Computational Optimization and Applications, 25, 17–38, 2003

A Simple Primal-Dual Feasible Interior-Point Method for Nonlinear Programming with Monotone Descent*

SASAN BAKHTIARI AND ANDRÉ L. TITS
Department of Electrical and Computer Engineering and Institute for Systems Research, University of Maryland, College Park, MD 20742, USA

Received February 3, 2002; Revised August 10, 2002; Accepted December 13, 2002

Dedicated to: (Tits) My research collaboration with E. Polak started in June 1978, at the end of my first academic year as a graduate student at Berkeley. For the next three years, till I left Berkeley in August 1981, I spent an enormous number of hours working with Lucien, mostly in his office but also, many a Sunday afternoon, at his house, up in the hills. While he did advise one or two other students at the time, and of course had many other responsibilities besides mentoring his research advisees, Lucien was available to meet with me whenever I needed his advice, or so it seemed. In such privileged setting I learned enormously from him. Countless times, when I was at a loss about how to modify a tentative algorithm to obtain certain desirable properties, he gave me enlightening geometric intuition, which often would render the issue all but trivial. Countless times, when I was unsuccessful in proving certain convergence properties, he steered me to a new, promising direction that I had not thought about. Over time, by osmosis, I slowly acquired some of these skills for myself. Above all, I inherited from Lucien a taste for uncompromising rigor and precision.

To me, Lucien Polak was the best research advisor I could have dreamed of. From the bottom of my heart, thank you Lucien!

Abstract. We propose and analyze a primal-dual interior point method of the "feasible" type, with the additional property that the objective function decreases at each iteration. A distinctive feature of the method is the use of different barrier parameter values for each constraint, with the purpose of better steering the constructed sequence away from non-KKT stationary points. Assets of the proposed scheme include relative simplicity of the algorithm and of the convergence analysis, strong global and local convergence properties, and good performance in preliminary tests. In addition, the initial point is allowed to lie on the boundary of the feasible set.

Keywords: constrained optimization, nonlinear programming, primal-dual interior-point methods, feasibility, monotone descent

1. Introduction

Primal-dual interior-point methods have enjoyed increasing popularity in recent years. In particular, many authors have investigated the extension of such methods—originally proposed in the context of linear programming, then in that of convex programming—to the solution of nonconvex, smooth constrained optimization problems. In this paper the main

*This work was supported in part by the National Science Foundation under Grant DMI9813057.

focus is on inequality constrained problems of the type

$$\begin{aligned} \min_{x \in \mathcal{R}^n} \quad & f(x) \\ \text{s.t.} \quad & d_j(x) \geq 0, \quad j = 1, \ldots, m \end{aligned} \tag{P}$$

where $f : \mathcal{R}^n \rightarrow \mathcal{R}$ and $d_j : \mathcal{R}^n \rightarrow \mathcal{R}$, $j = 1, \ldots, m$ are smooth, and where no convexity assumptions are made. For such problems, some of the proposed methods (e.g., [4, 5, 9]) are "feasible", in that, given an initial point that satisfies the constraints, they construct a sequence of points (hopefully converging to a solution) which all satisfy them as well. Others [1–3, 12–14] make the inequality constraints into equality constraints, by the introduction of slack variables; while the slack variables are forced to remain positive throughout, the original constraints may be occasionally violated. Such methods are often referred to as "infeasible".

In this paper, we proceed along the lines of the algorithm of [9],[1] a feasible primal-dual interior-point algorithm enjoying the additional property that the objective function value decreases at each iteration—a property we refer to here as "monotone descent". In [9] it is shown that, under certain assumptions, the proposed algorithm possesses attractive global as well as local convergence properties. An extension of this algorithm to problems with equality constraints was recently proposed [11]. That paper includes preliminary numerical results that suggest that the algorithm holds promise.

Our goal in the present paper is to improve on the algorithm of [9]. We contend that the new algorithm is simpler than that of [9], has better theoretical convergence properties, and performs better in practice. Specifically, a weakness of the algorithm of [9] is that, while it is proven to eventually converge to KKT points for (P), it may get bogged down over a significant number of iterations in the neighborhood of non-KKT stationary points, i.e., stationary points at which not all multipliers are nonnegative (constrained local maxima or constrained saddle points). This is because the norm of the search direction becomes arbitrarily small in the neighborhood of such points, a by-product of the fact that, in the algorithm of [9], the barrier parameter(s) must be small in the neighborhood of stationary points in order for the monotone descent property to hold.

The key idea in the algorithm proposed in this paper is the use of a vector barrier parameter, i.e., a different barrier parameter for each constraint. While this was already the case in [9], this feature does not play an essential role there, as all convergence results would still hold if an appropriate scalar barrier parameter were used instead. It turns out, however, that, close to stationary points of (P), careful assignment of distinct values to the components of the barrier parameter vector allows to guarantee the monotone descent property without forcing these components to be small, i.e., while allowing a reasonable norm for the search direction. This fact was exploited in [10] where a similar idea is used, albeit in a different context. Our implementation of the idea is such that, after a finite number of iterations, the resulting algorithm becomes identical to that of [9]. Overall though, the new algorithm is simpler than that of [9], the convergence results hold under milder assumptions, and preliminary numerical tests show an improvement in practical performance.

As a second improvement over the algorithm of [9], a significantly milder assumption is used on the sequence of approximate Hessians of the Lagrangian, as was done in [11]. In particular, like in [11], this allows for the use of exact second derivative information which, under second order sufficiency assumptions, induces quadratic convergence. Finally, the proposed algorithm does not require an interior initial point: the initial point may lie on the boundary of the feasible set.

The remainder of the paper is organized as follows. The new algorithm is motivated and stated in Section 2. Its global and local convergence properties are analyzed in Section 3. Numerical results on a few test problems are reported in Section 4. Finally, Section 5 is devoted to concluding remarks. Our notation is mostly standard. In particular, we denote by $\mathcal{R}^m_+$ the nonnegative orthant in $\mathcal{R}^m$, $\|\cdot\|$ denotes the Euclidean norm, and ordering relations such as $\geq$, when relating two vectors, are meant componentwise.

2. The algorithm

Let X and X_0 be the feasible and strictly feasible sets for (P), respectively, i.e., let

$$X = \{x \; : \; d_j(x) \geq 0, \forall j\}$$

and

$$X_0 = \{x \; : \; d_j(x) > 0, \forall j\}.$$

Also, for $x \in X$, let $I(x)$ denote the index set of active constraints at x, i.e.,

$$I(x) = \{j : d_j(x) = 0\}.$$

The following assumptions will be in force throughout.

Assumption 1. X is nonempty.

Assumption 2. f and d_j, $j = 1, \ldots, m$ are continuously differentiable.

Assumption 3. For all $x \in X$ the set $\{\nabla d_j(x) : j \in I(x)\}$ is linearly independent.

Note that these assumptions imply that X_0 is nonempty, and that X is its closure.

Let $g(x)$ denote the gradient of $f(x)$, $B(x)$ the Jacobian of $d(x)$, and let $D(x) = \mathrm{diag}(d_j(x))$. A point x is a *KKT point* for (P) if there exists $z \in \mathcal{R}^m$ such that

$$g(x) - B(x)^T z = 0 \tag{1}$$

$$d(x) \geq 0 \tag{2}$$

$$D(x) z = 0, \tag{3}$$

$$z \geq 0. \tag{4}$$

Following [9, 11] we term a point x *stationary* for (P) if there exists $z \in \mathcal{R}^m$ such that (1)–(3) hold (but possibly not (4)).

