in these methods:

$$\text{(FNP)} \quad \begin{array}{ll} \text{minimize}_{x \in \Re^n} & \|g(x)\|^2 \\ \text{subject to} & c_i(x) \geq 0, \ i \in \mathcal{I}. \end{array}$$

Similarly, the following feasibility problem is associated with infeasible methods:

$$\text{(INP)} \quad \text{minimize}_{x \in \Re^n} \quad \left\| \begin{bmatrix} \min\{c(x), 0\} \\ g(x) \end{bmatrix} \right\|^2.$$

A point $x \in \mathcal{F}$ and $\lambda \in \Re^m_+ \backslash \{0\}$ satisfy the KKT conditions for problem (FNP) if

$$\begin{array}{c} \sum_{i\in\mathcal{I}} \lambda_i \nabla c_i(x) = \sum_{i\in\mathcal{E}} g_i(x)\nabla g_i(x). \\ \lambda_i c_i(x) = 0,\ i \in \mathcal{I}. \end{array} \tag{2.4}$$

Definition 2.5. *We say that the MFCQ holds at a quasi-feasible point $x \in \mathcal{F}$ for problem (FNP) if there exists no $\lambda \in \Re^{m^o(x)}_+ \backslash \{0\}$ such that $\sum_{i\in\mathcal{I}^o(x)} \lambda_i \nabla c_i(x) = 0$.*

A quasi-feasible point $x \in \mathcal{F}$ is an infeasible FJ point of problem (FNP) if it is infeasible for problem (NP) and there exist $z \geq 0$ and $\lambda \in \Re^{m^o(x)}_+$ with $(z, \lambda) \neq 0$ such that

$$z \sum_{i\in\mathcal{E}} g_i(x)\nabla g_i(x) = \sum_{i\in\mathcal{I}^o(x)} \lambda_i \nabla c_i(x). \tag{2.5}$$

A point $x \in \Re^n$ is called an infeasible stationary point of the feasibility problem (INP), if x is infeasible for problem (NP) and

$$\sum_{i\in\mathcal{I}} \min\{c_i(x), 0\}\nabla c_i(x) + \sum_{i\in\mathcal{E}} g_i(x)\nabla g_i(x) = 0. \tag{2.6}$$

When solving a general nonlinear programming problem, if a quasi-feasible method is unable to obtain feasibility, it should at least be able to find a KKT point of problem (FNP) or a FJ point of it if the MFCQ for problem (FNP) fails at some quasi-feasible points, and an infeasible method should be able to find an infeasible stationary point of problem (INP). Such information indicates that problem (NP) appears to be locally infeasible. On the other hand, if problem (NP) is primal degenerate at some feasible points, a robust IPM should at least be able to find a FJ point of problem (NP) that fails to satisfy MFCQ, indicating that there may be no feasible KKT points.

2.2. Motivation and basic results. The focus of this paper is the development of globally convergent IPMs that do not suffer from two important drawbacks of many existing IPMs, especially barrier-SQP methods, for solving problem (NP): (i) iterates may converge to singular infeasible non-stationary points at which the equality constraint Jacobian is rank deficient; (ii) iterates may converge to singular feasible non-critical points at which LICQ fails to hold. To this end, let us have a closer look at barrier-SQP methods. To simplify the presentation, we only consider quasi-feasible methods in the following discussion.

The first order optimality conditions for the barrier subproblem (FP$_\mu$) with barrier parameter μ are

$$\mathcal{R}_\mu(x, u, v) = \begin{bmatrix} \nabla_x \mathcal{L}(x, u, v) \\ C(x)u - \mu e \\ g(x) \end{bmatrix} = 0, \tag{2.7}$$

where $x \in \mathcal{F}^o \cap \mathcal{X}$, $u \in \Re^m_+$, $v \in \Re^p$ and $\mathcal{L}(x, u, v)$ is the Lagrangian function of problem (NP). Note that (2.7) is written in primal-dual form as it incorporates the multiplier estimate u. For linear programming, IPMs apply a single iteration of Newton's method to (2.7) to determine the step direction and choose a step length that keeps the next iterate strictly feasible. In the nonlinear case, quasi-Newton or modified Newton methods are often used instead. Letting $\mathcal{H}$ be the Hessian of the Lagrangian $\nabla^2_{xx}\mathcal{L}(x,u,v)$ or an approximation to it, each step of these methods requires the solution of

$$\begin{bmatrix} \mathcal{H} & -\nabla c(x) & \nabla g(x) \\ \mathcal{U}\nabla c(x)^\top & C(x) & 0 \\ \nabla g(x)^\top & 0 & 0 \end{bmatrix} \begin{bmatrix} \triangle x \\ \triangle u \\ \triangle v \end{bmatrix} = \begin{bmatrix} \nabla c(x)u - \nabla g(x)v - \nabla f(x) \\ -C(x)u + \mu e \\ -g(x) \end{bmatrix} \tag{2.8}$$

where $\mathcal{U} = \mathrm{diag}(u)$. Since quasi-feasible methods keep $C(x) \succ 0$, the second block of equations in (2.8) can be solved for $\triangle u = \mu C(x)^{-1}e - u - C(x)^{-1}\mathcal{U}\nabla c(x)^\top \triangle x$, and (2.8) reduced to the smaller and denser symmetric linear system

$$\begin{bmatrix} \hat{\mathcal{H}} & \nabla g(x) \\ \nabla g(x)^\top & 0 \end{bmatrix} \begin{bmatrix} \triangle x \\ \triangle v \end{bmatrix} = \begin{bmatrix} \mu \nabla c(x)^\top C(x)^{-1}e - \nabla g(x)v - \nabla f(x) \\ -g(x) \end{bmatrix} \tag{2.9}$$

where $\hat{\mathcal{H}} = \mathcal{H} + \nabla c(x)C(x)^{-1}\mathcal{U}\nabla c(x)^\top$. The coefficient matrices of the linear systems (2.8) and (2.9) are nonsingular if $\hat{\mathcal{H}}$ is positive definite on the null space of the Jacobian $\nabla g(x)^\top$ and $\nabla g(x)$ has full column rank. For general nonconvex nonlinear programming, $\hat{\mathcal{H}}$ is usually indefinite especially when the iterate is far from a solution. Consequently, (2.8) may have no solution or it may have a solution that is not a descent direction for a given merit function. There are two popular ways for overcoming this difficulty. One approach is to modify $\hat{\mathcal{H}}$ to be positive definite (at least on the null space of the equality constraint Jacobian) either by adding a multiple of the identity matrix to the Hessian estimate [4] or by using an inertia controlling symmetric indefinite factorization [22, 25] of the matrix in (2.8). The other approach resorts to trust region techniques, which allows the use of the exact Hessian (e.g., see [7, 8]).

Now suppose that the linear system (2.8) is nonsingular and has a unique solution $(\triangle x, \triangle u, \triangle v)$. It is known that $\triangle x$ is a descent direction of the merit function $\Phi_{\mu,r}(x)$ if the penalty parameter r is large enough. Generally, r is required to be of at least the order of magnitude of $\|v\|$ to ensure descent. Consequently, if a singular infeasible point is approached, (2.8) becomes increasingly ill-conditioned and $\|v\|$ may become arbitrarily large. To ensure descent, r has to be increased to catch up with the growth of $\|v\|$. This leads to the phenomenon illustrated in [44] in which the search direction does not tend to zero while the line search step size shrinks to zero. Eventually, the merit function loses its ability to penalize infeasibility and the singular infeasible point is accepted. We refer the reader to [10] for a detailed discussion of such failures.

A common remedy for the above type of failure is to ensure uniform non-singularity of the coefficient matrix of (2.8) by perturbing its diagonal elements as follows,

$$\begin{bmatrix} \mathcal{H} & -\nabla c(x) & \nabla g(x) \\ \mathcal{U}\nabla c(x)^\top & C(x) + \varrho_{\mathcal{I}} I & 0 \\ \nabla g(x)^\top & 0 & -\varrho_{\mathcal{E}} I \end{bmatrix},$$

where $\varrho_{\mathcal{I}} \geq 0$ and $\varrho_{\mathcal{E}} \geq 0$. This remedy has been effective in practice for preventing convergence to singular infeasible non-stationary points and some heuristics have been suggested for choosing $\varrho_{\mathcal{I}}$ and $\varrho_{\mathcal{E}}$ (e.g., see [4, 47]). However, there are instances indicating that barrier-SQP methods with such diagonal perturbations may cycle back to a neighborhood of the difficult point infinitely often [4]. Griva, Shanno and Vanderbei [31] suggest choosing $\varrho_{\mathcal{I}} = 0$ and $\varrho_{\mathcal{E}}$ to be the reciprocal of the penalty parameter, which ensures a descent direction for an augmented Lagrangian quadratic-penalty-barrier function. For each $\varrho_{\mathcal{E}}$, their method which is an infeasible method, approximately solves a penalty subproblem, eventually attaining global convergence for the barrier subproblem when the penalty parameter tends to infinity.

We also choose $\varrho_{\mathcal{I}} = 0$, but $\varrho_{\mathcal{E}}$ is determined from the optimality conditions for the ℓ_2-exact penalty function $\Phi_{\mu,r}(\cdot)$. We prove that if this choice is accompanied by a certain feasibility control strategy, we obtain globally convergent algorithms. Specifically, we propose two line search primal-dual IPMs, one of the quasi-feasible type and the other of the infeasible type, that use a barrier method as an outer framework. To enforce convergence for the inner algorithm, we use ℓ_2-exact penalty functions, whose exactness eliminates the need to drive the corresponding penalty parameters to infinity when finite multipliers exist. The proposed methods enjoy theoretical robustness in the sense that they do not converge to singular infeasible or singular feasible non-critical points. In particular, we show that under standard assumptions, if the penalty parameter tends to infinity when solving a barrier subproblem, there is a limit point of the generated iterate sequence that is either an infeasible first order critical point of the feasibility problem (FNP) or (INP) or a FJ point of problem (NP) at which the MFCQ fails to hold; if the penalty parameter is bounded, any limit point of the iterate sequence is a FJ point of problem (NP) that fails to satisfy MFCQ, or a KKT point of the barrier subproblem.

3. A quasi-feasible IPM.

We now present a quasi-feasible IPM for solving problem (NP), whose basic iteration consists of an inner algorithm for solving barrier subproblem (FP_μ). The following standard assumptions are needed throughout the section.

A1. *$\mathcal{F}^o$ is nonempty.*

A2. *Functions f, c, g are real valued and twice continuously differentiable on $\mathcal{F}$.*

3.1. Solving (FP_μ).

The necessary conditions for a point x to be a local optimal solution of problem ($\ell_2\text{FP}_\mu$) can be written as

$$\begin{aligned} \nabla f(x) - \textstyle\sum_{i\in\mathcal{I}} u_i \nabla c_i(x) + \sum_{i\in\mathcal{E}} v_i \nabla g_i(x) &= 0, \\ \mu - u_i c_i(x) &= 0, \quad i \in \mathcal{I}, \\ g_i(x) - v_i \delta_x &= 0, \quad i \in \mathcal{E}, \end{aligned} \tag{3.1}$$

where $\delta_x = \|g(x)\|/r$. If $\delta_x = 0$, the conditions (3.1) become the KKT conditions (2.7) for problem (FP_μ); hence, (3.1) can be viewed as perturbed KKT conditions for problem (FP_μ). We take this approach here and treat δ_x as a constant. Applying Newton's method to (3.1) with δ_x constant yields the following linear system,

$$\begin{bmatrix} \mathcal{H} & -\nabla c(x) & \nabla g(x) \\ \mathcal{U}\nabla c(x)^\top & C(x) & 0 \\ \nabla g(x)^\top & 0 & -\delta_x I \end{bmatrix} \begin{bmatrix} \triangle x \\ \triangle u \\ \triangle v \end{bmatrix} = \begin{bmatrix} -\nabla f(x) + \nabla c(x) u - \nabla g(x) v \\ -C(x)u + \mu e \\ \delta_x v - g(x) \end{bmatrix}.$$

Furthermore, letting $\lambda = u + \triangle u$, $y = v + \triangle v$ in the above system and adding a super script to denote the iterate yields

$$\left\{ \begin{array}{llllll} \mathcal{H}^k \triangle x^k & -\nabla c(x^k)\lambda^k & +\nabla g(x^k) y^k & = & -\nabla f(x^k), \\ \mathcal{U}^k \nabla c(x^k)^\top \triangle x^k & +C(x^k)\lambda^k & & = & \mu e, \\ \nabla g(x^k)^\top \triangle x^k & & -\delta_{x^k} y^k & = & -g(x^k), \end{array} \right. \tag{3.2}$$

where $\delta_{x^k} = \|g(x^k)\|/r_k$ and r_k is the current penalty parameter. The system (3.2) forms the basis of our inner iteration. Note that v^k is now implicitly replaced by y^{k-1} and y^k will serve as our equality multiplier estimate. Both u^k and λ^k will serve as our inequality multiplier estimates. The difference between them is that at each iteration we keep u^k strictly dual feasible, i.e., $u^k > 0$, while we simply compute λ^k from the solution of (3.2) without requiring positivity.

Since $C(x^k) \succ 0$, we can eliminate λ^k in (3.2) to obtain the perturbed KKT system

$$\mathcal{M}^k \begin{bmatrix} \triangle x^k \\ y^k \end{bmatrix} = \begin{bmatrix} -\nabla f(x^k) + \mu \nabla c(x^k)^\top C(x^k)^{-1} e \\ -g(x^k) \end{bmatrix}, \tag{3.3}$$

where

$$\mathcal{M}^k = \begin{bmatrix} \hat{\mathcal{H}}^k & \nabla g(x^k) \\ \nabla g(x^k)^\top & -\delta_{x^k} I \end{bmatrix} \tag{3.4}$$

and

$$\hat{\mathcal{H}}^k = \mathcal{H}^k + \nabla c(x^k) C(x^k)^{-1} \mathcal{U}^k \nabla c(x^k)^\top. \tag{3.5}$$

The following lemma, which is easy to prove using Sylvester's law of inertia and is an immediate consequence of Proposition 2 in [21], gives conditions for (3.3) (and hence (3.2)) to both have unique solutions.

LEMMA 3.1. *Let $\delta \geq 0$ and*

$$\mathcal{M} = \begin{bmatrix} \mathcal{G} & \mathcal{J} \\ \mathcal{J}^\top & -\delta I \end{bmatrix},$$

where $\mathcal{G} \in \Re^{n \times n}$ is symmetric, $\mathcal{J} \in \Re^{n \times p}$ and I is the $p \times p$ identity matrix. Then $\mathrm{In}(\mathcal{M}) = (n, p, 0)$ *if , and only if, either (i) $\delta = 0$, $\mathcal{J}$ has rank p and $v^\top \mathcal{G} v > 0$ for all $v \neq 0$ such that $\mathcal{J}^\top v = 0$; or (ii) $\delta > 0$ and $\mathcal{G} + \frac{1}{\delta} \mathcal{J}\mathcal{J}^\top \succ 0$.*

We now show that $\triangle x^k$ is a descent direction of the merit function $\Phi_{\mu, r_k}(x)$ provided the inertia of $\mathcal{M}^k$ defined by (3.4) is $(n, p, 0)$. To this end, let us first study the continuity and directional differentiability of $\Phi_{\mu,r}(x)$.

