RC 22572 (W0209-111) September 27, 2002
Mathematics

IBM Research Report

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

Andre L. Tits[1], Andreas Waechter, Sasan Bakhtiari[1], Thomas J. Urban[2], Craig T. Lawrence[3]

IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

[1]University of Maryland
College Park, MD 20742

[2]Applied Physics Laboratory
Laurel, MD 20723

[3]Alphatech
Arlington, VA 22203

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

André L. Tits* Andreas Wächter† Sasan Bakhtiari*
Thomas J. Urban‡ Craig T. Lawrence§

July 16, 2002

Abstract

An exact-penalty-function-based 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 any given feasible interior-point method for inequality constrained problems. It is shown that the primal-dual interior-point framework allows for a simpler penalty parameter update rule than that discussed and analyzed by the originators of the scheme in the context of first order methods of feasible direction. Strong global and local convergence results are proved under mild assumptions. In particular, (i) the proposed algorithm does not suffer a common pitfall recently pointed out by Wächter and Biegler; and (ii) the positive definiteness assumption on the Hessian estimate, made in the original

*Department of Electrical and Computer Engineering and Institute for Systems Research, University of Maryland, College Park, MD 20742, USA
†IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA
‡Applied Physics Laboratory, Laurel, MD 20723, USA
§Alphatech, Arlington, VA 22203, USA

version of the algorithm, is relaxed, allowing for the use of exact Hessian information, resulting in local quadratic convergence. Promising numerical results are reported.

1 Introduction

Consider the problem

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

where $f : \mathcal{R}^n \to \mathcal{R}$, $c_j : \mathcal{R}^n \to \mathcal{R}$, $j = 1, ..., m_{\mathrm{e}}$ and $d_j : \mathcal{R}^n \to \mathcal{R}$, $j = 1, ..., m_{\mathrm{i}}$ 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, 8]. 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 [4, 5], while others, sometimes referred to as "infeasible", sidestep that requirement via the introduction of slack variables [1, 2, 3, 6, 7, 8]. As for equality constraints, some schemes include them "as is" in the perturbed optimality conditions [1, 2, 3, 4, 6, 7] while some soften this condition by making use of two sets of slack variables [8] 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, 7]) involve a trust region mechanism. In many cases (e.g. [2, 4, 8]), promising numerical results have been obtained. In some cases (e.g., [1, 2, 3, 6]), convergence properties have been proved under certain assumptions. Often, however, 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 [9] is not ruled out. An exception is [2], but rather strong assumptions are used there.

Recently, Wächter and Biegler [10] 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 can converge to spurious, infeasible points. They produced a simple, seemingly innocuous example where such behavior is observed when starting from rather arbitrary initial points. They pointed out that, where global convergence had been proved, it was under a linear independence assumption that often fails to hold in the case of problems with such number of constraints. One exception to this is [6], where the proposed trust-region-based method is proved to converge globally under fairly mild assumptions; another is the recent paper [11].

In this paper, we propose a line-search-based primal-dual interior-point algorithm of the "feasible" variety for which global and fast local convergence are proved to hold under rather mild assumptions. In particular, it involves a scheme to circumvent Maratos-like effects and is immune to the phenomenon observed in [10]. 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. The algorithm originates in two papers dating back more than one and two decades, respectively: [12] and [13]. The former proposed a feasible interior-point method for inequality constrained problems, proven to converge globally and locally superlineary, under appropriate assumptions. The latter offered a scheme for dealing with equality constraints in the context of a (largely arbitrary) algorithm for inequality constraint optimization.

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 for the solution of the equations in the KKT first order necessary conditions of optimality. Then solve again the same system of equations, but with the right-hand side appropriately perturbed so as to tilt the primal direction away from the constraint boundaries into the feasible set. The amount of perturbation is determined from the solution of the unperturbed system. Both the original and tilted primal directions are directions of descent for f. Decrease of f is then enforced by the line search to ensure global convergence. Maratos-like effects are avoided by means of a second order correction (adapted from an idea of Mayne and Polak [14]), allowing for fast local convergence to take place. These ideas were put forth in [12]. The central idea in the algorithm of [12] originated in earlier work by Herskovits and others [15, 16, 17]; see [18] for a detailed historical account. Ideas were also borrowed from [19] and [20].

In the mid-seventies Mayne and Polak proposed an ingenious scheme to incorporate equality constraints in methods of feasible directions [13].

The idea is to (1) relax each equality constraint ($c_j(x) = 0$) by replacing it with an inequality constraint ($c_j(x) \geq 0$); and (2) penalize departure from the constraint boundaries associated with these relaxed constraints by adding a simple penalty term ($\rho \sum c_j(x)$, $\rho > 0$) to the cost function. For fixed value of the penalty parameter ρ, the feasible direction method under consideration is used. 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 (for the original problem) at the solution. Accordingly, in [13] the penalty parameter is adaptively increased based on estimates of these multipliers. While [13] is concerned with classical first order feasible directions methods, it is pointed out in the introduction of that paper that the proposed scheme can convert "*any* [emphasis from [13]] 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. In [13], this is done by solving a linear least squares problem for all equality constraint multipliers, and all multipliers associated with ϵ-active inequality constraints. (That is, with inequality constraints whose current value is less than some fixed, prescribed ϵ—denoted ϵ' in [13].) The price to pay is that, if ϵ is "large", then (1) the computational overhead may become significant and, (2) the set of active constraints may be overestimated, leading to incorrect multiplier estimates. On the other hand, if ϵ is selected to be very small, the set of active constraints will be underestimated, again yielding incorrect multipler estimates. 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 [10], such assumption does not hold in the example discussed there, and it is typically violated on entire manifolds in problems with two or more equality constraints and a total number of constraints in excess of n.[1] In [17] it is suggested that the idea introduced in [13] could be readily applied to the interior-point algorithm proposed there, but no details are given. The Mayne-Polak idea was used in [21] in the context of feasible SQP. The ready availability of multiplier

[1]See discussion following the statement of Assumption 3 in Section 3 below.

estimates (for the penalized problem) in that context allows an improved multiplier estimation scheme (for the original problem), thus improving on the first shortcoming just pointed out; however, no attempt is made in [21] to dispense with the strong linear independence assumption.

In the 1980s and 1990s, other penalty parameter update rules have been proposed for ℓ_1 (as in [13]) or ℓ_∞ exact penalty functions, in the context of SQP and trust-region methods, among others. (See, e.g., [14, 22, 23, 24].) In most cases, just like in [13] and [21], the updating rule involves multiplier estimates whose computation requires the solution of a linear system of equations or even that of a linear program. An exception is [24] where the following simple rule is used: at iteration k, increase ρ if the constraint is far from being satisfied, specifically, if $\|c(x_k)\| > \upsilon_k$, where υ_k appropriately decreases to zero as k goes to infinity. This rule is proposed in the context of a trust region method, and υ_k involves the model decrease. A challenge when extending it to other contexts is that, if υ_k is chosen too small, ρ will increase unnecessarily, perhaps without bound.

The contributions of the present paper are as follows. First it is shown that all the convergence results proved in [12] for the algorithm proposed in that paper still hold if the positive definiteness assumption on the Hessian estimate is relaxed, and replaced with a significantly milder assumption. In particular, the new assumption allows for use of the exact Hessian. Subject to a minor modification of the algorithm, local quadratic convergence in the primal-dual space is proved when the exact Hessian is indeed used. Second, the algorithm is extended to general constrained problems by incorporating a modified Mayne-Polak scheme. Specifically, a new, simple penalty parameter update rule is introduced involving no additional computation. Such rule is made possible by the availability of multiplier estimates for the penalized problem through the primal-dual iteration. The resulting algorithm converges globally and locally superlinearly without requirement that a strong regularity assumption be satisfied, thus avoiding the pitfall observed in [10].

The balance of the paper is organized as follows. In Section 2 below, the algorithm from [12] is described in “modern” terms, from a barrier function perspective. It is shown how certain assumptions made in [12] can be relaxed, and quadratic convergence is shown for the case when the ”exact Hessian” is used. The overall algorithm is then motivated and described in Section 3. In Section 4, global and local superlinear convergence are proved. Preliminary numerical results are reported in Section 5, starting with results on the example discussed in [10]. Finally, Section 6 is devoted to concluding remarks. Throughout, $\|\cdot\|$ denotes the Euclidean norm or corresponding operator 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 [12], in the primal-dual interior-point formalism, then point out how the assumptions made in [12] can be relaxed without affecting the convergence theorems.

2.1 Brief review of [12]

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 [12] 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

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

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)

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

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

i.e.,

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

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, under the assumption made in [12] that W is positive definite[2] and given any strictly feasible primal-dual pair (x, z), the primal direction Δx^0 obtained by solving (6) with $\mu = 0$ is a descent direction for f at x. In [12], such a property is sought for the search direction and used in the line search. On the other hand, while any primal direction is "feasible" when starting from an interior point, Δx^0 is not necessarily a direction of ascent for "almost active" constraints, whereas when the components of μ are chosen to be strictly positive, such desirable ascent property is guaranteed but descent for f may be lost. Thus, the components of μ should

- be positive enough to prevent the primal step length from collapsing due to infeasibility,
- be small enough that significant descent for f is maintained, and
- go to zero fast enough to preserve the fast local convergence properties associated with the (quasi-)Newton iteration for (4)-(5) with $\mu = 0$.

