Newton-KKT Interior-Point Methods for Indefinite Quadratic Programming*

Pierre-Antoine Absil
School of Computational Science
Florida State University
Tallahassee, FL 32306-4120
http://www.csit.fsu.edu/~absil/

André L. Tits†
Department of Electrical and Computer Engineering
and Institute for Systems Research
University of Maryland
College Park, MD 20742
andre@umd.edu

Revised version, submitted to COAP on 4 August 2005

Abstract

Two interior-point algorithms are proposed and analyzed, for the (local) solution of (possibly) indefinite quadratic programming problems. They are of the Newton-KKT variety in that (much like in the case of primal-dual algorithms for linear programming) search directions for the "primal" variables and the Karush-Kuhn-Tucker (KKT)

*The work of the first author was supported in part by the School of Computational Science of Florida State University through a postdoctoral fellowship. Part of this work was done while this author was a Research Fellow with the Belgian National Fund for Scientific Research (Aspirant du F.N.R.S.) at the University of Liège. The work of the second author was supported in part by the National Science Foundation under Grants DMI-9813057 and DMI-0422931 and by the US Department of Energy under Grant DEFG0204ER25655.

†Corresponding author.

multiplier estimates are components of the Newton (or quasi-Newton) direction for the solution of the equalities in the first-order KKT conditions of optimality or a perturbed version of these conditions. Our algorithms are adapted from previously proposed algorithms for convex quadratic programming and general nonlinear programming. First, inspired by recent work by P. Tseng based on a "primal" affine-scaling algorithm (*à la* Dikin) [J. of Global Optimization, 30 (2004), no 2, 285–300], we consider a simple Newton-KKT *affine-scaling* algorithm. Then, a "*barrier*" version of the same algorithm is considered, which reduces to the affine-scaling version when the barrier parameter is set to zero at every iteration, rather than to the prescribed value. Global and local quadratic convergence are proved under nondegeneracy assumptions for both algorithms. Numerical results on randomly generated problems suggest that the proposed algorithms may be of great practical interest.

Key words. interior-point algorithms, primal-dual algorithms, indefinite quadratic programming, Newton-KKT

1 Introduction

Consider the quadratic programming problem

$$(P) \qquad \text{minimize } \frac{1}{2}\langle x, Hx\rangle + \langle c, x\rangle \;\text{ s.t. }\; Ax \leq b, \quad x \in \mathcal{R}^n,$$

with $A \in \mathcal{R}^{m\times n}$, $b \in \mathcal{R}^m$, $c \in \mathcal{R}^n$, and with $H \in \mathcal{R}^{n\times n}$ symmetric. In the past two decades, much research activity has been devoted to developing and analyzing interior-point methods for solving such problems in the convex case, i.e., when H is positive semidefinite. In particular, such algorithms were first proposed and analyzed in [Ye87, Ye89] and in [MA89]. The Interior Ellipsoid method of Ye and Tse [YT89] is a primal affine-scaling algorithm for convex quadratic programming in standard equality form. The primal affine scaling idea consists in minimizing the cost over a sequence of ellipsoids whose shape depends on the distance from the current interior feasible point to the faces of the feasible polyhedron. Dikin [Dik67] developed the method for linear problems and proved its global convergence under a primal nondegeneracy assumption. This method may alternatively be viewed as an interior trust-region method [CGT00] where the Dikin ellipsoid is used as the trust region.

Ye and Tse [YT89] showed that if the sequence generated by their Interior Ellipsoid algorithm converges, then the limit point is an optimal solution. In the strictly convex case, the algorithm generates sequences that converge to the optimal solution [Ye89]. In [YT89], a modified version is also proposed, derived as an extension of Karmarkar's linear programming algorithm, and complexity bounds are obtained. Monteiro and Tsuchiya [MT98] proved convergence of the algorithm without any nondegeneracy assumption. (We refer to [MT98] for further references on the convex quadratic programming problem.) Monteiro and Adler [MA89] proposed a barrier-based path-following algorithm and obtained complexity bounds.