LEMMA 3.2. *For any given $\mu > 0, r > 0$ and direction $\triangle x \in \Re^n$, the directional derivative $\Phi'_{\mu,r}(x; \triangle x)$ of function $\Phi_{\mu,r}(x)$ along $\triangle x$ exists on $\mathcal{F}^o$, and*

$$\Phi'_{\mu,r}(x; \triangle x) \leq \nabla \varphi_\mu(x)^\top \triangle x + r(\|g(x) + \nabla g(x)^\top \triangle x\| - \|g(x)\|). \tag{3.6}$$

Proof. First, for any $x \in \mathcal{F}^o$, function $\varphi_\mu(x)$ is continuously differentiable and $\varphi'_\mu(x; \triangle x) = \nabla \varphi_\mu(x)^\top \triangle x$. Let $\psi(x) = \|g(x)\|$. By using the triangle inequality, the convexity of the Euclidean norm and the continuous differentiability of g, we can derive as follows,

$$\begin{aligned}
& \psi'(x; \triangle x) \\
= \; & \lim_{t \downarrow 0} [\psi(x + t\triangle x) - \psi(x)]/t \\
= \; & \lim_{t \downarrow 0} [\|g(x + t\triangle x)\| - \|g(x)\|]/t \\
= \; & \lim_{t \downarrow 0} [\|g(x) + t\nabla g(x)^\top \triangle x + o(t)\| - \|g(x)\|]/t \\
\leq \; & \lim_{t \downarrow 0} [\|g(x) + t\nabla g(x)^\top \triangle x\| - \|g(x)\| + \|o(t)\|]/t \\
\leq \; & \lim_{t \downarrow 0} [(1-t)\|g(x)\| + t\|g(x) + \nabla g(x)^\top \triangle x\| - \|g(x)\| + \|o(t)\|]/t \\
= \; & \|g(x) + \nabla g(x)^\top \triangle x\| - \|g(x)\|.
\end{aligned}$$

The result follows immediately. ☐

The next lemma trivially follows from the continuous differentiability of $\ln c_i(x),\ i \in \mathcal{I}$ and $\|g(x)\|$ at points $x \in \bar{\mathcal{F}}^0$.

LEMMA 3.3. *For any given $\mu > 0, r > 0$ and $x \in \bar{\mathcal{F}}^o$, the merit function $\Phi_{\mu,r}$ is continuously differentiable at x with derivative*

$$\nabla\Phi_{\mu,r}(x) = \nabla f(x) - \mu\nabla c(x)C(x)^{-1}e + r\|g(x)\|^{-1}\nabla g(x)g(x). \tag{3.7}$$

Thus, for any direction $\triangle x \in \Re^n$, $\Phi'_{\mu,r}(x;\triangle x) = \nabla\Phi_{\mu,r}(x)^\top \triangle x$.

The next lemma together with Lemma 3.1 show that $\triangle x^k$ is a descent direction if $\mathrm{In}(\mathcal{M}^k) = (n, p, 0)$.

LEMMA 3.4. *Suppose $x^k \in \mathcal{F}^o$, $u_i^k > 0$ $(i = 1, ..., m)$ and the linear system (3.3) has a solution $(\triangle x^k, y^k)$. Then*

$$\Phi'_{\mu,r_k}(x^k;\triangle x^k) \le -(\triangle x^k)^\top \tilde{\mathcal{H}}^k \triangle x^k, \tag{3.8}$$

where

$$\tilde{\mathcal{H}}^k = \begin{cases} \hat{\mathcal{H}}^k, & \text{if } \|g(x^k)\| = 0, \\ \hat{\mathcal{H}}^k + \delta_{x^k}^{-1}\nabla g(x^k)\nabla g(x^k)^\top, & \text{if } \|g(x^k)\| > 0. \end{cases} \tag{3.9}$$

Proof. First suppose $\|g(x^k)\| > 0$. From (3.3) and (3.9) we have that

$$\nabla f(x^k) = -\tilde{\mathcal{H}}^k\triangle x^k + \mu\nabla c(x^k)C(x^k)^{-1}e - \delta_{x^k}^{-1}\nabla g(x^k)g(x^k), \tag{3.10}$$

and hence from Lemma 3.3 and (3.10) it follows that

$$\begin{aligned} & \Phi'_{\mu,r_k}(x^k;\triangle x^k) = (\triangle x^k)^\top\nabla\Phi_{\mu,r_k}(x^k) \\ = \; & (\triangle x^k)^\top\nabla f(x^k) - \mu(\triangle x^k)^\top\nabla c(x^k)C(x^k)^{-1}e + \delta_{x^k}^{-1}(\triangle x^k)^\top\nabla g(x^k)g(x^k) \\ = \; & -(\triangle x^k)^\top\tilde{\mathcal{H}}^k\triangle x^k. \end{aligned}$$

Now we consider the case $\|g(x^k)\| = 0$. From the second equation of (3.3) we have that $\nabla g(x^k)^\top\triangle x^k = 0$. Hence, it follows from the first equation of (3.3) that

$$(\triangle x^k)^\top\nabla f(x^k) = -(\triangle x^k)^\top\hat{\mathcal{H}}^k\triangle x^k + \mu(\triangle x^k)^\top\nabla c(x^k)C(x^k)^{-1}e. \tag{3.11}$$

From Lemma 3.2, (3.11) and (3.9) we obtain

$$\begin{aligned} & \Phi'_{\mu,r_k}(x^k;\triangle x^k) \\ \le \; & \nabla\varphi_\mu(x^k)^\top\triangle x^k + r_k(\|g(x^k) + \nabla g(x^k)^\top\triangle x^k\| - \|g(x^k)\|) \\ = \; & \nabla\varphi_\mu(x^k)^\top\triangle x^k \\ = \; & (\triangle x^k)^\top\nabla f(x^k) - \mu(\triangle x^k)^\top\nabla c(x^k)C(x^k)^{-1}e \\ = \; & -(\triangle x^k)^\top\hat{\mathcal{H}}^k\triangle x^k \\ = \; & -(\triangle x^k)^\top\tilde{\mathcal{H}}^k\triangle x^k, \end{aligned}$$

where the relation $\nabla g(x^k)^\top\triangle x^k = 0$ is used explicitly. ∎

Now suppose $\|g(x^k)\| > 0$ and $\mathcal{H}^k$ has been modified, if necessary, so that $\tilde{\mathcal{H}}^k \succ 0$. Then from the second equation of (3.3), we have

$$\|g(x^k)\|\varpi_i^k = r_k\nabla g_i(x^k)^\top\triangle x^k, \quad i \in \mathcal{E}, \tag{3.12}$$

where

$$\varpi_i^k = y_i^k - r_k\frac{g_i(x^k)}{\|g(x^k)\|}, \quad i \in \mathcal{E}, \tag{3.13}$$

which suggests that

if $\{\|\varpi^k\|\}$ can be kept bounded away from zero and $\{r_k\}$ remains bounded above, then $\{\|g(x^k)\|\}$ will be driven to zero when $\{\triangle x^k\}$ tends to zero.

This statement embodies the spirit of our feasibility control strategy; and, hence, our algorithm monitors ϖ^k. Three immediate questions come to mind.

◇ How can we detect if $\{\triangle x^k\}$ is tending to zero?

◇ How can we keep $\{\|\varpi^k\|\}$ bounded away from zero?

◇ What may happen if $\{r_k\}$ is not bounded?

In our algorithm we check the following conditions at each iteration,

$$\begin{array}{ll} \text{(C-1):} & \|g(x^k)\| > 0; \\ \text{(C-2):} & \|\triangle x^k\| \le \pi_\mu; \\ \text{(C-3):} & \kappa_1 \mu e \le C(x^k)\lambda^k \le \kappa_2 \mu e; \\ \text{(C-4):} & \|\varpi^k\| < \pi_\mu; \end{array} \tag{3.14}$$

where $\pi_\mu > 0$, $\kappa_1 \in (0,1)$, $\kappa_2 > 1$. Clearly, violation of condition (C-1) means that there is no need to employ feasibility control. (C-2) is a necessary condition for $\{\triangle x^k\}$ to tend to zero and under certain boundedness assumptions, so is (C-3). If these three conditions hold, it is probable that "$\{\triangle x^k\}$ is tending to zero" and according to the "outlined" statement below (3.13), $\|\varpi^k\|$ should be kept at a sufficiently positive level. Therefore, we further check condition (C-4). If it is met, we increase the value of the penalty parameter r_k to increase $\|\varpi^k\|$ and force the iterates toward feasibility.

The remaining question and perhaps, the most important one, is whether $\{r_k\}$ can tend to infinity and if it does, what happens. In section 3.3 below we will prove that under standard assumptions, $\{r_k\} \to \infty$ in only two cases. In the first case, the algorithm encounters difficulties in improving feasibility and has an accumulation point that is a FJ point of the inequality constrained feasibility problem (FNP). In the second case, the algorithm fails to achieve optimality and has an accumulation point that is a FJ point of problem (NP) failing to satisfy the MFCQ.

Finally, rather than solving the barrier subproblem (FP_μ) accurately, we will be content with an approximate solution (x^k, λ^k, y^k) satisfying

$$\|\mathcal{R}_\mu(x^k, \lambda^k, y^k)\| \le \epsilon_\mu \text{ and } \lambda^k \ge -\epsilon_\mu e, \tag{3.15}$$

where ϵ_μ is a μ-related tolerance parameter.

3.2. Inner algorithm. In this section, we describe a modified Newton method for solving the barrier subproblem (FP_μ) with a fixed parameter μ. In order to achieve robustness, our IPM must successfully address:

◇ how to obtain acceptable search directions when $\text{In}(\mathcal{M}^k) \ne (n, p, 0)$;

◇ how do this so that rapid convergence of Newton's method is not interfered with once the iterates get close to a local solution.

Moreover, to ensure global convergence we need the following condition to hold at each iterate

$$\text{(C-5):} \begin{cases} d^\top \tilde{\mathcal{H}}^k d \ge \nu \|d\|^2, \ \forall d \in \Re^n \backslash \{0\}, & \text{if } \|g(x^k)\| > 0, \\ d^\top \tilde{\mathcal{H}}^k d \ge \nu \|d\|^2, \ \forall d \ne 0 \text{ such that } \nabla g(x^k)^\top d = 0, & \text{if } \|g(x^k)\| = 0, \end{cases}$$

where ν is a small positive parameter and the matrices $\tilde{\mathcal{H}}^k$ are defined by (3.9) and (3.5). It follows from Lemma 3.1 that if (C-5) holds, then $\text{In}(\mathcal{M}^k) = (n, p, 0)$. Also, (C-5) holds in a neighborhood of a local minimizer of problem (FP_μ) that satisfies the second-order sufficient conditions. However, outside of such a neighborhood, it may be necessary to modify $\mathcal{H}^k$ so that (C-5) holds. One approach for doing this is based on solving the system (3.3) by computing the $LBL^\top$ factorization of the symmetric indefinite matrix $\mathcal{M}^k$ (e.g., see [27]). This gives $P\mathcal{M}^kP^\top = LBL^\top$, where L is a unit lower triangular matrix, P is a permutation matrix, and B is a symmetric block diagonal matrix whose diagonal blocks are either 1×1 or 2×2. Since $\text{In}(B) = \text{In}(\mathcal{M}^k)$, one can use the inertia-controlling pivot strategy derived in [21] to determine if the matrix $\mathcal{H}^k$ needs to be modified, and if so, by how much. If $\text{In}(\mathcal{M}^k) = (n, p, 0)$, no modification is required. If $\text{In}(\mathcal{M}^k) \neq (n, p, 0)$, this strategy determines a positive quantity $\rho > 0$ and those diagonal elements of $\mathcal{H}^k$ to which ρ should be added so that the resulting matrix $\bar{\mathcal{M}}^k$ has inertia $(n, p, 0)$. Since this only ensures that $d^\top \tilde{\mathcal{H}}^k d > 0$ for any $d \in \Re^n \backslash \{0\}$ if $\|g(x^k)\| > 0$ or for any $d \neq 0$ such that $\nabla g(x^k)^\top d = 0$ if $\|g(x^k)\| = 0$, it may be necessary to add νI to $\mathcal{H}^k$ so that (C-5) holds.

After the search direction $\triangle x^k$ is computed, a step size $t_k \in (0, 1]$ and the next primal iterate $x^{k+1} = x^k + t_k \triangle x^k$ are determined by using a backtracking line search procedure; i.e., a decreasing sequence of step sizes $t_{k,j} = \beta^j$ $(j = 0, 1, ...)$ with $\beta \in (0, 1)$ are considered until the following criteria are satisfied:

$$
\begin{aligned}
&\text{T-1:} \quad c_i(x^k + t_{k,j}\triangle x^k) > 0, \ \forall i \in \mathcal{I}; \\
&\text{T-2:} \quad \Phi_{\mu,r_k}(x^k + t_{k,j}\triangle x^k) - \Phi_{\mu,r_k}(x^k) \leq -\sigma t_{k,j} (\triangle x^k)^\top \tilde{\mathcal{H}}^k \triangle x^k.
\end{aligned} \tag{3.16}
$$

Condition T-1 ensures quasi-feasibility. Condition T-2 guarantees a sufficient reduction in the merit function in view of condition (C-5). Note that T-2 is a modification of the standard Armijo rule applied to the merit function $\Phi_{\mu,r}(x)$ in that $\Phi'_{\mu,r_k}(x^k, \triangle x^k)$ is replaced by $-(\triangle x^k)^\top \tilde{\mathcal{H}}^k \triangle x^k$. For the dual iterates, we set for $i \in \mathcal{I}$

$$
u_i^{k+1} = \begin{cases} \lambda_i^k, & \text{if } \mu\gamma_{\min}/c_i(x^k) \leq \lambda_i^k \leq \mu\gamma_{\max}/c_i(x^k), \\ \mu\gamma_{\min}/c_i(x^k), & \text{if } \lambda_i^k < \mu\gamma_{\min}/c_i(x^k), \\ \mu\gamma_{\max}/c_i(x^k), & \text{if } \lambda_i^k > \mu\gamma_{\max}/c_i(x^k), \end{cases} \tag{3.17}
$$

where the parameters $\gamma_{\min}$ and $\gamma_{\max}$ satisfy $0 < \gamma_{\min} < 1 < \gamma_{\max}$. We require u_i^{k+1} to be less than $\mu\gamma_{\max}/c_i(x^k)$ to avoid ill-conditioned growth of u^k when the problem is well posed, i.e., $\{c(x^k)\}$ is uniformly bounded away from zero. The truncation value $\mu\gamma_{\min}/c_i(x^k)$ guarantees the strict dual feasibility of u_i^{k+1}.