This is achieved in [12] by selecting

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

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

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

[2]Below (Section 2.2) we show that this assumption can be relaxed.

where $\nu > 2$ and $\theta \in (0,1)$ are prespecified;[3] condition (8) ensures that Δx is still a descent direction for f.

In [12] primal and dual strict feasibility is enforced at each iteration. An arc search is performed to select a next primal iterate x^+. The search criterion includes decrease of f and strict primal feasibility. It involves a second order correction $\Delta\tilde{x}$ to allow a full Newton (or 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

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

with $\tau \in (2,3)$ and $\kappa \in (0,1)$ prespecified. If $J \neq \emptyset$ or (9) is infeasible or unbounded or $\|\Delta\tilde{x}\| > \|\Delta x\|$, $\Delta\tilde{x}$ is set to 0. The rationale for the first of these three conditions is that computing the Maratos correction involves some cost, and it is known to be of help only close to a solution: when $J \neq \emptyset$, the correction is not computed. 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 $\{1, \eta, \eta^2, \ldots\}$ such that

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

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

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

[3]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

$$z^{+,(j)} = \min\{\max\{\|\Delta x\|, z^{(j)} + \Delta z^{(j)}\}, z_{\max}\} \tag{14}$$

where $z_{\max} > 0$ is prespecified. 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 [12, Section 5] that the total work per iteration (in addition to function evaluations) is essentially one Cholesky decomposition of size $m_{\rm i}$ and one Cholesky decomposition of size equal to the number of active constraints at the solution.[4]

On the issue of global convergence, it is shown in [12] 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)[5] is finite.

Superlinear convergence of the primal sequence—in particular, eventual acceptance of the full step of one by the arc search—is also proved in [12] under appropriate second order assumptions, provided that none of the KKT multipliers at the solution are larger than $z_{\max}$ and that, asymptotically, W_k suitably approximates 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 [18] for details.

[4] There are two misprints in [12, 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 $B_k^{-1} M_k$.

[5] Such points are referred to in [12] as *stationary points.*

2.2 Global convergence under milder assumptions

Two assumptions made in [12] can be relaxed without affecting the convergence results proved there. First, Assumption A4 (x_0 is the initial point),

> The set $X \cap \{x \text{ s.t. } f(x) \leq f(x_0)\}$ is compact

can be eliminated altogether. Indeed, this assumption is invoked only in the proof of Lemma 3.8 and 3.9 of [12]. The former is readily proved without such assumption: convergence of $\{x_{k-1}\}$ on K directly follows from the assumed convergence on K of $\{x_k\}$ and $\{d_{k-1}\}$ (in the notation of [12]) and from the last displayed equation in the proof. As for the latter, a weaker statement by which K is selected under the additional restriction that $\{x_k\}$ converges on K is sufficient for the use made of that lemma, in Proposition 3.10 and Theorem 3.11.

Second and more significantly, Assumption A6 of [12], (in the notation of this paper)

> There exist $\sigma_1, \sigma_2 > 0$ such that $\sigma_1 \|v\|^2 \leq \langle v, W_k v \rangle \leq \sigma_2 \|v\|^2$, for all k, for all $v \in \mathcal{R}^n$

can be replaced with the following milder assumption.

Assumption PTH-A6*. *Given any index set K such that $\{x_k\}_{k \in K}$ is bounded, there exist $\sigma_1, \sigma_2 > 0$, such that, for all $k \in K$,*

$$\|W_k\| \leq \sigma_2$$

and

$$\langle v, \left(W_k + \sum_{i=1}^{m_{\rm i}} \frac{z_k^{(i)}}{d_i(x_k)} \nabla d_i(x_k) \nabla d_i(x_k)^{\rm T} \right) v \rangle \geq \sigma_1 \|v\|^2 \ \forall v \in R^n.$$

(Here $\{x_k\}$, $\{z_k\}$, and $\{W_k\}$ are the sequences of values of x, and z and W generated by the algorithm outlined above. The restriction of this assumption to bounded subsequences of $\{x_k\}$ is made in connection with our dropping Assumption A4.)

The difference with Assumption A6 of [12] is significant because, as is readily verified the (exact) Hessian of the Lagrangian satisfies the relaxed assumption in the neighborhood of any solution of (1) at which strong second order sufficiency conditions of optimality hold. It is shown in the appendix that all the results proved in [12] still hold under the new assumption. In particular, it is proven that the direction Δx^0 (in the notation of this paper) is still well defined and is a direction of descent for f. It should be noted that, when W is not positive definite, there are two ways (rather than one) in which (9) can fail to have a solution: when its feasible set is nonempty, its cost function could possibly be unbounded from below. As observed in the appendix, the analysis of [12] still implies that, locally around a "strong" minimizer, (9) still has a solution.

2.3 Local quadratic convergence

As noted at the end of Subsection 2.1, superlinear convergence of $\{x_k\}$ is proved in [12] under appropriate local assumptions. Here we show that, under the further assumption that, eventually, W_k is the Hessian evaluated at (x_k, z_k) of the Lagrangian associated with problem (1), the pair (x_k, z_k) converges Q-quadratically provided the following minor modification is made to the algorithm of [12]: replace (14) with

$$z^{+,(j)} = \min\{\max\{\|\Delta x\|^2, z^{(j)} + \Delta z^{(j)}\}, z_{\max}\},$$

i.e., allow z_k to go to zero like $\|\Delta x_k\|^2$ rather then merely $\|\Delta x_k\|$. It can be checked that this modification does not affect the analysis carried out in [12].

The proof is based on Proposition 3.10 of [18], which we restate here for ease of reference. (Related result are obtained in [3] and [25].)

Lemma 1 *Let $F : \mathcal{R}^\ell \to \mathcal{R}^\ell$ be twice continuously differentiable and let $w^* \in \mathcal{R}^\ell$ and $r > 0$ be such that $F(w^*) = 0$ and $\frac{\partial F}{\partial w}(w)$ is nonsingular whenever $w \in B(w^*, r) := \{w : \|w^* - w\| \leq r\}$. Let $v^N : B(w^*, r) \to \mathcal{R}^\ell$ be defined by $v^N(w) = -\left(\frac{\partial F}{\partial w}(w)\right)^{-1} F(w)$. Then given any $\Gamma_1 > 0$ there exists $\Gamma_2 > 0$ such that*

$$\|w^+ - w^*\| \leq \Gamma_2 \|w - w^*\|^2 \tag{15}$$

for every $w \in B(w^*, r)$ and $w^+ \in \mathcal{R}^\ell$ for which, for each $i \in \{1, \ldots, \ell\}$, either

(i) $|w^{+,(i)} - w^{*,(i)}| \leq \Gamma_1 \|v^N(w)\|^2$

or

(ii) $|w^{+,(i)} - (w^{(i)} + v_i^N(w))| \leq \Gamma_1 \|v^N(w)\|^2$.

Let $w := [x^{\mathrm{T}}, z^{\mathrm{T}}]^{\mathrm{T}}$, $w_k := [x_k^{\mathrm{T}}, z_k^{\mathrm{T}}]^{\mathrm{T}}$, etc., let

$$\Phi(w,\mu) = \begin{bmatrix} -(g(x) - B(x)^{\mathrm{T}} z) \\ D(x)z - \mu \end{bmatrix} \tag{16}$$

and let $M(w)$ denote the matrix in the left-hand side of (6) with W the "exact Hessian", i.e.,

$$W = \nabla_{xx}^2 f(x) - \sum_{i=1}^{m} z^{(i)} \nabla_{xx}^2 d_i(x)$$

Thus $M(w)$ is the Jacobian of $\Phi(w,\mu)$ with respect to w. (Note that $M(w)$ does not depend on μ.) We will invoke Lemma 1 with $F := \Phi(\cdot, 0)$. Observe that

$$\Delta w_k^0 = -M(w_k)^{-1} F(w_k, 0)$$

and

$$\Delta w_k = -M(w_k)^{-1} F(w_k, \mu_k),$$

and, since $\mu_k = O(\|\Delta x_k^0\|^\nu)$ and $M(w_k)^{-1}$ is bounded (in view of Lemma PTH-3.5* in the appendix), that

$$\Delta w_k - \Delta w_k^0 = O(\|\Delta x_k^0\|^\nu). \tag{17}$$

Next, we observe that, with a simple additional observation, the proof of Lemma 4.4 in [12] establishes that

$$\|\Delta \tilde{x}\| = O(\|\Delta w\|^2). \tag{18}$$

Indeed, in connection with the last displayed equation in that proof, since under our strict complementarity assumption $z_k^{(i)} + \Delta z_k^{(i)}$ ($\mu_{k,i}$ in the notation of [12]) is bounded away from zero for large k, we can write

$$\frac{z_k^{(i)}}{z_k^{(i)} + \Delta z_k^{(i)}} - 1 = O(|\Delta z_k^{(i)}|) = O(\|\Delta w_k\|).$$

and the claim follows.