The starting point of the primal-dual feasible interior-point iteration is the system of equations in (x, z)

$$g(x) - B(x)^T z = 0, \tag{5}$$

$$D(x)z = \mu. \tag{6}$$

The vector μ is the "barrier parameter" vector, typically taken to be a scalar multiple of the vector $e = [1, \dots, 1]^T$; in [9] though, the components of μ are distinct, and this will be the case here as well. System (5)–(6) can be viewed as a perturbed (by μ) version of the equations in the KKT conditions of optimality for (P). The idea is then to attempt to solve this nonlinear system of equations by means of a Newton or quasi-Newton iteration, while driving μ to zero and enforcing primal and dual (strict) feasibility at each iteration. Specifically, the following linear system in $(\Delta x^\mu, \Delta z^\mu)$ is considered:

$$-W\Delta x^\mu + \sum_{j=1}^{m} \Delta z^{\mu j} \nabla d_j(x) = g(x) - \sum_{j=1}^{m} z^j \nabla d_j(x) \tag{7}$$

$$z^j \langle \nabla d_j(x), \Delta x^\mu \rangle + d_j(x)\Delta z^{\mu j} = \mu^j - d_j(x) z^j, \quad j = 1, \dots, m, \tag{8}$$

or, equivalently,

$$M(x, z, W)\begin{pmatrix} \Delta x^\mu \\ \Delta z^\mu \end{pmatrix} = \begin{pmatrix} g(x) - B(x)^T z \\ \mu - D(x)z \end{pmatrix}. \tag{$L(x, z, W, \mu)$}$$

Here W equals, or appropriately approximates, the Hessian of the Lagrangian associated to (P), i.e.,

$$W \approx \nabla^2_{xx}\mathcal{L}(x, z) = \nabla^2 f(x) - \sum_{j=1}^{m} z^j \nabla^2 d_j(x),$$

and $M(x, z, W)$ is given by

$$M(x, z, W) = \begin{pmatrix} -W & B(x)^T \\ ZB(x) & D(x) \end{pmatrix},$$

where $Z = \text{diag}(z^j)$.

In [9],[2] given $x \in X_0$, $z > 0$, and W symmetric and positive definite, which insures that $M(x, z, W)$ is nonsingular, $L(x, z, W, 0)$ is solved first, yielding $(\Delta x^0, \Delta z^0)$. When $x \in X$ is stationary, it is readily checked that Δx^0 is zero, and vice-versa. Consequently, $\|\Delta x^0\|$ is small whenever x is close to a stationary point. It follows that the norm of Δx^0 is a measure of the proximity to a stationary point of (P). On the basis of this observation, in [9, 11], $L(x, z, W, \mu)$ is then solved with $\mu = \|\Delta x^0\|^\nu z$, where $\nu > 0$ is prescribed, yielding $(\Delta x^1, \Delta z^1)$. (For fast local convergence, ν is selected to be larger than 2.) Since Δx^1 is not necessarily a direction of descent for f, the search direction Δx is selected on the line segment from Δx^0 to Δx^1, as close as possible to Δx^1 subject to "sufficient" descent for f.

Clearly, $\|\Delta x\|$ is then small whenever x is close to a stationary point. Since the aim is to construct a sequence that converges to KKT points (stationary points with nonnegative multipliers) of (P), this state of affairs is unfortunate when x is not a KKT point, as it will slow down the constructed sequence. (Still, it is shown in [9] that, provided stationary points are isolated, convergence will occur to KKT points only.) The main contribution of the algorithm proposed in this paper is to address this issue.

Before pursuing this discussion further, we establish some preliminary results. The first lemma below, a restatement of Lemma PTH-3.1* from the appendix of [9], provides specific conditions under which $M(x, z, W)$ is invertible. The conditions on W are as in [11], and are much less restrictive than those used in [9]. As a result, already stressed in [11], a much wider choice of Hessian estimates can be used than that portrayed in [9]. This includes the possible use of exact second derivative information.

Given $x \in X$, let $T(x) \subset \mathcal{R}^n$ be the tangent plane to the active constraint boundaries at x, i.e.,

$$T(x) = \{u : \langle \nabla d_j(x), u \rangle = 0, \forall j \in I(x)\}.$$

We denote by $\mathcal{S}$ the set of triples (x, z, W) such that $x \in X$, $z \in \mathcal{R}^m_+$, $z^j > 0$ for all $j \in I(x)$, $W \in \mathcal{R}^{n\times n}$, symmetric, and the condition

$$\left\langle u, \left(W + \sum_{j \notin I(x)} \frac{z^j}{d_j(x)} \nabla d_j(x) \nabla d_j(x)^T \right) u \right\rangle > 0 \quad \forall u \in T(x) \backslash 0 \tag{9}$$

holds.

Lemma 1. *Suppose Assumptions* 1–3 *hold. Let* $(x, z, W) \in \mathcal{S}$. *Then* $M(x, z, W)$ *is nonsingular.*

Since we seek descent for f, in order for a Hessian estimate W to be acceptable, it must yield a descent direction. The next result, also proved in the appendix of [9] (see (67) in [9]), shows that Δx is indeed such a direction when $(x, z, W) \in \mathcal{S}$ and x is not stationary.

Lemma 2. *Suppose Assumptions* 1–3 *hold. Let* $(x, z, W) \in \mathcal{S}$, *let* $\sigma > 0$ *satisfy*

$$\left\langle u, \left(W + \sum_{j \notin I(x)} \frac{z^j}{d_j(x)} \nabla d_j(x) \nabla d_j(x)^T \right) u \right\rangle \geq \sigma \|u\|^2 \quad \forall u \in T(x), \tag{10}$$

and let $(\Delta x^0, \Delta z^0)$ *solve* $L(x, z, W, 0)$. *Then*

$$\langle g(x), \Delta x^0 \rangle \leq -\sigma \|\Delta x^0\|^2$$

and Δx^0 *vanishes if and only if* x *is a stationary point for* (P), *in which case* $z + \Delta z^0$ *is the associated multiplier vector.*

Traditionally, the role of the barrier-parameter μ is to "tilt" the search direction towards the interior of the feasible set. Here, in addition, μ is to be chosen in such a way that, even when Δx^0 is small (or vanishes; close to or at a non-KKT stationary point for (P)), a direction Δx^μ of significant decrease for f is generated. Addressing this issue is the main contribution of the present paper. The key to this is the identity provided by the next lemma:[3] it gives an exact characterization of the set of values of the vector barrier parameter for which the resulting primal direction is a direction of descent for f. In particular, it shows that, in the neighborhood of a non-KKT stationary point, $\|\mu\|$ can be chosen large, yielding a large $\|\Delta x^\mu\|$.

Lemma 3. *Suppose Assumptions* 1–3 *hold. Let* $(x, z, W) \in \mathcal{S}$. *Moreover, let* $\mu \in \mathcal{R}^m$ *be such that* $\mu^j = 0$ *for all* j *for which* $z^j = 0$ *and let* $(\Delta x^0, \Delta z^0)$ *and* $(\Delta x^\mu, \Delta z^\mu)$ *be the solutions to* $L(x, z, W, 0)$ *and* $L(x, z, W, \mu)$, *respectively. Then*

$$\langle g(x), \Delta x^\mu \rangle = \langle g(x), \Delta x^0 \rangle + \sum_{j \text{ s.t. } z^j > 0} \frac{z^j + \Delta z^{0j}}{z^j} \mu^j. \tag{11}$$

Proof: First suppose that $x \in X_0$, i.e, suppose $D(x)$ is nonsingular. Let S denote the Schur complement of $D(x)$ in $M(x, z, W)$, i.e.,

$$S = -(W + B(x)^T D(x)^{-1} Z B(x)).$$

Since, in view of Lemma 1, $M(x, z, W)$ is nonsingular, so is S. Taking this into account and solving $L(x, z, W, 0)$ for $z + \Delta z^0$ yields, after some algebra,

$$z + \Delta z^0 = -Z D(x)^{-1} B(x) S^{-1} g(x).$$

Thus

$$\operatorname{diag}(\mu^j)(z + \Delta z^0) = -Z \operatorname{diag}(\mu^j) D(x)^{-1} B(x) S^{-1} g(x).$$