We now describe our algorithm for solving the barrier subproblem (FP_μ). The main computational effort comes from solving the linear system (3.2) and modifying $\mathcal{H}^k$, if necessary, at each iteration.

ALGORITHM I. INNER ALGORITHM FOR SOLVING PROBLEM (FP_μ).

STEP 0: INITIALIZATION.

Parameters: $\epsilon_\mu > 0$, $\sigma \in (0, \frac{1}{2})$, $\beta \in (0, 1)$, $\chi > 1$, $\kappa_1 \in (0, 1)$, $\kappa_2 > 1$, $\pi_\mu = \max\{\mu, 0.1\}$, $\nu > 0$, $0 < \gamma_{\min} < 1 < \gamma_{\max}$.

Data: $x^0 \in \mathcal{F}^o$, $u^0 > 0$, $r_0 > 0$, $\mathcal{H}^0 \in \Re^{n\times n}$.

Set $k \leftarrow 0$.

REPEAT

STEP 1: SEARCH DIRECTION.

Modify $\mathcal{H}^k$, if necessary, so that (C-5) holds.

Compute $(\triangle x^k, \lambda^k, y^k)$, the solution of the linear system (3.2).

IF no solution is found THEN STOP with MFCQ FAILURE;
ELSE
STEP 2: TERMINATION.
IF (3.15) holds THEN STOP with SUCCESS;
ELSE
STEP 3: PENALTY PARAMETER UPDATE.
IF all conditions (C-1)$\sim$(C-4) hold THEN set $r_{k+1} \leftarrow \chi r_k$,
$(x^{k+1}, u^{k+1}, \mathcal{H}^{k+1}) \leftarrow (x^k, u^k, \mathcal{H}^k)$, $k \leftarrow k+1$;
ELSE
STEP 4: LINE SEARCH.
Compute the line search step size t_k according to (3.16).
STEP 5: UPDATE.
Set u^{k+1} as in (3.17), $x^{k+1} \leftarrow x^k + t_k \triangle x^k$, $r_{k+1} \leftarrow r_k$.
Compute the new Hessian or its estimate $\mathcal{H}^{k+1}$.
Set $k \leftarrow k+1$.
END IF
END IF
END IF
END

REMARK 3.1. *Algorithm I can terminate at Step 1 only if the coefficient matrix of the linear system (3.2) is singular. Since condition (C-5) holds, we must have that* $\delta_{x^k} = 0$, *(i.e.,* $\|g(x^k)\| = 0$*) and that the equality constraint gradients* $\{\nabla g_i(x^k),\ i \in \mathcal{E}\}$ *are linearly dependent according to Lemma 3.1. This implies that* x^k *is a FJ point of problem (NP) that does not satisfy the MFCQ.*

REMARK 3.2. *Algorithm I gives a very general framework for what to do at least on the theoretic level, but it does not consider many practical matters. For example, choices of* δ_x *other than* $\|g(x^k)\|/r_k$ *might lead to better performance. Also, it might not be good to directly return to Step 1 if the penalty parameter* r_k *is increased in Step 3. Instead, since the satisfaction of conditions (C-1)$\sim$(C-4) implies the need to improve feasibility, one might invoke a feasibility restoration phase as in filter methods in addition to increasing* r_k*. Such issues will be discussed in a future paper about implementation.*

We now show that Algorithm I is well defined. First, assumption A1 guarantees the existence of a strict quasi-feasible starting point $x^0 \in \mathcal{F}^o$. By Remark 3.1, condition (C-5) ensures that a unique solution $(\triangle x^k, \lambda^k, y^k)$ of the linear system (3.2) exists unless a FJ point that fails to satisfy the MFCQ is encountered and then Algorithm I terminates at Step 1. The following two lemmas establish that the line search step is well-defined.

LEMMA 3.5. *Suppose Algorithm I is at Step 4 of the kth iteration. Then we have* $\triangle x^k \neq 0$.

Proof. The proof is by contradiction. Assume to the contrary that $\triangle x^k = 0$. It follows from (3.2) that $C(x^k)\lambda^k = \mu e$ and $\nabla_x \mathcal{L}(x^k, \lambda^k, y^k) = 0$; hence, from (2.7) $\|\mathcal{R}_\mu(x^k, \lambda^k, y^k)\| = \|g(x^k)\|$. Since in the kth iteration Algorithm I hasn't terminated at Step 2 and is now at Step 4, we must have $\|g(x^k)\| > 0$ in view of (3.15). Furthermore, having successfully passed through Step 3 means that at least one condition among (C-1)$\sim$(C-4) is violated. Since conditions (C-1)$\sim$(C-3) are clearly satisfied by the above analysis, it follows that (C-4) is violated, i.e., $\|\varpi^k\| \geq \pi_\mu$. Hence it follows from (3.12) that $\|g(x^k)\| = 0$, a contradiction. ∎

LEMMA 3.6. *Suppose Algorithm I is at Step 4 of the kth iteration. Then there exists a* $\bar{t}_k \in (0,1]$ *such that for all* $t_{k,j} \in (0, \bar{t}_k]$, *(3.16) holds.*

Proof. Since $c_i(x^k) > 0$ and $c_i(x)$ is continuously differentiable for each $i \in \mathcal{I}$, we know there exists a $\tilde{t}_k \in (0,1]$ such that for all $t_{k,j} \in (0, \tilde{t}_k]$, condition T-1 of (3.16) holds. By Lemma 3.5 we have $\triangle x^k \neq 0$, which implies that $\Phi'_{\mu,r_k}(x^k; \triangle x^k) < 0$ in view of (3.8) and (C-5). Hence from Lemma 3.4, we can conclude that there exists a $\hat{t}_k \in (0, \tilde{t}_k]$ such that for all $t_{k,j} \in (0, \hat{t}_k]$,

$$
\begin{aligned}
& \Phi_{\mu,r_k}(x^k + t_{k,j}\triangle x^k) - \Phi_{\mu,r_k}(x^k) \\
\leq \; & \sigma t_{k,j} \Phi'_{\mu,r_k}(x^k; \triangle x^k) \\
\leq \; & -\sigma t_{k,j} (\triangle x^k)^\top \hat{\mathcal{H}}^k \triangle x^k,
\end{aligned}
$$

where the first inequality follows from the fact that $\sigma \in (0, \frac{1}{2})$. □

Lemma 3.6 indicates that the line search step is well defined. Consequently, at any iteration k Algorithm I either terminates at Step 1 with a FJ point of problem (NP) or Step 2 with an approximate KKT point of problem (FP$_\mu$) or readily generates the next iterate at Step 3 or Step 5. Therefore we have proved:

PROPOSITION 3.7. *Under assumptions A1 and A2, Algorithm I is well defined.*

3.3. Global convergence of Algorithm I. To establish the global convergence of Algorithm I when applied to the barrier subproblem (FP$_\mu$) for a fixed μ, we make the following additional assumptions.

A3. *The primal iterate sequence* $\{x^k\}$ *lies in a bounded set.*

A4. *The Hessian estimate sequence* $\{\mathcal{H}^k\}$ *is bounded.*

Assumption A3 can be ensured by adding simple bound constraints to the problem. This assumption and A2 guarantee that the merit function is bounded below. Assumptions A3 and A4 are standard; i.e., these assumptions or similar ones are required to prove convergence of other IPMs.

Let us assume for the present that Algorithm I generates an infinite sequence of iterates, i.e., it does not stop at Step 1 or Step 2 even if the termination criteria (3.15) are met. The next result analyzes the possible outcomes when the penalty parameter r_k tends to infinity.

THEOREM 3.8. *Under assumptions A1∼A4, if* r_k *is increased infinitely often then there exists a limit point* x^* *that is either an infeasible FJ point of the inequality constrained feasibility problem (FNP) or a FJ point of problem (NP) that fails to satisfy the MFCQ.*

Proof. Since r_k is increased infinitely many times, there exists an infinite index set $\mathcal{K}$ such that $r_{k+1} = \chi r_k$ for $k \in \mathcal{K}$. The fact $\chi > 1$ implies that $\{r_k\} \to \infty$ as $k \to \infty$. Define $\alpha_k = \max\{r_k, \|\lambda^k\|_\infty\}$ and $\bar{r}_k = \alpha_k^{-1} r_k$, $\bar{\lambda}_i^k = \alpha_k^{-1}\lambda_i^k$, $i \in \mathcal{I}$. By construction $\max\{\bar{r}_k, \|\bar{\lambda}^k\|_\infty\} = 1$ for all $k \in \mathcal{K}$. Moreover, by assumption A3, $\{x^k\}$ lies in a bounded set. Hence, there exists an infinite subset $\mathcal{K}' \subseteq \mathcal{K}$ such that $\{(\bar{r}_k, \bar{\lambda}^k)\}_{\mathcal{K}'} \to (\bar{r}, \bar{\lambda}) \neq 0$ and $\{x^k\}_{\mathcal{K}'} \to x^*$ with $x^* \in \mathcal{F}$. The criteria that trigger increasing of r_k must be satisfied, i.e., conditions (C-1)∼(C-4) must hold for all $k \in \mathcal{K}'$. Since $\alpha_k \to \infty$ for $k \in \mathcal{K} \to \infty$ and hence for $k \in \mathcal{K}' \to \infty$, condition (C-3) gives that $\bar{\lambda} \geq 0$ and $C(x^*)\bar{\lambda} = 0$. Hence, we have $\bar{\lambda}_i = 0$, $i \in \mathcal{I} \backslash \mathcal{I}^o(x^*)$. Furthermore, it follows from (3.2) and (3.13) that

$$
-\mathcal{H}^k \triangle x^k + \sum_{i \in \mathcal{I}} \lambda_i^k \nabla c_i(x^k) - \sum_{i \in \mathcal{E}} \left(r_k \frac{g_i(x^k)}{\|g(x^k)\|} + \varpi_i \right) \nabla g_i(x^k) = \nabla f(x^k) \tag{3.18}
$$

since $\|g(x^k)\| > 0$ for all $k \in \mathcal{K}$ in view of the satisfaction of condition (C-1). Note that $\{\|\triangle x^k\|\}_{\mathcal{K}}$ and $\{\|\varpi^k\|\}_{\mathcal{K}}$ are bounded as conditions (C-2) and (C-4) hold on $\mathcal{K}$. $\{\nabla g(x^k)\}$, $\{\nabla c(x^k)\}$ and $\{\nabla f(x^k)\}$ are bounded by assumptions A2 and A3. $\{\mathcal{H}^k\}$ is bounded by assumption A4. There are two cases to analyze.

Case 1. $\|g(x^*)\| > 0$. Dividing both sides of (3.18) by α_k and letting $k \in \mathcal{K}' \to \infty$ yields

$$\sum_{i\in\mathcal{I}^o(x^*)} \bar\lambda_i \nabla c_i(x^*) - \frac{\bar r}{\|g(x^*)\|}\sum_{i\in\mathcal{E}} g_i(x^*)\nabla g_i(x^*) = 0.$$

Since $(\bar r, \bar\lambda) \geq 0$ but $(\bar r, \bar\lambda) \neq 0$ and $x \in \mathcal{F}$, we conclude from (2.5) that x^* is a FJ point of problem (FNP).

Case 2. $\|g(x^*)\| = 0$. Then we have $x^* \in \mathcal{X}$. For all $k \in \mathcal{K}$ we have

$$\frac{1}{\sqrt{p}} \leq \max\left\{\frac{|g_i(x^k)|}{\|g(x^k)\|},\ i\in\mathcal{E}\right\} \leq 1.$$

Thus, there exists an infinite subset $\mathcal{K}' \subseteq \mathcal{K}$ and a vector $\bar y \neq 0$ such that

$$\left\{\frac{g_i(x^k)}{\|g(x^k)\|}\right\}_{\mathcal{K}'} \to \bar y_i, \quad i \in \mathcal{E}.$$

Dividing both sides of (3.18) by α_k and letting $k \in \mathcal{K}' \to \infty$ yields

$$\sum_{i\in\mathcal{I}^o(x^*)} \bar\lambda_i \nabla c_i(x^*) - \sum_{i\in\mathcal{E}} \bar r \bar y_i \nabla g_i(x^*) = 0.$$

Since $(\bar r, \bar\lambda) \neq 0$ and $\bar y \neq 0$, we must have $(\bar r\bar y, \bar\lambda) \neq 0$. Hence, since $\bar\lambda \geq 0$, it follows from Definitions 2.3 and 2.4 that x^* is a FJ point of problem (NP) that fails to satisfy the MFCQ. ∎

The following results analyze the convergence behavior of Algorithm I when the penalty parameter is increased only finitely many times.

LEMMA 3.9. *Suppose there exist $\bar k > 0$ and $\bar r > 0$ such that $r_k = \bar r$ for all $k \geq \bar k$. Then $\{c(x^k)\}$ and $\{u^k\}$ are componentwise bounded away from zero and $\{u^k\}$ is bounded above.*

Proof. The proof is by contradiction. Assume to the contrary that there exists an infinite subset $\mathcal{K}$ and an index $j \in \mathcal{I}$ such that $\{c_j(x^k)\}_{\mathcal{K}} \downarrow 0$. We know from the line search step that for all $k \geq \bar k$,

$$\Phi_{\mu,r_k}(x^k) = \Phi_{\mu,\bar r}(x^k) \leq \Phi_{\mu,\bar r}(x^{\bar k}). \tag{3.19}$$

Since f, c, g are continuous real valued functions on $\mathcal{F}$, assumption A3 implies that $\{f(x^k)\}$, $\{c(x^k)\}$ and $\{g(x^k)\}$ are all bounded both above and below. Thus, we have that $\{\Phi_{\mu,\bar r}(x^k)\}_{\mathcal{K}} \to \infty$, which contradicts (3.19). Hence $\{c(x^k)\}$ is componentwise bounded away from zero. Thus, (3.17) implies that $\{u^k\}$ is componentwise bounded away from zero and bounded above. ∎

LEMMA 3.10. *Suppose there exist $\bar k > 0$ and $\bar r > 0$ such that $r_k = \bar r$ for all $k \geq \bar k$. If there exists an infinite subset $\mathcal{K}$ such that the whole sequence $\{\|(\triangle x^k, \lambda^k, y^k)\|\}_{\mathcal{K}}$ tends to infinity, any limit point of the sequence $\{x^k\}_{\mathcal{K}}$ is a FJ point of problem (NP) that fails to satisfy the MFCQ.*

Proof. By assumptions A3, A4 and Lemma 3.9, there exists an infinite subset $\mathcal{K}' \subseteq \mathcal{K}$ such that $\{x^k\}_{\mathcal{K}'} \to x^*$, $\{u^k\}_{\mathcal{K}'} \to u^* > 0$, $\{\mathcal{H}^k\}_{\mathcal{K}'} \to \mathcal{H}^*$ and $C(x^*) \succ 0$ by Lemma 3.9. Thus, as $k \in \mathcal{K}' \to \infty$,

$$\mathcal{M}^k \to \mathcal{M}^* = \begin{bmatrix} \hat{\mathcal{H}}^* & \nabla g(x^*) \\ \nabla g(x^*)^\top & -\delta_{x^*} I \end{bmatrix},$$

where $\hat{\mathcal{H}}^* = \mathcal{H}^* + \nabla c(x^*) C(x^*)^{-1} \mathcal{U}^* \nabla c(x^*)^\top$, $\mathcal{U}^* = \text{diag}(u^*)$ and $\delta_{x^*} = \frac{\|g(x^*)\|}{\bar{r}}$. Since the right hand side of (3.2) is bounded and continuous, the unboundedness of $\{\triangle x^k, \lambda^k, y^k\}_{\mathcal{K}'}$ implies that the limit of the coefficient matrices of (3.2) and, hence, $\mathcal{M}^*$ are singular. If $\|g(x^*)\| > 0$, we can choose $\mathcal{K}'$ so that $\|g(x^k)\| > 0$ for all $k \in \mathcal{K}'$. Therefore, by continuity and condition (C-5), $\tilde{\mathcal{H}}^* = \hat{\mathcal{H}}^* + \frac{1}{\delta_{x^*}} \nabla g(x^*) \nabla g(x^*)^\top \succ 0$. But then by Lemma 3.1, $\mathcal{M}^*$ must be nonsingular, a contradiction. Hence, it follows that $\|g(x^*)\| = 0$.