Now, proceed analogously to the proof of Theorem 3.11 in [18]. Thus, with reference to Lemma 1, let $r > 0$ be such that $M(w)$ is nonsingular for all $w \in B(w^*, r)$ (in view of Lemma PTH-3.5* in the appendix, such r exists). Since $\{w_k\} \to w^*$ as $k \to \infty$, there exists k_0 such that $w_k \in B(w^*, r)$ for all $k \geq k_0$. Now let us first consider $\{z_k\}$. For $i \in I(x^*)$, in view of strict complementarity, $z_{k+1}^{(i)} = z_k^{(i)} + \Delta z_k^{(i)}$ for k large enough, so that, in view of (17), condition (ii) in Lemma 1 holds for k large enough. Next, for $i \notin I(x^*)$, for each k either again $z_{k+1}^{(i)} = z_k^{(i)} + \Delta z_k^{(i)}$ or (in view of our modified updating formula for z_k) $z_{k+1}^{(i)} = \|\Delta x_k\|^2$. In the latter case, since $z^{*,(i)} = 0$, noting again (17), we conclude that condition (i) in Lemma 1 holds. Next, consider $\{x^k\}$. Since $\alpha_k = 1$ for k large enough, we have

$$\|x_{k+1} - (x_k + \Delta x_k^0)\| = \|\Delta x_k - \Delta x_k^0 + \Delta \tilde{x}_k\|$$

which, in view of (17) and of (18) implies that condition (ii) again holds. Thus the conditions of Lemma 1 hold, and Q-quadratic convergence follows.

3 Overall Algorithm

Suppose now that m_e is not necessarily zero. Denote by X the feasible set for (P), i.e., let

$$X := \{x \in \mathcal{R}^n : c_j(x) = 0,\ j = 1, ..., m_e,\ d_j(x) \geq 0,\ j = 1, ..., m_i\}. \tag{19}$$

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

In [13], Mayne and Polak proposed a scheme to convert (P) to a sequence of inequality constrained optimization problems of the type

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

where $f_\rho(x) = f(x) + \rho \sum\limits_{j=1}^{m_e} c_j(x)$, 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 a feasible-iterate 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. In other words, 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

$$\tilde{X} := \{x \in \mathcal{R}^n : c_j(x) \geq 0,\ j = 1, ..., m_{\rm e}, \quad d_j(x) \geq 0,\ j = 1, \ldots, m_{\rm i}\}, \tag{20}$$

$$\tilde{X}_0 := \{x \in \mathcal{R}^n : c_j(x) > 0,\ j = 1, ..., m_{\rm e}, \quad d_j(x) > 0,\ j = 1, \ldots, m_{\rm i}\}. \tag{21}$$

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, $i = 1 \ldots, m_{\rm e}$ and d_i, $i = 1, \ldots, m_{\rm i}$ are continuously differentiable.*

Assumption 3 *For all $x \in \tilde{X}$, (i) the set $\{\nabla c_j(x) : j \in I^{\rm e}(x)\} \cup \{\nabla d_j(x) : j \in I^{\rm i}(x)\}$ is linearly independent; (ii) if $x \notin X$, then no scalars $y^{(j)} \geq 0$, $j \in I^{\rm e}(x)$, and $z^{(j)} \geq 0$, $j \in I^{\rm i}(x)$ exist such that*

$$\sum_{j=1}^{m_{\rm e}} \nabla c_j(x) = \sum_{j \in I^{\rm e}(x)} y^{(j)} \nabla c_j(x) + \sum_{j \in I^{\rm i}(x)} z^{(j)} \nabla d_j(x). \tag{22}$$

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

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. As observed in [10], 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 stated in the next lemma, Assumption 3(ii) is equivalent to the mere existence at every $x \in \tilde{X} \setminus X$ of a feasible (with respect to $\tilde{X}$) direction of strict descent for the ℓ_1 norm of $c(x)$. (Indeed Assumption 3(ii) simply states that the sum in the left-hand side of (22) does not belong to the closed convex cone generated by the listed constraint gradients and existence of such strict descent direction amounts to strict separation of that sum from this cone.)

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

$$\begin{aligned}\langle \sum_{j=1}^{m_\mathrm{e}} \nabla c_j(x), v\rangle &< 0,\\ \langle \nabla c_j(x), v\rangle > 0 &\quad \forall j \in I^\mathrm{e}(x),\\ \langle \nabla d_j(x), v\rangle > 0 &\quad \forall j \in I^\mathrm{i}(x).\end{aligned}$$

In [10], a simple optimization problem was exhibited, on which many recently proposed interior-point methods converge to infeasible points at which such a 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 2 do exist at all spurious limit points identified in [10]. Indeed, in the problem from [10], for some a, b, $c_1(x) = (x^{(1)})^2 - x^{(2)} + a$, $c_2(x) = -x^{(1)} + x^{(3)} + b$,[6] $d_1(x) = x^{(2)}$, and $d_2(x) = x^{(3)}$ and the spurious limit points are points of the form $[\zeta, 0, 0]^\mathrm{T}$, with $\zeta < 0$, at which *both c_1 and c_2 are nonzero*; $v = [1, 0, 0]^\mathrm{T}$ meets our conditions at such points. In fact, it is readily verified that Assumption 3(ii) is satisfied whenever $a \geq 0$, or $a < 0$ with $b \leq -\sqrt{|a|}$, and that, when $a < 0$ and $b > -\sqrt{|a|}$, the only point $x \in \tilde{X} \setminus X$ at which the condition in Assumption 3(ii) is violated is $(-\sqrt{|a|}, 0, 0)^\mathrm{T}$, at which $c_1(x) = 0$. In Section 5 we will discuss the behavior on this example of the algorithm proposed below.

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

$$g(x) - A(x)^\mathrm{T} y - B(x)^\mathrm{T} z = 0, \tag{23}$$

$$c(x) = 0, \tag{24}$$

$$d(x) \geq 0, \tag{25}$$

$$z^{(j)} d_j(x) = 0, \ j = 1, ..., m_\mathrm{i}, \tag{26}$$

$$z \geq 0. \tag{27}$$

[6] Our $c_2(x)$ is the negative of that in [10] because in our framework equality constraints must take on positive values at the initial point, while at the initial points of interest (as per Theorem 1 in [10]) $c_2(x)$ as defined in [10] is negative.

Following [12] 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 (23)-(26) hold (but possibly not (27)). 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

$$\begin{aligned}
g(x) + A(x)^{\mathrm{T}}(\rho \mathrm{e}) - A(x)^{\mathrm{T}} y - B(x)^{\mathrm{T}} z &= 0, && (28)\\
c(x) &\geq 0, && (29)\\
d(x) &\geq 0, && (30)\\
y^{(j)} c_j(x) &= 0, \; j = 1, ..., m_{\mathrm{e}}, && (31)\\
y &\geq 0, && (32)\\
z^{(j)} d_j(x) &= 0, \; j = 1, ..., m_{\mathrm{i}}, && (33)\\
z &\geq 0, && (34)
\end{aligned}$$

where $\mathrm{e} \in \mathcal{R}^{m_{\mathrm{e}}}$ is a vector whose components are all 1. A point 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 (28)-(31) and (33) hold (but possibly not (32) and (34)). The following proposition, found in [13], is crucial to the development and is repeated here for ease of reference.

Proposition 3 *Suppose Assumptions 1 and 2 hold. Let ρ be given. 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. Furthermore, if $z \geq 0$, then x is a KKT point for (P).*

Proof: Using the fact that $c(x) = 0$, equations (28)-(31) and (33) imply

$$\begin{aligned}
g(x) - A(x)^{\mathrm{T}}(y - \rho \mathrm{e}) - B(x)^{\mathrm{T}} z &= 0 && (35)\\
c(x) &= 0 && (36)\\
d(x) &\geq 0 && (37)\\
z^{(j)} d_j(x) &= 0 && (38)
\end{aligned}$$

Thus x is stationary 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 outlined in Section 2. The key issue will then be to determine how to 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

$$\nabla_x \beta(x,\rho,\mu) = g(x) + A(x)^\mathrm{T}(\rho \mathrm{e}) - A(x)^\mathrm{T} C(x)^{-1}\mu_\mathrm{e} - B(x)^\mathrm{T} D(x)^{-1}\mu_\mathrm{i}. \quad (39)$$

Proceeding as in Section 2, define

$$y = C(x)^{-1}\mu_\mathrm{e}, \quad (40)$$
$$z = D(x)^{-1}\mu_\mathrm{i}, \quad (41)$$

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

$$g(x) + A(x)^\mathrm{T}(\rho \mathrm{e} - y) - B(x)^\mathrm{T} z = 0, \quad (42)$$
$$\mu_\mathrm{e} - C(x)y = 0, \quad (43)$$
$$\mu_\mathrm{i} - D(x)z = 0, \quad (44)$$

by means of the (quasi-)Newton iteration

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

where $Y = \mathrm{diag}(y^{(j)})$, $Z = \mathrm{diag}(z^{(j)})$ and W is equal to, or approximates, the Hessian with respect to x, at (x,y,z), of the Lagrangian associated with (P_ρ).

System $L(x,y,z,W,\rho,\mu_\mathrm{e},\mu_\mathrm{i})$ 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 [13] 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 [21] 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 [12] 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. We avoid this by requiring that the following three 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 and γ_3 are prescribed positive constants): (a) $\|\Delta x_k^0\| \leq \gamma_1$, indicating the proximity of a stationary point for (P_{ρ_k}); (b) $y_k + \Delta y_k^0 \not\geq \gamma_2 \mathrm{e}$, i.e., not all c_js become strongly binding as the limit point is approached; (c) $y_k + \Delta y_k^0 \geq -\gamma_3 \mathrm{e}$ and $z_k + \Delta z_k^0 \geq -\gamma_3 \mathrm{e}$, i.e., no components of y_k or z_k is diverging to $-\infty$ due to ρ_k being increased too fast (i.e., if ρ_k is growing large, either y_k and z_k become nonnegative or their negative components become negligible compared to ρ_k), violation of which would indicate that the limit point is not KKT.