Interior-point methods have also been proposed for the computation of local solutions to general, nonlinear programming problems (see, e.g., [FM68, Her82, Her86, PTH88, ETTZ96, FG98, GOW98, Yam98, VS99, BGN00, QQ00, BT03, TWB$^+$03, YLQ03, GZ05], and the recent survey [FGW02]), and these of course can be used for tackling (P). However, only limited attention has been devoted to exploiting the quadratic programming structure of (P) in the nonconvex case. Notable exceptions include the work of Ye and of Tseng ([Ye89, Ye92, Tse04, TY02]) on Dikin's algorithm [Dik67], and that of Bonnans and Bouhtou [BB95], of Coleman and Liu [CL99], and of [Ye98], as we discuss next.

Interior-point methods for the general (indefinite) quadratic programming problem were first considered in [Ye89], where numerical experiments with the Interior Ellipsoid method in the indefinite case were mentioned. The Interior Ellipsoid method was formally extended to indefinite quadratic programming in [Ye92]. In that paper, under some nondegeneracy assumptions, the sequences produced by that algorithm are shown to converge to a point satisfying the first and second order necessary conditions of optimality. This algorithm was further analyzed by Bonnans and Bouhtou [BB95], who proposed an extended algorithm allowing inexact solution of the trust-region subproblems and the possibility of a line search in the direction obtained from the subproblem. Under nondegeneracy assumptions and an assumption on the step length, they show that this algorithm converges to a first-order optimal point. These ideas were further generalized by Bonnans and Pola [BP97] to nonlinear optimization under linear constraints.

Coleman and Liu [CL99] proposed an "Interior-Newton" algorithm for (indefinite) quadratic programming in standard equality form. Each step involves two directions: the solution to a projected trust-region subproblem and a projected steepest descent direction. The trust-region-based step is prefered over the steepest-descent-based step when it produces a sufficient decrease of a test function. In the trust-region subproblems of [CL99], the model and the ellipsoid are different from the ones of the Interior Ellipsoid method [Ye92, BB95], and are chosen in such a way that the subproblem solution is ultimately the primal part of the Newton direction for the KKT equations. Under standard assumptions (compact sublevel sets, primal nondegeneracy and strict complementarity), the authors show that their Interior-Newton algorithm converges (globally) to a single point satisfying the second-order necessary conditions, and that the rate of convergence is 2-step quadratic if the limit point is a strong local minimizer. However, the result requires that a trust-region subproblem be solved exactly at each iteration, a computationally costly task.

Tseng and Ye [TY02] showed how path-following and affine-scaling type methods, applied to nonconvex quadratic optimization, may fail to converge to a local minimizer. Strategies are proposed in [TY02] to overcome this difficulty, based on increasing the size of the ellipsoid or partitioning the feasible region.

Recently, focusing on the box-constrained case, Tseng [Tse04] produced a global and local convergence analysis of a primal affine scaling algorithm, similar to the ones of [Ye92, BB95], that does not require degeneracy assumptions: under the assumption that H is rank dominant with respect to its maximally-positive-semidefinite principal submatrices, the sequences generated by the algorithm converge globally and linearly to a solution satisfying the first and weak second order optimality conditions.

Few interior-point algorithms have been considered for indefinite quadratic programming that are not of the primal affine-scaling type. For the box-constrained case, a barrier function method was proposed by Dang and Xu [DX00].

In this paper, we propose and analyze two interior-point methods, one of the affine scaling type, the other of the barrier type, for the solution of problem (P). Like the algorithms of [Ye89, Ye92, Tse04, TY02, BB95, CL99, Ye98], they both construct feasible primal iterates (and require an initial feasible point). Strong global and local convergence properties are established for both, and a numerical comparison with the Dikin-type algorithm studied in [Ye92] and [Tse04] (and tested in [Tse04]) is reported, that shows clear promise.

The proposed algorithms do not involve trust regions. Much like in the case of primal-dual algorithms for linear programming, search directions for the "primal" variables and the Karush-Kuhn-Tucker (KKT) multiplier estimates are components of the Newton (or quasi-Newton) direction for the solution of the equalities in the first-order KKT conditions of optimality or a perturbed version of these conditions. (KKT points and KKT multipliers are formally defined in Section 2.) While in the nonlinear programming literature such algorithms are often referred to as primal-dual, mindful of the stricter tradition in the linear/quadratic programming literature, we choose to refer to the proposed schemes as *Newton-KKT*.