Now we show that $\nabla g(x^*)$ is rank deficient. Suppose $\nabla g(x^*)$ has rank p. Then it follows that $\nabla g(x^k)$ has rank p for all sufficiently large $k \in \mathcal{K}'$. Hence, for any $d \neq 0$ such that $\nabla g(x^*)^\top d = 0$ and large enough $k \in \mathcal{K}'$, we can define

$$d^k = (I - \nabla g(x^k)(\nabla g(x^k)^\top \nabla g(x^k))^{-1} \nabla g(x^k)^\top) d.$$

For such k, $\nabla g(x^k)^\top d^k = 0$ and hence it follows from condition (C-5) that $(d^k)^\top \tilde{\mathcal{H}}^k d^k = (d^k)^\top \hat{\mathcal{H}}^k d^k \geq \nu \|d^k\|^2$. Since clearly, $\{d^k\}_{\mathcal{K}'} \to d$, letting $k \in \mathcal{K}' \to \infty$ yields that $d^\top \hat{\mathcal{H}}^* d \geq \nu \|d\|^2$. Thus, we again know from Lemma 3.1 that $\mathcal{M}^*$ is nonsingular, a contradiction. Therefore, x^* is a FJ point of problem (NP) that fails to satisfy the MFCQ. ∎

Lemma 3.10 implies that provided the penalty parameter is bounded, the multiplier sequence remains uniformly bounded if the MFCQ holds on the feasible region $\mathcal{X}$.

LEMMA 3.11. *Suppose there exist $\bar{k} > 0$ and $\bar{r} > 0$ such that $r_k = \bar{r}$ for all $k \geq \bar{k}$. If $\{\triangle x^k, \lambda^k, y^k\}_{\mathcal{K}}$ is a bounded subsequence then $\{\triangle x^k\}_{\mathcal{K}} \to 0$.*

Proof. Assume to the contrary that there exists an infinite subset $\hat{\mathcal{K}} \subseteq \mathcal{K}$ such that $\{\triangle x^k\}_{\hat{\mathcal{K}}} \to \triangle x^* \neq 0$. Since $\{x^k, \mathcal{H}^k, u^k, \lambda^k, y^k\}_{\mathcal{K}}$ are all bounded, $\{x^k\} \to x^*$, $\{\mathcal{H}^k\} \to \mathcal{H}^*$, $\mathcal{U}^k \to \mathcal{U}^* > 0$ and $\{(\lambda^k, y^k)\} \to (\lambda^*, y^*)$ as $k \in \mathcal{K}' \to \infty$, where $\mathcal{K}' \subseteq \hat{\mathcal{K}}$. Clearly,

$$\{(\triangle x^k)^\top \tilde{\mathcal{H}}^k \triangle x^k\}_{\mathcal{K}'} \to (\triangle x^*)^\top \tilde{\mathcal{H}}^* \triangle x^*, \tag{3.20}$$

as long as either $\|g(x^*)\| > 0$ or $g(x^k) = 0$ for all $k \in \mathcal{K}'$, where

$$\tilde{\mathcal{H}}^* = \begin{cases} \hat{\mathcal{H}}^*, & \text{if } \|g(x^*)\| = 0, \\ \hat{\mathcal{H}}^* + \delta_{x^*}^{-1} \nabla g(x^*) \nabla g(x^*)^\top, & \text{if } \|g(x^*)\| > 0, \end{cases} \tag{3.21}$$

and

$$\hat{\mathcal{H}}^* = \mathcal{H}^* + \nabla c(x^*) C(x^*)^{-1} \mathcal{U}^* \nabla c(x^*)^\top. \tag{3.22}$$

If $\|g(x^k)\| > 0$ for all $k \in \mathcal{K}'$ and $\|g(x^*)\| = 0$, we have that

$$\begin{aligned} (\triangle x^k)^\top \tilde{\mathcal{H}}^k \triangle x^k &= (\triangle x^k)^\top \hat{\mathcal{H}}^k \triangle x^k + \frac{1}{\delta_{x^k}} (\triangle x^k)^\top (\nabla g(x^k) \nabla g(x^k)^\top) \triangle x^k \\ &= (\triangle x^k)^\top \hat{\mathcal{H}}^k \triangle x^k + \delta_{x^k} \sum_{i \in \mathcal{E}} \left(\bar{r} \frac{g_i(x^k)}{\|g(x^k)\|} - y_i^k \right)^2 \end{aligned} \tag{3.23}$$

for all $k \in \mathcal{K}'$. Hence, (3.20) then follows from the boundedness of $\{y^k\}_{\mathcal{K}'}$ and $\{g(x^k)\}_{\mathcal{K}'}$ and the fact that $\{\delta_{x^k}\}_{\mathcal{K}'} \to 0$. Since c is continuously differentiable, we know there exists a $\tilde{t} \in (0, 1]$ such that for all $t \in (0, \tilde{t}]$,

$$c_i(x^* + t\triangle x^*) > 0, \ \forall i \in \mathcal{I}. \tag{3.24}$$

By continuity, letting $k \in \mathcal{K}' \to \infty$ in (3.2) yields

$$\left\{\begin{array}{llllll} \mathcal{H}^*\triangle x^* & -\nabla c(x^*)\lambda^* & +\nabla g(x^*)y^* & = & -\nabla f(x^*), \\ \mathcal{U}^*\nabla c(x^*)^\top \triangle x^* & +C(x^*)\lambda^* & & = & \mu e, \\ \nabla g(x^*)^\top \triangle x^* & & -\delta_{x^*}y^* & = & -g(x^*), \end{array}\right. \tag{3.25}$$

where $\delta_{x^*} = \frac{\|g(x^*)\|}{r_*}$. Since $\triangle x^* \neq 0$, we have from (3.20) and condition (C-5) that $-(\triangle x^*)^\top \tilde{\mathcal{H}}^* \triangle x^* < 0$. By following the same argument as in the proof of Lemma 3.4, we have

$$\Phi'_{\mu,\bar{r}}(x^*; \triangle x^*) \leq -(\triangle x^*)^\top \tilde{\mathcal{H}}^* \triangle x^* < 0. \tag{3.26}$$

Thus, we can conclude that there exists a $\hat{t} \in (0, \tilde{t}]$ such that for all $t \in (0, \hat{t}]$,

$$\begin{array}{ll} & \Phi_{\mu,\bar{r}}(x^* + t\triangle x^*) - \Phi_{\mu,\bar{r}}(x^*) \\ \leq & 1.1\sigma t \Phi'_{\mu,\bar{r}}(x^*; \triangle x^*) \\ < & -\sigma t (\triangle x^*)^\top \tilde{\mathcal{H}}^* \triangle x^*, \end{array} \tag{3.27}$$

since $0 < 1.1\sigma < 1$ as $\sigma \in (0, \frac{1}{2})$. Let $m_* = \min\{j | \beta^j \in (0, \hat{t}], j = 0, 1, 2...\}$. Due to the continuity of $\Phi_{\mu,\bar{r}}$, we know from (3.20), (3.24) and (3.27) that for $t_* = \beta^{m_*}$ and $k \in \mathcal{K}'$ large enough,

$$\left\{\begin{array}{l} c_i(x^k + t_*\triangle x^k) > 0, \ \forall i \in \mathcal{I}, \\ \Phi_{\mu,r_k}(x^k + t_*\triangle x^k) - \Phi_{\mu,r_k}(x^k) < -\sigma t_* (\triangle x^k)^\top \tilde{\mathcal{H}}^k \triangle x^k. \end{array}\right.$$

This implies the step size $t_k \geq t_*$ for all $k \in \mathcal{K}'$ large enough according to the backtracking line search rule. Consequently, it follows from condition (C-5) and (3.16) that for large enough $k \in \mathcal{K}'$

$$\begin{array}{ll} & \Phi_{\mu,\bar{r}}(x^{k+1}) \\ \leq & \Phi_{\mu,\bar{r}}(x^k) - \sigma t_k (\triangle x^k)^\top \tilde{\mathcal{H}}^k \triangle x^k \\ \leq & \Phi_{\mu,\bar{r}}(x^k) - \frac{1}{2}\sigma t_* \nu \|\triangle x^*\|. \end{array}$$

But since $\Phi_{\mu,\bar{r}}(x^k)$ is monotone decreasing, this implies that $\{\Phi_{\mu,\bar{r}}(x^k)\}_{\mathcal{K}'} \to -\infty$, a contradiction to assumptions A2 and A3. ∎

LEMMA 3.12. *Suppose there exist $\bar{k} > 0$ and $\bar{r} > 0$ such that $r_k = \bar{r}$ for all $k \geq \bar{k}$. If $\{\triangle x^k, \lambda^k, y^k\}_{\mathcal{K}}$ is a bounded subsequence, any limit point of $\{(x^k, \lambda^k, y^k)\}_{\mathcal{K}}$ satisfies the first order optimality conditions (2.7).*

Proof. The proof is by contradiction. Assume to the contrary that there exists an infinite subset $\mathcal{K}' \subseteq \mathcal{K}$ such that $\{(x^k, \lambda^k, y^k)\}_{\mathcal{K}'}$ tends to (x^*, λ^*, y^*) that fails to satisfy the first order optimality conditions (2.7). Since all components in (3.2) are bounded on $\mathcal{K}$ and $\{\triangle x^k\}_{\mathcal{K}} \to 0$ by Lemma 3.11, if we let $k \in \mathcal{K}' \to \infty$, (3.2) gives

$$\begin{array}{rcl} \nabla c(x^*)\lambda^* - \nabla g(x^*)y^* & = & \nabla f(x^*), \\ C(x^*)\lambda^* & = & \mu e, \\ \|g(x^*)\|y^* & = & r_* g(x^*). \end{array} \tag{3.28}$$

Thus, we must have $\|g(x^*)\| > 0$ due to the failure of (2.7) at (x^*, λ^*, y^*). Hence, by continuity we know $\|g(x^k)\| > 0$ for all $k \in \mathcal{K}'$ large enough, i.e., condition (C-1) holds. Moreover, $\{\triangle x^k\}_{\mathcal{K}} \to 0$ and the second equation of (3.28) imply the satisfaction of conditions (C-2) and (C-3), respectively, for $k \in \mathcal{K}'$ large enough. Finally, the last equation of (3.28) implies for $k \in \mathcal{K}'$ large enough condition (C-4) holds. But this satisfaction of conditions (C-1)$\sim$(C-4) contradicts the fact that $r_k = \bar{r}$ for all sufficiently large k. ☐

Combining the results of Theorem 3.8, Lemma 3.10 and Lemma 3.12, we are now in a position to state a global convergence theorem for Algorithm I for solving the barrier subproblem (FP_μ) with a fixed μ.

THEOREM 3.13. *Suppose Algorithm I generates an infinite sequence of iterates, i.e., it does not terminate at Step 1 or Step 2 even if the termination criteria have been met. Suppose assumptions A1$\sim$A4 hold. Then*

(i) if the penalty parameter $\{r_k\}$ is unbounded, there exists a limit point of $\{x^k\}$ that is either an infeasible FJ point of the inequality constrained feasibility problem (FNP), or a FJ point of problem (NP) failing to satisfy MFCQ;

(ii) if $\{r_k\}$ is bounded and $\{\triangle x^k, \lambda^k, y^k\}_{\mathcal{K}}$ is an unbounded subsequence containing no bounded infinite subsequences, any limit point of $\{x^k\}$ is a FJ point of problem (NP) failing to satisfy MFCQ;

(iii) if $\{r_k\}$ is bounded and $\{\triangle x^k, \lambda^k, y^k\}_{\mathcal{K}}$ is a bounded subsequence, any limit point of $\{(x^k, \lambda^k, y^k)\}_{\mathcal{K}}$ satisfies the first order optimality condition (2.7).

THEOREM 3.14. *Suppose assumptions A1$\sim$A4 hold. The outcome of applying Algorithm I is one of the following.*

(A) *A FJ point of problem (NP) failing to satisfy MFCQ is found at Step 1 in a finite number of iterations.*

(B) *An approximate KKT point of problem (FP_μ) satisfying the termination criteria (3.15) is found at Step 2 in a finite number of iterations.*

(C) *Finite termination does not occur and there exists a limit point that is either an infeasible FJ point of the inequality constrained feasibility problem (FNP), or a FJ point of problem (NP) failing to satisfy MFCQ.*

3.4. Overall algorithm. Our quasi-feasible IPM for solving problem (NP) successively solves the barrier subproblem (FP_μ) for a decreasing sequence $\{\mu\}$ by applying Algorithm I. Ideally, one would choose a sequence $\{\mu\}$ that converges to zero at a superlinear rate. This was shown by Zhang, Tapia and Dennis [16] to be a necessary condition for superlinear convergence of primal iterates. In practice, however, a fast linear reduction of $\{\mu\}$ will serve just as well, because the cost of computing the final iterates is typically much less than the cost of computing the initial iterates. We refer the reader to [16] for further discussion about updating $\{\mu\}$.

The tolerance ϵ_μ in (3.15), which determines the accuracy in the solution of the barrier subproblems, is decreased from one barrier subproblem to the next and must converge to zero. In this paper we use the simple strategy of reducing both ϵ_μ and μ by a constant factor $\alpha \in (0, 1)$. Moreover, another μ-related parameter π_μ is used in each barrier subproblem as a threshold value for conditions (C-2) and (C-4). Although there are no restrictive requirements on π_μ to prove global convergence, care must be taken in a practical implementation to avoid both too infrequent and too frequent increases of the penalty parameter. Finally, we test optimality for problem (NP) by means of the residual norm $\|\mathcal{R}(x, \lambda, y)\|$.