We are now ready to state the algorithm.

Algorithm A.

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

Data. $x_0 \in \tilde{X}_0$, $\rho_0 > 0$, $y_0^{(i)} \in (0, w_{\max}]$, $i = 1, \ldots, m_{\mathrm{e}}$, $z_0^{(i)} \in (0, w_{\max}]$, $i = 1, \ldots, m_{\mathrm{i}}$; $W_0 \in \mathcal{R}^{n \times n}$, such that

$$W_0 + \begin{bmatrix} A(x_0)^{\mathrm{T}} & B(x_0)^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} C(x_0)^{-1} Y_0 & \\ & D(x_0)^{-1} Z_0 \end{bmatrix} \begin{bmatrix} A(x_0) \\ B(x_0) \end{bmatrix} \tag{45}$$

is positive definite.

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 $L(x_k, y_k, z_k, W_k, \rho_k, 0, 0)$. If $\Delta x_k^0 = 0$ and $m_{\mathrm{e}} = 0$ then stop.

ii. Check the following three conditions: (i) $\|\Delta x_k^0\| \leq \gamma_1$, (ii) $y_k + \Delta y_k^0 \not\geq \gamma_2 \mathrm{e}$, (iii) $y_k + \Delta y_k^0 \geq -\gamma_3 \mathrm{e}$ and $z_k + \Delta z_k^0 \geq -\gamma_3 \mathrm{e}$. If all three conditions hold, then set $\rho_{k+1} = \delta \rho_k$, $x_{k+1} = x_k$, $y_{k+1} = y_k$, $z_{k+1} = z_k$, $W_{k+1} = W_k$, set $k = k + 1$ and go back to Step 1i. Otherwise, proceed to Step 1iii.

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

iv. Set

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

v. Set

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

vi. Set

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

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

$$\min \tfrac{1}{2}\langle \Delta \tilde{x}, W_k \Delta \tilde{x} \rangle \ \text{s.t.} \ \begin{array}{l} c_j(x_k + \Delta x_k) + \langle \nabla c_j(x_k), \Delta \tilde{x}_k \rangle = \psi_k, \ \forall j \in I_k^{\mathrm{e}} \\ d_j(x_k + \Delta x_k) + \langle \nabla d_j(x_k), \Delta \tilde{x}_k \rangle = \psi_k, \ \forall j \in I_k^{\mathrm{i}} \end{array} \tag{46}$$

where

$$\psi_k = \max\left\{ \|\Delta x_k\|^\tau, \max_{j \in I_k^{\mathrm{e}}} \left| \frac{\Delta y_k^{(j)}}{y_k^{(j)} + \Delta y_k^{(j)}} \right|^\kappa \|\Delta x_k\|^2, \max_{j \in I_k^{\mathrm{i}}} \left| \frac{\Delta z_k^{(j)}}{z_k^{(j)} + \Delta z_k^{(j)}} \right|^\kappa \|\Delta x_k\|^2 \right\}.$$

If $J_k^{\mathrm{e}} \cup J_k^{\mathrm{i}} \ne \emptyset$ or (46) is infeasible or unbounded 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

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

Step 3. Updates.
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

$$\begin{aligned} y_{k+1}^{(j)} &= \min\{\max\{\|\Delta x_k\|^2, y_k^{(j)} + \Delta y_k^{(j)}\}, w_{\max}\} \ j = 1, \ldots, m_{\mathrm{e}}, \\ z_{k+1}^{(j)} &= \min\{\max\{\|\Delta x_k\|^2, z_k^{(j)} + \Delta z_k^{(j)}\}, w_{\max}\} \ j = 1, \ldots, m_{\mathrm{i}}; \end{aligned}$$

otherwise, set $y_{k+1} = y_0$ and $z_{k+1} = z_0$. Set $\rho_{k+1} = \rho_k$ and select W_{k+1} such that

$$W_{k+1} + \begin{bmatrix} A(x_{k+1}) & B(x_{k+1}) \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} C(x_{k+1})^{-1} Y_{k+1} & \\ & D(x_{k+1})^{-1} Z_{k+1} \end{bmatrix} \begin{bmatrix} A(x_{k+1}) \\ B(x_{k+1}) \end{bmatrix}$$

is positive definite. Set $k = k + 1$ and go back to Step 1.

Remark 1 *The values assigned to y_{k+1} and z_{k+1} in Step 1ii are of no consequence as far as the theoretical properties of the algorithm are concerned provided dual feasibility is preserved. Rather than re-using the previous values as stated in the algorithm, it may be advisable to make use of the just computed corrections Δy_k^0 and Δz_k^0, e.g., by setting*

$$\begin{aligned} y_{k+1}^{(j)} &= \min\{\max\{y_k^{(j)}, y_k^{(j)} + \Delta y_k^{0,(j)}\}, w_{\max}\}, \ j = 1, \ldots, m_{\mathrm{e}}, \\ z_{k+1}^{(j)} &= \min\{\max\{z_k^{(j)}, z_k^{(j)} + \Delta z_k^{0,(j)}\}, w_{\max}\}, \ j = 1, \ldots, m_{\mathrm{i}}, \end{aligned}$$

which still insures dual feasibility. (A side effect of such rule is that the components of y_k and z_k are possibly increased but never decreased when ρ_k is increased, which makes some intuitive sense.)

Remark 2 *Similarly, variations can be considered for the dual variable update rule in Step 3 in the case when $J_k^{\mathrm{e}} \cup J_k^{\mathrm{i}} \neq \emptyset$. Indeed the convergence analysis of [12] remains unaffected as long as the components of y_{k+1} and*

z_{k+1} stay bounded away from zero (and bounded) over the set of iterates k at which $J_k^{\rm e} \cup J_k^{\rm i} \neq \emptyset$. A possible choice would be

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

where $w_{\min} \in (0, w_{\max})$ is prescribed. Unlike the update rule used in the algorithm statement (taken from [12]) this rule attempts to make use of some of the multiplier estimates even when $J_k^{\rm e} \cup J_k^{\rm i} \neq \emptyset$.

Remark 3 *If an initial point $x_0 \in \tilde{X}_0$ is not readily available, a point $x_0 \in \tilde{X}$ can be constructed as follows: (i) Perform a "Phase I" search by maximizing $\min_j d_j(x)$ without constraints. This can be done, e.g., by applying Algorithm A to the problem*

$$\max_{(x,\zeta)\in\mathcal{R}^{n+1}} \zeta \text{ s.t. } d_j(x) - \zeta \geq 0 \;\forall j.$$

A point x_0 satisfying $\min_j d_j(x) \geq 0$ will eventually be obtained (or the constructed sequence $\{x_k\}$ will be unbounded) provided $\min_j d_j(x)$ has no stationary point with negative value, i.e., provided that, for all x such that $\zeta := \min_j d_j(x) < 0$, the origin does not belong to the convex hull of $\{\nabla d_j(x) : d_j(x) = \zeta\}$. (ii) Redefine $c_j(x)$ to take values $-c_j(x)$ for every j such that the original $c_j(x)$ is negative. As a result, x_0 will be in $\tilde{X}$ for the reformulated problem. If it is on the boundary of $\tilde{X}$ rather than in its interior $\tilde{X}_0$, it can be readily perturbed into a point in $\tilde{X}_0$ (under Assumption 3(i)).

4 Convergence Analysis

We first show that Algorithm A is well defined. First of all, the conditions imposed on W_0 and (in Step 3) on W_k in Algorithm A are identical, for every fixed k, to the second condition in Assumption PTH-A6*. Thus the matrix in our (quasi-)Newton iteration is nonsingular, and it follows from Proposition 3.4 of [12] that, if $\Delta x_k^0 = 0$ for some k, then $\nabla f_{\rho_k}(x_k) = 0$, i.e., x_k is an unconstrained KKT point for (P_ρ); and it is readily checked that, in such case, $y_k + \Delta y_k^0$ and $z_k + \Delta z_k^0$ are the associated KKT multiplier vectors, i.e., are both zero. Thus, if finite termination occurs at Step 1i (i.e.,

$m_{\mathrm{e}} = 0$) then $\nabla f(x_k) = 0$, i.e., x_k is an unconstrained KKT point for (P); and if $\Delta x_k^0 = 0$ but finite termination does not occur (i.e., $m_{\mathrm{e}} > 0$) then Conditions (i) through (iii) in Step 1ii are satisfied, and the algorithm loops back to Step 1i. Thus Step 1iii is never executed when Δx_k^0 is zero. It then follows from Proposition 3.3 of [12] that, under Assumptions 1, 2, and 3(i), Algorithm A is well defined. (Assumptions A4 through A6 of [12] are not needed in that proposition.)

From now on, we assume that the algorithm never stops, i.e., that an infinite sequence $\{x_k\}$ is constructed. Our next task will be to show that, unless $\{x_k\}$ itself is unbounded, ρ_k is increased at most finitely many times. Assumption 3(ii) will be crucial here. An additional assumption, adapted from PTH-A6*, will be needed as well.