Inspired by [Tse04], the present work first focuses on affine scaling. In contrast with [Tse04] though, a Newton-KKT (rather than purely primal) algorithm is considered. It is an improved, affine-scaling version of the barrier-based general nonlinear programming algorithm of [PTH88] and [TWB$^+$03], refined to take advantage of the structure of (P). (A related affine-scaling algorithm was considered in [TZ94] for the case of convex quadratic programming.) Following [TWB$^+$03], in early iterations, the Newton-KKT direction is replaced by a quasi-Newton direction obtained by substituting for H (or the Hessian of the Lagrangian in the general case of [TWB$^+$03]) a carefully chosen matrix $W = H + E$, with E positive semidefinite. The reason for doing this is that, in the absence of convexity, the Newton-KKT system may be singular or, when it is not, may yield a primal direction that is a direction of ascent rather than descent for the objective function. In the present context however, the quadratic programming structure allows for a more efficient computation of W and such computation can even often be skipped (and W reused from the previous iteration). As another enhancement, applicable to general nonlinear programming problems, a simpler update rule than the one used in [PTH88] and [TWB$^+$03] is adopted for the KKT multiplier estimates. Global convergence as well as local q-quadratic convergence of the constructed sequence to KKT points is proved under nondegeneracy assumptions.

While affine scaling algorithms have the advantage of simplicity, it has been observed in various contexts that comparatively faster convergence is often achieved by certain barrier-based interior-point methods. The search direction generated by such algorithms can be thought of as consisting of an affine scaling component and a centering component. When the barrier parameter is set to zero, the centering component vanishes, and the direction reduces to the affine scaling direction. As a second contribution of this paper, we propose a Newton-KKT barrier-based interior-point method for the solution of (P). The proposed algorithm is, again, strongly inspired from [PTH88] and [TWB$^+$03] and indeed, reduces to our Newton-KKT affine scaling algorithm if the rule assigning a value to the barrier-parameter μ^k at each iteration is replaced by the rule $\mu^k := 0$. Apart from modifications to exploit the quadratic programming structure of (P) and from the simplified KKT multiplier estimate update rule mentioned above, the main difference between the proposed algorithm and that of [TWB$^+$03] is that the former uses a scalar barrier parameter, as is done traditionally in interior-point methods, whereas the latter employs a "vector" barrier parameter, i.e., a different barrier parameter value for each constraints. Specifically, in [TWB$^+$03] (and [PTH88]), these values are selected to be proportional to the corresponding components of the current KKT multiplier vector estimate $z > 0$. The proof of superlinear convergence given in [PTH88] (and invoked in [TWB$^+$03]) relies on this selection, specifically in Proposition 4.5 of [PTH88] where it is shown that, under appropriate assumptions, close to KKT points, the full (quasi-)Newton step of one is always accepted by the line search. A secondary contribution of the present paper is to establish local quadratic convergence (in particular, acceptability of a stepsize asymptotically close to one) with a modified, *scalar* barrier parameter: it is proportional to the *smallest* among the components of z. Like for the affine scaling algorithm, global convergence as well as local q-quadratic convergence of the constructed sequence to KKT points is proved under nondegeneracy assumptions.

An additional, side constribution of the present paper is that the analysis is self-contained (while the analysis in [TWB$^+$03] makes references to that in [PTH88], where a different notation is used), and, we believe, significantly more transparent than that in [PTH88] and [TWB$^+$03].

The remainder of the paper is organized as follows. Section 2 includes a full statement of the proposed affine scaling algorithm, and a discussion of its main features. Section 3 is devoted to a careful analysis of the global and local convergence properties of this algorithm. The proposed barrier-type algorithm is stated, discussed and analyzed in Section 4. Implementation issues are considered in Section 5 and numerical experiments are reported in Section 6. Finally, Section 7 is devoted to concluding remarks. Our notation is standard. In particular, $\|\cdot\|$ denotes the Euclidean norm, and $A \succeq B$ indicates that matrix $A - B$ is positive semidefinite (and so does $B \preceq A$).