In view of our assumption on μ this implies that

$$-\mu^j [D(x)^{-1} B(x) S^{-1} g(x)]^j = \begin{cases} \dfrac{z^j + \Delta z^{0j}}{z^j} \mu^j & \text{if } z^j > 0 \\ 0 & \text{otherwise,} \end{cases}$$

yielding

$$\sum_{j \text{ s.t. } z^j > 0} \frac{z^j + \Delta z^{0j}}{z^j} \mu^j = -\langle \mu, D(x)^{-1} B(x) S^{-1} g(x) \rangle. \tag{12}$$

Now, it is readily verified that $\Delta x^\mu - \Delta x^0$ and $\Delta z^\mu - \Delta z^0$ satisfy the equation

$$M(x, z, W)\begin{bmatrix} \Delta x^\mu - \Delta x^0 \\ \Delta z^\mu - \Delta z^0 \end{bmatrix} = \begin{bmatrix} 0 \\ \mu \end{bmatrix}, \tag{13}$$

which yields, after some algebra,

$$\Delta x^\mu - \Delta x^0 = -S^{-1}B(x)^T D(x)^{-1}\mu. \tag{14}$$

Substituting (14) into the right-hand side of (12) proves that (11) holds when $x \in X_0$. To complete the proof, we use a continuity argument to show that the claim holds even when $D(x)$ is singular, i.e., when $x \in X$ is arbitrary. In view of Assumption 3, there exists a sequence $\{x_k\} \subset X_0$ converging to x. We show that, for k large enough, $(x_k, z, W) \in \mathcal{S}$. Since $I(x_k) = \emptyset$ for all $k > 0$, this amounts to showing that, for k large enough, and all $u \neq 0$,

$$\left\langle u, \left(W + \sum_{j \notin I(x)} \frac{z^j}{d_j(x_k)} \nabla d_j(x_k) \nabla d_j(x_k)^T\right) u\right\rangle + \sum_{j \in I(x)} \frac{z^j}{d_j(x_k)} \langle \nabla d_j(x_k), u\rangle^2 > 0, \tag{15}$$

where we have split the sum into two parts. Two cases arise. When $u \in T(x)$, in view of Assumption 2, since $(x, z, W) \in \mathcal{S}$, the first term in (15) is strictly positive for k large enough. Since the second term in (15) is nonnegative for all k, (15) holds. When $u \notin T(x)$, on the other hand, Assumption 2 implies that the first term in (15) is bounded, and, by virtue of (x, z, W) belonging to $\mathcal{S}$, since $z^j > 0$ for all $j \in I(x)$, the second term tends to $+\infty$ when k tends to ∞. Hence, again, (15) holds. Thus $(x_k, z, W) \in \mathcal{S}$ for all k large enough, and, in view of the first part of this proof, for all k large enough,

$$\langle g(x_k), \Delta x_k^\mu\rangle = \langle g(x_k), \Delta x_k^0\rangle + \sum_{j \text{ s.t. } z^j > 0} \frac{z^j + \Delta z_k^{0j}}{z^j} \mu^j, \tag{16}$$

where $(\Delta x_k^0, \Delta z_k^0)$ and $(\Delta x_k^\mu, \Delta z_k^\mu)$ denote the solutions of $L(x_k, z, W, 0)$ and $L(x_k, z, W, \mu)$, respectively. Finally, it follows from Lemma 1 that, as $k \to \infty$, $\Delta x_k^0 \to \Delta x^0$ and $\Delta x_k^\mu \to \Delta x^\mu$. Taking the limit as k tends to ∞ in both sides of (16) then completes the proof. □

Now let $(x, z, W) \in \mathcal{S}$ be our current primal and dual iterates and current estimate of the Hessian of the Lagrangian, and let Δx^μ be the primal component of the solution to $L(x, z, W, \mu)$. We aim at determining a value μ such that Δx^μ satisfies the following requirements: Δx^μ is a direction of (significant) descent for f, and $\|\Delta x^\mu\|$ is small only in the neighborhood of KKT points. From Lemmas 2 and 3 we infer that, given $(x, z, W) \in \mathcal{S}$, whenever $x \in X$ is not close to a KKT point for (P), $\langle g(x), \Delta x^\mu\rangle$ can be made significantly negative by an appropriate choice of a positive μ. To investigate this further, let $\phi \in \mathcal{R}_+^m$ be such that $\phi^j = 0$ whenever $z^j + \Delta z^{0j} \geq 0$ and, whenever x is a stationary point for (P), $\phi^j > 0$ for all j such that $z^j + \Delta z^{0j} < 0$. Let $\delta := \langle g(x), \Delta x^{\phi} \rangle$, i.e., in view of Lemmas 2 and 3,

$$\delta := \langle g(x), \Delta x^0 \rangle + \sum_{j=1}^{m} \frac{z^j + \Delta z^{0j}}{z^j} \phi^j \leq \langle g(x), \Delta x^0 \rangle. \tag{17}$$

We note the following:

(i) In view of Lemma 2,

$$\delta < 0 \tag{18}$$

whenever x is not a KKT point for (P).

(ii) In the spirit of interior-point methods, all the components of μ should be significantly positive away from a KKT point for (P).

(iii) As per the analysis in [9], for fast local convergence, μ should go to zero like $\|\Delta x^0\|^\nu$, with $\nu > 2$, as a KKT point for (P) is approached.

In view of (ii), the assignment $\mu := \phi$ is not appropriate in general, since some (or all) of the components of ϕ may vanish. Making μ proportional to $\|\Delta x^0\|^\nu$, motivated by (iii), is not appropriate either, since it is small near non-KKT stationary points, i.e., (ii) fails to hold. The choice $(\|\Delta x^0\|^\nu + \|\phi\|)z$ shows promise since all the components of this expression are positive away from KKT points for (P) and since, if, as intended, ϕ becomes identically zero near KKT points, the value used in [9] will be recovered near such points. However, with such choice, descent for f is not guaranteed. Since such descent *is* guaranteed with $\mu := \phi$ though ((i) above), this leads to the following choice: select μ to be on the line segment between ϕ and $(\|\Delta x^0\|^\nu + \|\phi\|)z$, as close as possible to the latter subject to sufficient descent for f. Specifically,

$$\mu := (1 - \varphi)\phi + \varphi(\|\Delta x^0\|^\nu + \|\phi\|)z \tag{19}$$

with $\varphi \in (0, 1]$ as close as possible to 1 subject to

$$\langle g(x), \Delta x^\mu \rangle \leq \theta\delta \tag{20}$$

where $\theta \in (0, 1)$ is prescribed. It is readily verified that the appropriate value of φ is given by

$$\varphi = \begin{cases} 1 & \text{if } E \leq 0 \\ \min\left\{\dfrac{(1-\theta)|\delta|}{E}, 1\right\} & \text{otherwise} \end{cases} \tag{21}$$

where

$$E = \sum_{j=1}^{m} \frac{z^j + \Delta z^{0j}}{z^j}((\|\Delta x^0\|^\nu + \|\phi\|)z^j - \phi^j). \tag{22}$$

Note that it follows from (20), (17), Lemma 2 and the conditions imposed on ϕ that

$$\langle g(x), \Delta x^\mu \rangle \leq \theta\delta \leq \theta \langle g(x), \Delta x^0 \rangle \leq 0; \tag{23}$$

that, in view of (18), whenever x is not a KKT point,

$$\langle g(x), \Delta x^\mu \rangle < 0; \tag{24}$$

and that, in view of (8) and strict positivity of all components of μ,

$$\langle \nabla d_j(x), \Delta x^\mu \rangle = \frac{\mu^j}{z^j} > 0 \quad \forall j \in I(x). \tag{25}$$

In the sequel, we leave ϕ largely unspecified. We merely require that it be assigned the value $\phi(x, z + \Delta z^0)$, with $\phi : X \times \mathcal{R}^m \rightarrow \mathcal{R}^m$ continuous and such that, for some $M > 0$ and for all $x \in X$,

$$\phi^j(x, \zeta) \begin{cases} >0 & \text{if} \quad \zeta^j < -M d_j(x) \\ =0 & \text{otherwise} \end{cases} \quad j = 1, \ldots, m. \tag{26}$$

Note that this condition meets the requirements stipulated above. In addition, it insures that, if $(x, z + \Delta z^0)$ converges to a KKT pair where strict complementarity slackness holds, ϕ will eventually become identically zero, and the algorithm will revert to that of [9], with the immediate benefit that rate of convergence results proved there can be invoked. An example of a function ϕ satisfying the above is given by $\phi^j(x, \zeta) = \max\{0, -\zeta^j - d_j(x)\}$.