Assumption 4 *Given any index set K such that the sequence $\{x_k\}$ constructed by Algorithm A is bounded, there exist $\sigma_1, \sigma_2 > 0$ such that, for all $k \in K$,*

$$\|W_k\| \leq \sigma_2$$

and

$$\langle v, \left(W_k + \sum_{i=1}^{m_{\mathrm{e}}} \frac{y_k^{(i)}}{c_i(x_k)} \nabla c_i(x_k) \nabla c_i(x_k)^{\mathrm{T}} + \sum_{i=1}^{m_{\mathrm{i}}} \frac{z_k^{(i)}}{d_i(x_k)} \nabla d_i(x_k) \nabla d_i(x_k)^{\mathrm{T}} \right) v \rangle$$

$$\geq \sigma_1 \|v\|^2 \ \forall v \in \mathcal{R}^n.$$

Lemma 4 *Suppose Assumptions 1–4 hold. If the infinite sequence $\{x_k\}$ generated by Algorithm A is bounded, then ρ_k is increased at most 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., with $y_k' = y_k + \Delta y_k^0$ and $z_k' = z_k + \Delta z_k^0$,

$$\|\Delta x_k^0\| \leq \gamma_1, \ \forall k \in \mathcal{K} \tag{47}$$

$$y_k' \not\geq \gamma_2 \mathrm{e}, \ \forall k \in \mathcal{K} \tag{48}$$

$$y_k' \geq -\gamma_3 \mathrm{e}, \ \forall k \in \mathcal{K} \tag{49}$$

$$z_k' \geq -\gamma_3 \mathrm{e}, \ \forall k \in \mathcal{K}. \tag{50}$$

As per Step 1i of the algorithm, we have:

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

Since $\{\rho_k\}$ tends to infinity, it follows from (48) that $\{\|\rho_k \mathrm{e} - y_k'\|_\infty\}$ tends to infinity on $\mathcal{K}$. Consequently, the sequence $\{\alpha_k\}$, with

$$\alpha_k = \max\{\|\rho_k \mathrm{e} - y_k'\|_\infty, \|z_k'\|_\infty, 1\},$$

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

$$
\begin{aligned}
\hat{y}_k' &= \alpha_k^{-1}(\rho_k \mathrm{e} - y_k'), \quad j = 1, \ldots, m_{\mathrm{e}}, \qquad (54)\\
\hat{z}_k' &= \alpha_k^{-1} z_k', \quad j = 1, \ldots, m_{\mathrm{i}} \qquad (55)
\end{aligned}
$$

for $k \in \mathcal{K}$. By construction $\max\{\|\hat{y}_k\|_\infty, \|\hat{z}_k\|_\infty\} = 1$ for all $k \in \mathcal{K}$, k large enough. 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 $\hat{z}^* \in \mathcal{R}^{m_i}$, with $\hat{y}^*$ and $\hat{z}^*$ not both zero, such that

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

Boundedness of $\{x_k\}$ and the continuity assumptions imply that $\{A(x_k)\}$ and $\{B(x_k)\}$ are bounded. Further, $\{Y_k\}$ and $\{Z_k\}$ are bounded by construction. Dividing both sides of (52) by α_k, letting $k \to \infty$, $k \in \mathcal{K}'$ and using (47) shows that

$$\alpha_k^{-1} C(x_k) y_k' \to 0 \quad \text{as } k \to \infty,\ k \in \mathcal{K}'$$

implying that, for every $j \notin I^{\mathrm{e}}(x^*)$,

$$\alpha_k^{-1} y_k'^{,(j)} \to 0 \quad \text{as } k \to \infty,\ k \in \mathcal{K}'.$$

Together with (54), this implies that $\{\rho_k/\alpha_k\}$ converges to some limit $\omega \geq 0$ as $k \to \infty$, $k \in \mathcal{K}'$, with

$$\hat{y}^{*,(j)} = \omega, \quad \forall\, j \notin I^{\mathrm{e}}(x^*).$$

Next it follows from (49), (50), (54) and (55) that

$$\hat{y}^{*,(j)} \leq \omega \ \forall j \tag{56}$$

and that

$$\hat{z}^* \geq 0. \tag{57}$$

Further, dividing (53) by α_k and taking the limit as $k \to \infty$, $k \in \mathcal{K}'$ yields:

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

Thus $\hat{z}^{*,(j)} = 0$ for all $i \notin I^{\mathrm{i}}(x^*)$. Finally, in view of (47) and of Assumption 4, dividing (51) by α_k and taking the limit as $k \to \infty$, $k \in \mathcal{K}'$, yields

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

i.e.,

$$\sum_{j=1}^{m_{\mathrm{e}}} \hat{y}^{*,(j)}\nabla c_j(x^*) - \sum_{j\in I^{\mathrm{i}}(x^*)} \hat{z}^{*,(j)}\nabla d_j(x^*) = 0. \tag{58}$$

Since $\hat{y}^*$ and $\hat{z}^*$ are not both zero, (58) together with Assumption 3(i) implies that $I^{\mathrm{e}}(x^*) \neq \{1,\ldots,m_{\mathrm{e}}\}$ (i.e., $x^* \notin X$) and $\omega > 0$. Dividing both sides of (58) by ω and adding and subtracting $\sum_{j\in I^{\mathrm{e}}(x^*)} \nabla c_j(x^*)$ then yields

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

where we have defined $y^{(j)} = 1 - \frac{\hat{y}^{*,(j)}}{\omega}$ and $z^j = \frac{\hat{z}^{*,(j)}}{\omega}$. In view of (56) and (57) and since $x^* \notin X$, this 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 [12] 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. Recall that $\bar{\rho}$ is of the form $\rho_0\delta^\ell$ for some nonnegative integer ℓ.

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

Thus, under Assumptions 1–5, all accumulation points of $\{x_k\}$ are KKT point for $(P_{\bar{\rho}})$. Now, since ρ_k eventually stops increasing, at least one of the conditions in Step 1ii of Algorithm A is not eventually always satisfied. For convergence to KKT points of (P) to be guaranteed, the fact that Condition (ii) in Step 1ii of Algorithm A must eventually be violated if ρ_k stops increasing is crucial, since this would imply that $c(x_k)$ goes to zero. A glance at the three conditions in that step suggests that this will be the case if the dual variables converge to the KKT multipliers for $(P_{\bar{\rho}})$ (since in such case Conditions (i) and (iii) will eventually hold). To prove that the latter indeed occurs, one more assumption is used.

Assumption 6 *The sequence $\{x_k\}$ generated by Algorithm A has an accumulation point which is an isolated KKT point for $(P_{\bar{\rho}})$ and at which strict complementarity holds.*

Proposition 5 *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 vectors corresponding respectively to the "c" and "d" constraints,*

(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 [12], noting that our Assumption 6 is all that is needed from Assumption A8 of [12] in the proofs of Lemma 4.1 of [12] and Proposition 4.2 of [12]. △

Theorem 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). 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 5 that (i) $\{x_k\} \to x^*$, a KKT point for $(P_{\bar\rho})$; (ii) $\{\Delta x_k\} \to 0$; (iii) $\{y_k + \Delta y_k\} \to y^* \geq 0$, the multiplier vector associated with the "c" constraints, and (iv) $\{z_k + \Delta z_k\} \to z^* \geq 0$, the multiplier vector associated with the "d" constraints. Further, in view of strict complementarity, it follows from Lemma PTH-3.1* in the appendix that the matrix in $L(x^*, y^*, z^*, W^*, \bar\rho, 0, 0)$ is nonsingular given any accumulation point W^* of $\{W_k\}$. Together with (i), (iii) and (iv) above, and since $L(x^*, y^*, z^*, W^*, \bar\rho, 0, 0)$ admits $(0, y^*, z^*)$ as its unique solution, this implies that, on every subsequence on which $\{W_k\}$ converges, $\{\Delta x_k^0\}$ goes to 0, $\{y_k + \Delta y_k^0\}$ goes to y^*, and $\{z_k + \Delta z_k^0\}$ goes to z^*. As a consequence (invoking Assumption 4 and a simple contradiction argument), without need to go down to a subsequence, $\{\Delta x_k^0\} \to 0$, $\{y_k + \Delta y_k^0\} \to y^*$ and $\{z_k + \Delta z_k^0\} \to z^*$. Thus conditions (i) and (iii) in Step 1(ii) of Algorithm A are all satisfied for k large enough. Since $\rho_k = \bar\rho$ for k large enough, it follows from Step 1(ii) of Algorithm A that condition (ii) must fail for k large enough, i.e., $y_k + \Delta y_k^0 \geq \gamma_2 \mathrm{e}$ for k large enough, implying that $y^* \geq \gamma_2 \mathrm{e}$. Since $\gamma_2 > 0$, 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 3 concludes the proof of the first claim. The second claim then follows from Proposition 3 and Proposition 5(i). $\triangle$

Rate of convergence results are inherited from the results in [12]. We report them here for ease of reference. As above, let $\bar y^*$ and z^* be the multipliers associated with KKT point x^* of (P). The Lagrangian associated with (P) is given by

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

With the correspondence $\bar y = y - \bar\rho \mathrm{e}$, it is identical to the Lagrangian associated with $(P_{\bar\rho})$, i.e.,

$$\mathcal{L}_{\bar\rho}(x, y, z) = f(x) + \bar\rho \sum_{j=1}^{m_\mathrm{e}} c_j(x) - \langle y, c(x) \rangle - \langle z, d(x) \rangle.$$

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

$$\frac{\|N_k(W_k - \nabla^2_{xx}\mathcal{L}(x^*, \bar{y}^*, z^*))N_k\Delta x_k\|}{\|\Delta x_k\|} \to 0 \quad \text{as } k \to \infty \tag{59}$$

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 7 *Suppose Assumptions 1–8 hold and suppose that $y^{*,(j)} \le w_{\max}$, $j = 1, \ldots, m_{\mathrm{e}}$, and $z^{*,(j)} \le w_{\max}$, $j = 1, \ldots, m_{\mathrm{i}}$. 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.,*

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

Finally, it is readily verified that, under Assumption 7, for k large enough, Assumption 4 holds when W_k is selected to be *equal* to the Hessian of the Lagrangian $\mathcal{L}_{\bar{\rho}}$. In view of the discussion in Section 2.3, Q-quadratic convergence follows.