We are ready to state the overall IPM, in which the index k denotes an outer iteration, while j denotes the last inner iteration of Algorithm I.

ALGORITHM II. OUTER ALGORITHM FOR SOLVING PROBLEM (NP).

Specify parameters $\mu_0 > 0$, $\epsilon_{\mu_0} > 0$, $\alpha \in (0,1)$ and the final step tolerance $\epsilon_{\inf}$.
Choose the starting point $x^0 \in \mathcal{F}^o$, $u^0 \in \Re^m_+$, $\lambda^0 \in \Re^m$, $y^0 \in \Re^p$, the initial penalty parameter $r_0 > 0$ and compute the initial Hessian or its estimate $\mathcal{H}^0$.
$k \leftarrow 0$.
REPEAT until $\|\mathcal{R}(x^k, \lambda^k, y^k)\| \le \epsilon_{\inf}$:

1. Apply Algorithm I, starting from $(x^k, u^k, r_k, \mathcal{H}^k)$ with parameters $(\mu_k, \epsilon_{\mu_k}, \pi_{\mu_k})$, to find an approximate solution $(x^{k,j}, \lambda^{k,j}, y^{k,j})$ of the barrier subproblem (FP$_{\mu_k}$), which satisfies termination criteria (3.15).
2. Set $\mu_{k+1} \leftarrow \alpha\mu_k$, $\epsilon_{\mu_{k+1}} \leftarrow \alpha\epsilon_{\mu_k}$, $x^{k+1} \leftarrow x^{k,j}$, $u^{k+1} \leftarrow u^{k,j}$, $\lambda^{k+1} \leftarrow \lambda^{k,j}$, $y^{k+1} \leftarrow y^{k,j}$, $r_{k+1} \leftarrow r_{k,j}$, $\mathcal{H}^{k+1} \leftarrow \mathcal{H}^{k,j}$.
Set $k \leftarrow k+1$.

END

The next result gives the global convergence of Algorithm II.

THEOREM 3.15. *Suppose Algorithm II is applied by ignoring its termination condition. Suppose assumptions A1 and A2 hold, A3 holds with the same bound for every μ_k and A4 holds for each μ_k. Then, one of the following outcomes occurs.*

(A) *For some parameter μ_k, the termination criteria (3.15) of Algorithm I are never met, in which case there exists a limit point of the inner iterates that is either an infeasible FJ point of problem (FNP) or a FJ point of problem (NP) failing to satisfy the MFCQ.*

(B) *Algorithm I successfully terminates for each μ_k, in which case the limit point of any bounded subsequence $\{x^k, \lambda^k, y^k\}_{\mathcal{K}}$ satisfies the first order optimality condition (2.1); while the limit point of any subsequence $\{x^k\}_{\mathcal{K}}$ with unbounded multiplier subsequence $\{(\lambda^k, y^k)\}_{\mathcal{K}}$ is a FJ point of problem (NP) failing to satisfy the MFCQ.*

Proof. Outcome (A) follows immediately from Theorem 3.14. We now consider outcome (B). There are two cases.

Case 1. Suppose $\{x^k, \lambda^k, y^k\}_{\mathcal{K}}$ is a bounded subsequence with a limit point (x^*, λ^*, y^*). Since $\{\mu_k, \epsilon_{\mu_k}\} \to 0$ in Algorithm II, it follows from (3.15) that

$$\{\mathcal{R}_{\mu_k}(x^k, \lambda^k, y^k)\}_{\mathcal{K}} \to \mathcal{R}(x^*, \lambda^*, y^*) = 0 \text{ and } \{\lambda^k\}_{\mathcal{K}} \to \lambda^* \ge 0,$$

and $x^* \in \mathcal{F}$. Hence, the KKT conditions (2.1) hold at (x^*, λ^*, y^*).

Case 2. Suppose $\{x^k, \lambda^k, y^k\}_{\mathcal{K}}$ is an unbounded subsequence. By assumption A3, $\{x^k\}$ is bounded and WLOG, we can assume $\{x^k\}_{\mathcal{K}} \to x^*$. Let

$$\xi_k = \max\{\|\lambda^k\|_\infty, \|y^k\|_\infty, 1\}.$$

Define

$$\bar\lambda^k = \xi_k^{-1}\lambda^k \text{ and } \bar y^k = \xi_k^{-1} y^k.$$

Obviously, $\{(\bar\lambda^k, \bar y^k)\}$ is bounded. Hence, there exists an infinite subset $\mathcal{K}' \subseteq \mathcal{K}$ such that $\{(\bar\lambda^k, \bar y^k)\}_{\mathcal{K}'} \to (\bar\lambda, \bar y) \ne 0$ as $\max\{\|\bar\lambda^k\|_\infty, \|\bar y^k)\|_\infty\} = 1$ for large $k \in \mathcal{K}$. From (3.15) we know $\bar\lambda \ge 0$ and $g(x^*) = 0$. Dividing both sides of the first inequality of (3.15) by ξ_k and letting $k \in \mathcal{K}' \to \infty$ and observing that $\xi_k \to \infty$ yields

$$\begin{aligned} \nabla c(x^*)\bar\lambda - \nabla g(x^*)\bar y &= 0, \\ C(x^*)\bar\lambda &= 0. \end{aligned}$$

Since x^* is a feasible point, from Definitions 2.3 and 2.4 we know that x^* is a FJ point of problem (NP) failing to satisfy MFCQ. ∎

Theorem 3.15 implies that if Algorithm I successfully terminates for each μ_k, Algorithm II terminates finitely with either an approximate KKT point of problem (NP) or an approximate FJ point of problem (NP) failing to satisfy the MFCQ.

The next result analyzes the behavior of the penalty parameter for the overall algorithm.

THEOREM 3.16. *Suppose assumptions A1 and A2 hold, A3 holds with the same bound for every μ_k and A4 holds for each μ_k. Moreover, suppose the MFCQ holds on the feasible region and problem (NP) is not locally infeasible on $\mathcal{F}$; i.e., there are no FJ points of problem (FNP) on $\mathcal{F}$ that are infeasible to (NP). Then ignoring the termination condition of Algorithm II, the penalty parameter is uniformly bounded and any limit point of the sequence $\{x^k, \lambda^k, y^k\}$ generated by Algorithm II satisfies the first order optimality conditions (2.1).*

Proof. From Theorem 3.14 we know that Algorithm I successfully terminates for each μ_k since the MFCQ holds on the feasible region and problem (NP) is not locally infeasible on $\mathcal{F}$. Now assume to the contrary that there exists an infinite subset $\mathcal{K} = \{k, j\}$ in which the penalty parameter is increased, where k denotes an outer iteration and j denotes an inner iteration. There exists a limit point x^* of $\{x^{k,j}\}_{\mathcal{K}}$ in view of assumption A3. By following the same argument as in the proof of Theorem 3.8, we obtain that either the MFCQ is violated at x^* or problem (NP) is locally infeasible at x^*. This gives a contradiction. Hence, the penalty parameter is uniformly bounded. The last statement of the theorem then follows from Theorem 3.15. ∎

4. An infeasible IPM. In this section we present an infeasible IPM for solving the slack-based problem (SP) and hence problem (NP). This method is essentially our quasi-feasible IPM in which advantage is taken of the special nature and structure of the slack variables. We first describe the inner algorithm for solving the barrier subproblem (IP_μ) with a fixed μ.

4.1. Inner algorithm. The first order optimality conditions for problem (IP_μ) are

$$\mathcal{R}_\mu(x, w, u, v) = \begin{bmatrix} \nabla_x \mathcal{L}(x, u, v) \\ \mathcal{W}u - \mu e \\ c(x) - w \\ g(x) \end{bmatrix} = 0, \tag{4.1}$$

where $\mathcal{W} = \text{diag}(w)$ and $w > 0$. Throughout this section, we assume

A2'. *Functions f, c, g are real valued and twice continuously differentiable on $\Re^n$.*

Letting $\delta_{x,w} = \frac{\psi(x,w)}{r}$, the necessary conditions for (x, w) to be a local optimal solution for problem ($\ell_2\text{IP}_\mu$), or equivalently a perturbation of the KKT conditions (4.1) are,

$$\begin{aligned} \nabla f(x) - \textstyle\sum_{i\in\mathcal{I}} u_i \nabla c_i(x) + \sum_{i\in\mathcal{E}} v_i \nabla g_i(x) &= 0, \\ u_i w_i - \mu &= 0, \quad i \in \mathcal{I}, \\ w_i - c_i(x) - u_i \delta_{x,w} &= 0, \quad i \in \mathcal{I}, \\ g_i(x) - v_i \delta_{x,w} &= 0, \quad i \in \mathcal{E}. \end{aligned} \tag{4.2}$$

Treating $\delta_{x,w}$ as a constant and applying Newton's method to (4.2) yields the following linear system, that has to be solved at the kth step after defining $\lambda^k = u^k + \triangle u^k$ and

$y^k = v^k + \triangle v^k$,

$$\text{(4.3)} \quad \begin{cases} \mathcal{H}^k \triangle x^k & & -\nabla c(x^k)\lambda^k & +\nabla g(x^k) y^k & = & -\nabla f(x^k), \\ & \mathcal{U}^k \triangle w^k & +\mathcal{W}^k \lambda^k & & = & \mu e, \\ -\nabla c(x^k)^\top \triangle x^k & +\triangle w^k & -\delta_{x^k,w^k}\lambda^k & & = & c(x^k) - w^k, \\ \nabla g(x^k)^\top \triangle x^k & & & -\delta_{x^k,w^k} y^k & = & -g(x^k), \end{cases}$$

where $\mathcal{H}^k$ is the Lagrangian Hessian estimate, $\mathcal{U}^k = \text{diag}(u^k)$, $\delta_{x^k,w^k} = \frac{\psi(x^k,w^k)}{r_k}$ and r_k is the current penalty parameter. We maintain the strict positivity of w^k and, as in Algorithm I, the strict dual feasibility of u^k for all k.

If at the current iteration $\psi(x^k, w^k)$ is positive, from the third and forth equations of (4.3), we have that

$$\text{(4.4)} \qquad \psi(x^k, w^k)\varpi_i^k = \begin{cases} r_k(\triangle w_i^k - \nabla c_i(x^k)^\top \triangle x^k), & i \in \mathcal{I}, \\ r_k \nabla g_i(x^k)^\top \triangle x^k, & i \in \mathcal{E}, \end{cases}$$

where

$$\text{(4.5)} \qquad \varpi_i^k = \begin{cases} \lambda_i^k - \frac{r_k}{\psi(x^k,w^k)}(c_i(x^k) - w_i^k), & i \in \mathcal{I}, \\ y_i^k - \frac{r_k}{\psi(x^k,w^k)} g_i(x^k), & i \in \mathcal{E}. \end{cases}$$

Therefore as in Algorithm I, the following conditions, whose purpose and meaning are analogous to those of conditions (C-1)$\sim$(C-4), are checked at each iteration,

$$\text{(4.6)} \qquad \begin{array}{ll} \text{(D-1):} & \psi(x^k, w^k) > 0; \\ \text{(D-2):} & \|\triangle x^k\| \le \pi_\mu \text{ and } \|\triangle w^k\| \le \pi_\mu; \\ \text{(D-3):} & \kappa_1 \mu e \le \mathcal{W}^k \lambda^k \le \kappa_2 \mu e; \\ \text{(D-4):} & \|\varpi^k\| < \pi_\mu. \end{array}$$

Since $\mathcal{U}^k \succ 0$ and $\mathcal{W}^k \succ 0$, from (4.3) we obtain that

$$\text{(4.7)} \qquad \begin{array}{l} \triangle w^k = (\mathcal{U}^k)^{-1}(\mu e - \mathcal{W}^k \lambda^k), \\ \lambda^k = -\mathcal{D}^k(\nabla c(x^k)^\top \triangle x^k + c(x^k) - w^k - (\mathcal{U}^k)^{-1}\mu e), \end{array}$$

where

$$\text{(4.8)} \qquad \mathcal{D}^k = ((\mathcal{U}^k)^{-1}\mathcal{W}^k + \delta_{x^k,w^k} I)^{-1}$$

Substituting (4.7) into (4.3) yields a smaller and denser symmetric linear system

$$\text{(4.9)} \quad \mathcal{M}^k \begin{bmatrix} \triangle x^k \\ y^k \end{bmatrix} = \begin{bmatrix} -\nabla f(x^k) + \nabla c(x^k)\mathcal{D}^k(w^k - c(x^k) + \mu(\mathcal{U}^k)^{-1}e) \\ -g(x^k) \end{bmatrix},$$

where

$$\text{(4.10)} \qquad \mathcal{M}^k = \begin{bmatrix} \hat{\mathcal{H}}^k & \nabla g(x^k) \\ \nabla g(x^k)^\top & -\delta_{x^k,w^k} I \end{bmatrix},$$

$$\text{(4.11)} \qquad \hat{\mathcal{H}}^k = \mathcal{H}^k + \nabla c(x^k)\mathcal{D}^k \nabla c(x^k)^\top.$$

Again, we know from Lemma 3.1 that the inertia of the coefficient matrix of (4.9) is $(n, p, 0)$ if, and only if, either $\delta_{x^k,w^k} > 0$ and $\tilde{\mathcal{H}}^k$ is positive definite, where

$$\text{(4.12)} \qquad \tilde{\mathcal{H}}^k = \begin{cases} \hat{\mathcal{H}}^k, & \text{if } \psi(x^k, w^k) = 0, \\ \hat{\mathcal{H}}^k + \delta_{x^k,w^k}^{-1} \nabla g(x^k) \nabla g(x^k)^\top, & \text{if } \psi(x^k, w^k) > 0; \end{cases}$$

or $\psi(x^k, w^k) = 0$, $\nabla g(x^k)$ has rank p and $\hat{\mathcal{H}}^k$ is positive definite on the null space of $\nabla g(x^k)^\top$. By using the same strategy as in our quasi-feasible method, we can ensure that for a given $\nu > 0$, at each iteration

$$\text{(D-5):} \quad \begin{cases} d^\top \tilde{\mathcal{H}}^k d \geq \nu \|d\|^2, \ \forall d \in \Re^n \backslash \{0\}, & \text{if } \psi(x^k, w^k) > 0, \\ d^\top \tilde{\mathcal{H}}^k d \geq \nu \|d\|^2, \ \forall d \neq 0 \text{ such that } \nabla g(x^k)^\top d = 0, & \text{if } \psi(x^k, w^k) = 0, \end{cases}$$

and hence that $(\triangle x^k, \triangle w^k)$ is a descent direction of the merit function Φ_{μ, r_k}, since we have

LEMMA 4.1. *Suppose at the kth iteration, $u^k > 0$, $w^k > 0$, and the linear system (4.3) has a solution $(\triangle x^k, \triangle w^k, \lambda^k, y^k)$. Then*

$$\Phi'_{\mu, r_k}(x^k, w^k; \triangle x^k, \triangle w^k) \leq -(\triangle x^k)^\top \tilde{\mathcal{H}}^k \triangle x^k - \Theta_k(\triangle x^k, \triangle w^k), \tag{4.13}$$

where

$$\Theta_k(\triangle x^k, \triangle w^k) = \begin{cases} \delta^{-1}_{x^k, w^k} (p^k)^\top \mathcal{V}^k p^k, & \textit{if } \psi(x^k, w^k) > 0, \\ 0, & \textit{if } \psi(x^k, w^k) = 0, \end{cases} \tag{4.14}$$

$\mathcal{V}^k = (\mathcal{U}^k)^{-1} \mathcal{W}^k \mathcal{D}^k$ and $p^k = (\mathcal{V}^k)^{-1} \triangle w^k - \nabla c(x^k)^\top \triangle x^k$.

