TECHNICAL RESEARCH REPORT

A Primal-Dual Interior-Point Method for Nonlinear Programming with Strong Global and Local Convergence Properties

by A.L. Tits, T.J. Urban, S. Bakhtiari, C.T. Lawrence

TR 2001-3

A primal-dual interior-point method for nonlinear programming with strong global and local convergence properties

André L. Tits Thomas J. Urban Sasan Bakhtiari
Craig T. Lawrence
Department of Electrical and Computer Engineering
and Institute for Systems Research
University of Maryland, College Park, MD 20742, USA

February 19, 2001

Abstract

A scheme—inspired from an old idea due to Mayne and Polak (Math. Prog., vol. 11, 1976, pp. 67–80)—is proposed for extending to general smooth constrained optimization problems a previously proposed feasible interior-point method for inequality constrained problems. It is shown that the primal-dual interior point framework allows for a significantly more effective implementation of the Mayne-Polak idea than that discussed an analyzed by the originators in the context of first order methods of feasible direction. Strong global and local convergence results are proved under mild assumptions. In particular, the proposed algorithm does not suffer the Wächter-Biegler effect.

1 Introduction

Consider the problem

$$\begin{array}{lll} \min\limits_{x\in\mathcal{R}^n} & f(x) & \\ \text{s.t.} & c_j(x) = 0, & j = 1,...,m_{\text{e}} \\ & d_j(x) \geq 0, & j = 1,...,m_{\text{i}} \end{array} \qquad (P)$$

where $f : \mathcal{R}^n \to \mathcal{R}$, $c_j : \mathcal{R}^n \to \mathcal{R}$, $j = 1,...,m_{\text{e}}$ and $d_j : \mathcal{R}^n \to \mathcal{R}$, $j = 1,...,m_{\text{i}}$ are smooth. No convexity assumptions are made. A number of primal-dual interior-point methods have been proposed to tackle such problems; see, e.g., [1, 2, 3, 4, 5, 6, 7]. While all of these methods make use of a search direction generated by a Newton or quasi-Newton iteration on a perturbed version of some first order necessary conditions of optimality, they differ in many respects. For example, some algorithms enforce feasibility of all iterates with respects to inequality constraints [1, 2, 4, 5], while others, sometimes referred to as "infeasible", sidestep that requirement via the introduction of slack variables [3, 6, 7]; as for equality constraints, some schemes include them "as is" in the perturbed optimality conditions [1, 2, 3, 4, 6] while some soften this condition by making use of two sets of slack variables [7] or by introducing a quadratic penalty function, yielding optimality conditions involving a perturbed version of "$c(x) = 0$" [5]; also, some proposed algorithms (e.g., [2, 6]) involve a trust region mechanism. In many cases (e.g. [2, 4, 7], promising numerical results have been obtained. In some cases (e.g., [1, 2, 3]), convergence properties have been proved under certain assumptions. Often however global and fast local convergence are proved for two different versions of the algorithm; in particular, it is not proved that the line search eventually accepts a step size close enough to one to allow fast local convergence, i.e., a Maratos-like effect is not ruled out. An exception is [2], but rather strong assumptions are used there.

Recently, Wächter-Biegler [8] showed that many of the proposed algorithms suffer a major drawback in that, for problems with two or more equality constraints and a total number of constraints in excess of the dimension of the space, the constructed primal sequence is likely to converge to spurious, infeasible points. They produced a simple, seemingly innocuous example where such behavior is observed when starting from rather arbitrary initial points. Where global convergence had been proved, they pinpointed rather strong assumptions that essentially make the proposed algorithm unfit for the solution of problems with such number of constraints.

In this paper, we propose a primal-dual interior point algorithm of the "feasible" type for which global and fast local convergence are proved to hold (i) for one, single version of the algorithm, and (ii) under rather mild assumptions. In particular, it involves a scheme to circumvent Maratos-like effects and does not suffer the Wächter-Biegler effect. A distinguishing feature of the proposed algorithm is that it makes use of both a barrier parameter and an "exterior" penalty parameter, both of which are adaptively adjusted to insure global and fast local convergence. (In [2] two such parameters are use as well, but the latter is kept fixed and assumed to be large enough.) The algorithm originates in two papers dating back more than one and two decades, respectively: [9] and [10]. The former proposed a feasible interior point method for inequality constrained problems, with strong convergence properties; the latter offered a scheme for extending to problems with equality constraints algorithms that generate feasible iterates for inequality constrained problems. These two ideas are discussed next.

In the 1980s, a feasible-iterate algorithm for solving (P) was proposed for the case without equality constraints, based on the following idea. First, given strictly feasible estimates $\hat{x}$ of a solution and $\hat{z}$ of the corresponding Karush-Kuhn-Tucker (KKT) multiplier vector, compute the Newton (or a quasi-Newton) direction $(\Delta x, \Delta z)$ for the solution of the equalities in the KKT first order necessary conditions of optimality. Note that, if the Hessian (or Hessian estimate) is positive definite, the primal direction Δx is a direction of descent for f but that it may not allow a reasonably long step to be taken inside the feasible set. Second, motivated by this observation, solve again the same system of equations, but with the right-hand side perturbed so as to tilt the primal direction away from the constraint boundaries into the feasible set. The perturbation should be small enough that the tilted primal direction remains a descent direction for f and its size should decrease as a solution is approached, i.e., as $\|\Delta x\|$ decreases, so that a solution point located on the constraint boundaries can be reached. Third, bend the primal direction by means of a second order correction, and perform a search on the resulting arc, with f as a merit function. Bending is necessary if a Maratos-like effect is to be avoided, i.e., if a full step of one is to be allowed by the line search criterion close to the solution. These ideas were put forth in [9]. It was shown there that, under standard assumptions, global convergence as well as local superlinear convergence can be achieved if the amounts of tilting and bending are appropriately chosen. The central idea in the algorithm of [9] originated in earlier work by Herskovits and others [11, 12, 13]; see [14] for a detailed historical account. Ideas were also borrowed from [15] and [16].

In the mid-seventies Mayne and Polak proposed an ingenious scheme to incorporate equalities constraints in methods of feasible directions [10]. The idea is to keep the iterates on one side of the constraint manifolds corresponding to each equality constraint by replacing those constraints with corresponding inequality constraints which the iterates are forced to satisfy by the feasible direction paradigm; and to penalize departure from the constraint boundaries associated with these fictive inequality constraints by means of a simple, exact differentiable penalty function. It is readily shown that, locally, convergence to KKT points of the original problem takes place provided the penalty parameter is increased to a value larger than the magnitude of the most negative equality constraint multiplier at the solution. Accordingly, in [10] the penalty parameter is adaptively increased based on estimates of these multipliers. While [10] is concerned with classical first order feasible directions methods, interestingly and somewhat premonitorily, it is pointed out in the introduction of that paper that the proposed scheme can convert "*any* [emphasis from [10]] interior point algorithm for inequality constrained optimization problems into an algorithm for optimization subject to combined equality and inequality constraints." A careful examination of the proposed algorithm however reveals two shortcomings. The first one concerns the computation of multiplier estimates, which in [10] is done by solving a linear least squares problem for all equality constraint multipliers and all multipliers associated with ϵ-active inequality constraints; i.e., inequality constraints with current value less than some fixed, prescribed ϵ (denoted ϵ' in [10]). The price to pay is that, if ϵ is "large", the computational overhead may become significant and, moreover, the set of active constraints may be overestimated, even in the neighborhood of a solution, leading to incorrect multiplier estimates; while if ϵ is selected to be very small, progress will be slow in early iterations. The second shortcoming is that global convergence is proved under the strong assumption that at every point in the extended feasible set (where one-side violation of equality constraints is allowed) the gradients of *all* equality constraints and of the active inequality constraints are linearly independent. Indeed, as pointed out in [8], such assumption does not hold in the Wächter-Biegler example, and indeed it all but rules out problems with two or more equality constraints and a total number of constraints in excess of n. In [13] it is suggested that the idea introduced in [10] could be readily applied to the interior-point algorithm proposed there, but no details are given. The Mayne-Polak idea was used in [17] in the context of feasible

SQP. The ready availability of multiplier estimates in that context allows an improved multiplier estimation scheme, thus improving on the first shortcoming just pointed out; however no attempt is made in [17] to dispense with the strong regularity assumption.