We are now ready to state the algorithm. Much of it is borrowed from [9]. The lower bound in the update formula (35) for the dual variable z is as in [11], allowing for possible quadratic convergence. As discussed in [9], the purpose of the arc search, which employs a second order correction $\Delta\tilde{x}_k$ computed in Step 1v, is avoidance of Maratos-like effects.

Algorithm A

Parameters. $\xi \in (0, 1/2)$, $\eta \in (0, 1)$, $\nu > 2$, $\theta \in (0, 1)$, $z_{\max} > 0$, $\tau \in (2, 3)$, $\kappa \in (0, 1)$.
Data. $x_0 \in X$, $z_0^j \in (0, z_{\max}]$, $j = 1, \ldots, m$, W_0 such that $(x_0, z_0, W_0) \in \mathcal{S}$.

Step 0: Initialization. Set $k = 0$.
Step 1: Computation of search arc:

i. Compute $(\Delta x_k^0, \Delta z_k^0)$ by solving $L(x_k, z_k, W_k, 0)$. If $\Delta x_k^0 = 0$ and $z_k + \Delta z_k^0 \geq 0$ then stop.

ii. Set $\phi_k = \phi(x_k, z_k + \Delta z_k^0)$, and set

$$\delta_k = \langle g(x_k), \Delta x_k^0 \rangle + \sum_{j=1}^{m} \frac{z_k^j + \Delta z_k^{0j}}{z_k^j} \phi_k^j. \tag{27}$$

If $\sum_{j=1}^m \frac{z_k^j + \Delta z_k^{0j}}{z_k^j}((\|\Delta x_k^0\|^\nu + \|\phi_k\|) z_k^j - \phi_k^j) \leq 0$, set $\varphi_k = 1$; otherwise, set

$$\varphi_k = \min\left\{ \frac{(1-\theta)|\delta_k|}{\sum_{j=1}^m \frac{z_k^j + \Delta z_k^{0j}}{z_k^j}\left(\left(\left\|\Delta x_k^0\right\|^\nu + \|\phi_k\|\right) z_k^j - \phi_k^j\right)}, 1 \right\} \tag{28}$$

Finally set

$$\mu_k = (1 - \varphi_k)\phi_k + \varphi_k\left(\left\|\Delta x_k^0\right\|^\nu + \|\phi_k\|\right) z_k. \tag{29}$$

iii. Compute $(\Delta x_k, \Delta z_k)$ by solving $L(x_k, z_k, W_k, \mu_k)$.

iv. Set

$$I_k = \left\{ j : d_j(x_k) \leq z_k^j + \Delta z_k^j \right\} \tag{30}$$

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

$$\min \frac{1}{2}\langle \Delta\tilde{x}, W_k \Delta\tilde{x}\rangle \quad \text{s.t.} \quad d_j(x_k + \Delta x_k) + \langle \nabla d_j(x_k), \Delta\tilde{x}\rangle = \psi_k \quad \forall j \in I_k \tag{31}$$

where

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

If (31) is infeasible or unbounded or $\|\Delta\tilde{x}_k\| > \|\Delta x_k\|$, then set $\Delta\tilde{x}_k = 0$.

Step 2. Arc search. Compute α_k, the first number α in the sequence $\{1, \eta, \eta^2, \ldots\}$ satisfying

$$f(x_k + \alpha\Delta x_k + \alpha^2\Delta\tilde{x}_k) \leq f(x_k) + \xi\alpha\langle g(x_k), \Delta x_k\rangle \tag{32}$$
$$\forall d_j(x_k + \alpha\Delta x_k + \alpha^2\Delta\tilde{x}_k) > 0, \quad j = 1, \ldots, m. \tag{33}$$

Step 3. Updates.

Set

$$x_{k+1} = x_k + \alpha_k\Delta x_k + \alpha_k^2\Delta\tilde{x}_k \tag{34}$$
$$z_{k+1}^j = \min\left\{\max\left\{\|\Delta x_k\|^2, z_k^j + \Delta z_k^j\right\}, z_{\max}\right\}, \quad j = 1, \ldots, m. \tag{35}$$

Select W_{k+1} such that $(x_{k+1}, z_{k+1}, W_{k+1}) \in \mathcal{S}$. Set $k = k + 1$ and go back to Step 1.

Remark 1. Algorithm A is simpler than the algorithm of [9, 11], in that no special care is devoted to constraints with multiplier estimates of "wrong sign". In [9, 11], in the absence of the "ϕ" component of μ, such care was needed in Steps 1v, 2, and 3, in order to avoid convergence to non-KKT stationary points.

Remark 2. A minor difference between Algorithm A and the algorithm as stated in [9, 11], as well as with most published descriptions of "interior-point" algorithms, is that we do not require that the initial point be strictly feasible. Ability to use boundary points as initial points is beneficial in some contexts where, for instance, an interior-point scheme is used repeatedly in an inner loop.

Remark 3. At the end of [9] it is pointed out that the convergence analysis carried out in that paper is unaffected if the right-hand side of (32) is replaced with, e.g., $\min\{d_j(x_k)/2, \psi_k/2\}$. It can be checked that, similarly, such modification does not affect any of the results proved in the present paper. Of course, enforcing such feasibility margin, possibly with "2" replaced with, say, "1000", is very much in the spirit of the interior-point approach to constrained optimization. (With "2" in the denominator though, the step would be unduly restricted in early iterations.) In the present context however, in most cases, this would hardly make a difference, since nonlinearity of the constraints prevents us from first computing the exact step to the boundary of the feasible set, as is often done in linear programming: as a result, chances are that the first feasible trial step will also satisfy such required (small) feasibility margin. This was indeed our observation when we tried such variation in the context of the numerical experiments reported in Section 4 below. (Also, recall that the present algorithm has no difficulty dealing with iterates that lie on the boundary of the feasible set.)

Remark 4. If an initial point $x_0 \in X$ is not readily available, it can be constructed by performing a "Phase I" search, i.e., by maximizing $\min_j d_j(x)$ without constraints. This can be done, e.g., by applying Algorithm A to the problem

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

for which initial feasible points are readily available. A point x_0 satisfying $\min_j d_j(x_0) \geq 0$ 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 $\omega := \min_j d_j(x) < 0$, the origin does not belong to the convex hull of $\{\nabla d_j(x) : d_j(x) = \omega\}$.

In view of Lemma 1, $L(x_k, z_k, W_k, 0)$ and $L(x_k, z_k, W_k, \mu_k)$ have unique solutions for all k. Also, taking (24) and (25) into account, it is readily verified that the arc search is well defined, and produces a strictly feasible next iterate $x_{k+1} \in X_0$, whenever x_k is not a KKT point. On the other hand, the algorithm stops at Step 1i if and only if x_k is a KKT point. Thus, Algorithm A is well defined and, unless it stops at Step 1i, it constructs infinite sequences $\{x_k\}$ and $\{z_k\}$, with $x_k \in X_0$ for all $k \geq 1$; and if it does stop at Step 1i for some k_0, then x_{k_0} is a KKT point and (see Lemma 2) $z_k + \Delta z_k^0$ is the associated KKT multiplier vector.

3. Convergence analysis

We now consider the case when Algorithm A constructs infinite sequences $\{x_k\}$ and $\{z_k\}$. In proving global convergence, an additional assumption will be used, which places restrictions on how the W_k matrices should be constructed.