Proof. If $\psi(x^k, w^k) > 0$, it follows from (4.3) that

$$\begin{aligned} \lambda^k &= \delta^{-1}_{x^k, w^k} (\triangle w^k - \nabla c(x^k)^\top \triangle x^k - c(x^k) + w^k), \\ y^k &= \delta^{-1}_{x^k, w^k} (g(x^k) + \nabla g(x^k)^\top \triangle x^k). \end{aligned} \tag{4.15}$$

Substituting (4.15) into (4.3) yields

$$\begin{bmatrix} -\mathcal{B}^k & \delta^{-1}_{x^k, w^k} \nabla c(x^k) \\ \delta^{-1}_{x^k, w^k} \nabla c(x^k)^\top & -\mathcal{J}^k \end{bmatrix} \begin{bmatrix} \triangle x^k \\ \triangle w^k \end{bmatrix} = \begin{bmatrix} \nabla_x \Phi_{\mu, r_k}(x^k, w^k) \\ \nabla_w \Phi_{\mu, r_k}(x^k, w^k) \end{bmatrix}, \tag{4.16}$$

where

$$\begin{aligned} &\mathcal{B}^k = \mathcal{H}^k + \delta^{-1}_{x^k, w^k} (\nabla c(x^k) \nabla c(x^k)^\top + \nabla g(x^k) \nabla g(x^k)^\top), \\ &\mathcal{J}^k = (\mathcal{W}^k)^{-1} \mathcal{U}^k + \delta^{-1}_{x^k, w^k} I, \\ &\nabla_x \Phi_{\mu, r_k}(x^k, w^k) = \nabla f(x^k) + r_k \psi(x^k, w^k)^{-1} (\nabla c(x^k)(c(x^k) - w^k) + \nabla g(x^k) g(x^k)), \\ &\nabla_w \Phi_{\mu, r_k}(x^k, w^k) = -\mu (\mathcal{W}^k)^{-1} e + r_k \psi(x^k, w^k)^{-1} (w^k - c(x^k)). \end{aligned}$$

Since $\mathcal{V}^k = I - \delta_{x^k, w^k} \mathcal{D}^k = \delta^{-1}_{x^k, w^k} (\mathcal{J}^k)^{-1}$, we have that

$$\mathcal{B}^k = \widetilde{\mathcal{H}}^k + \delta^{-2}_{x^k, w^k} \nabla c(x^k) (\mathcal{J}^k)^{-1} \nabla c(x^k)^\top,$$

and hence, that

$$\begin{aligned} & \Phi'_{\mu, r_k}(x^k, w^k; \triangle x^k, \triangle w^k) \\ = \ & \nabla_x \Phi_{\mu, r_k}(x^k, w^k)^\top \triangle x^k + \nabla_w \Phi_{\mu, r_k}(x^k, w^k)^\top \triangle w^k \\ \leq \ & -(\triangle x^k)^\top \mathcal{B}^k \triangle x^k + 2\delta^{-1}_{x^k, w^k} (\triangle w^k)^\top \nabla c(x^k)^\top \triangle x^k - (\triangle w^k)^\top \mathcal{J}^k \triangle w^k \\ = \ & -(\triangle x^k)^\top \tilde{\mathcal{H}}^k \triangle x^k - \delta^{-2}_{x^k, w^k} (\triangle x^k)^\top \nabla c(x^k) (\mathcal{J}^k)^{-1} \nabla c(x^k)^\top \triangle x^k \\ & -(\triangle w^k)^\top \mathcal{J}^k (\mathcal{J}^k)^{-1} \mathcal{J}^k \triangle w^k + 2\delta^{-1}_{x^k, w^k} (\triangle w^k)^\top \mathcal{J}^k (\mathcal{J}^k)^{-1} \nabla c(x^k)^\top \triangle x^k \\ = \ & -(\triangle x^k)^\top \tilde{\mathcal{H}}^k \triangle x^k - \Theta_k(\triangle x^k, \triangle w^k). \end{aligned}$$

On the other hand, if $\psi(x^k, w^k) = 0$, we have from (4.3) that $\nabla c(x^k)^\top \triangle x^k = \triangle w^k$ and $\nabla g(x^k)^\top \triangle x^k = 0$. Hence, it follows from an obvious extension of Lemma 3.2 and (4.3) that

$$\begin{array}{rl} & \Phi'_{\mu,r_k}(x^k, w^k; \triangle x^k, \triangle w^k) \\ \leq & \nabla f(x^k)^\top \triangle x^k - \mu e^\top (\mathcal{W}^k)^{-1} \triangle w^k \\ = & -(\triangle x^k)^\top \mathcal{H}^k \triangle x^k + (\lambda^k)^\top \nabla c(x^k)^\top \triangle x^k \\ & -(y^k)^\top \nabla g(x^k)^\top \triangle x^k - \mu e^\top (\mathcal{W}^k)^{-1} \triangle w^k \\ = & -(\triangle x^k)^\top \mathcal{H}^k \triangle x^k + ((\lambda^k)^\top \mathcal{W}^k - \mu e^\top)(\mathcal{W}^k)^{-1} \triangle w^k \\ = & -(\triangle x^k)^\top \mathcal{H}^k \triangle x^k - (\triangle w^k)^\top (\mathcal{W}^k)^{-1} \mathcal{U}^k \triangle w^k \\ = & -(\triangle x^k)^\top \hat{\mathcal{H}}^k \triangle x^k \\ = & -(\triangle x^k)^\top \tilde{\mathcal{H}}^k \triangle x^k. \end{array}$$

Since $\Theta_k(\triangle x^k, \triangle w^k) = 0$, this completes the proof. ☐

After a search direction $(\triangle x^k, \triangle w^k)$ is computed, we obtain the next iterate $(x^{k+1}, w^{k+1}) = (x^k, w^k) + t_k(\triangle x^k, \triangle w^k)$ by a basic backtracking line search procedure; i.e., we consider a decreasing sequence of step sizes $t_{k,j} = \tau\beta^j$ $(j = 0, 1, ...)$ with $\beta \in (0, 1)$ until the following criteria are satisfied,

$$\begin{array}{ll} \text{S-1:} & w_i^k + t_{k,j} \triangle w_i^k > 0, \ \forall i \in \mathcal{I}; \\ \text{S-2:} & \Phi_{\mu,r_k}(x^k + t_{k,j} \triangle x^k, w^k + t_{k,j} \triangle w^k) - \Phi_{\mu,r_k}(x^k, w^k) \\ & \leq -\sigma t_{k,j} ((\triangle x^k)^\top \tilde{\mathcal{H}}^k \triangle x^k + \Theta_k(\triangle x^k, \triangle w^k)). \end{array} \tag{4.17}$$

Since the maximum value τ of $t_{k,j}$ that satisfies S-1 can easily be determined by the standard minimum ratio test $\tau = \min\{-\frac{w_i^k}{\triangle w_i^k} | \triangle w_i^k < 0, i \in \mathcal{I}\}$, the backtracking procedure could be started with a step size that is smaller than $\beta^0 = 1$. For the dual iterates, we set

$$u_i^{k+1} = \begin{cases} \lambda_i^k, & \text{if } \mu\gamma_{\min}/w_i^k \leq \lambda_i^k \leq \mu\gamma_{\max}/w_i^k, \\ \mu\gamma_{\min}/w_i^k, & \text{if } \lambda_i^k < \mu\gamma_{\min}/w_i^k, \\ \mu\gamma_{\max}/w_i^k, & \text{if } \lambda_i^k > \mu\gamma_{\max}/w_i^k, \end{cases} \tag{4.18}$$

where $0 < \gamma_{\min} < 1 < \gamma_{\max}$.

Finally, rather than solving the barrier subproblem (IP$_\mu$) accurately, we will be content with an approximate solution $(x^k, w^k, \lambda^k, y^k)$ satisfying

$$\|\mathcal{R}_\mu(x^k, w^k, \lambda^k, y^k)\| \leq \epsilon_\mu \text{ and } \lambda^k \geq -\epsilon_\mu e. \tag{4.19}$$

We are now ready to state the inner algorithm for our infeasible IPM.

Algorithm III. Inner Algorithm for Solving Problem (IP$_\mu$).

STEP 0: INITIALIZATION.
Parameters: $\epsilon_\mu > 0$, $\sigma \in (0, \frac{1}{2})$, $\beta \in (0, 1)$, $\chi > 1$, $\kappa_1 \in (0, 1)$, $\kappa_2 > 1$, $\pi_\mu = \max\{\mu, 0.1\}$, $\nu > 0$, $0 < \gamma_{\min} < 1 < \gamma_{\max}$.
Data: $(x^0, w^0) \in \Re^{n+m}$ with $w^0 > 0$, $u^0 > 0$, $r_0 > 0$, $\mathcal{H}^0 \in \Re^{n \times n}$.
$k \leftarrow 0$.
REPEAT
STEP 1: SEARCH DIRECTION.
Modify $\mathcal{H}^k$, if necessary, so that (D-5) holds.

Compute $(\triangle x^k, \triangle w^k, \lambda^k, y^k)$, the solution of the linear system (4.3).
IF no solution is found THEN STOP with MFCQ FAILURE;
ELSE
STEP 2: TERMINATION.
IF (4.19) holds THEN STOP with SUCCESS;
ELSE
STEP 3: PENALTY PARAMETER UPDATE.
IF all conditions (D-1)∼(D-4) hold THEN set $r_{k+1} \leftarrow \chi r_k$, $(x^{k+1}, w^{k+1}, u^{k+1}, \mathcal{H}^{k+1}) \leftarrow (x^k, w^k, u^k, \mathcal{H}^k)$, $k \leftarrow k+1$;
ELSE
STEP 4: LINE SEARCH.
Compute the line search step size t_k according to (4.17).
STEP 5: UPDATE.
Set u^{k+1} as in (4.18), $x^{k+1} \leftarrow x^k + t_k \triangle x^k$, $w^{k+1} \leftarrow w^k + t_k \triangle w^k$, $r_{k+1} \leftarrow r_k$.
Compute the new Hessian or its estimate $\mathcal{H}^{k+1}$.
Set $k \leftarrow k+1$.
END IF
END IF
END IF
END

By an argument analogous to those given at the end of Section 3.2 we have:

PROPOSITION 4.2. *Under assumption A2', Algorithm III is well defined.*

4.2. Global convergence of Algorithm III. To analyze the application of Algorithm III to problem (IP_μ), we shall first show that this is essentially equivalent to the application of Algorithm I to that problem. It is easy to verify that the latter results in the following Newton system (see (3.2)) for determining the step at iteration k

$$\begin{bmatrix} \mathcal{H}^k & 0 & 0 & -\nabla c(x^k) & \nabla g(x^k) \\ 0 & 0 & -I & I & 0 \\ 0 & \mathcal{U}^k & \mathcal{W}^k & 0 & 0 \\ -\nabla c(x^k)^\top & I & 0 & -\delta_{x^k,w^k} I & 0 \\ \nabla g(x^k)^\top & 0 & 0 & 0 & -\delta_{x^k,w^k} I \end{bmatrix} \begin{bmatrix} \triangle x^k \\ \triangle w^k \\ v^k \\ \lambda^k \\ y^k \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \\ \mu e \\ c(x^k) - w^k \\ -g(x^k) \end{bmatrix}, \tag{4.20}$$

where u_i^k and v_i^k are the dual variables associated with the ith equality constraint $c_i(x) - w_i = 0$ and ith nonnegativity constraint $w_i \geq 0$, respectively, in (IP_μ). Since the second block of equations in (4.20) gives $v^k = \lambda^k$, eliminating v^k then gives the system (4.3) used by Algorithm III to determine the step at iteration k.

It is easy to see that all other aspects of Algorithm III are equivalent to those of Algorithm I applied to problem (IP_μ). In practice, however, the algorithms may follow different paths depending on how the line search is carried out (recall that in Algorithm III a minimum ratio test could be used to determine a trial step length) and how condition (D-5) is ensured. However, if sufficient care is taken in applying Algorithm I, then the paths can be the same. For example, it can be shown that the matrix $\mathcal{M}^k$ given by (4.10), (4.11) and (4.8) is equal to the Schur complement of the submatrix

$$\begin{bmatrix} \mathcal{U}^k(\mathcal{W}^k)^{-1} & I \\ I & -\delta_{x^k,w^k} I \end{bmatrix}$$

in the matrix $\mathcal{M}^k$ given by (3.4) and (3.5) that arises in this case from the coefficient matrix in (4.20). Thus a carefully implemented inertia-controlling pivot strategy can result in the same perturbation to the matrix $\mathcal{H}^k$ in both algorithms.

In spite of these implementational differences, the global convergence properties of Algorithm III follow directly from those of Algorithm I after two issues are clarified. First we will make assumptions A3 and A4; however, we will not assume that the slack variable sequence $\{w^k\}$ lies in a bounded set. Also, when the penalty parameter r_k tends to infinity in Algorithm III, we need to be concerned with the unconstrained infeasibility problem (INP) rather than the constrained feasibility feasibility problem (FNP). Concerning the sequence $\{w^k\}$ we have

LEMMA 4.3. *Under assumptions A2' and A3, the sequence $\{w^k\}$ lies in a bounded set.*

Proof. First we consider the case that $r_k = \bar{r}$ for all sufficiently large k. Since $\{c(x^k)\}$ is bounded under assumptions A2' and A3, if $\{\|w^k\|\}_{\mathcal{K}} \to \infty$, where $\mathcal{K}$ is an infinite index set, then $\{\psi(x^k, w^k)\}_{\mathcal{K}} \to \infty$ and hence $\{\Phi_{\mu,\bar{r}}(x^k, w^k)\}_{\mathcal{K}} \to \infty$ since the term $\bar{r}\psi(x^k, w^k)$ dominates the term $-\mu\sum_{i\in\mathcal{I}} \ln w_i^k$ as $k \in \mathcal{K} \to \infty$ and $\{f(x^k)\}$ is bounded. But this contradicts the fact that $\Phi_{\mu,\bar{r}}(x^k, w^k)$ is monotonically decreased by Algorithm III.