The algorithm proposed in the present paper extends that of [9] to general constrained problems by incorporating a modified Mayne-Polak scheme that suffers neither of the shortcomings pointed out above. Specifically, (i) it completely dispenses with a multiplier estimation scheme, and (ii) it converges globally and locally superlinearly without requirement that a strong regularity assumption be satisfied, thus avoiding the Wächter-Biegler effect.

The balance of the paper is organized as follows. In Section 2 below, the algorithm from [9] is described in "modern" terms, from a barrier function perspective. The overall algorithm is then motivated and described in Section 3. In Section 4, global and local superlinear convergence are proved. Finally, Section 5 is devoted to concluding remarks. Throughout, $\|\cdot\|$ denotes the Euclidean norm and, given two vectors v_1 and v_2, inequalities such as $v_1 \leq v_2$ and $v_1 < v_2$ are to be understood component-wise. Much of our notation is borrowed from [4].

2 Problems Without Equality Constraints

We briefly review the algorithm of [9], in the primal-dual interior-point formalism.

Consider problem (P) with $m_\mathrm{e} = 0$, i.e.,

$$\begin{array}{ll} \min\limits_{x \in \mathcal{R}^n} & f(x) \\ \text{s.t.} & d_j(x) \geq 0, \quad j = 1, ..., m_\mathrm{i}\ . \end{array} \tag{1}$$

The algorithm proposed in [9] for problems such as (1) can equivalently be stated based on the logarithmic barrier function

$$\beta(x, \mu) = f(x) - \sum_{j=1}^{m_\mathrm{i}} \mu_j \log d_j(x) \tag{2}$$

where $\mu = [\mu_1, \ldots, \mu_{m_\mathrm{i}}]^\mathrm{T} \in \mathcal{R}^{m_\mathrm{i}}$ and the μ_js are positive. The barrier gradient is given by

$$\nabla\beta(x, \mu) = g(x) - B(x)^\mathrm{T} D(x)^{-1} \mu, \tag{3}$$

where g denotes the gradient of f, B the Jacobian of d and $D(x)$ the diagonal matrix $\mathrm{diag}(d_j(x))$.

Problem (1) can be tackled via a sequence of unconstrained minimizations of $\beta(x, \mu)$ with $\mu \to 0$. In view of (3), $z = D(x)^{-1}\mu$ can be viewed as an approximation to the KKT multiplier vector associated with a solution of (1) and the right-hand side of (3) as the value at (x, z) of the gradient (w.r.t. x) of the Lagrangian

$$\mathcal{L}(x, z) = f(x) - \langle z, d(x) \rangle.$$

Accordingly, and in the spirit of primal-dual interior-point methods, consider using a (quasi-)Newton iteration for the solution of the system of equations in (x, z)

$$g(x) - B(x)^T z = 0, \tag{4}$$

$$D(x)z = \mu, \tag{5}$$

i.e.,

$$\begin{bmatrix} -W & B(x)^T \\ ZB(x) & D(x) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z \end{bmatrix} = \begin{bmatrix} g(x) - B(x)^T z \\ \mu - D(x)z \end{bmatrix} \tag{6}$$

where $Z = \mathrm{diag}(z_j)$ and where W is equal to, or approximates, the Hessian (w.r.t. x) of the Lagrangian $\mathcal{L}(x, z)$. When $\mu = 0$, a primal-dual feasible solution to (4)-(5) is a KKT point for (1). Moreover, it turns out that, under the assumption made in [9] that W is positive definite, given any (x, z) primal-dual feasible, the primal direction Δx^0 obtained by setting $\mu = 0$ is a descent direction for f at x. In [9], such a property is sought for the search direction and used in the line search. On the other hand, the components of μ should be positive enough to prevent the primal step length from collapsing, but small enough that the fast local convergence properties associated with the (quasi-)Newton iteration for (4)-(5) with $\mu = 0$ are preserved. This is achieved in [9] by selecting

$$\mu = \varphi \|\Delta x^0\|^\nu z, \tag{7}$$

with $\varphi \in (0, 1]$ as large as possible subject to the constraint

$$\langle g(x), \Delta x \rangle \leq \theta \langle g(x), \Delta x^0 \rangle, \tag{8}$$

where $\nu > 2$ and $\theta \in (0,1)$ are prespecified;[1] condition (8) ensures that Δx is still a descent direction for f.

In [9] the line search criterion includes a decrease of f and strict feasibility. It involves a second order correction $\Delta\tilde{x}$ to allow a full (quasi-)Newton step to be taken near the solution. With index sets I and J defined by

$$I = \{j : d_j(x) \leq z^j + \Delta z^j\},$$

$$J = \{j : z^j + \Delta z^j \leq -d_j(x)\},$$

$\Delta\tilde{x}$ is the solution of the linear least squares problem

$$\min \frac{1}{2}\langle \Delta\tilde{x}, W\Delta\tilde{x}\rangle \text{ s.t. } d_j(x+\Delta x) + \langle \nabla d_j(x), \Delta\tilde{x}\rangle = \psi, \quad \forall j \in I \tag{9}$$

where

$$\psi = \max\left\{ \|\Delta x\|^\tau, \max_{j\in I}\left|\frac{\Delta z^j}{z^j+\Delta z^j}\right|^\kappa \|\Delta x\|^2 \right\}, \tag{10}$$

where $\tau \in (2,3)$ and $\kappa \in (0,1)$ are prespecified. If $J \neq \emptyset$ or (9) has no solution or $\|\Delta\tilde{x}\| > \|\Delta x\|$, $\Delta\tilde{x}$ is set to 0. Note that I estimates the active index set and that J (multipliers of "wrong" sign) should be empty near the solution when strict complementarity holds. An (Armijo-type) arc search is then performed as follows: given $\eta \in (0,1)$, compute the first number α in the sequence $\{1, \eta, \eta^2, \dots\}$ such that

$$\begin{aligned}
f\left(x+\alpha\Delta x+\alpha^2\Delta\tilde{x}\right) &\leq f(x) + \xi\alpha\langle\nabla f(x), \Delta x\rangle && (11)\\
d_j\left(x+\alpha\Delta x+\alpha^2\Delta\tilde{x}\right) &> 0, \quad \forall j && (12)\\
d_j\left(x+\alpha\Delta x+\alpha^2\Delta\tilde{x}\right) &\geq d_j(x), \quad j \in J && (13)
\end{aligned}$$

where $\xi \in (0,1/2)$ is prespecified and where the third inequality is introduced to prevent convergence to points with negative multipliers. The next primal iterate is then set to

$$x^+ = x + \alpha\Delta x + \alpha^2\Delta\tilde{x}.$$

[1] Note that Δx depends on φ affinely and thus Δx is computed at no extra cost once (6) has been solved with, say, $\mu = \|\Delta x^0\|^\nu z$.

Finally, the dual variable z is reinitialized whenever $J \neq \emptyset$; if $J = \emptyset$ the new value $z^{+,j}$ of z^j is set to

$$z^{+,j} = \min\{z_{\max}, \max\{z^j + \Delta z^j, \|\Delta x\|\},$$

where $z_{\max}$ is a prespecified (large) number. Thus $z^{+,j}$ is allowed to be close to 0 only if $\|\Delta x\|$ is small, indicating proximity to a solution.

It is observed in [9, Section 5] that the total work per iteration (in addition to function evaluations) is essentially one Cholesky decomposition of size m_{i} and one Cholesky decomposition of size equal to the number of active constraints at the solution.[2]

In [9] it is shown that, given an initial strictly feasible primal-dual pair (x_0, z_0) and given a sequence of symmetric matrices $\{W_k\}$, uniformly bounded and uniformly positive definite, the primal sequence $\{x_k\}$ constructed by the algorithm just described (with W_k used as W at the kth iteration) converges to KKT points for (1), provided the following assumptions hold: (i) $\{x : f(x) \leq f(x_0), d(x) \geq 0\}$ is bounded, so that the primal sequence remains bounded, (ii) for all feasible x the vectors $\nabla d_j(x)$, $j \in \{j : d_j(x) = 0\}$ are linearly independent, and (iii) the set of feasible points x for which (4)-(5) hold for some z (with no restriction on the sign of the components of z) is finite.[3]

Superlinear convergence—in particular, eventual acceptance of the full step of one by the arc search—is also proved in [9] under appropriate second order assumptions, provided that, asymptotically, W_k suitably approximate the Hessian of the Lagrangian at the solution on the tangent plane to the active constraints.