2 Problem Definition and Algorithm Statement

Let $I = \{1, \cdots, m\}$, where m is the number of rows of A, and, for $i \in I$, let a_i be the *transpose* of the ith *row* of A, let b_i be the ith entry of b, and let $g_i(x) := \langle a_i, x\rangle - b_i$. Also let $f(x) := \frac{1}{2}\langle x, Hx\rangle + \langle c, x\rangle$, and let $\nabla f(x)$ denote its gradient, $Hx + c$. Of course, for any Δx,

$$f(x + \Delta x) = f(x) + \langle \nabla f(x), \Delta x\rangle + \frac{1}{2}\langle \Delta x, H\Delta x\rangle. \tag{1}$$

The feasible set $\mathcal{F}$ is given by

$$\mathcal{F} := \{x \in \mathcal{R}^n : g_i(x) \leq 0 \;\; \forall i \in I\},$$

and the strictly feasible $\mathcal{F}^o$ set by

$$\mathcal{F}^o := \{x \in \mathcal{R}^n : g_i(x) < 0 \;\; \forall i \in I\}.$$

A point $x^* \in \mathcal{F}$ is said to be *stationary*[1] for (P) if there exists an associated *multiplier (vector)* $z^* \in \mathcal{R}^m$ such that

$$\begin{aligned} \nabla f(x^*) + A^{\mathrm{T}} z^* &= 0 \\ z_i^* g_i(x^*) &= 0 \quad \forall i \in I. \end{aligned} \tag{2}$$

(In particular, all vertices of $\mathcal{F}$ are stationary.) If furthermore $z^* \geq 0$, then x^* is a *KKT point* for (P). (z^* is then an associated *KKT multiplier (vector)*, and (x^*, z^*) a *KKT pair.*) Given $x \in \mathcal{F}$, we let $I(x)$ denote the index set of active constraints at x, i.e.

$$I(x) := \{i \in I : \;\; g_i(x) = 0\}.$$

Let (x, z) be an estimate of a KKT pair (x^*, z^*) for (P) and substitute for the left-hand side of (2) its first order expansion around (x, z) evaluated at $(x + \Delta x, z + \Delta z)$, i.e., consider the linear system of equations in $(\Delta x, \Delta z)$

$$\begin{aligned} \nabla f(x) + H\Delta x + A^{\mathrm{T}}(z + \Delta z) &= 0 \\ z_i\langle a_i, \Delta x\rangle + g_i(x)(z_i + \Delta z_i) &= 0 \quad \forall i \in I, \end{aligned} \tag{3}$$

[1] Such terminology has been used by a number of authors in the context of inequality constrained optimization at least as far back as [Her82]. Consistent with its age-old use in unconstrained and equality constrained contexts, with this definition, the term "stationary" applies equally to minimizers and maximizers.

which we refer to as the *Newton-KKT equations*. This system is equivalently written as

$$M(x, z, H) \begin{bmatrix} \Delta x \\ \zeta \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}, \tag{4}$$

where $\zeta := z + \Delta z$ and where, for any given $n \times n$ symmetric matrix W,

$$M(x, z, W) := \begin{bmatrix} W & A^{\mathrm{T}} \\ \mathrm{diag}(z_i)A & \mathrm{diag}(g_i(x)) \end{bmatrix}. \tag{5}$$

It will be shown that, under mild assumptions, and after possible adjustment of H, if $x \in \mathcal{F}^o$ (*strict primal feasibility*) and $z_i > 0$ for all $i \in I$ (*strict "dual feasibility"*), then the solution Δx of (3), if it is nonzero, is a feasible direction which is also a direction of descent for f, a useful property when seeking global convergence to KKT points. Note that a favorable effect of strict primal and dual feasibility is that it implies that $\langle a_i, \Delta x \rangle < 0$ whenever $z_i + \Delta z_i < 0$, so that the iterate will tend to move away from stationary points that are not KKT points.

A pure Newton iteration for the solution of (2) amounts to selecting $x^+ = x + \Delta x$ and $z^+ = z + \Delta z$ as next iterates, where $(\Delta x, \Delta z)$ solves (3). Under appropriate nondegeneracy assumptions, such iteration yields a local q-quadratic rate of convergence in (x, z). However, even close to a solution of (2), this iteration may not preserve primal and dual feasibility. Fortunately, it is possible to define next iterates x^+ and z^+ that are strictly feasible and close enough to $x + \Delta x$ and $z + \Delta z$ that the quadratic convergence of Newton's method is preserved. Such iterates are used in Algorithm A1, which we now state. Note that the algorithm statement implicitly requires that $\mathcal{F}^o$ be nonempty.

Algorithm A1.