Theorem 8 *Suppose Assumptions 1–7 hold, suppose that, at every iteration except possibly finitely many, W_k is selected as*

$$W_k = \nabla^2_{xx}\mathcal{L}_{\rho_k}(x_k, y_k, z_k),$$

and suppose that $y^{*,(j)} \le w_{\max}$, $j = 1, \ldots, m_{\mathrm{e}}$, and $z^{*,(j)} \le w_{\max}$, $j = 1, \ldots, m_{\mathrm{i}}$. Then (x_k, y_k, z_k) converges to (x^*, y^*, z^*), equivalently $(x_k, y_k - \rho_k\mathrm{e}, z_k)$ converges to $(x^*, \bar{y}^*, z^*)$, Q-quadratically; i.e., there exists a constant $\Gamma > 0$ such that

$$\left\|\begin{bmatrix} x_{k+1} - x^* \\ y_{k+1} - y^* \\ z_{k+1} - z^* \end{bmatrix}\right\| \le \Gamma \left\|\begin{bmatrix} x_k - x^* \\ y_k - y^* \\ z_k - z^* \end{bmatrix}\right\|^2 \quad \text{for all } k, \tag{60}$$

equivalently,

$$\left\| \begin{bmatrix} x_{k+1} - x^* \\ y_{k+1} - \rho_{k+1}\mathrm{e} - \bar{y}^* \\ z_{k+1} - z^* \end{bmatrix} \right\| \leq \Gamma \left\| \begin{bmatrix} x_k - x^* \\ y_k - \rho_k \mathrm{e} - \bar{y}^* \\ z_k - z^* \end{bmatrix} \right\|^2 \quad \text{for all } k.$$

5 Numerical Examples

We tested a MATLAB 6.1 implementation of Algorithm A with the following differences in comparison with the algorithm statement of Section 3:

- The suggestion made in Remark 2 was adopted.
- In the update formulae for the multipliers in Step 3, $\|\Delta x_k\|^2$ was changed to $\min\{w_{\min}, \|\Delta x_k\|^2\}$ in both places. The motivation is that $\|\Delta x_k\|$ is meaningful only when it is small. This change does not affect the convergence analysis.

The following parameter values were used: $\xi = 10^{-4}$, $\eta = 0.8$, $\gamma_1 = 1$, $\gamma_2 = 1$, $\gamma_3 = 1$, $\nu = 3$, $\theta = 0.8$, $w_{\min} = 10^{-4}$ (see Remark 2), $w_{\max} = 10^{20}$, $\delta = 2$, $\tau = 2.5$, and $\kappa = 0.5$.

In our tests, we allowed for the initial point x_0 to lie on the boundary of the feasible set $\tilde{X}$. It is readily checked that in such case, under our assumptions, $L(x_0, y_0, z_0, W_0, \rho_0, 0, 0)$ is still uniquely solvable and, unless $\Delta x_0^0 = 0$, the initial iteration is still well-defined and yields a strictly feasible second iterate. (When $\Delta x_0^0 = 0$ and $c(x_0) = 0$, x_0 is stationary for (P). When $\Delta x_0^0 = 0$ but $c(x_0) \neq 0$, x_0 is stationary for (P_{ρ_0}) but not for (P), and unless the sum of the gradients of inactive $c_j(x_0)$'s belongs to the span of the gradients of all active constraints, increasing ρ forces a nonzero Δx_0^0 and the iteration can proceed.) When x_0 is not in the interior of $\tilde{X}$, the condition to be satisfied by W_0 must be modified by replacing in (45) the infinite (diagonal) entries of $C(x_0)^{-1}$ and $D(x_0)^{-1}$ by 0, and by requiring positive definiteness of the modified expression merely on the tangent plane to the active constraints, i.e., on

$$\left\{ v \in \mathcal{R}^n : \langle \nabla c_i(x_0), v \rangle = 0,\ \langle \nabla d_j(x_0), v \rangle = 0,\ \forall i \in I^{\mathrm{e}}(x_0), j \in I^{\mathrm{i}}(x_0) \right\}.$$

In the numerical tests reported below, the initial value x_0 was selected in each case as specified in the source of the test problem. Initial values y_0, z_0

and ρ_0 were selected as follows. Let y'_0 and z'_0 be the (linear least squares) solution of

$$\min_{y'_0, z'_0} \|g(x_0) - A(x_0)y'_0 - B(x_0)z'_0\|^2.$$

Then ρ_0 was set to the smallest power of δ that is no less than $\max\left\{1, \max_j\{\gamma_2 - y_0^{(j)}\}\right\}$, and, for all j, $y_0^{(j)}$ was set to $y_0'^{(j)} + \rho_0$ and $z_0^{(j)}$ to $\max\{0.1, z_0'^{(j)}\}$. In all the tests, y_0 and z_0 thus defined satisfied the condition specified in the algorithm that their components should all be no larger than $w_{\max}$.

Next, for $k = 0, 1, \ldots$, W_k was constructed as follows, from second order derivative information. Let $\lambda_{\min}$ be the leftmost eigenvalue of the restriction of the matrix

$$\nabla^2_{xx}\mathcal{L}_{\rho_k}(x_k, y_k, z_k) + \sum_{i \in I_k^{e'}} \frac{y_k^{(i)}}{c_i(x_k)} \nabla c_i(x_k) \nabla c_i(x_k)^{\mathrm{T}} + \sum_{i \in I_k^{i'}} \frac{z_k^{(i)}}{d_i(x_k)} \nabla d_i(x_k) \nabla d_i(x_k)^{\mathrm{T}},$$

where $I_k^{e'}$ and $I_k^{i'}$ are the sets of indices of "c" and "d" constraints with value *larger* that 10^{-10}, to the tangent plane to the constraints left out of the sum, i.e., to the subspace

$$\left\{v \in \mathcal{R}^n : \langle \nabla c_i(x_k), v\rangle = 0,\ \langle \nabla d_j(x_k), v\rangle = 0,\ \forall i \notin I_k^{e'}, j \notin I_k^{i'}\right\}.$$

Then, set

$$W_k = \nabla^2_{xx}\mathcal{L}_{\rho_k}(x_k, y_k, z_k) + h_k I$$

where

$$h_k = \begin{cases} 0 & \text{if } \lambda_{\min} > 10^{-5} \\ -\lambda_{\min} + 10^{-5} & \text{if } |\lambda_{\min}| \le 10^{-5} \\ 2|\lambda_{\min}| & \text{otherwise.} \end{cases}$$

Note that, under our regularity assumptions (which imply that W_k is bounded whenever x_k is bounded), this insures that Assumption 4 holds. The motivation for the third alternative is to preserve the order of magnitude of the eigenvalues and condition number.

The stopping criterion (inserted at the end of Step 1i) was as follows, with $\epsilon_{\text{stop}} = 10^{-8}$. First, accounting for the fact that, in our tests, we allowed the initial point to lie on the boundary of $\tilde{X}$, we stopped with the error message "initial point stationary for (P)" if $\|\Delta x_0^0\| < 0.001\epsilon_{\text{stop}}$ and $\|c(x_0)\|_\infty <$

ϵ_{stop}. Second, the run was deemed to have terminated successfully if at any iteration k

$$\max\left\{\|c(x_k)\|_\infty, \max_j\left\{-\left(y_k^{(j)} + \Delta y_k^{0(j)}\right)\right\}, \max_j\left\{-\left(z_k^{(j)} + \Delta z_k^{0(j)}\right)\right\}\right\} < \epsilon_{\text{stop}}$$

and either

$$\|\Delta x_k^0\|_\infty < \epsilon_{\text{stop}}$$

or

$$\max\left\{\|\nabla_x L(x_k, y_k, z_k)\|_\infty, \max_j\left\{z_k^{(j)} d_j(x_k)\right\}\right\} < \epsilon_{\text{stop}}.$$

Iterations at which only Steps 1i an 1ii are executed were not included in our iteration counts. The reason is that the computational cost of these iterations is dramatically less than that of "regular" iterations: no additional function evaluations and no additional matrix factorization—the same factorization is later used at the next regular iteration. All tests were run within the CUTEr testing environment [26], on a SUN Enterprise 250 with two UltraSparc-II 400MHz processors, running Solaris 2.7.