Finally, stronger convergence results hold for a variation of the present algorithm, under weaker assumptions, in the LP and convex QP cases. In particular, global convergence to the solution set X^* takes place whenever X^* is nonempty and bounded, the feasible set X has a nonempty interior, and for every $x \in X$ the gradients of the active constraints at x are linearly independent. See [14] for details.

[2] There are two misprints in [9, Section 5]: in equation (5.3) (statement of Proposition 5.1) as well as in the last displayed equation in the proof of Proposition 5.1 (expression for λ_k^0), $M_k B_k^{-1}$ should be $B_k^{-1} M_k$.

[3] Such points are referred to in [9] as *stationary points*.

3 Overall Algorithm

Suppose now that $m_{\rm e} > 0$. Denote by X the feasible set for (P), i.e., let

$$X := \{x \in \mathcal{R}^n : c(x) = 0,\ d_j(x) \geq 0,\ j = 1, ..., m_{\rm i}\}. \tag{14}$$

Further, let A denote the Jacobian of c, let $C(x) = \text{diag}(c_j(x))$ and, just as above, let B denote the Jacobian of d and let $D(x) = \text{diag}(d_j(x))$.

In [10], Mayne and Polak proposed a scheme to convert (P) to a sequence of inequality constrained optimization problems of the type

$$\begin{array}{lll} \min\limits_{x \in \mathcal{R}^n} & f_\rho(x) & \\ \text{s.t} & c_j \geq 0 & j = 1, ..., m_{\rm e}, \\ & d_j \geq 0 & j = 1, ..., m_{\rm i}, \end{array} \tag{P_ρ}$$

where $f_\rho(x) = f(x) + \rho \sum\limits_{j=1}^{m_{\rm e}} c_j(x)$, and where $\rho > 0$. Examination of (P_ρ) shows that large values of ρ penalize iterates satisfying $c_j(x) > 0$ for any j while feasibility for the modified problem insures that $c_j(x) \geq 0$. Thus, intuitively, for large values of ρ, iterates generated by the feasible iterates algorithm will tend towards feasibility for the original problem (P). In fact, the penalty function is "exact" in that convergence to a solution of (P) is achieved without need to drive ρ to infinity. I.e., it turns out that, under mild assumptions, for large enough but finite values of ρ, solutions to (P_ρ) are solutions to (P).

Let $\tilde{X}$ and $\tilde{X}_0$ be the feasible and strictly feasible sets for Problems (P_ρ), i.e., let

$$\tilde{X} := \{x \in \mathcal{R}^n : c_j(x) \geq 0,\ j = 1, ..., m_{\rm e}, \quad d_j(x) \geq 0,\ j = 1, \ldots, m_{\rm i}\}, \tag{15}$$

$$\tilde{X}_0 := \{x \in \mathcal{R}^n : c_j(x) > 0,\ j = 1, ..., m_{\rm e}, \quad d_j(x) > 0,\ j = 1, \ldots, m_{\rm i}\}. \tag{16}$$

Also, for $x \in \tilde{X}$, let $I^{\rm e}(x)$ and $I^{\rm i}(x)$ be the active index sets corresponding to c and d, i.e.,

$$I^{\rm e}(x) = \{j : c_j(x) = 0\}; \quad I^{\rm i}(x) = \{j : d_j(x) = 0\}.$$

Before proceeding, we state some basic assumptions.

Assumption 1 *X is nonempty.*

Assumption 2 *f, c_i, $i = 1 \ldots, m_{\mathrm{e}}$ and d_i, $i = 1, \ldots, m_{\mathrm{i}}$ are continuously differentiable.*

Assumption 3 [4] *For all $x \in \tilde{X}$, (i) the set $\{\nabla c_j(x) : j \in I^{\mathrm{e}}(x)\} \cup \{\nabla d_j(x) : j \in I^{\mathrm{i}}(x)\}$ is linearly independent; (ii) if for scalars y^j, $j = 1, \ldots, m_{\mathrm{e}}$, with $y^j \geq 0$ for all $j \notin I^{\mathrm{e}}(x)$, and $z^j \geq 0$ for all $j \in I^{\mathrm{i}}(x)$, it holds that*

$$\sum_{j=1}^{m_{\mathrm{e}}} y^j \nabla c_j(x) - \sum_{j \in I^{\mathrm{i}}(x)} z^j \nabla d_j(x) = 0, \tag{17}$$

then it must be the case that $y^j = 0$, $j = 1, \ldots, m_{\mathrm{e}}$ and $z^j = 0$, $j \in I^{\mathrm{i}}(x)$.

Note that Assumption 1 implies that $\tilde{X}$ is nonempty and, together with Assumptions 2 and 3(i), that $\tilde{X}_0$ is nonempty.

Our regularity assumption, Assumption 3, is considerably milder than linear independence of the gradients of all c_i's and all active d_i's. Wächter and Biegler observed in [8] that the latter assumption is undesirable, in that whenever there are two or more equality constraints and the total number of constraints exceeds n, it is typically violated over entire submanifolds of $\tilde{X} \setminus X$. On the other hand, as shown in the next lemma (proved in the appendix), Assumption 3(ii) is equivalent to the existence at every $x \in \tilde{X}$ of a direction of strict descent for the violated equality and active inequality constraints in the tangent space to the non-violated equality constraints. (Note that at points $x \in \tilde{X}$ where such direction does not exist, the problem is "locally infeasible" (as observed in [8]) in which case failure of any local search algorithm is to be expected when the initial point is in the "region of attraction" of such x.)

Lemma 1 *Suppose Assumption 2 holds. Then Assumption 3(ii) is equivalent to the following statement (S): for every $x \in \tilde{X}$, there exists $v \in \mathcal{R}^n$ such that*

$$\langle \nabla c_i(x), v \rangle = 0 \quad \forall i \in I^{\mathrm{e}}(x),$$

$$\langle \nabla c_i(x), v \rangle < 0 \quad \forall i \notin I^{\mathrm{e}}(x),$$

$$\langle \nabla d_i(x), v \rangle > 0 \quad \forall i \in I^{\mathrm{i}}(x).$$

[4] It is readily checked that, under Assumption 3(i), Assumption 3(ii) is unaffected if "$z^j = 0$, $j \in I^{\mathrm{i}}(x)$" is deleted.

Wächter and Biegler [8] exhibited a simple optimization problem on which many recently proposed interior-point methods converge to infeasible points at which such direction v exists, in effect showing that convergence of these algorithms to KKT points cannot be proved unless a strong assumption is used that rules out such seemingly innocuous problems. On the other hand, it is readily checked that directions v as in Lemma 1 do exist at all spurious limit points in that example. Indeed, in the problem from [8], for some a, b, $c_1(x) = x_1^2 - x_2 + a$, $c_2(x) = -x_1 + x_3 + b$, $d_1(x) = x_2$, and $d_2(x) = x_3$ and the spurious limit points are of the form $[\zeta, 0, 0]^{\mathrm{T}}$ with $\zeta < 0$; $v = [2, 1, 1]^{\mathrm{T}}$ meets the conditions at such points. (Our $c_2(x)$ is the negative of that in [8] because in our framework the c_i's are nonnegative in $\tilde{X}$.) We will see that, indeed, the algorithm proposed below performs satisfactorily on this example.