Assumption 4. Given any index set K such that the sequence $\{x_k\}_{k\in K}$ is bounded, there exist $\sigma_1, \sigma_2 > 0$ such that, for all $k \in K$, $k \geq 1$,

$$\left\langle u, \left(W_k + \sum_j \frac{z_k^j}{d_j(x_k)} \nabla d_j(x_k) \nabla d_j(x_k)^T \right) u \right\rangle \geq \sigma_1 \|u\|^2, \quad \forall u \tag{36}$$

and

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

Remark 5. This assumption, already used in [11], is significantly milder than the uniform positive definiteness assumption used in [9]. In particular, it is readily checked (using an argument similar to that used in the second part of the proof of Lemma 3) that, close to a KKT point of (P) where second order sufficiency conditions with strict complementarity hold, Assumption 4 is satisfied when W_k is the exact Hessian of the Lagrangian.

Our analysis is patterned after that of [9], but is simpler in its logic, and differs in the details. Roughly, given an infinite index set K such that, for some x^*, $\{x_k\}_{k\in K}$ converges to x^*, we show that (i) if $\{\Delta x_k\}_{k\in K} \to 0$ then x^* is KKT for (P), (ii) if $\{\Delta x_{k-1}\}_{k\in K} \to 0$ then x^* is KKT for (P), and (iii) there exists an infinite subset K' of K such that either $\{\Delta x_k\}_{k\in K'} \to 0$ or $\{\Delta x_{k-1}\}_{k\in K'} \to 0$. The first step is accomplished with the following lemma, which in addition provides a characterization of the limit points of $\{z_k + \Delta z_k^0\}$ and $\{z_k + \Delta z_k\}$, which will be of use in a later part of the analysis. The proof of this lemma borrows significantly from that of Lemma 3.7 in [9].

Lemma 4. *Suppose Assumptions* 1–4 *hold. Suppose* $\{x_k\}_{k\in K} \to x^*$ *for some infinite index set* K. *If* $\{\Delta x_k\}_{k\in K} \to 0$, *then* x^* *is a* KKT *point for* (P). *Moreover, in such case,* (i) $\{z_k + \Delta z_k^0\}_{k\in K} \to z^*$ *and* (ii) $\{z_k + \Delta z_k\}_{k\in K} \to z^*$, *where* z^* *is the KKT multiplier vector associated with* x^*.

Proof: We first use contradiction to prove that $\{z_k + \Delta z_k^0\}_{k\in K}$ is bounded. If it is not, then there exists an infinite index set $K' \subseteq K$ such that $z_k + \Delta z_k^0 \neq 0$ for all $k \in K'$ and $\max_j |z_k^j + \Delta z_k^{0j}| \to \infty$ as $k \to \infty$, $k \in K'$. Define

$$\zeta_k = \left\{ \max_j \left| z_k^j + \Delta z_k^{0j} \right| : \quad k \in K' \right\}.$$

Multiplying both sides of $L(x_k, z_k, W_k, 0)$, for $k \in K'$, by ζ_k^{-1} yields, for $k \in K'$,

$$-\frac{1}{\zeta_k} W_k \Delta x_k^0 + \sum_j \hat{z}_k^j \nabla d_j(x_k) = \frac{1}{\zeta_k} g(x_k) \tag{37}$$

$$\frac{1}{\zeta_k} z_k^j \langle \nabla d_j(x_k), \Delta x_k^0 \rangle + \hat{z}_k^j d_j(x_k) = 0 \tag{38}$$

where we have defined

$$\hat{z}_k^j = \frac{1}{\zeta_k}\left(z_k^j + \Delta z_k^{0j}\right).$$

Since $\max_j |\hat{z}_k^j| = 1$ for all $k \in K'$, there exists an infinite index set $K'' \subseteq K'$ such that $\{\hat{z}_k\}_{k\in K''}$ converges to $\hat{z}^* \neq 0$. Now, since $\{\Delta x_k\}_{k\in K}$ goes to zero, it follows from (23) and from Lemma 2 that $\{\Delta x_k^0\}_{k\in K}$ goes to zero as well. Thus, letting $k \to \infty$, $k \in K''$ in (37) and (38) and invoking boundedness of W_k (Assumption 4), we obtain

$$\sum_j \hat{z}^{*j} \nabla d_j(x^*) = 0 \tag{39}$$

$$\hat{z}^{*j} d_j(x^*) = 0 \tag{40}$$

which implies that $\hat{z}^{*j} = 0$ for all $j \notin I(x^*)$ and

$$\sum_{j\in I(x^*)} \hat{z}^{*j} \nabla d_j(x^*) = 0.$$

Since $\hat{z}^* \neq 0$, this contradicts Assumption 3, proving boundedness of $\{z_k + \Delta z_k^0\}_{k\in K}$, which implies the existence of accumulation points for that sequence. Now, let z^* be such an accumulation point, i.e., suppose that, for some infinite index set $K''' \subseteq K$, $\{z_k + \Delta z_k^0\}_{k\in K'''}$ converges to z^*. Letting $k \to \infty$, $k \in K'''$ in $L(x_k, z_k, W_k, 0)$ and using Assumptions 2 and 4 yields

$$g(x^*) - \sum_{j\in I(x^*)} z^{*j} \nabla d_j(x^*) = 0$$

$$z^{*j} d_j(x^*) = 0, \quad j = 1, \dots, m$$

showing that x^* is a stationary point of (P) with multiplier z^*. Next, we show that $z^{*j} \geq 0$ for all $j \in I(x^*)$, thus proving that x^* is a KKT point for (P). Indeed, since $\{\Delta x_k\}_{k\in K}$ goes to zero, it follows from (23) and (17) that

$$\sum_j \frac{z_k^j + \Delta z_k^{0j}}{z_k^j} \phi_k^j \to 0 \quad \text{as } k \to \infty, k \in K,$$

which, since, in view of (26), all terms in this sum are nonnegative, and since, by construction, $\{z_k\}$ is bounded, implies that

$$\left(z_k^j + \Delta_k^{0j}\right)\phi_k^j \to 0 \quad \text{as } k \to \infty, k \in K. \tag{41}$$

By the assumed continuity of $\phi(\cdot,\cdot)$, this implies that

$$z^{*j}\phi^j(x^*, z^*) = 0.$$

Since $d_j(x^*) \geq 0$ for all j, it follows from (26) that $\phi(x^*, z^*) = 0$ and that $z^{*j} \geq 0$ for all $j \in I(x^*)$. Thus x^* is a KKT point for (P) and every accumulation point of $\{z_k + \Delta z_k^0\}_{k\in K}$ is an associated KKT multiplier vector. Since Assumption 3 implies uniqueness of the KKT multiplier vector, boundedness implies that $\{z_k + \Delta z_k^0\}_{k\in K}$ converges to z^*, proving (i). To prove (ii), observe that, since ϕ is continuous and $\phi(x^*, z^*) = 0$, it must hold that $\{\phi_k\}_{k\in K}$ goes to zero and, in view of (29), that $\{\mu_k\}_{k\in K}$ goes to zero. Using this fact and repeating the argument used above for $\{z_k + \Delta z_k^0\}_{k\in K}$ proves that $\{z_k + \Delta z_k\}_{k\in K}$ converges to some z^{**}, and x^* is a stationary point of (P) with multiplier z^{**}. Assumption 3 implies that $z^{**} = z^*$, completing the proof. □

Now on to the second step.

Lemma 5. *Suppose Assumptions* 1–4 *hold. Suppose* $\{x_k\}_{k\in K} \to x^*$ *for some infinite index set* K. *If* $\{\Delta x_{k-1}\}_{k\in K} \to 0$, *then* x^* *is a* KKT *point for* (P).

Proof: In view of Step 3 in Algorithm A, we can write

$$\|x_k - x_{k-1}\| = \left\|\alpha_{k-1}\Delta x_{k-1} + \alpha_{k-1}^2 \Delta\tilde{x}_{k-1}\right\| \leq \alpha_{k-1}\|\Delta x_{k-1}\| + \alpha_{k-1}^2\|\Delta\tilde{x}_{k-1}\|,$$

yielding, since $\|\Delta\tilde{x}_{k-1}\| \leq \|\Delta x_{k-1}\|$ and $\alpha_{k-1} \leq 1$,

$$\|x_k - x_{k-1}\| \leq 2\|\Delta x_{k-1}\|.$$

Since $\{x_k\}_{k\in K} \to x^*$ and $\{\Delta x_{k-1}\}_{k\in K} \to 0$, it follows that $\{x_{k-1}\}_{k\in K} \to x^*$. The claim then follows from Lemma 4. □

The third step relies on "continuity" results proved in the next lemma.