We first considered two instances of the example from [10] briefly discussed in Section 3 (immediately following Lemma 2), specifically $(a, b) = (1, 1)$ with $(-3, 1, 1)^{\mathrm{T}}$ as initial guess, and $(a, b) = (-1, 1/2)$ with $(-2, 1, 1)^{\mathrm{T}}$ as initial guess, both of which satisfy the conditions in Theorem 1 of [10]. In both cases we used $f(x) = x^{(1)}$ as objective function (as in the example of Section 4 of [10]). Recall that, under those conditions, all methods of type "Algorithm I" in [10] construct sequences that converge to points of the form $(\zeta, 0, 0)^{\mathrm{T}}$, with $\zeta < 0$, where both c_1 and c_2 are nonzero. As noted in our earlier discussion, Assumption 3(ii) is satisfied in the first instance while, in the second instance, the condition in that assumption is violated only at $\hat{x} := (-1, 0, 0)^{\mathrm{T}}$ (with $c_1(\hat{x}) = 0$). Thus, at $\hat{x}$, there is no direction of strict descent for $c_1(x) + c_2(x)$ (the ℓ_1 norm of $c(x)$ when $x \in \tilde{X}$) that is feasible for $c_1(x) \geq 0$, $d_1(x) \geq 0$, and $d_2(x) \geq 0$.

In the first instance, our Algorithm A was observed to converge to the unique global solution $(1, 2, 0)^{\mathrm{T}}$ in 13 iterations, with a final penalty parameter value $\bar{\rho}$ of 4. In the second instance, Algorithm A failed in that the constructed sequence converged to the infeasible point $\hat{x}$. Interestingly, it can be checked that, at $\hat{x}$, there is not even a descent direction for $\|c(x)\|_1$ that is feasible for the mere bound constraints $d_1(x) \geq 0$ and $d_2(x) \geq 0$.

Remark 4 *For the second instance of the example of [10] just discussed, directions do exist at $\hat{x}$ that are of strict descent for the* Euclidean *norm of $c(x)$ and are feasible for the bound constraints. The existence of such directions allows the algorithm proposed in [6] to proceed from such point. (Also see the related discussion in the penultimate paragraph of [10].)*

We then ran the MATLAB code on all but three of those problems from [27] for which the initial point provided in [27] satisfies all inequality constraints. (While a phase I-type scheme could be used on the other problems—see Remark 3—testing such scheme is outside the main scope of this paper.) Problems 68, 69 and 87 were left out: the first two because of numerical difficulties in connection with the use of Chebyshev approximations in function evaluations, and the last one because the objective function in that problem is nonsmooth. In problems 31, 35, 44, 55, 71, and 86, the given x_0 is stationary for problem (P) and in problem 74, the given x_0 is stationary for (P_ρ) for every ρ. Results on the remaining 63 problems are reported in Table 1. The first column in the table gives the problem number from [27], the second column the total number of iterations, the third column the final value $\bar{\rho}$ of the penalty parameter, and the last column the final value of the objective function.

On three of the problems (66, 107, and 111) our stopping criterion was not met after one thousand iterations. However in all three cases the final objective value was equal, with three or more figures of accuracy, to the optimal value given in [27]. (Indeed, four figures of accuracy were obtained on problems 66 and 111, after 120 and 887 iterations, respectively; and three figures of accuracy were reached on problem 107 after 95 iterations.) A large number of iterations was needed on problem 85, on which the algorithm failed in the last iteration to produce an acceptable step size due to numerical difficulties. When the algorithm stopped, near $x = (704.41, 68.60, 102.90, 282.03, 37.46)^{\mathrm{T}}$, the value of $\|\Delta x_k^0\|$ was less than $2 \cdot 10^{-8}$, and the objective value obtained was lower than the (locally) optimal value given in [27]. Overall, comparison with published results obtained on the same problems with other interior-point methods suggests that Algorithm A has promise. In particular, on 39 of the 63 problems listed in Table 1, our results in terms of number of iterations are better than those reported in [8] (on three other problems they are identical, and problem 67 is not listed in [8]). More extensive testing on larger size problems is in order for a more definite assessment of the value of the proposed approach. Such testing will require a more elaborate implementation

Prob.	#Itr	$\bar{\rho}$	f_{final}	Prob.	#Itr	$\bar{\rho}$	f_{final}
HS1	24	1	6.5782e-27	HS52	4	8	5.3266e+00
HS3	4	1	8.5023e-09	HS53	5	8	4.0930e+00
HS4	4	1	2.6667e+00	HS54	23	4	-1.6292e-54
HS5	6	1	-1.9132e+00	HS56	12	4	-3.4560e+00
HS6	7	2	0.0000e+00	HS57	15	1	2.8460e-02
HS7	9	2	-1.7321e+00	HS60	7	1	3.2568e-02
HS8	14	1	-1.0000e+00	HS61	44	128	-1.4365e+02
HS9	10	1	-5.0000e-01	HS62	5	1	-2.6273e+04
HS12	5	1	-3.0000e+01	HS63	5	2	9.6172e+02
HS24	14	1	-1.0000e+00	HS66	$1000^{\dagger}$	1	5.1817e-01
HS25	62	1	1.8185e-16	HS67	14	1	-1.1621e+03
HS26	19	2	2.8430e-12	HS70	22	1	1.7981e-01
HS27	14	32	4.0000e-02	HS73	16	1	2.9894e+01
HS28	6	1	0.0000e+00	HS75	28	16	5.1744e+03
HS29	8	1	-2.2627e+01	HS77	13	1	2.4151e-01
HS30	7	1	1.0000e+00	HS78	4	4	-2.9197e+00
HS32	24	4	1.0000e+00	HS79	7	2	7.8777e-02
HS33	29	1	-4.5858e+00	HS80	6	2	5.3950e-02
HS34	30	1	-8.3403e-01	HS81	9	8	5.3950e-02
HS36	10	1	-3.3000e+03	HS84	30	1	-5.2803e+06
HS37	7	1	-3.4560e+03	HS85	$296^{\ddagger}$	1	-2.2156e+00
HS38	37	1	3.1594e-24	HS93	12	1	1.3508e+02
HS39	19	4	-1.0000e+00	HS99	8	2	-8.3108e+08
HS40	4	2	-2.5000e-01	HS100	9	1	6.8063e+02
HS42	6	4	1.3858e+01	HS107	$1000^{\dagger}$	8192	5.0545e+03
HS43	9	1	-4.4000e+01	HS110	6	1	-4.5778e+01
HS46	25	2	6.6616e-12	HS111	$1000^{\dagger}$	1	-4.7760e+01
HS47	25	16	8.0322e-14	HS112	11	1	-4.7761e+01
HS48	6	4	0.0000e+00	HS113	10	1	2.4306e+01
HS49	69	64	3.5161e-12	HS114	39	256	-1.7688e+03
HS50	11	512	4.0725e-17	HS117	25	1	3.2349e+01
HS51	8	4	0.0000e+00				

Table 1: Results on Test Problems from [27]

of the algorithm.

6 Concluding Remarks

An interior-point algorithm for the solution of general nonconvex constrained optimization problems has been proposed and analyzed. The algorithm involves a novel, simple exact penalty parameter updating rule. Global convergence as well as local superlinear and quadratic convergence have been proved under mild assumptions. In particular, it was pointed out that the proposed algorithm does not suffer a common pitfall recently pointed out in [10]. Promising preliminary numerical results were reported.

While the present paper focussed on applying a version of the Mayne-Polak scheme to the algorithm of [12], there should be no major difficulty in similarly extending other feasible interior-point algorithms for inequality constrained problems to handle general constrained problems.

7 Appendix

We discuss the implications of substituting Assumption PTH-A6*, as stated in Section 2, for Assumption A6 of [12]. For the reader's ease of reference, throughout this appendix, we use the notation of [12]; Assumption PTH-A6* then reads as follows.

Assumption PTH-A6*. *Given any index set K such that $\{x_k\}_{k\in K}$ is bounded, there exist $\sigma_1, \sigma_2 > 0$ such that, for all $k \in K$,*

$$\|H_k\| \leq \sigma_2$$

and

$$\langle d, \left(H_k - \sum_{i=1}^{m} \frac{\mu_{k,i}}{g_i(x_k)} \nabla g_i(x_k) \nabla g_i(x_k)^T \right) d \rangle \geq \sigma_1 \|d\|^2 \ \forall d \in R^n.$$

First of all, under this weaker assumption, a stronger version of Lemma 3.1 of [12] is needed.

Lemma PTH-3.1*. *Let $x \in X$, let $\mu \in R^m$ be such that $\mu_i \geq 0$ for all i and $\mu_i > 0$ for all $i \in I(x)$, and let $H \in R^{n \times n}$, symmetric, satisfy the condition*

$$\langle d, \left(H - \sum_{i \notin I(x)} \frac{\mu_i}{g_i(x)} \nabla g_i(x) \nabla g_i(x)^T \right) d \rangle > 0 \quad \forall d \in \mathcal{T}(x) \setminus \{0\} \tag{61}$$

where

$$\mathcal{T}(x) = \{d \in \mathcal{R}^n : \langle \nabla g_i(x), d \rangle = 0 \ \forall i \in I(x)\}.$$

Then the matrix $F(x, H, \mu)$ as defined by

$$F(x, H, \mu) = \begin{bmatrix} H & \nabla g_1(x) & \dots & \nabla g_m(x) \\ \mu_1 \nabla g_1(x)^{\mathrm{T}} & g_1(x) & & \bigcirc \\ \vdots & & \ddots & \\ \mu_m \nabla g_m(x)^{\mathrm{T}} & \bigcirc & & g_m(x) \end{bmatrix}$$

is nonsingular.