Before presenting our algorithm, we briefly explore a connection between Problems (P) and (P_ρ). A point x is a KKT point of (P) if there exist $y \in \mathcal{R}^{m_{\mathrm{e}}}$, $z \in \mathcal{R}^{m_{\mathrm{i}}}$ such that

$$\begin{aligned} g(x) - A(x)^{\mathrm{T}} y - B(x)^{\mathrm{T}} z &= 0, && (18)\\ c(x) &= 0, && (19)\\ d(x) &\geq 0, && (20)\\ z^j d_j(x) &= 0,\ j = 1, ..., m_{\mathrm{i}}, && (21)\\ z^j &\geq 0,\ j = 1, ..., m_{\mathrm{i}}. && (22) \end{aligned}$$

Following [9] we term a point x *stationary* for (P) if there exist $y \in \mathcal{R}^{m_{\mathrm{e}}}$, $z \in \mathcal{R}^{m_{\mathrm{i}}}$ such that (18)-(21) hold (but possibly not (22)). Next, for given ρ, a point $x \in \tilde{X}$ is a KKT point of (P_ρ) if there exist $y \in \mathcal{R}^{m_{\mathrm{e}}}$, $z \in \mathcal{R}^{m_{\mathrm{i}}}$ such that

$$\begin{aligned} g(x) + A(x)^{\mathrm{T}}(\rho \mathrm{e}) - A(x)^{\mathrm{T}} y - B(x)^{\mathrm{T}} z &= 0, && (23)\\ c(x) &\geq 0, && (24)\\ d(x) &\geq 0, && (25)\\ y_j c_j(x) &= 0,\ j = 1, ..., m_{\mathrm{e}} && (26)\\ y_j &\geq 0,\ j = 1, ..., m_{\mathrm{e}} && (27)\\ z_j d_j(x) &= 0,\ j = 1, ..., m_{\mathrm{i}} && (28)\\ z_j &\geq 0,\ j = 1, ..., m_{\mathrm{i}}, && (29) \end{aligned}$$

where $\mathrm{e} \in \mathcal{R}^{m_{\mathrm{e}}}$ is a vector whose components are all 1. x is stationary for (P_ρ) if there exist $y \in \mathcal{R}^{m_{\mathrm{e}}}$, $z \in \mathcal{R}^{m_{\mathrm{i}}}$ such that (23)-(26) and (28) hold (but possibly not (27) and (29)).

The following proposition, found in [10], is crucial to the development and is repeated here for ease of reference.

Proposition 2 *Let ρ be given. If x is a KKT point for (P_ρ) with multiplier vectors y and z and $c(x) = 0$, then x is a KKT point for (P) with multipliers vectors $y - \rho \mathrm{e}$ and z. If x is stationary for (P_ρ) with multiplier vectors y and z and $c(x) = 0$, then it is stationary for (P) with multipliers vectors $y - \rho \mathrm{e}$ and z.*

Proof: Using the fact that $c(x) = 0$, equations (23)-(29) imply

$$
\begin{aligned}
g(x) - A(x)^{\mathrm{T}}(y - \rho e) - B(x)^{\mathrm{T}} z &= 0 && (30)\\
c(x) &= 0 && (31)\\
d(x) &\geq 0 && (32)\\
z_j d_j(x) &= 0 && (33)\\
z_j &\geq 0 && (34)
\end{aligned}
$$

Thus x is a KKT point for (P) with multipliers $y - \rho \mathrm{e} \in \mathcal{R}^{m_{\mathrm{e}}}$ and $z \in \mathcal{R}^{m_{\mathrm{i}}}$. The second assertion follows similarly. $\triangle$

The proposed algorithm is based on solving Problem (P_ρ) for fixed values of $\rho > 0$ using the interior-point method outined in Section 2. The key issue will then be how to adaptively adjust ρ to force the iterate to asymptotically satisfy $c(x) = 0$.

For problem (P_ρ), the barrier function (2) becomes

$$
\beta(x, \rho, \mu) = f(x) + \rho \sum_{j=1}^{m_{\mathrm{e}}} c_j(x) - \sum_{j=1}^{m_{\mathrm{e}}} {\mu_{\mathrm{e}}}^j \ln(c_j(x)) - \sum_{j=1}^{m_{\mathrm{i}}} {\mu_{\mathrm{i}}}^j \ln(d_j(x)).
$$

Its gradient is given by

$$
\nabla \beta(x, \rho, \mu) = g(x) + A(x)^{\mathrm{T}}(\rho \mathrm{e}) - A(x)^{\mathrm{T}} C(x)^{-1} \mu_{\mathrm{e}} - B(x)^{\mathrm{T}} D(x)^{-1} \mu_{\mathrm{i}}. \quad (35)
$$

Proceeding as in Section 2, define

$$
\begin{aligned}
y &= C(x)^{-1} \mu_{\mathrm{e}}, && (36)\\
z &= D(x)^{-1} \mu_{\mathrm{i}}, && (37)
\end{aligned}
$$

and consider solving the nonlinear system in (x, y, z):

$$g(x) + A(x)^{\mathrm{T}}(\rho \mathrm{e} - y) - B(x)^{\mathrm{T}} z = 0, \quad (38)$$

$$\mu_{\mathrm{e}} - C(x) y = 0, \quad (39)$$

$$\mu_{\mathrm{i}} - D(x) z = 0, \quad (40)$$

by means of the quasi-Newton iteration

$$\begin{bmatrix} -W & A(x)^{\mathrm{T}} & B(x)^{\mathrm{T}} \\ YA(x) & C(x) & 0 \\ ZB(x) & 0 & D(x) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \begin{bmatrix} g(x) + A(x)^{\mathrm{T}}(\rho \mathrm{e} - y) - B(x)^{\mathrm{T}} z \\ \mu_{\mathrm{e}} - C(x) y \\ \mu_{\mathrm{i}} - D(x) z \end{bmatrix}, \quad (L(x, y, z, \rho, \mu_{\mathrm{e}}, \mu_{\mathrm{i}}))$$

where $Y = \mathrm{diag}(y_i)$, $Z = \mathrm{diag}(z_i)$ and W approximates the Hessian at (x, y, z) of the Lagrangian associated with (P_ρ). Properties of this linear system of equations are summarized in the next lemma, which is standard.

Lemma 3 *Suppose Assumptions 1, 2, and 3(i) hold. Given any vector $x \in \tilde{X}$, any positive definite matrix $W \in \mathcal{R}^{n \times n}$ and any nonnegative vectors $y \in \mathcal{R}^{m_e}$ and $z \in \mathcal{R}^{m_i}$ such that $y^j > 0$ if $c_j(x) = 0$ and $z^j > 0$ if $d_j(x) = 0$, the matrix*

$$\begin{bmatrix} -W & A(x)^{\mathrm{T}} & B(x)^{\mathrm{T}} \\ YA(x) & C(x) & 0 \\ ZB(x) & 0 & D(x) \end{bmatrix}$$

is nonsingular. Furthermore, given $\rho > 0$, x is stationary for (P_ρ) if and only if there exist $\Delta y^0 \in \mathcal{R}^{m_e}$, $\Delta z^0 \in \mathcal{R}^{m_i}$ such that $(0, \Delta y^0, \Delta z^0)$ solves $L(x, y, z, \rho, 0, 0)$; in such case, $y + \Delta y^0$ and $z + \Delta z^0$ are the (unique) multiplier vectors associated with x.

System $L(x, y, z, \rho, \mu_{\mathrm{e}}, \mu_{\mathrm{i}})$ is solved first with $(\mu_{\mathrm{e}}, \mu_{\mathrm{i}}) = (0, 0)$, then with $(\mu_{\mathrm{e}}, \mu_{\mathrm{i}})$ set analogously to (7). Following that, a correction $\Delta \tilde{x}$ is computed by solving the appropriate linear least squares problem, and new iterates x^+, y^+ and z^+ are obtained as in Section 2.