Parameters. $\beta \in (0, 1)$, $\underline{z} > 0$, $z_{\mathrm{u}} > 0$, $\sigma > 0$, $\gamma > 1$.

Data. $x^0 \in \mathcal{F}^o$, $z_i^0 > 0 \ \forall i \in I$.

Step 0. Initialization. Set $k := 0$. Set $\overline{I} := \emptyset$. Set $\overline{\alpha}_i := 0$, $i = 1, \ldots, m$. Set $\overline{E} := \mathrm{I}$.[2]

Step 1. Computation of modified Hessian. Set $W^k := H + E^k$, where $E^k \succeq 0$ is computed as follows.

[2] The initial values assigned to the components of $\overline{\alpha}$ and to $\overline{E}$ are immaterial as long as $\overline{E} \neq 0$.

- If $H \succeq \sigma \mathrm{I}$, then set $E^k := 0$.
- Else
  - If $\frac{z_i^k}{|g_i(x^k)|} \leq \overline{\alpha}_i$ for some $i \in \overline{I}$ or ($\overline{E} \neq 0$ and $\overline{I} = \emptyset$) or ($\overline{E} \neq 0$ and $\frac{z_i^k}{|g_i(x^k)|} \geq \gamma^2 \overline{\alpha}_i$ for some $i \in \overline{I}$) then (i) set $\overline{I} := \{i : \frac{z_i^k}{|g_i(x^k)|} \geq 1\}$ and $\overline{\alpha}_i := \frac{1}{\gamma} \frac{z_i^k}{|g_i(x^k)|}$, $i \in \overline{I}$; and (ii)
    - if $H + \sum_{\overline{I}} \overline{\alpha}_i a_i a_i^{\mathrm{T}} \succeq \sigma \mathrm{I}$, set $\overline{E} := 0$;
    - else pick $\overline{E} \succeq 0$, with $\overline{E} \preceq (\|H\|_{\mathrm{F}} + \sigma)\mathrm{I}$, such that $H + \sum_{\overline{I}} \overline{\alpha}_i a_i a_i^{\mathrm{T}} + \overline{E} \succeq \sigma \mathrm{I}$. (Here, $\|H\|_{\mathrm{F}}$ is the Frobenius norm of H.)
  - Set $E^k := \overline{E}$.

Step 2. Computation of a search direction. Let $(\Delta x^k, \zeta^k)$ solve the linear system in $(\Delta x, \zeta)$

$$W^k \Delta x + A^{\mathrm{T}} \zeta = -\nabla f(x^k) \tag{6a}$$

$$z_i^k \langle a_i, \Delta x \rangle + g_i(x^k) \zeta_i = 0 \quad \forall i \in I. \tag{6b}$$

If $\Delta x^k = 0$, stop.

Step 3. Updates.
(i) Set

$$\overline{t}^k := \begin{cases} \infty & \text{if } \langle a_i, \Delta x^k \rangle \leq 0 \;\; \forall i \in I, \\ \min \left\{ \frac{|g_i(x^k)|}{\langle a_i, \Delta x^k \rangle} : \langle a_i, \Delta x^k \rangle > 0, i \in I \right\} & \text{otherwise.} \end{cases} \tag{7}$$

Set

$$t^k := \min \left\{ \max\{\beta \overline{t}^k, \overline{t}^k - \|\Delta x^k\|\},\ 1 \right\}. \tag{8}$$

Set $x^{k+1} := x^k + t^k \Delta x^k$.
(ii) Set $\left(\zeta_-^k\right)_i := \min\{\zeta_i^k, 0\}$, $\forall i \in I$. Set

$$z_i^{k+1} := \min\left\{\max\left\{\min\{\|\Delta x^k\|^2 + \|\zeta_-^k\|^2, \underline{z}\}, \zeta_i^k\right\},\ z_{\mathrm{u}}\right\},\ \forall i \in I. \tag{9}$$

(iii) Set $k := k + 1$. Go to Step 1.