Lemma 6. *Suppose Assumptions* 1–4 *hold. Let* K *be an infinite index set such that for some* x^*, z^*, W^*,

$$\{x_k\}_{k\in K} \to x^*, \tag{42}$$

$$\{z_k\}_{k\in K} \to z^*, \tag{43}$$

$$\{W_k\}_{k\in K} \to W^*. \tag{44}$$

Moreover, suppose that $z^{*j} > 0$ *for all* j *such that* $d_j(x^*) = 0$. *Then* $M(x^*, z^*, W^*)$ *is nonsingular. Furthermore*

(i) $\{\Delta x_k^0\}_{k\in K} \to \Delta x^{0*}$ *and* $\{\Delta z_k^0\}_{k\in K} \to \Delta z^{0*}$ *where* $(\Delta x^{0*}, \Delta z^{0*})$ *is the solution to the nonsingular system* $L(x^*, z^*, W^*, 0)$;
(ii) $\{\phi_k\}_{k\in K} \to \phi^*$, $\{\delta_k\}_{k\in K} \to \delta^*$, $\{\varphi_k\}_{k\in K} \to \varphi^*$, *and* $\{\mu_k\}_{k\in K} \to \mu^*$, *for some* $\phi^* \geq 0$, $\delta^* \leq 0$, $\varphi^* \in [0, 1]$, *and* $\mu^* \geq 0$;
(iii) $\{\Delta x_k\}_{k\in K} \to \Delta x^*$ *and* $\{\Delta z_k\}_{k\in K} \to \Delta z^*$ *where* $(\Delta x^*, \Delta z^*)$ *is the solution to the nonsingular system* $L(x^*, z^*, W^*, \mu^*)$;
(iv) $\delta^* = 0$ *if and only if* $\Delta x^* = 0$.

Proof: Nonsingularity of $M(x^*, z^*, W^*)$ is established in Lemma PTH-3.5* in the appendix of [11]. Claim (i) readily follows. Now consider Claim (ii). First, by continuity of $\phi(\cdot,\cdot)$, $\{\phi_k\}_{k\in K}$ converges to $\phi^* := \phi(x^*, z^* + \Delta z^{0*}) \geq 0$. Next, the second term in (27) converges as $k \to \infty$, $k \in K$. In particular, for j such that $z^{*j} = 0$, convergence of $(z_k^j + \Delta z_k^{0j})/z_k^j$ on K follows from the identity, derived from the second equation in $L(x_k, z_k, W_k, 0)$,

$$\frac{z_k^j + \Delta z_k^{0j}}{z_k^j} = -\frac{\langle \nabla d_j(x_k), \Delta x_k^0 \rangle}{d_j(x_k)} \tag{45}$$

and our assumption that $d_j(x^*) > 0$ whenever $z^{*j} = 0$, which implies convergence of the right-hand side of (45). In view of (18), it follows that $\{\delta_k\}_{k\in K}$ converges to some $\delta^* \leq 0$. It also follows that the denominator in (28) converges on K, implying that $\{\varphi_k\}_{k\in K}$ converges to φ^* for some $\varphi^* \in [0, 1]$. (In particular, if the denominator in (28) converges to zero, then $\varphi^* = 1$.) And it follows from (29) that $\{\mu_k\}_{k\in K}$ converges to some $\mu^* \geq 0$. Next, Claim (iii) directly follows from the above (since $M(x^*, z^*, W^*)$ is nonsingular) and the "if" part of Claim (iv) follows from (23). To conclude we prove the "only if" part of Claim (iv). First, (23) and Lemma 2 imply that, if $\delta^* = 0$, then $\Delta x^{0*} = 0$. Then it follows from (27) that

$$\sum_j \frac{z_k^j + \Delta z_k^{0j}}{z_k^j} \phi_k^j \to 0 \quad \text{as } k \to \infty, k \in K.$$

An argument identical to that used in the proof of Lemma 4 then implies that $\phi(x^*, z^* + \Delta z^{0*}) = 0$. Finally, it follows from (29) that $\mu^* = 0$ and, since $\Delta x^{0*} = 0$, this implies that $\Delta x^* = 0$. □

The third and final step can now be taken.

Lemma 7. *Suppose Assumptions* 1–4 *hold. Suppose that, for some infinite index set* K, $\{x_k\}_{k\in K}$ *converges and* $\inf_{k\in K} \|\Delta x_{k-1}\| > 0$. *Then* $\{\Delta x_k\}_{k\in K} \to 0$.

Proof: Since z_k is bounded by construction and W_k is bounded by Assumption 4, there is no loss of generality in assuming that $\{z_k\}_{k\in K}$ and $\{W_k\}_{k\in K}$ converge to some z^* and W^*, respectively. Since $\inf_{k\in K} \|\Delta x_{k-1}\| > 0$, it follows from the construction rule for z_k in Step 3 of Algorithm A that all components of z^* are strictly positive. Thus Lemma 6 applies. Let Δz^{0*}, ϕ^*, δ^*, φ^* and μ^* be as guaranteed by that lemma.

Now proceed by contradiction, i.e., suppose that there exists an infinite subset $K' \subseteq K$ such that $\inf_{k\in K'} \|\Delta x_k\| > 0$. We show that there exists some $\underline{\alpha} > 0$ such that $\alpha_k \geq \underline{\alpha}$ for all $k \in K'$, k sufficiently large. First, since $\inf_{k\in K'} \|\Delta x_k\| > 0$, it follows from Lemma 6(iv) that $\delta^* < 0$. In view of (20) it follows that, for all $k \in K'$, k large enough,

$$\langle g(x_k), \Delta x_k \rangle < \frac{\theta\delta^*}{2} < 0. \tag{46}$$

Next, it follows from (27) that Δx^{0*} and ϕ^* cannot both be zero, and from (28) that $\varphi^* > 0$. In view of (29), it follows that $\mu^{*j} > 0$ for all j. Also, it follows from (8) that, for $j \in I(x^*)$, $\langle \nabla d_j(x_k), \Delta x_k \rangle$ converges to μ^{*j}/z^{*j} as $k \to \infty, k \in K'$. As a consequence, for all $k \in K'$, k large enough,

$$\langle \nabla d_j(x_k), \Delta x_k \rangle \geq \frac{\mu^{*j}}{2z^{*j}} > 0 \quad \forall j \in I(x^*). \tag{47}$$

Finally, with (46) and (47) in hand, it is routine to show that there exists $\underline{\alpha} > 0$ such that $\alpha_k \geq \underline{\alpha}$ for all $k \in K'$ (see, e.g., the proof of Lemma 3.9 in [9]). It follows from (32) and (46) that, for $k \in K'$, k large enough,

$$f\left(x_k + \alpha_k \Delta x_k + \alpha_k^2 \Delta \tilde{x}_k\right) - f(x_k) < -\underline{\alpha}\xi\theta\underline{\delta}.$$

In view of continuity of f (Assumption 2) and of the monotonic decrease of f (see (32)), this contradicts the convergence of $\{x_k\}$ on K'. □

Global convergence thus follows.

Theorem 8. *Suppose Assumptions* 1–4 *hold. Then every accumulation point of* $\{x_k\}$ *is a* KKT *point for* (P).

Our next goal is to show that the entire sequence $\{x_k\}$ converges to a KKT point of (P) and that $\{z_k\}$ converges to the associated KKT multiplier vector. An additional assumption will be used, as well as a lemma.

Assumption 5. The sequence $\{x_k\}$ generated by Algorithm A has an accumulation point x^* which is an *isolated* KKT point for (P). Furthermore, strict complementarity holds at x^*.