Now for the central issue of how ρ is updated. As noted in the introduction, Mayne and Polak [10] adaptively increase ρ to keep it above the magnitude of the most negative equality constraint multiplier estimate. They use a rather expensive estimation scheme, which was later improved upon in [17] in a different context. A simpler update rule is used here, which involves no computational overhead. It is based on the observation that ρ should be increased whenever convergence is detected to a point—a KKT point for (P_ρ), in view of the convergence properties established in [9] and reviewed in Section 2—where some equality constraints is violated. Care must be exercised because, if such convergence is erroneously signaled (false alarm), a runaway phenomenon may be triggered, with ρ increasing uncontrollably without a KKT point of (P) being approached. That this does not take place is ensured by requiring that the following four conditions—all of which are needed in the convergence proof—be all satisfied in order for an increase of ρ to be triggered (here γ_1, γ_2, γ_3 and γ_4 are prescribed positive constants): (a) $\rho_k\|\Delta x_k^0\| < \gamma_3$, so that ρ_k is allowed to become large only when Δx_k^0 becomes small, indicating that a stationary point for (P_{ρ_k}) is being approached; (b) $\min_j\{y_k^j + \Delta y_k^j\} < \gamma_1$, i.e., not all c_js are clearly becoming active as the limit point is approached; (c) $z_k^j + \Delta z_k^j \geq -\gamma_2$ for all j, i.e., no z_k^j is diverging to $-\infty$ due to ρ_k being increased too fast (i.e., if ρ_k is growing large, either z_k^j is positive or it is becoming negligible compared to ρ_k) violation of which would indicate that the limit point is not KKT; (d) for every $j \in \{1,\dots,m_{\rm e}\}$, either $y_k^j + \Delta y_k^j < \rho_k$ or $\rho_k c_j(x_k) < \gamma_4$, i.e., if ρ_k is becoming large and c_j is not becoming active then the associated multiplier has already settled down to a reasonably small value.

We are now ready to state the algorithm. Note that, departing from [9], we allow the initial guess x_0 to lie on the boundary of $\tilde{X}$.

Algorithm A.

Parameters. $\xi \in (0,1/2)$, $\eta \in (0,1)$, $\gamma_1 > 0$, $\gamma_2 > 0$, $\gamma_3 > 0$, $\gamma_4 > 0$, $\nu > 2$, $\theta \in (0,1)$, $w_{\max} > 0$, $\delta > 1$, $\tau \in (2,3)$, $\kappa \in (0,1)$.

Data. $x_0 \in \tilde{X}$, $W_0 \in \mathcal{R}^{n\times n}$, positive definite, $\rho_0 > 0$, $y_0^i \in (0, w_{\max})$, $i = 1,\dots,m_{\rm e}$, $z_0^i \in (0, w_{\max})$, $i = 1,\dots,m_{\rm i}$

Step 0: Initialization. Set $k = 0$.

Step 1: Computation of search arc:

i. Compute $(\Delta x_k^0, \Delta y_k^0, \Delta z_k^0)$ by solving $L(x_k, y_k, z_k, \rho_k, 0, 0)$. If $\Delta x_k^0 = 0$ then: if $x_k \in \tilde{X} \setminus \tilde{X}_0$, stop, else set $\Delta y_k = \Delta y_k^0$, $\Delta z_k = \Delta z_k^0$, $x_{k+1} = x_k$, and go to Step 3ii.

ii. Compute $(\Delta x_k^1, \Delta y_k^1, \Delta z_k^1)$ by solving $L(x_k, y_k, z_k, \rho_k, \|\Delta x_k^0\|^\nu y_k, \|\Delta x_k^0\|^\nu z_k)$.

iii. Set

$$\varphi_k = \begin{cases} 1 & \text{if } \langle \nabla f_{\rho_k}(x_k), \Delta x_k^1 \rangle \le \theta \langle \nabla f_{\rho_k}(x_k), \Delta x_k^0 \rangle \\ (1-\theta)\frac{\langle \nabla f_{\rho_k}(x_k), \Delta x_k^0 \rangle}{\langle \nabla f_{\rho_k}(x_k), \Delta x_k^0 - \Delta x_k^1 \rangle} & \text{otherwise.} \end{cases}$$

iv. Set

$$\begin{aligned} \Delta x_k &= (1-\varphi_k)\Delta x_k^0 + \varphi_k \Delta x_k^1, \\ \Delta y_k &= (1-\varphi_k)\Delta y_k^0 + \varphi_k \Delta y_k^1, \\ \Delta z_k &= (1-\varphi_k)\Delta z_k^0 + \varphi_k \Delta z_k^1. \end{aligned}$$

v. Set

$$\begin{aligned} I_k^{\mathrm{e}} &= \{j : c_j(x_k) \le y_k^j + \Delta y_k^j\}, \\ I_k^{\mathrm{i}} &= \{j : d_j(x_k) \le z_k^j + \Delta z_k^j\}, \\ J_k^{\mathrm{e}} &= \{j : y_k^j + \Delta y_k^j \le -c_j(x_k)\}, \\ J_k^{\mathrm{i}} &= \{j : z_k^j + \Delta z_k^j \le -d_j(x_k)\}. \end{aligned}$$

vi. Set $\Delta\tilde{x}_k$ to be the solution of the linear least squares problem

$$\min \frac{1}{2}\langle \Delta\tilde{x}, W_k \Delta\tilde{x} \rangle \text{ s.t. } \quad c_j(x_k + \Delta x_k) + \nabla c_j(x_k)^{\mathrm{T}} \Delta\tilde{x}_k = \psi_k, \quad \forall j \in I_k^{\mathrm{e}} \tag{41}$$

$$d_j(x_k + \Delta x_k) + \nabla d_j(x_k)^{\mathrm{T}} \Delta\tilde{x}_k = \psi_k, \quad \forall j \in I_k^{\mathrm{i}} \tag{42}$$

where

$$\psi_k = \max \left\{ \|\Delta x_k\|^\tau, \max_{j \in I_k^{\mathrm{e}}} \left| \frac{\Delta y^j}{y^j + \Delta y^j} \right|^\kappa \|\Delta x_k\|^2, \max_{j \in I_k^{\mathrm{i}}} \left| \frac{\Delta z^j}{z^j + \Delta z^j} \right|^\kappa \|\Delta x_k\|^2 \right\}. \tag{43}$$

If $J_k^{\mathrm{e}} \cup J_k^{\mathrm{i}} \ne \emptyset$ or (41)-(42) has no solution or $\|\Delta\tilde{x}_k\| > \|\Delta x_k\|$, set $\Delta\tilde{x}_k$ to 0.
Step 2. Arc search. Compute α_k, the first number α in the sequence $\{1, \eta, \eta^2, ...\}$ satisfying

$$\begin{aligned} f_{\rho_k}(x_k + \alpha\Delta x_k + \alpha^2 \Delta\tilde{x}_k) &\le f_{\rho_k}(x_k) + \xi\alpha\langle \nabla f_{\rho_k}(x_k), \Delta x_k \rangle \\ c_j(x_k + \alpha\Delta x_k + \alpha^2 \Delta\tilde{x}_k) &> 0, \quad \forall j \\ d_j(x_k + \alpha\Delta x_k + \alpha^2 \Delta\tilde{x}_k) &> 0, \quad \forall j \\ c_j(x_k + \alpha\Delta x_k + \alpha^2 \Delta\tilde{x}_k) &> c_j(x_k), \quad \forall j \in J_k^{\mathrm{e}} \\ d_j(x_k + \alpha\Delta x_k + \alpha^2 \Delta\tilde{x}_k) &> d_j(x_k), \quad \forall j \in J_k^{\mathrm{i}} \end{aligned}$$

Step 3. Updates.
i. Set

$$x_{k+1} \;=\; x_k + \alpha_k \Delta x_k + \alpha_k^2 \Delta \tilde{x}_k.$$

If $J_k^{\mathrm{e}} \cup J_k^{\mathrm{i}} = \emptyset$, set

$$\begin{aligned} y_{k+1}^j &= \min\{w_{\max}, \max\{y_k^j + \Delta y_k^j, \|\Delta x_k\|\}\},\ j = 1, \ldots, m_{\mathrm{e}}, \\ z_{k+1}^j &= \min\{w_{\max}, \max\{z_k^j + \Delta z_k^j, \|\Delta x_k\|\}\},\ j = 1, \ldots, m_{\mathrm{i}}; \end{aligned}$$

otherwise, set $y_{k+1} = y_0$ and $z_{k+1} = z_0$. Select W_{k+1}, positive definite.
ii. If the following four conditions are satisfied: (i) $\min_j\{y_k^j + \Delta y_k^j\} < \gamma_1$, (ii) $\rho_k \|\Delta x_k^0\| < \gamma_3$, (iii) $z_k^j + \Delta z_k^j \geq -\gamma_2$ for all j, (iv) for every $j \in \{1, \ldots, m_{\mathrm{e}}\}$, either $y_k^j + \Delta y_k^j < \rho_k$ or $\rho_k c_j(x_k) < \gamma_4$; then set

$$\rho_{k+1} = \delta \rho_k;$$

else set $\rho_{k+1} = \rho_k$.
iii. Set $k = k + 1$ and go back to Step 1.
Note that, in view of the stopping criterion in Step 1i and of the arc search rule in Step 2, with the only possible exception of the initial guess x_0, all iterates belong to $\tilde{X}_0$, the strictly feasible set for all (P_ρ)s.