Now we consider the case that $\{r_k\} \to \infty$. Suppose there exists an infinite index set $\mathcal{K}$ for which r_k is increased such that $\{\|w^k\|\}_{\mathcal{K}} \to \infty$. Then by assumptions A2' and A3,

$$\frac{\|w^k\|}{\psi(x^k, w^k)} \to 1 \quad \text{as} \quad k \in \mathcal{K} \to \infty. \tag{4.21}$$

On the other hand, it follows from the third equation of (4.3) and condition (D-3) that

$$\begin{aligned} & w_i^k - (\nabla c_i(x^k)^\top \triangle x^k - \triangle w_i^k + c_i(x^k)) \\ = \; & \tfrac{1}{r_k}\psi(x^k, w^k)\lambda_i^k \le \tfrac{1}{r_k}\psi(x^k, w^k)\kappa_2\mu/w_i^k, \end{aligned} \quad i \in \mathcal{I}. \tag{4.22}$$

Since the sequence $\{(\triangle x^k, \triangle w^k)\}_{\mathcal{K}}$ is bounded by condition (D-2), dividing (4.22) by $\psi(x^k, w^k)$ yields that

$$\frac{(w_i^k)}{\psi(x^k, w^k)} \to 0 \quad \text{as} \quad k \in \mathcal{K} \to \infty, \quad \forall i \in \mathcal{I},$$

which contradicts (4.21). Hence, there exist $M > 0$ such that $\|w^k\| \le M$ and hence an $\bar{M} > 0$ such that $f(x^k) - \mu\sum_{i\in\mathcal{I}} \ln w_i^k + \bar{M} \ge 0$ for all k in which r_k is increased. Therefore, the sequence $\{(\Phi_{\mu,r_k}(x^k, w^k) + \bar{M})/r_k\}$ is monotonically decreasing by Algorithm III. This implies that $\{\psi(x^k, w^k)\}$ and hence $\{w^k\}$ are bounded. ∎

The next result is the analog of Theorem 3.8, which gives the possible outcomes when $\{r_k\} \to \infty$. Its proof is similar, but not identical to the proof of Theorem 3.8.

THEOREM 4.4. *Under assumptions A2', A3 and A4, if the penalty parameter r_k is unbounded, there exists a limit point $\bar{x}$ that is either an infeasible stationary point of the feasibility problem (INP), or a FJ point of problem (NP) that fails to satisfy the MFCQ.*

Proof. Since $\{r_k\}$ is unbounded, there exists an infinite index set $\mathcal{K}$ set such that r_k is increased for all $k \in \mathcal{K}$. From the first equation of the linear system (4.3) and (4.5) we obtain

$$\begin{aligned} &\mathcal{H}^k \triangle x^k + \textstyle\sum_{i\in\mathcal{I}} \left(\varpi_i^k + r_k \frac{c_i(x^k)-w_i^k}{\psi(x^k,w^k)}\right) \nabla c_i(x^k) \\ &+ \textstyle\sum_{i\in\mathcal{E}} \left(\varpi_i^k + r_k \frac{g_i(x^k)}{\psi(x^k,w^k)}\right) \nabla g_i(x^k) + \nabla f(x^k) = 0, \end{aligned} \tag{4.23}$$

where $\psi(x^k, w^k) > 0$ for all $k \in \mathcal{K}$ in view of condition (D-1). Since from assumption A3 and Lemma 4.3 $\{(x^k, w^k)\}_{\mathcal{K}}$ is bounded, $\{(x^k, w^k)\}_{\mathcal{K}'} \to (\bar{x}, \bar{w})$ with $\bar{w} \geq 0$ for some infinite subset $\mathcal{K}' \subseteq \mathcal{K}$. By the continuity of c and g, we know $\{\psi(x^k, w^k)\}_{\mathcal{K}'} \to \psi(\bar{x}, \bar{w})$.

It follows from the first line of (4.5) that for $k \in \mathcal{K}'$

$$\lambda_i^k = \varpi_i^k + r_k \frac{w_i^k - c_i(x^k)}{\psi(x^k, w^k)}, \quad \forall i \in \mathcal{I}. \tag{4.24}$$

Condition (D-3) gives that $\lambda_i^k > 0$, $\forall i \in \mathcal{I}$ for all $k \in \mathcal{K}'$. Since $\{\|\varpi^k\|\}_{\mathcal{K}'}$ is bounded by condition (D-4) and $\{r_k\}$ tends to infinity, (4.24) implies that

$$\{w_i^k - c_i(x^k)\}_{\mathcal{K}'} \to \bar{w}_i - c_i(\bar{x}) \geq 0, \quad \forall i \in \mathcal{I}, \tag{4.25}$$

and $\bar{w}_i - c_i(\bar{x}) > 0$ only if $\{\lambda_i^k\}_{\mathcal{K}} \to \infty$. Moreover, from condition (D-3) we conclude that for any $i \in \mathcal{I}$, $\bar{w}_i = 0$ if and only if $\{\lambda_i^k\}_{\mathcal{K}} \to \infty$. Consequently, we have

$$\bar{w}_i - c_i(\bar{x}) > 0 \;\text{ only if }\; \bar{w}_i = 0, \quad \forall i \in \mathcal{I}. \tag{4.26}$$

To complete our analysis when the penalty parameter is unbounded. We consider two cases. First recall that $\{\mathcal{H}^k\}$ is bounded by assumption A4, $\{(\triangle x^k, \triangle w^k)\}_{\mathcal{K}'}$ is bounded by condition (D-2) and $\{\varpi^k\}_{\mathcal{K}'}$ is bounded by condition (D-4).

Case 1. $\psi(\bar{x}, \bar{w}) > 0$. Dividing both sides of (4.23) by r_k and letting $k \in \mathcal{K}' \to \infty$ gives

$$\sum_{i\in\mathcal{I}} \frac{c_i(\bar{x}) - \bar{w}_i}{\psi(\bar{x}, \bar{w})} \nabla c_i(\bar{x}) + \sum_{i\in\mathcal{E}} \frac{g_i(\bar{x})}{\psi(\bar{x}, \bar{w})} \nabla g_i(\bar{x}) = 0.$$

This together with (4.25) and (4.26) indicates that

$$\sum_{i\in\mathcal{I}: c_i(\bar{x})<0} c_i(\bar{x}) \nabla c_i(\bar{x}) + \sum_{i\in\mathcal{E}} g_i(\bar{x}) \nabla g_i(\bar{x}) = 0. \tag{4.27}$$

In view of (2.6), we conclude that $\bar{x}$ is a stationary point of the feasibility problem (INP).

Case 2. $\psi(\bar{x}, \bar{w}) = 0$. Then $g(\bar{x}) = 0$ and $c(\bar{x}) = \bar{w} \geq 0$. Hence $\bar{x}$ is a feasible point of problem (NP). By the property of Euclidean norm, we have for each $k \in \mathcal{K}'$

$$\frac{1}{\sqrt{m+p}} \leq \max\left\{ \frac{|c_i(x^k) - w_i^k|}{\psi(x^k, w^k)},\ i \in \mathcal{I};\ \ \frac{|g_i(x^k)|}{\psi(x^k, w^k)},\ i \in \mathcal{E} \right\} \leq 1.$$

Thus, there exist an infinite index set $\mathcal{K}'' \subseteq \mathcal{K}'$ and a vector $(\bar{\lambda}, \bar{y}) \neq 0$ such that

$$\left\{ \frac{w_i^k - c_i(x^k)}{\psi(x^k, w^k)} \right\}_{\mathcal{K}''} \to \bar{\lambda}_i, \quad i \in \mathcal{I} \quad \text{and} \quad \left\{ \frac{g_i(x^k)}{\psi(x^k, w^k)} \right\}_{\mathcal{K}''} \to \bar{y}_i, \quad i \in \mathcal{E}.$$

Since $\lambda^k > 0$ for all $k \in \mathcal{K}''$ in view of condition (D-3), dividing both sides of (4.24) by r_k and letting $k \in \mathcal{K}'' \to \infty$ gives that $\bar{\lambda}_i \geq 0$ for all $i \in \mathcal{I}$. Again, from condition (D-3) we know for each $i \in \mathcal{I}$, $\{\lambda_i^k\}_{\mathcal{K}''}$ is bounded if $\bar{w}_i > 0$. Hence, dividing both sides of (4.24) by r_k and letting $k \in \mathcal{K}'' \to \infty$ yields that $\bar{\lambda}_i = 0$ for $i \in \mathcal{I}$ with $\bar{w}_i > 0$, i.e., for $i \in \mathcal{I} \backslash \mathcal{I}^o(\bar{x})$. Now dividing both sides of (4.23) by r_k and letting $k \in \mathcal{K}'' \to \infty$ yields that

$$\sum_{i \in \mathcal{I}^o(\bar{x})} \bar{\lambda}_i \nabla c_i(\bar{x}) = \sum_{i \in \mathcal{E}} \bar{y}_i \nabla g_i(\bar{x}). \tag{4.28}$$

Since $(\bar{\lambda}, \bar{y}) \neq 0$ and $\bar{\lambda}_i \geq 0, \ \forall i \in \mathcal{I}$, (4.28) together with Definitions 2.3 and 2.4 implies that $\bar{x}$ is a FJ point of problem (NP) that fails to satisfy the MFCQ. ∎

Because of the essential equivalence between Algorithm I and III, appropriate versions of Lemmas 3.9∼3.12 and Theorems 3.13 and 3.14 hold for Algorithm III. Here we only state the final theorem (analogue of Theorem 3.14).

THEOREM 4.5. *Suppose assumptions A2', A3 and A4 hold. The outcome of applying Algorithm III is one of the following.*

(A) *A FJ point of problem (NP) failing to satisfy the MFCQ is found at Step 1 in a finite number of iterations.*

(B) *An approximate KKT point of problem (IP_μ) satisfying termination criteria (4.19) is found at Step 2 in a finite number of iterations.*

(C) *Finite termination does not occur and there exists at least one limit point of $\{x^k\}$ that is either an infeasible stationary point of the feasibility problem (INP), or a FJ point of problem (NP) failing to satisfy the MFCQ.*

4.3. Overall algorithm. Our infeasible IPM for solving problem (NP) successively solves the barrier subproblem (IP_μ) for a decreasing sequence $\{\mu\}$ by applying Algorithm III. We use the index k to denote an outer iteration and j to denote the last inner iteration of Algorithm III.

ALGORITHM IV. INFEASIBLE OUTER ALGORITHM FOR SOLVING (NP).

Specify parameters $\mu_0 > 0$, $\epsilon_{\mu_0} > 0$, $\alpha \in (0, 1)$ and the final step tolerance ϵ_{inf}.
Choose the starting point $(x^0, w^0) \in \Re^{n+m}$ with $w^0 > 0$, $u^0 \in \Re^m$ with $u^0 > 0$, $\lambda^0 \in \Re^m$, $y^0 \in \Re^p$, $r_0 > 0$ and the initial Hessian or its estimate $\mathcal{H}^0$.
$k \leftarrow 0$.
REPEAT until $\|\mathcal{R}(x^k, \lambda^k, y^k)\| \leq \epsilon_{\text{inf}}$:

1. Apply Algorithm III, starting from $(x^k, w^k, u^k, r_k, \mathcal{H}^k)$ with parameters $(\mu_k, \epsilon_{\mu_k}, \pi_{\mu_k})$, to find an approximate solution $(x^{k,j}, w^{k,j}, \lambda^{k,j}, y^{k,j})$ of the barrier subproblem (IP_{μ_k}), which satisfies termination criteria (4.19).
2. Set $\mu_{k+1} \leftarrow \alpha\mu_k$, $\epsilon_{\mu_{k+1}} \leftarrow \alpha\epsilon_{\mu_k}$, $(x^{k+1}, w^{k+1}) \leftarrow (x^{k,j}, w^{k,j})$, $u^{k+1} \leftarrow u^{k,j}$, $\lambda^{k+1} \leftarrow \lambda^{k,j}$, $y^{k+1} \leftarrow y^{k,j}$, $r_{k+1} \leftarrow r_{k,j}$, $\mathcal{H}^{k+1} \leftarrow \mathcal{H}^{k,j}$. Set $k \leftarrow k + 1$.

END

From our discussion in the previous section, it is clear that Algorithm IV is essentially equivalent to Algorithm II applied to problem (SP). The following two theorems are analogues of Theorems 3.15 and 3.16 and their proofs are almost identical.

THEOREM 4.6. *Suppose Algorithm IV is run ignoring its termination criterion. Suppose assumption A2' holds, A3 holds with the same bound for every* μ_k *and A4 holds for each* μ_k*. Then one of the following outcomes occurs.*

(A) *For some parameter* μ_k*, the termination criteria (4.19) of Algorithm III are never met, in which case there exists a limit point of the inner iterates that is either an infeasible stationary point of problem (INP) or a FJ point of problem (NP) failing to satisfy the MFCQ.*

(B) *Algorithm III successfully terminates for each* μ_k*, in which case the limit point of any bounded subsequence* $\{x^k, \lambda^k, y^k\}_{\mathcal{K}}$ *satisfies the first order optimality condition (2.1); while the limit point of any subsequence* $\{x^k\}_{\mathcal{K}}$ *with unbounded multiplier subsequence* $\{(\lambda^k, y^k)\}_{\mathcal{K}}$ *is a FJ point of problem (NP) failing to satisfy MFCQ.*

THEOREM 4.7. *Suppose assumption A2' holds, A3 holds with the same bound for every* μ_k *and A4 holds for each* μ_k*. Moreover, suppose the MFCQ holds on the feasible region and problem (NP) is not locally infeasible; i.e., there are no stationary points of problem (INP) on the closure of the bounded set containing all primal iterates that are infeasible to (NP). Then ignoring the termination condition of Algorithm IV, the penalty parameter is uniformly bounded and any limit point of the sequence* $\{x^k, \lambda^k, y^k\}$ *generated by Algorithm IV satisfies the first order optimality conditions (2.1).*

5. Numerical results. We only present numerical results on some small irregular problems from various sources in the literature that corroborate our global convergence theorems. In a subsequent paper we will establish that appropriate versions of our methods that incorporate some local procedures exhibit fast local convergence, and report on numerical tests involving large scale problems.

Our basic quasi-feasible and infeasible methods were implemented in MATLAB 6.1 and tested on an IBM ThinkPad X40 notebook with an Intel-Pentium-M 1.3GHz processor. We set the parameters of the inner algorithms I and III as follows:

$$\sigma = 0.1,\ \beta = 0.9,\ \kappa_1 = 0.1,\ \kappa_2 = 5,\ \chi = 2,\ \nu = 10^{-6},\ \gamma_{\min} = 0.1,\ \gamma_{\max} = 5.$$

We set the parameters and initial data of the outer algorithms II and IV as follows:

$$\mu_0 = 1,\ \epsilon_{\mu_0} = 5,\ \alpha = 0.01,\ \epsilon_{\text{inf}} = 10^{-8},\ u^0 = e,\ \lambda^0 = e,\ y^0 = e,\ r_0 = 2.$$

For the infeasible method we set $w^0 = e$. We set $\mathcal{H}^k$ to the exact Hessian of the Lagrangian unless condition (C-5) or (D-5) failed to hold, in which case we simply added a suitable multiple of the identity to $\mathcal{H}^k$ so that the condition held.