Let z^* denote the (unique, in view of Assumption 3) KKT multiplier vector associated to x^*. By strict complementarity, $z^{*j} > 0$ for all $j \in I(x^*)$.

Lemma 9. *Suppose Assumptions* 1–5 *hold. If there exists an infinite index set K such that* $\{x_{k-1}\}_{k\in K} \to x^*$ *and* $\{x_k\}_{k\in K} \to x^*$, *where* x^* *is as in Assumption* 5, *then* $\{\Delta x_k\}_{k\in K} \to 0$.

Proof: Proceeding by contradiction, suppose that, for some infinite index set $K' \subseteq K$, $\inf_{k\in K'} \|\Delta x_k\| > 0$. Then, in view of Lemma 7, $\inf_{k\in K'} \|\Delta x_{k-1}\| = 0$, i.e., there exists an infinite index set $K'' \subseteq K'$ such that $\{\Delta x_{k-1}\}_{k\in K''} \to 0$. In view of Lemma 4 and of the multiplier update formula in Step 3 of Algorithm A, it follows that, for all j, $\{z_k^j\}_{k\in K''}$ converges to $\min\{z^{*j}, z_{\max}\}$, which, in view of Assumption 5, is strictly positive for all $j \in I(x^*)$. Also, in view of Assumption 4, there is no loss of generality in assuming that $\{W_k\}_{k\in K''} \to W^*$ for some matrix W^*. In view of Lemma 6, $M(x^*, z^*, W^*)$ is nonsingular. From Lemmas 2 and 6(i) it follows that $\{\Delta x_k^0\}_{k\in K''} \to 0$ and $\{z_k + \Delta z_k^0\}_{k\in K''} \to z^*$. It follows that $\{\phi_k\}_{k\in K''} \to 0$ and $\{\mu_k\}_{k\in K''} \to 0$. Thus $L(x^*, z^*, W^*, \mu^*)$ is identical to

$L(x^*, z^*, W^*, 0)$ and, in view of Lemmas 2 and 6(iii), $\{\Delta x_k\}_{k\in K''} \to 0$. Since $K'' \subseteq K'$, this is a contradiction. □

Proposition 10. *Suppose Assumptions* 1–5 *hold. Then the entire sequence* $\{x_k\}$ *constructed by Algorithm* A *converges to* x^*. *Moreover,* (i) $\Delta x_k \to 0$, (ii) $z_k + \Delta z_k \to z^*$, (iii) $z_k^j \to \min\{z^{*j}, z_{\max}\}$ *for all* j, (iv) $\mu_k \to 0$ *and, for all k large enough,* (v) $I_k = I(x^*)$ *and* (vi) $\phi_k = 0$.

Proof: The main claim as well as claims (i) through (iii) are proved as in the proof of Proposition 4.2 of [9], with our Lemmas 4, 7, and 9 in lieu of Lemmas 3.7, 3.9, and 4.1 of [9]. (Note that the second order sufficiency condition is not used in that proof.) Claims (iv), (v) and (vi) readily follow, in view of strict complementarity, continuity of $\phi(\cdot,\cdot)$, and convergence of $\{x_k\}$ and $\{z_k\}$. □

It follows that, for k large enough, Algorithm A is identical to the algorithm of [9], with dual update formula modified as in [11] to allow for quadratic convergence. Thus the rate of convergence results enjoyed by the algorithm of [11] hold here. Three more assumptions are needed. The first two are stronger versions of Assumptions 2 and 5.

Assumption 6. The functions f and d_j, $j = 1, \dots, m$ are three times continuously differentiable.

Assumption 7. The second order sufficiency conditions of optimality with strict complementary slackness hold at x^*. Furthermore, the associated KKT multiplier vector z^* satisfies $z^{*j} \le z_{\max}$ for all j.

Assumption 8.

$$\frac{\left\|N_k\left(W_k - \nabla^2_{xx}\mathcal{L}(x^*, z^*)\right) N_k \Delta x_k\right\|}{\|\Delta x_k\|} \to 0 \quad \text{as } k \to \infty \tag{48}$$

where

$$N_k = I - \hat{B}_k^\dagger \hat{B}_k,$$

with

$$\hat{B}_k = [\nabla d_j(x_k),\ j \in I(x^*)]^T \in \mathcal{R}^{|I(x^*)| \times n}.$$

Here, $\hat{B}_k^\dagger$ denotes the Moore-Penrose pseudo-inverse of $\hat{B}_k$.

Like in [9], two-step superlinear convergence ensues.

Theorem 11. *Suppose Assumptions* 1–8 *hold. 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, as observed in [11], under Assumption 7, Assumption 4 holds here for k large enough when W_k is selected to be *equal* to the Hessian of the Lagrangian $\mathcal{L}$, and such choice results in quadratic convergence.

Theorem 12. *Suppose Assumptions* 1–7 *hold, suppose that, at every iteration except possibly finitely many,* W_k *is selected as*

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

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

$$\left\| \begin{bmatrix} x_{k+1} - x^* \\ z_{k+1} - z^* \end{bmatrix} \right\| \le \Gamma \left\| \begin{bmatrix} x_k - x^* \\ z_k - z^* \end{bmatrix} \right\|^2 \quad \text{for all } k. \tag{49}$$

4. Numerical examples

In order to assess the practical value of the proposed "non-KKT stationary point avoidance" scheme, we performed preliminary comparative tests of MATLAB 6.1 implementations of the algorithm of [9] and Algorithm A. Implementation details for the former and for the portion of the latter that is common to both were as specified in [11]. Those that are relevant to Algorithm A are repeated here for the reader's convenience. (Part of the following text is copied verbatim from [11].) First, a difference with Algorithm A as stated in Section 2 was that in the update formula for the multipliers in Step 3, $\|\Delta x_k\|^2$ was changed to $\min\{z_{\min}, \|\Delta x_k\|^2\}$, where $z_{\min} \in (0, z_{\max})$ is an additional algorithm parameter. The motivation is that $\|\Delta x_k\|$ is meaningful only when it is small. This change does not affect the convergence analysis. Second the following parameter values were used: $\xi = 10^{-4}$, $\eta = 0.8$, $\nu = 3$, $\theta = 0.8$, $z_{\min} = 10^{-4}$, $z_{\max} = 10^{20}$, $\tau = 2.5$, and $\kappa = 0.5$. Next, the initial value x_0 was selected in each case as specified in the source of the test problem, and z_0^j was assigned the value $\max\{0.1, z_0'^j\}$, where z_0' is the (linear least squares) solution of

$$\min_{z_0'} \|g(x_0) - B(x_0)z_0'\|^2.$$

In all the tests, z_0 thus defined satisfied the condition specified in the algorithm that its components should all be no larger than $z_{\max}$. 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

$$\nabla^2_{xx}\mathcal{L}(x_k, z_k) + \sum_{i \in I'_k} \frac{z_k^i}{d_i(x_k)} \nabla d_i(x_k) \nabla d_i(x_k)^T,$$

where I'_k is the set of indices of constraints with value *larger* than 10^{-10}, to the tangent plane to the constraints left out of the sum, i.e., to the subspace

$$\{v \in \mathcal{R}^n : \langle \nabla d_j(x_k), v \rangle = 0, \ \forall j \notin I'_k\}.$$

Then, set

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

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. Finally,[4] the stopping criterion (inserted at the end of Step 1i) was as follows, with $\epsilon_{\text{stop}} = 10^{-8}$. The run was deemed to have terminated successfully if at any iteration k

$$\max_j\{-(z_k^j + \Delta z_k^{0j})\} < \epsilon_{\text{stop}}$$

and either

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

or

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

Concerning the implementation of the "non-KKT stationary point avoidance" scheme, $\phi(\cdot,\cdot)$ was given by

$$\phi^j(x, \zeta) = \min\{\max\{0, -\zeta^j - 10^3 d_j(x)\}, 1\} \quad j = 1, \ldots, m,$$

which satisfies the conditions set forth in Section 2. Note that, with such $\phi(\cdot,\cdot)$, as a result, ϕ_k^j is never large (bounded by 1) and it can be nonzero only very close to the corresponding constraint boundary. The motivation for this is that the algorithm from [9] should not be "corrected" whenever it performs well, i.e., away from non-KKT stationary points. All tests were run within the CUTEr testing environment [6], on a SUN Enterprise 250 with two UltraSparc-II 400 MHz processors, running Solaris 2.7.