Proof: It is enough to show that the only solution (d, λ) of the homogeneous system

$$Hd + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = 0 \tag{62}$$

$$\mu_i \langle \nabla g_i(x), d \rangle + \lambda_i g_i(x) = 0, \quad i = 1, \dots, m \tag{63}$$

is (0,0). Scalar multiplication of both sides of (62) by d yields

$$\langle d, Hd \rangle + \sum_{i=1}^{m} \lambda_i \langle \nabla g_i(x), d \rangle = 0. \tag{64}$$

On the other hand, it follows from (63) and the assumption on μ that

$$\langle \nabla g_i(x), d \rangle = 0 \quad \forall\ i \in I(x). \tag{65}$$

Now, from (64) and (65), we get

$$\langle d, Hd \rangle + \sum_{i \notin I(x)} \lambda_i \langle \nabla g_i(x), d \rangle = 0. \tag{66}$$

Solving (63) for λ_i (for $i \notin I(x)$) and substituting in (66) yields

$$\langle d, Hd \rangle - \sum_{i \notin I(x)} \langle \nabla g_i(x), d \rangle \frac{\mu_i}{g_i(x)} \langle \nabla g_i(x), d \rangle = 0.$$

In view of (65), it follows from (61) that $d = 0$. It then follows from (63) that $\lambda_i = 0$ for all $i \notin I(x)$. Assumption A5 together with (62) then implies that $(d, \lambda) = (0, 0)$. $\triangle$

Next, the first inequality in Eq. (3.6) of [12] is unaffected. While the second inequality in that equation still holds as well, it is not of much value now that H_k is no longer assumed to be positive definite. However we note that, denoting by S_k the Schur complement of $G_k := \mathrm{diag}(g_i(x_k))$ in $F(x_k, H_k, \mu_k)$ (see page 794 of [12]), i.e.,

$$S_k := H_k - A_k G_k^{-1} M_k A_k^{\mathrm{T}},$$

with A_k and M_k defined on p. 808 in [12], we get

$$d_k^0 = -S_k^{-1} \nabla f(x_k)$$

yielding

$$\langle \nabla f(x_k), d_k^0 \rangle = -\langle d_k^0, S_k d_k^0 \rangle \leq -\sigma_1 \|d_k^0\|^2, \tag{67}$$

where we have invoked Assumption PTH-A6*. Where Eqn. (3.6) (of [12]) is used in the analysis of [12], Eqn. (67) must sometimes be used instead.

Propositions 3.3 and 3.4 of [12] then readily follow. The only remaining notable issue is that a stronger version of Lemma 3.5 [12] is needed, as follows.

Lemma PTH-3.5*. *Let K be an infinite index set such that, for some x^* and μ^**

$$\lim_{\substack{k\to\infty \\ k\in K}} x_k = x^* \text{ and } \lim_{\substack{k\to\infty \\ k\in K}} \mu_k = \mu^*.$$

Suppose moreover that $\mu_i^* > 0$ if $g_i(x^*) = 0$. Then, given any accumulation point H^* of $\{H_k\}_{k\in K}$, $F(x^*, H^*, \mu^*)$ is nonsingular. Moreover there exists C such that for all $k \in K$,

$$\|d_k - d_k^0\| \leq C \|d_k^0\|^\nu.$$

Proof: Let $K' \subseteq K$ be an infinite index set such that $H_k \to H^*$ as $k \to \infty$, $k \in K'$. We first show that (x^*, H^*, μ^*) satisfies the assumptions of Lemma PTH-3.1*. Thus let $v \neq 0$ be such that

$$\langle \nabla g_i(x^*), v \rangle = 0 \quad \forall i \in I(x^*).$$

In view of our linear independence assumption, there exists a sequence $\{v_k\}_{k\in K'}$ converging to v on K', such that for all $k \in K'$

$$\langle \nabla g_i(x_k), v_k\rangle = 0 \quad \forall i \in I(x^*).$$

It then follows from Assumption PTH-6* (by adding zero terms) that for all $k \in K'$

$$\langle v_k, \left(H_k - \sum_{i \notin I(x^*)} \frac{\mu_{k,i}}{g_i(x_k)} \nabla g_i(x_k) \nabla g_i(x_k)^T \right) v_k\rangle \geq \sigma_1 \|v_k\|^2.$$

Letting $k \to \infty$, $k \in K'$ shows that

$$\langle v, \left(H^* - \sum_{i \notin I(x^*)} \frac{\mu_i^*}{g_i(x^*)} \nabla g_i(x^*) \nabla g_i(x^*)^T \right) v\rangle \geq \sigma_1 \|v\|^2 > 0.$$

Thus the assumptions of Lemma PTH-3.1* are satisfied. It follows that $F(x^*, H^*, \mu^*)$ is nonsingular. Since $F(x_k, H_k, \mu_k)$ is nonsingular for all k, boundedness of $\{H_k\}$ and our continuity assumptions imply that $F(x_k, H_k, \mu_k)^{-1}$ is uniformly bounded on K. The remainder of the proof is as in [12]. $\triangle$

With these strengthened results, the remainder of the analysis in [12] is essentially unaffected by the weakening of the assumption on H_k. Specifically, Lemma 3.6 of [12] (where the "old" Assumption A6 is invoked) still follows, using the stronger Lemmas PTH-3.1* and PTH-3.5*. While the "boundedness of H_k" portion of Assumption A6 is used at many other places in the analysis of [12], the "positive definiteness" portion of that assumption (which is the only portion that is relaxed in Assumption PTH-A6*) is not used anywhere else. The strengthened Lemmas PTH-3.1* and PTH-3.5* are needed in the proof of Lemma 4.1 of [12]: Lemma 3.1 of [12] is implicitly used in the last sentence of that proof, to conclude that the limit system (4.1)–(4.2) of [12] is invertible.

Finally, Lemma 4.4 still holds under the milder Assumption PTH-A6* (and so do Proposition 4.5 and Theorem 4.6), but again the strengthened Lemmas PTH-3.1* and PTH-3.5* are needed in its proof. In particular, for k large enough, the second order sufficiency condition still holds at the solution of (LS3) and thus solving (LS3) is still equivalent to solving the stated linear system of equations (in the proof of Lemma 4.4). (The notation $\|d\|^2_{H_k}$ used in (LS3) is now inappropriate though, and should be replaced with $\langle d, H_k d\rangle$.)

Furthermore, it follows from Lemma PTH-3.5* and the fact that $\mu_{k,i}$ tends to zero for $i \notin I(x^*)$ that, for k large enough, this linear system still has a unique solution, i.e., (LS3) still has a well defined (unique) minimizer.

8 Acknowledgment

The authors wish to thank Jean-Charles Gilbert and Paul Armand for discussions in connection with an early version of Assumption 3(ii); as a result the current Assumption 3(ii) is significantly less restrictive. They also wish to thank Bill Woessner for his help with the numerical experiments, as well as an anonymous referee and the associate editor for their feedback that helped improve the paper. This work was supported in part by the National Science Foundation under Grant DMI-9813057.

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, J.C. Gilbert, and J. Nocedal. A trust region method based on interior point techniques for nonlinear programming. *Mathematical Programming*, 89:149–185, 2000.

[7] 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.

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

[9] N. Maratos. *Exact Penalty Function Algorithms for Finite Dimensional and Optimization Problems.* PhD thesis, Imperial College of Science and Technology, University of London, 1978.

[10] 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.

[11] A. Wächter and L.T. Biegler. Global and local convergence of line search filter methods for nonlinear programming. Technical Report CAPD B-01-09, Carnegie Mellon University, 2001.

[12] 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.

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

[14] D. Q. Mayne and E. Polak. A superlinearly convergent algorithm for constrained optimization problems. *Math. Programming Stud.*, 16:45–61, 1982.

[15] 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.

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

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

[18] 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.

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

[20] 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.

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

[22] M. Sahba. Globally convergent algorithm for nonlinearly constrained optimization problems. *J. of Optimization Theory and Applications*, 52(2):291–309, 1987.

[23] J.V. Burke. A robust trust region method for constrained nonlinear programming problems. *SIAM J. Optimization*, 2:325–347, 1992.

[24] Y. Yuan. On the convergence of a new trust region algorithm. *Numerische Mathematik*, 70:515–539, 1995.

[25] H. Yabe abd H. Yamashita. Q-superlinear convergence of primal-dual interior point quasi-newton methods for constrained optimization. *J. Oper. Res. Soc. Japan*, 40(3):415–436, 1997.

[26] N.I.M. Gould, D. Orban, and Ph.L. Toint. CUTEr (and SifDec), a constrained and unconstrained testing environment, revisited. Technical Report TR/PA/01/04, CERFACS, Toulouse, France, 2001.

[27] W. Hock and K. Schittkowski. *Test Examples for Nonlinear Programming Codes*, volume 187 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, 1981.