(A) Convergence to KKT points with bounded penalty parameters.

EXAMPLE 1. This test example is from [44] and [10]. Although it is well-posed, many barrier-SQP methods ("Type-I Algorithms" in [44]) fail to obtain feasibility for a range of infeasible starting points.

$$\begin{array}{ll} \min & x_1 \\ \text{s.t.} & x_1^2 - x_2 + a = 0, \\ & x_1 - x_3 - b = 0, \\ & x_2 \geq 0,\ x_3 \geq 0. \end{array}$$

Our quasi-feasible method applied to this problem with $(a, b) = (-1, 1/2)$ and $x_0 = (-2, 1, 1)$ terminated at an approximate solution $(1, 0, 0.5)$ in 17 iterations with residual $\|\mathcal{R}(x, \lambda, y)\| = 1.7592\text{E} - 10$ and $r = 2$. An approximate solution $(1, 2, 0)$ of the problem with $(a, b) = (1, 1)$ and $x^0 = (-3, 1, 1)$ was found by our infeasible method in 11 iterations with $\|\mathcal{R}(x, \lambda, y)\| = 5.9818\text{E} - 10$ and $r = 2$.

EXAMPLE 2. This is a pathological problem used in [20] to illustrate a possible failure of the filter-SQP method, caused by the incompatibility of the QP subproblem for any infeasible points. The starting point $(1, 0)$ is taken from [20].

$$\begin{array}{ll} \min & (x_2 - 1)^2 \\ \text{s.t.} & x_1^2 = 0;\ x_1^3 = 0. \end{array}$$

Our quasi-feasible method found the solution $(0, 1)$ in 12 iterations with final residual $6.0193\text{E} - 10$ and penalty parameter 2.

(B) Convergence to infeasible stationary points of the feasibility problem with unbounded penalty parameters.

EXAMPLE 3. This problem is problem 17 from [32].

$$\begin{array}{ll} \min & 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \\ \text{s.t.} & x_2^2 - x_1 \geq 0,\ x_1^2 - x_2 \geq 0, \\ & -0.5 \leq x_1 \leq 0.5,\ x_2 \leq 1. \end{array}$$

In an early version of this paper, we applied Algorithm III to the barrier subproblem with $\mu = 0.1$, $\epsilon_\mu = 10^{-6}$ and used a damped BFGS formula to update $\mathcal{H}^k$ instead of using the exact Hessian. Using the starting point given in [32], in 251 inner iterations Algorithm III converged to a point (0.6298,0.6571) with $r = 262144$. It is easy to check that this is an approximate infeasible stationary point of the infeasibility measure. When the exact Hessian was employed, Algorithm IV found the solution $(0, 0)$ in 28 iterations with final residual $2.8222\text{E} - 10$ and penalty parameter 4.

EXAMPLE 4. This problem is from [5]. It minimizes any objective function on an infeasible set defined by the constraints

$$x^2 + 1 \leq 0, \quad x \leq 0.$$

Here we chose to minimize x. Setting $\mu = 0.01$ and $x^0 = 10$, Algorithm III converged to $x^* = -8.3487\text{E} - 05$ in 59 iterations with slack variables $w_1^* = 2.8992\text{E} - 08$ and $w_2^* = 3.2376\text{E} - 05$ and $r = 131072$. It is easy to check that x^* is an approximate stationary point of the infeasibility measure.

(C) Convergence to FJ points that fail to satisfy the MFCQ with unbounded penalty parameters.

EXAMPLE 5. This test example is problem 13 in [32].

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

At its solution (1,0), the MFCQ fails to hold. For $\mu = 0.01$, using the starting point given in [32], Algorithm III converged to an approximate FJ point $(1.0009\text{E} + 00, 3.5519\text{E} - 08)$ in 138 iterations with $r = 32768$.

6. Concluding remarks. Under very mild assumptions (the existence of a not necessarily feasible interior point, smoothness of the objective and constraint functions and boundedness of the primal iterates and Hessian estimates), we have proved that our proposed interior-point ℓ_2-penalty methods converge either to a KKT point or a FJ point that fails to satisfy the MFCQ of either the original nonlinear program (NP) or the feasibility problem associated with the NP. As far as we are aware, these convergence results are as strong as or stronger than any that have been obtained for other barrier-SQP or barrier-modified Newton methods. These convergence results are similar to those that have been obtained for the filter-SQP method of Fletcher and Leyffer [20]. However, the latter method is only weakly convergent (i.e., a subsequence of the iterates converges) and is more expensive to implement.

Another method that has provably strong convergence properties is the barrier-ℓ_1-penalty method of Gould, Orban and Toint [29]. This method adds slack variables to the equality as well as the inequality constraints resulting in a sequence of unconstrained minimization problems min $\phi_{\mu,r}(x,s)$ depending on the barrier and penalty parameters μ and r, where

$$\phi_{\mu,r}(x,s) = f(x) + r\sum_{i\in\mathcal{E}}(c_i(x)+2s_i) + r\sum_{i\in\mathcal{I}} s_i - \mu\sum_{i\in\mathcal{I}\cup\mathcal{E}}\log(c_i(x)+s_i) - \mu\sum_{i\in\mathcal{I}\cup\mathcal{E}}\log s_i.$$

Unfortunately, this method can attain feasibility only as $\mu \downarrow 0$.

Our methods share a number of features with several existing barrier-SQP methods. In particular, they use line search and an ℓ_2-exact penalty function as does the Liu and Sun method [33], and they compute step directions for their line searches using perturbed KKT systems as does the method implemented in the LOQO software package [4, 38, 43, 31]. We note that NITRO also uses an ℓ_2-merit function but employs trust-regions to solve the barrier subproblems that arise at each iteration.

Acknowledgments. We thank the referees for their comments and suggestions which helped us improve the paper.

REFERENCES

[1] M. Argáez and R. A. Tapia, *On the global convergence of a modified augmented Lagrangian linesearch interior-point Newton method for Nonlinear Programming*, J. Optim. Theory Appl., 114 (2002), pp. 1-25.
[2] S. Bakhtiari and A. L. Tits, *A Simple primal-dual feasible interior-point method for nonlinear programming with monotone descent,* Comput. Optim. Appl., 25 (2003), pp. 17-38.
[3] 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), pp. 257-272.
[4] ———, *Interior-point methods for nonconvex nonlinear programming: Jamming and comparative numerical testing*, Math. Programming, 99 (2004), pp. 35-48.
[5] J. V. Burke and S. P. Han, *A robust sequential quadratic programming method*, Math. Programming, 43 (1989), 277-303.
[6] R. H. Byrd, *Robust trust-region method for constrained optimization*, Paper presented at the SIAM Conference on Optimization, Houston, TX, 1987.
[7] R. H. Byrd, J. C. Gilbert and J. Nocedal, *A trust region method based on interior point techniques for nonlinear programming*, Math. Programming, 89 (2000), pp. 149-185.
[8] R. H. Byrd, M. E. Hribar and J. Nocedal, *An interior point algorithm for large-scale nonlinear programming*, SIAM J. Optim., 11 (1999), pp. 877-900.
[9] R. H. Byrd, G. Liu and J. Nocedal, *On the local behavior 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.
[10] R. H. Byrd, M. Marazzi and J. Nocedal, *On the convergence of Newton iterations to nonstationary points*, Math. Programming, 99 (2004), pp. 127-148.

[11] R. H. Byrd, J. Nocedal and A. Waltz, *Feasible interior methods using slacks for nonlinear optimization*, Comput. Optim. Appl., 26 (2003), pp. 35-61.
[12] A. R. Conn, N. I. M. Gould, D. Orban and P. L. Toint, *A primal-dual trust region algorithm for non-convex nonlinear programming*, Math. Programming, 87 (2000), pp. 215-249.
[13] A. R. Conn, N. I. M. Gould and P. L. Toint, *A primal-dual algorithm for minimizing a non-convex function subject to bound and linear equality constraints*, In Nonlinear optimization and related topics (Erice, 1998), Kluwer Acad. Publ., Dordrecht, 2000, pp. 15-49.
[14] 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), pp. 1750-1794.
[15] C. Durazzi and V. Ruggiero, *Global convergence of the Newton interior-point method for nonlinear programming*, J. Optim. Theory Appl., 120 (2004), pp. 199-208.
[16] A. S. El-Bakry, R. A. Tapia, T. Tsuchiya and Y. Zhang, *On the formulation and theory of the Newton interior-point method for nonlinear programming*, J. Optim. Theory Appl., 89 (1996), pp. 507-541.
[17] A. V. Fiacco and G. P. McCormick, *Nonlinear programming: Sequential Unconstrained Minimization Techniques*, Classics Appl. Math. 4, SIAM, Philadelphia, PA, 1990.
[18] R. Fletcher, N. I. M. Gould, S. Leyffer, P. L. Toint and A. Wächter, *Global convergence of a trust-region SQP-filter algorithm for general nonlinear programming*, SIAM J. Optim., 13 (2002), pp. 635-659.
[19] R. Fletcher and S. Leyffer, *Nonlinear Programming without a penalty function*, Math. Programming, 91 (2002), pp. 239-269.
[20] R. Fletcher, S. Leyffer and P. L. Toint, *On the global convergence of a filter-SQP algorithm*, SIAM J. Optim., 13 (2002), pp. 44-59.
[21] A. Forsgren, *Inertia-controlling factorizations for optimization algorithms*, Appl. Numer. Math., 43 (2002), pp. 91-107.
[22] A. Forsgren and P. E. Gill, *Primal-dual interior method for nonconvex nonlinear programming*, SIAM J. Optim., 8 (1998), pp. 1132-1152.
[23] A. Forsgren, P. E. Gill and M. H. Wright, *Interior methods for nonlinear optimization*, SIAM Rev., 44 (2002), pp. 525-597.
[24] D. M. Gay, M. L. Overton and M. H. Wright, *A primal-dual interior method for nonconvex nonlinear programming*, In Advances in Nonlinear Programming (Beijing, 1996), Y. Yuan, ed., Kluwer Acad. Publ., Dordrecht, 1998, pp. 31-56.
[25] E. M. Gertz and P. E. Gill, *A primal-dual trust region algorithm for nonlinear optimization*, Math. Programming, 100 (2004), pp. 49-94.
[26] D. Goldfarb, R. Polyak, K. Scheinberg and I. Yuzefovich, *A modified barrier-augmented Lagrangian method for constrained minimization*, Comput. Optim. Appl., 14 (1999), pp. 55-74.
[27] G. H. Golub and C. F. Van Loan, *Matrix computations*, 3rd ed., The Johns Hopkins Univ. Press, Baltimore, 1996.
[28] N. I. M. Gould, D. Orban, A. Sartenaer and Ph. L. Toint, *Superlinear convergence of primal-dual interior point algorithms for nonlinear programming*, SIAM J. Optim., 11 (2001), pp. 974-1002.
[29] N. I. M. Gould, D. Orban and Ph. L. Toint, *An interior-point l_1-penalty method for nonlinear optimization*, RAL-TR-2003-022, Computational Science and Engineering Department, Rutherford Appleton Laboratory, Chilton, Oxfordshire, England, 2003.
[30] C. C. Gonzaga, E. Karas and M. Vanti, *A globally convergent filter method for nonlinear programming*, SIAM J. Optim., 14 (2003), pp. 646-669.
[31] I. Griva, D. F. Shanno and R. J. Vanderbei, *Convergence analysis of a primal-dual interior-point method for nonlinear programming*, Optimization Online (http://www.optimization-online.org/DB_HTML/2004/07/913.html), July, 2004.
[32] W. Hock and K. Schittkowski, *Tests examples for nonlinear programming codes,* volume 187 of *Lecture Notes in Economics and Mathematical Systems*, Springer-Verlag, 1981.
[33] X. Liu and J. Sun, *A robust primal-dual interior point algorithm for nonlinear programs*, SIAM J. Optim., 14 (2004), pp. 1163-1186.
[34] ———, *Global convergence analysis of line search interior point methods for nonlinear programming without regularity assumptions*, J. Optim. Theory Appl., to appear.
[35] D. Q. Mayne and E. Polak, *Feasible direction algorithms for optimization problems with equality and inequality constraints*, Math. Programming, 11 (1976), pp. 67-80.
[36] J. M. Moguerza and F. J. Prieto, *An augmented Lagrangian interior-point method using directions of negative curvature*, Math. Programming, 95 (2003), pp. 573-616.
[37] E. O. Omojokun, *Trust region algorithms for nonlinear equality and inequality constraints*,

PhD thesis, Department of Computer Science, University of Colorado, Boulder, 1989.

[38] D. F. Shanno and R. J. Vanderbei, *Interior-point methods for nonconvex nonlinear programming: orderings and high-order methods*, Math. Programming, 87 (2000), pp. 303-316.

[39] G. Sporre and A. Forsgren, *Relations between divergence of multipliers and convergence to infeasible points in primal-dual interior methods for nonconvex nonlinear programming*, Tech. Report TRITA-MAT-02-OS7, Department of Mathematics, Royal Institute of Technology (KTH), Stockholm, Sweden, 2002.

[40] 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), pp. 173-199.

[41] P. Tseng, *A convergent infeasible interior-point trust-region method for constrained minimization*, SIAM J. Optim., 13 (2002), pp. 432-469.

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

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

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

[45] ———, *Line search filter methods for nonlinear programming: Motivation and global convergence*, Res. Report RC 23036, IBM T. J. Watson Research Center, Yorktown, accepted by SIAM J. Optim., 2004.

[46] ———, *Line search filter methods for nonlinear programming: Local convergence*, Res. Report RC 23033, IBM T. J. Watson Research Center, Yorktown, accepted by SIAM J. Optim., 2004.

[47] ———, *On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming*, Res. Report RC 23149, IBM T. J. Watson Research Center, Yorktown, accepted by Math. Programming, 2004.

[48] R. A. Waltz, J. L. Morales, J. Nocedal and D. Orban, *An Interior Algorithm for Nonlinear Optimization that Combines Line Search and Trust Region Steps,* Tech. Report, Optimization Technology Center, Northwestern University, Evanston, IL, 2003, to appear in Math. Programming.

[49] H. Yamashita, *A globally convergent primal-dual interior-point method for constrained optimization*, Optim. Methods Softw., 10 (1998), pp. 443-469.

[50] H. Yamashita and H. Yabe, *An interior point method with a primal-dual quadratic barrier penalty function for nonlinear optimizaiton*, SIAM J. Optim., 14 (2003), pp. 479-499.

[51] 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. Programming, 102 (2004), pp. 111-151.