We then ran the MATLAB code on 29 of the 30 problems[5] from [7] that do not include equality constraints and for which the initial point provided in [7] satisfies all inequality constraints. (While a phase I-type scheme could be used on the other problems—see Remark 4—testing such scheme is outside the main scope of this paper.) The results are compared in Table 1. In that table the first three columns indicate the problem number,

Table 1. Results.

			Algorithm A		Algorithm from [11]	
Prob.	n	m	#Itr	f_{final}	#Itr	f_{final}
HS1	2	1	24	6.5782e−27	24	6.5782e−27
HS3	2	1	4	8.5023e−09	4	8.5023e−09
HS4	2	2	4	2.6667e+00	4	2.6667e+00
HS5	2	4	6	−1.9132e+00	6	−1.9132e+00
HS12	2	1	5	−3.0000e+01	5	−3.0000e+01
HS24	2	5	14	−1.0000e+00	14	−1.0000e+00
HS25	3	6	62	1.8185e−16	62	1.8185e−16
HS29	3	1	8	−2.2627e+01	8	−2.2627e+01
HS30	3	7	7	1.0000e+00	7	1.0000e+00
HS31	3	7	7	6.0000e+00	0	1.9000e+01
HS33	3	6	29	−4.5858e+00	29	−4.5858e+00
HS34	3	8	19	−8.3403e−01	30	−8.3403e−01
HS35	3	4	8	1.1111e−01	0	9.0000e+00
HS36	3	7	10	−3.3000e+03	10	−3.3000e+03
HS37	3	8	7	−3.4560e+03	7	−3.4560e+03
HS38	4	8	37	3.1564e−24	37	3.1594e−24
HS43	4	3	9	−4.4000e+01	9	−4.4000e+01
HS44	4	10	16	−1.5000e+01	0	0.0000e+00
HS57	2	3	15	2.8460e−02	15	2.8460e−02
HS66	3	8	11	5.1816e−01	1001	5.1817e−01
HS70	4	9	22	1.7981e−01	22	1.7981e−01
HS84	5	16	29	−5.2803e+06	30	−5.2803e+06
HS85	5	48	296	−2.2156e+00	296	−2.2156e+00
HS86	5	15	14	−3.2349e+01	0	2.0000e+01
HS93	6	8	12	1.3508e+02	12	1.3508e+02
HS100	7	4	9	6.8063e+02	9	6.8063e+02
HS110	10	20	6	−4.5778e+01	6	−4.5778e+01
HS113	10	8	10	2.4306e+01	10	2.4306e+01
HS117	15	20	19	3.2349e+01	25	3.2349e+01

the number of variables, and the number of (inequality) constraints, the next two indicate the number of iterations and final objective value for Algorithm A and the last two show the same information for the algorithm from [11]. A value of 0 for the number of iterations indicates that the algorithm was unable to move from the initial point, and a value of 1001 indicates that the stopping criterion was not satisfied after 1000 iterations.

Algorithm A solved successfully all problems. In all cases, the final objective value is identical to that listed in [7]. (For problem HS38, the value given in [7] is 0.) On most of these problems, the algorithm from [11] and Algorithm A performed similarly. It was observed that, in fact, in most of the problems, ϕ_k is zero throughout, in which case a similar behavior of the two algorithms can indeed be expected. The benefit provided by the scheme introduced in the present paper is most visible in the results obtained on HS31, HS35, HS44, and HS86, where the algorithm of [11] fails because the initial point is stationary for (P), and Algorithm A performs satisfactorily. Worth noting is one other problem where Algorithm A performs well while the algorithm of [11] experiences difficulties: problem HS66.

We anticipate that the advantages of the new scheme will be clearer on problems of larger size since the possibility that at least one multiplier estimate be negative (and thus that ϕ_k be nonzero) will then be much stronger. Numerical testing on such problems is underway.

5. Concluding remarks

A simple primal-dual feasible interior-point method has been presented and analyzed. Preliminary numerical tests suggest that the new algorithm holds promise. While we have focussed on the case of problems not involving any equality constraints, the proposed algorithm is readily extended to such problems using the scheme analyzed in [11], which is based on an idea proposed decades ago by D.Q. Mayne and E. Polak [8]. We feel that, by its simplicity (and the simplicity of the convergence analysis), its strong convergence properties under mild assumptions, and its promising behavior in preliminary numerical tests, the proposed algorithm is a valuable addition to the arsenal of primal-dual interior-point methods.

Acknowledgment

The authors wish to thank two anonymous referees for their valuable suggestions, which helped improve the paper.

Notes

1. There are two misprints in [9, Section 5]: in Eq. (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 $B_k^{-1} M_k$.
2. See [11] for a description of the algorithm of [9] in the notation used in the present paper. As we already stressed in the introduction, in [11], the positive definiteness assumption on W is significantly relaxed.
3. Inspired from [10].
4. Another minor difference between the implementations and the formal statement of the algorithms was that we required nonstrict feasibility rather than strict feasibility in the line search. This only made a difference on

a few problems in which, in the final iterates, due to round-off, the first trial point lay on the boundary of the feasible set, even though the theory guarantees that it is strictly feasible.

5. We left out problem HS67 which is termed "irregular" in [7]: computation of the cost and constraint functions involves an iterative procedure with variable number of iterations, rendering these functions discontinuous.

References

1. R.H. Byrd, J.C. Gilbert, and J. Nocedal, "A trust region method based on interior point techniques for nonlinear programming," Mathematical Programming, vol. 89, pp. 149–185, 2000.
2. R.H. Byrd, M.E. Hribar, and J. Nocedal, "An interior point algorithm for large-scale nonlinear programming," SIAM J. on Optimization, vol. 9, no. 4, pp. 877–900, 1999.
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., vol. 89, pp. 507–541, 1996.
4. A. Forsgren and P.E. Gill, "Primal-dual interior methods for nonconvex nonlinear programming," SIAM J. on Optimization, vol. 8, no. 4, pp. 1132–1152, 1998.
5. D.M. Gay, M.L. Overton, and M.H. Wright, "A primal-dual interior method for nonconvex nonlinear programming," in Y. Yuan (Ed.), Advances in Nonlinear Programming, Kluwer Academic Publisher, 1998, pp. 31–56.
6. 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.
7. W. Hock and K. Schittkowski, "Test examples for nonlinear programming codes," vol. 187 of Lecture Notes in Economics and Mathematical Systems. Springer-Verlag, 1981.
8. D.Q. Mayne and E. Polak, "Feasible direction algorithms for optimization problems with equality and inequality constraints," Math. Programming, vol. 11, pp. 67–80, 1976.
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., vol. 26, no. 4, pp. 788–811, 1988.
10. H.-D. Qi and L. Qi, "A new QP-free, globally convergent, locally superlinearly convergent algorithm for inequality constrained optimization," SIAM J. on Optimization, vol. 11, no. 1, pp. 113–132, 2000.
11. A.L. Tits, A. Wächter, S. Bakhtiari, T.J. Urban, and C.T. Lawrence, "A primal-dual interior-point method for nonlinear programming with strong global and local convergence properties," SIAM J. on Optimization 2003 Technical Report TR 2002-29, Institute for Systems Research, University of Maryland, College Park, Maryland, July 2002. http://www.isr.umd.edu/~andre/pdiprev.ps.
12. R.J. Vanderbei and D.F. Shanno, "An interior-point algorithm for nonconvex nonlinear programming," Computational Optimization and Applications, vol. 13, pp. 231–252, 1999.
13. H. Yamashita, "A globally convergent primal-dual interior point method for constrained optimization," Optimization Methods and Software, vol. 10, pp. 443–469, 1998.
14. 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.