4 Convergence Analysis

We just observed that $x_k \in \tilde{X}_0$ for all $k > 0$. Thus the algorithm can stop at Step 1i only for $k = 0$. In view of Lemma 3, under Assumptions 1, 2, and 3(i), this can happen only if x_0 is stationary for Problem (P_{ρ_0}). It may or may not be KKT for (P). If it is not, a different starting point should be used. In the sequel, we implicitly assume that x_0 does not meet the stopping criterion in Step 1i.

Proposition 4 *Under Assumptions 1–3, Algorithm A is well defined and constructs an infinite sequence.*

Proof: We first show that the algorithm cannot loop indefinitely through the cycle Step 1i–Step 3ii–iii. Indeed, in view of Lemma 3 and of the stopping criterion in Step 1i, branching from Step 1i to Step 3ii at iteration k can take place only if x_k is stationary for (P_{ρ_k}) and $x_k \notin \tilde{X}_0$, implying that

$$(\nabla f_{\rho_k}(x_k) =)\nabla f(x_k) + \rho_k \sum_{j=1}^{m_e} \nabla c_j(x_k) = 0, \tag{44}$$

and $\Delta x_k^0 = 0$, $y_k + \Delta y_k = 0$, $z_k + \Delta z_k = 0$. The latter implies that all conditions in Step 3ii are met, so that $\rho_{k+1} > \rho_k$. Also, $x_{k+1} = x_k$. Cycling can take place only if, again, $\Delta x_{k+1}^0 = 0$, which would imply

$$\nabla f(x_k) + \rho_{k+1} \sum_{j=1}^{m_e} \nabla c_j(x_k) = 0. \tag{45}$$

However, since $\rho_{k+1} > \rho_k$, (44) together with (45) imply

$$\sum_{j=1}^{m_e} \nabla c_j(x_k) = 0$$

in contradiction with Assumption 3(ii), proving the claim. The proposition then directly follows from Proposition 3.3 of [9]. (Assumptions A4 through A6 of [9] are not needed for that proposition.) $\triangle$

The next lemma is central to this analysis. Its proof relies on an additional assumption, on the sequence $\{W_k\}$.

Assumption 4 *If the sequence $\{x_k\}$ generated by Algorithm A is bounded, then there exists constants $\sigma_2 \geq \sigma_1 > 0$ such that, for all k,*

$$\sigma_1 \|x\|^2 \leq \langle x, W_k x \rangle \leq \sigma_2 \|x\|^2 \quad \forall x \in \mathcal{R}^n.$$

Lemma 5 *Suppose Assumptions 1–4 hold. If the infinite sequence $\{x_k\}$ generated by Algorithm A is bounded, then ρ_k is increased only finitely many times.*

Proof: By contradiction. Suppose ρ_k is increased infinitely many times, i.e., there exists an infinite index set $\mathcal{K}$ such that $\rho_{k+1} > \rho_k$ for all $k \in \mathcal{K}$. The criteria that trigger ρ_k to increase must thus be satisfied for all $k \in \mathcal{K}$, i.e.,

$$\begin{aligned} \min_j \{y_k^j + \Delta y_k^j\} &< \gamma_1, \ \forall k \in \mathcal{K} && (46) \\ \rho_k \|\Delta x_k^0\| &< \gamma_3, \ \forall k \in \mathcal{K} && (47) \\ z_k^j + \Delta z_k^j &\geq -\gamma_2 \ \forall j, \ \forall k \in \mathcal{K}, && (48) \end{aligned}$$

and for every $j \in \{1, \dots, m_\mathrm{e}\}$, either

$$y_k^j + \Delta y_k^j < \rho_k \ \forall k \in \mathcal{K} \qquad (49)$$

or

$$\rho_k c_j(x_k) < \gamma_4 \ \forall k \in \mathcal{K}. \qquad (50)$$

As per Step 1i of the algorithm, the following holds, with $y_k' = y_k + \Delta y_k^0$ and $z_k' = z_k + \Delta z_k^0$:

$$\begin{aligned} W_k \Delta x_k^0 + g(x_k) + A(x_k)^\mathrm{T}(\rho_k \mathrm{e} - y_k') - B(x_k)^\mathrm{T} z_k' &= 0 && (51) \\ Y_k A(x_k) \Delta x_k^0 + C(x_k) y_k' &= 0 && (52) \\ Z_k B(x_k) \Delta x_k^0 + D(x_k) z_k' &= 0 && (53) \end{aligned}$$

Since $\{\rho_k\}$ tends to infinity, it follows from (46) that $\{\rho_k \mathrm{e} - y_k'\}$ tends to infinity on $\mathcal{K}$. Consequently, the sequence $\{\alpha_k\}$, with

$$\alpha_k = \max\left\{\left|\rho_k - y_{1,k}'\right|, \dots, \left|\rho_k - y_{m_\mathrm{e},k}'\right|, \left|z_{1,k}'\right|, \dots, \left|z_{m_\mathrm{i},k}'\right|, 1\right\},$$

tends to infinity on $\mathcal{K}$ as well. Define

$$\begin{aligned} \hat{y}_k^j &= \frac{\rho_k - y_{j,k}'}{\alpha_k}, \quad j = 1, \dots, m_\mathrm{e}, \\ \hat{z}_k^j &= \frac{z_{j,k}'}{\alpha_k}, \quad j = 1, \dots, m_\mathrm{i} \end{aligned}$$

for $k \in \mathcal{K}$. By construction $\left|\hat{y}_k^j\right| \leq 1$ for $j = 1, ..., m_e$ and $\left|\hat{z}_k^j\right| \leq 1$ for $j = 1, \dots, m_i$ for all $k \in \mathcal{K}$. Since in addition the sequence $\{x_k\}_{k \in \mathcal{K}}$ is bounded by assumption, there must exist an infinite index set $\mathcal{K}' \subseteq \mathcal{K}$and vectors $x^* \in \mathcal{R}^n$, $\hat{y}^* \in \mathcal{R}^{m_e}$, and $\hat{z}^* \in \mathcal{R}^{m_i}$ such that

$$\begin{aligned} \lim_{\substack{k\to\infty \\ k\in\mathcal{K}'}} x_k &= x^* \\ \lim_{\substack{k\to\infty \\ k\in\mathcal{K}'}} \hat{y}_k &= \hat{y}^* \\ \lim_{\substack{k\to\infty \\ k\in\mathcal{K}'}} \hat{z}_k &= \hat{z}^*. \end{aligned}$$

Further it follows from (48)–(50) that $z^* \geq 0$ and, for every $j \in \{1, \ldots, m_e\}$, either $\hat{y}^{*,j} \geq 0$ or $c_j(x^*) = 0$ (or both), implying that $\hat{y}^{*,j} \geq 0$ for all $j \notin I^{\mathrm{e}}(x^*)$. Boundedness of $\{x_k\}$ and the continuity assumptions imply that $\{A^{\mathrm{T}}(x_k)\}$ and $\{B^{\mathrm{T}}(x_k)\}$ are bounded. In view of (47) and of Assumption 4, dividing (51) by α_k and taking the limit as k goes to infinity, $k \in \mathcal{K}'$, yields

$$A(x^*)^{\mathrm{T}}\hat{y}^* - B(x^*)^{\mathrm{T}}\hat{z}^* = 0. \tag{54}$$

Since $\mathcal{K}' \subseteq \mathcal{K}$, (47) implies that $\Delta x_k^0 \overset{k\in\mathcal{K}'}{\to} 0$. Further, $\{Y_k\}$ and $\{Z_k\}$ are bounded by construction. Thus dividing (53) by α_k and taking the limit as $k \overset{k\in\mathcal{K}'}{\to} \infty$ yields:

$$D\,(x^*)\,\hat{z}^* \;=\; 0$$

Thus $\hat{z}_j^* = 0$ for all $i \notin I^{\mathrm{i}}(x^*)$. Since $\hat{y}^*$ is nonzero, (54) contradicts Assumption 3(ii). $\triangle$

In the sequel, we denote by $\bar{\rho}$ the final value of ρ_k.

Algorithm A now reduces to the algorithm described in Section 2 applied to Problem $(P_{\bar{\rho}})$. It is shown in [9] that, under Assumptions 1–4, if the sequence $\{x_k\}$ constructed by Algorithm A is bounded, then all its accumulation points are stationary for $(P_{\bar{\rho}})$. To conclude that they are KKT points for $(P_{\bar{\rho}})$, an additional assumption is used, under which it also follows that the entire sequence $\{x_k\}$ converges. (See the proof of Theorem 3.11 in [9] for the latter fact, which was inadvertently omitted from the statement of that theorem.) Recall that $\bar{\rho}$ is of the form $\rho_0\delta^\ell$ for some nonnegative integer ℓ.

Assumption 5 *For $\rho \in \{\rho_0\delta^\ell : \ \ell$ a nonnegative integer$\}$, all stationary points of (P_ρ) are isolated.*

For convergence to KKT points of (P) to follow, the fact that condition (i) in Step 3ii of Algorithm A must eventually be violated if ρ_k stops increasing is crucial. A glance at the four conditions in that step suggests that this will be the case if the dual variables converge to the KKT multipliers for $(P_{\bar{\rho}})$. To prove that this indeed occurs, one more assumption is used.

Assumption 6 *For $\rho \in \{\rho_0 \delta^\ell : \ \ell \text{ a nonnegative integer}\}$, strict complementarity holds at all KKT points of (P_ρ).*

Proposition 6 *Suppose Assumptions 1–6 hold. If the infinite sequence $\{x_k\}$ generated by Algorithm A is bounded, then it converges to a KKT point x^* of $(P_{\bar{\rho}})$. Moreover, with y^* and z^* the associated KKT multiplier vector for $c(x)$ and $d(x)$,*

(i) $\{\Delta x_k\} \to 0$ as $k \to \infty$, $\{y_k + \Delta y_k\} \to y^*$ as $k \to \infty$ and $\{z_k + \Delta z_k\} \to z^*$ as $k \to \infty$;

(ii) for k large enough, $J_k^{\mathrm{e}} = \emptyset = J_k^{\mathrm{i}}$, $I_k^{\mathrm{e}} = I^{\mathrm{e}}(x^*)$, and $I_k^{\mathrm{i}} = I^{\mathrm{i}}(x^*)$;

(iii) if $y^{*,j} \le w_{\max}$ for all j, then $\{y_k\} \to y^*$ as $k \to \infty$; if $z^{*,j} \le w_{\max}$ for all j, then $\{z_k\} \to z^*$ as $k \to \infty$.

Proof: Follows from Proposition 4.2 in [9], noting that, except for strict complementarity which is guaranteed by our Assumption 6, Assumption A8 of [9] is used in the proofs of Lemma 4.1 of [9] and Proposition 4.2 of [9] only to infer that x^* is an *isolated* KKT point, a fact which in the present situation follows from Assumption 5. $\triangle$

Theorem 7 *Suppose Assumptions 1–6 hold. If the infinite sequence $\{x_k\}$ generated by Algorithm A is bounded, then it converges to a KKT point x^* of (P). Moreover, in such case, $\{y_k + \Delta y_k - \rho \mathrm{e}\}$ converges to $\bar{y}^*$ and $\{z_k + \Delta z_k\}$ converges to z^*, where $\bar{y}^*$ and z^* are the multiplier vectors associated to x^* for problem (P).*

Proof: We know from Proposition 6 that (i) $\{x_k\} \to x^*$, a KKT point for $(P_{\bar{\rho}})$; (ii) $\{\Delta x_k\} \to 0$; (iii) $\{y_k + \Delta y_k\} \to y^* \ge 0$, the multiplier vector associated with $c(x) \ge 0$, and (iv) $\{z_k + \Delta z_k\} \to z^* \ge 0$, the multiplier vector associated with $d(x) \ge 0$. Thus conditions (ii), (iii) and (iv) in Step 3(ii) if Algorithm A are all satisfied for k large enough. Since $\rho_k = \bar{\rho}$ for k large enough, it follows from Step 3(ii) of Algorithm A that condition (i) must fail for k large enough, i.e., $y_k + \Delta y_k \geq \gamma_1 \mathrm{e}$ for k large enough, implying that $y^* \geq \gamma_1 \mathrm{e}$. On the other hand, since $\gamma_1 > 0$, it follows from complementary slackness that $c(x^*) = 0$. Since the algorithm generates feasible iterates, we are guaranteed that $d_j(x^*) \geq 0$, $j = 1, ..., m_i$. Application of Proposition 2 concludes the proof of the first claim. The second claim then follows from Proposition 2 and Proposition 6(i). $\triangle$

Finally, rate of convergences results are inherited from the results in [9]. We report them here for ease of reference. Let x^* be the KKT point for (P) and $(P_{\bar\rho})$ to which, in view of Theorem 7, the sequence $\{x_k\}$ is guaranteed to converge under Assumptions 1–6, and let y^* and z^* be the associated multipliers for $(P_{\bar\rho})$. The Lagrangian associated with (P) is given by

$$\mathcal{L}(x, \bar{y}, z) = f(x) - \langle \bar{y}, c(x) \rangle - \langle z, c(x) \rangle.$$

Assumption 7 *f, c_j, $j = 1, \ldots, m_\mathrm{e}$, and d_j, $j = 1, \ldots, m_\mathrm{i}$ are three times continuously differentiable. Furthermore, the second order sufficiency condition holds (with strict complementarity under Assumption 6) for (P) at x^*, i.e., $\nabla^2 \mathcal{L}_{xx}(x^*, \bar{y}^*, z^*)$ is positive definite on the subspace $\{v \text{ s.t. } \langle \nabla c_j(x^*), v \rangle = 0\ \forall j, \langle \nabla d_j(x^*), v \rangle = 0\ \forall j \in I^\mathrm{i}(x^*)\}$.*

It is readily checked that the second order sufficiency condition for $(P_{\bar\rho})$ is identical to that for (P).

As a final assumption, superlinear convergence requires that the sequence $\{W_k\}$ asymptotically carry appropriate second order information.

Assumption 8

$$\frac{\|N_k(W_k - \nabla^2_{xx}\mathcal{L}(x^*, \bar{y}^*, z^*))N_k \Delta x_k\|}{\|\Delta x_k\|} \to 0 \quad \text{as } k \to \infty \tag{55}$$

where

$$N_k = I - \hat{G}_k^\mathrm{T} \left(\hat{G}_k \hat{G}_k^\mathrm{T} \right)^{-1} \hat{G}_k$$

with

$$\hat{G}_k = [\nabla c_j(x_k),\ j = 1, \ldots, m_\mathrm{e}, \nabla d_j(x_k),\ j \in I^\mathrm{i}(x^*)]^\mathrm{T} \in \mathcal{R}^{(m_\mathrm{e} + |I(x^*)|) \times n},$$

Theorem 8 *Suppose Assumption 1–8 holds and suppose that* $y^{j,*} \leq w_{\max}$*,* $j = 1, \ldots, m_{\rm e}$*, and* $z^{j,*} \leq w_{\max}$*,* $j = 1, \ldots, m_{\rm i}$*. Then the arc search in Step 2 of Algorithm A eventually accepts a full step of one, i.e.,* $\alpha_k = 1$ *for all* k *large enough, and* $\{x_k\}$ *converges to* x^* *two-step superlinearly, i.e.,*

$$\lim_{k\to\infty} \frac{\|x_{k+2} - x^*\|}{\|x_k - x^*\|} = 0.$$

As can be expected, if Assumption 8 is strengthened to

$$\frac{\|N_k(W_k - \nabla^2_{xx}\mathcal{L}(x^*, y^*, z^*))\Delta x_k\|}{\|\Delta x_k\|} \to 0 \quad \text{as } k \to \infty,$$

Q-superlinear convergence follows.

5 Concluding Remarks

An interior-point algorithms for the solution of general nonconvex constrained optimization problems has been proposed an analyzed. Global convergence and local superlinear convergence have been proved under mild assumptions. In particular, it was pointed out that the proposed algorithm does not suffer the Wächter-Biegler effect. (This indeed has been verified on the Wächter-Biegler example, with $a = b = 1$, using a preliminary MATLAB implementation the algorithm. Fast convergence was observed to the global solution $[1, 2, 0]^{\rm T}$.) Numerical testing is underway.

While the present paper focussed on applying a version of the Mayne-Polak scheme to the algorithm of [9], there should be no major difficulty in similarly extending other feasible interior-point algorithms for inequality constrained problems to handle general constrained problems.

6 Appendix

Proof of Lemma 1

Proof: First suppose that (S) holds but Assumption 3(ii) does not. From the latter, there exist y^j, $j = 1 \ldots, m_{\rm e}$ and z^j, $j \in I^{\rm i}(x)$, not all zero, with $y^j \geq 0$ for all $j \notin I^{\rm e}(x)$ and $z^j \geq 0$ for all j, such that (17) holds. With v as implied by the former, it follows that

$$\sum_{j\notin I^{\rm e}(x)} y^j \langle \nabla c_j(x), v\rangle - \sum_{j\in I^{\rm i}(x)} z^j \langle \nabla d_j(x), v\rangle = 0,$$

where we have used the fact that $\langle \nabla c_j(x), v\rangle = 0$ for all $j \in I^{\mathrm{e}}(x)$. Every term in the sum is nonpositive, which implies that all must be zero. The strict inequalities in (S) then contradicts the fact that not all y_j's and z_j's are zero, proving the claim. To prove converse, let us suppose (S) does not hold and proceed to proof that Assumption 3(ii) must be violated. Let P denote projection onto the subspace $\mathcal{N} = \{v : \langle \nabla c_j(x), v\rangle = 0 \ \forall j \in I^{\mathrm{e}}(x)\}$. Since (S) does not hold, there cannot exist $\hat{v} \in \mathcal{R}^n$ such that $P\hat{v}$ (which is in $\mathcal{N}$) satisfies

$$\langle \nabla c_j(x), P\hat{v}\rangle < 0 \quad \forall j \notin I^{\mathrm{e}}(x)$$

and

$$-\langle \nabla d_j(x), P\hat{v}\rangle < 0 \quad \forall j \in I^{\mathrm{i}}(x).$$

or equivalently,

$$\langle P\nabla c_j(x), \hat{v}\rangle < 0 \quad \forall j \notin I^{\mathrm{e}}(x)$$

and

$$-\langle P\nabla d_j(x), \hat{v}\rangle < 0 \quad \forall j \in I^{\mathrm{i}}(x).$$

In other words, the set $\{P\nabla c_j(x), j \notin I^{\mathrm{e}}(x)\} \cup \{-P\nabla d_j(x), j \in I^{\mathrm{i}}(x)\}$ cannot be strictly separated from the origin, i.e., it is positively linearly dependent. Thus the origin is contained in the convex hull of that set, i.e., there exist $y^j \geq 0$, $j \notin I^{\mathrm{e}}(x)$, and $z^j \geq 0$, $j \in I^{\mathrm{i}}(x)$, not all zero, such that

$$\sum_{j \notin I^{\mathrm{e}}(x)} y^j P\nabla c_j(x) - \sum_{j \in I^{\mathrm{i}}(x)} z^j P\nabla d_j(x) = 0$$

or equivalently, such that

$$\sum_{j \notin I^{\mathrm{e}}(x)} y^j \nabla c_j(x) - \sum_{j \in I^{\mathrm{i}}(x)} z^j \nabla d_j(x) \perp \mathcal{N},$$

i.e.,

$$\sum_{j \notin I^{\mathrm{e}}(x)} y^j \nabla c_j(x) - \sum_{j \in I^{\mathrm{i}}(x)} z^j \nabla d_j(x) \in \mathrm{span}(\{\nabla c_j(x) : j \in I^{\mathrm{e}}(x)\}).$$

This contradicts Assumption 3(ii), and the proof is complete. $\triangle$

References

[1] H. Yamashita. A globally convergent primal-dual interior point method for constrained optimization. *Optimization Methods and Software*, 10:443–469, 1998.

[2] H. Yamashita, H. Yabe, and T. Tanabe. A globally and superlinearly convergent primal-dual interior point trust region method for large scale constrained optimization. Technical report, Mathematical Systems, Inc., Tokyo, Japan, July 1998.

[3] 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. Opt. Theory Appl.*, 89:507–541, 1996.

[4] D. M. Gay, M. L. Overton, and M. H. Wright. A primal-dual interior method for nonconvex nonlinear programming. In Y. Yuan, editor, *Advances in Nonlinear Programming*, pages 31–56. Kluwer Academic Publisher, 1998.

[5] A. Forsgren and P.E. Gill. Primal-dual interior methods for nonconvex nonlinear programming. *SIAM J. on Optimization*, 8(4):1132–1152, 1998.

[6] R.H. Byrd, M.E. Hribar, and J. Nocedal. An interior point algorithm for large-scale nonlinear programming. *SIAM J. on Optimization*, 9(4):877–900, 1999.

[7] R.J. Vanderbei and D.F. Shanno. An interior-point algorithm for nonconvex nonlinear programming. *Computational Optimization and Applications*, 13:231–252, 1999.

[8] A. Wächter and L.T. Biegler. Failure of global convergence for a class of interior point methods for nonlinear programming. *Mathematical Programming*, 88:565–574, 2000.

[9] E.R. Panier, A.L. Tits, and J.N. Herskovits. A QP-free, globally convergent, locally superlinearly convergent algorithm for inequality constrained optimization. *SIAM J. Contr. and Optim.*, 26(4):788–811, July 1988.

[10] D. Q. Mayne and E. Polak. Feasible direction algorithms for optimization problems with equality and inequality constraints. *Math. Programming*, 11:67–80, 1976.

[11] S. Segenreich, N. Zouain, and J.N. Herskovits. An optimality criteria method based on slack variables concept for large structural optimization. In *Proceedings of the Symposium on Applications of Computer Methods in Engineering*, pages 563–572, Los Angeles, California, 1977.

[12] J.N. Herskovits. *Développement d'une Méthode Numérique pour l'Optimization Non-Linéaire.* PhD thesis, Université Paris IX - Dauphine, Paris, France, January 1982.

[13] J.N. Herskovits. A two-stage feasible directions algorithm for nonlinear constrained optimization. *Math. Programming*, 36(1):19–38, 1986.

[14] A.L. Tits and J.L. Zhou. A simple, quadratically convergent algorithm for linear and convex quadratic programming. In W.W. Hager, D.W. Hearn, and P.M. Pardalos, editors, *Large Scale Optimization: State of the Art*, pages 411–427. Kluwer Academic Publishers, 1994.

[15] A.V. Fiacco and G.P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques.* Wiley, New-York, 1968.

[16] E. Polak and A.L. Tits. On globally stabilized quasi-newton methods for inequality constrained optimization problems. In R.F. Drenick and E.F. Kozin, editors, *Proceedings of the 10th IFIP Conference on System Modeling and Optimization — New York, NY, August-September 1981*, volume 38 of *Lecture Notes in Control and Information Sciences*, pages 539–547. Springer-Verlag, 1982.

[17] C.T. Lawrence and A.L. Tits. Nonlinear equality constraints in feasible sequential quadratic programming. *Optimization Methods and Software*, 6:265–282, 1996.