□

Let S^k denote the Schur complement of $\mathrm{diag}(g_i(x^k))$ in $M(x^k, z^k, W^k)$, i.e., since $x^k \in \mathcal{F}^o$,

$$S^k := W^k + \sum_{i=1}^{m} \frac{z_i^k}{|g_i(x^k)|} a_i a_i^{\mathrm{T}}. \tag{10}$$

Step 1 in Algorithm A1 ensures that $S^k \succeq \sigma\mathrm{I}$, as we now explain. The case when $H \succeq \sigma\mathrm{I}$ is clear. Assume that $H \not\succeq \sigma\mathrm{I}$. Then, after completion of Step 1, the relation $W^k + \sum_{\overline{I}} \overline{\alpha}_i a_i a_i^{\mathrm{T}} \succeq \sigma\mathrm{I}$ is satisfied, and $\frac{z_i^k}{|g_i(x^k)|} \geq \overline{\alpha}_i$ for all $i \in \overline{I}$. Consequently, $S^k \succeq W^k + \sum_{\overline{I}} \overline{\alpha}_i a_i a_i^{\mathrm{T}} \succeq \sigma\mathrm{I}$, hence the desired conclusion. Next, under appropriate assumptions, $\{E^k\}$ is eventually zero (see Lemma 3.12 below) and Newton-related superlinear convergence can set in. Step 1 also ensures that $\{W^k\}$ is bounded, since $W^k \preceq H + (\|H\|_{\mathrm{F}} + \sigma)\mathrm{I}$ for all k. (Note that, since it always holds that $H + \|H\|_{\mathrm{F}}\mathrm{I} \succeq 0$, the conditions imposed on $\overline{E}$ can always be met.) Further, an interesting feature of Step 1 is that $\overline{E}$ need not be updated at each step: update occurs in particular if one of the ratios $\frac{z_i^k}{|g_i(x^k)|}$, $i \in \overline{I}$, leaves the interval $(\overline{\alpha}_i, \gamma^2\overline{\alpha}_i)$ through its lower end, or if $\overline{E} \neq 0$ and one of these ratios leave the corresponding interval through its upper end. Finally, while Step 1 leaves open the option to always select the Hessian correction E^k to be either 0 or $(\|H\|_{\mathrm{F}} + \sigma)\mathrm{I}$, a "small", intermediate value is of course preferable when 0 is not allowed, so Newton's method is approximated more closely.

In Step 3(i), borrowed from [TZ94], $\bar{t}^k$ is the maximum step preserving primal feasibility ($x^k + \bar{t}^k \Delta x^k \in \mathcal{F}$) and the term $\beta\bar{t}^k$ ensures that t^k is positive even when $\|\Delta x^k\|$ is large. It will be shown that, close to a solution of (P), t^k is close enough to 1 for quadratic convergence to take place.

Step 3(ii) is partly new, though strongly inspired from the multiplier update rule used in [PTH88, TZ94, TWB$^+$03]. In contrast to usual practice, z_i^{k+1} is not obtained by a step in the direction of $\zeta^k - z^k$. The reason for this is that the lower bound $\|\Delta x^k\|^2 + \|\zeta_-^k\|^2$ in (9) is key to our global convergence analysis, in that it forces the components of z^k to remain away from zero, which will be shown, in turn, to force $M(x^k, z^k, W^k)$ to remain away from singularity, unless a KKT point is approached—$\|\Delta x^k\|^2 + \|\zeta_-^k\|^2$ goes to zero near such points. (The reason for the "square" is that—together with the bound $\bar{t}^k - \|\Delta x^k\|$ in (8)—it allows for local quadratic convergence to take place.) Updating z^k by means of a step in direction $\zeta^k - z^k$ would preclude enforcement of such lower bound. Indeed, for some components, the lower bound might unduly reduce the step size, and may even necessitate a "backwards" step.

Like in [PTH88, TZ94, TWB$^+$03], the upper bound z_{u} is needed in our global convergence analysis (though in our experience not needed in practice); note however that the analysis guarantees appropriate convergence of ζ^k regardless of the value of z_{u} (see Proposition 3.11). (In practice, z_{u} should be set to a "large" value.) As for $\underline{z}$, it was used in the numerical experiments in [TWB$^+$03] (see second bullet at the bottom of page 191 in that paper) but was not included in the formal statement of the algorithm there. Removing it from (9) (i.e., setting it to $+\infty$) would not affect the theoretical convergence properties of the algorithm. However, allowing small values of z_i^{k+1} even when $\|\Delta x^k\|^2 + \|\zeta_-^k\|^2$ is large proved beneficial in practice, especially in early iterations. (Accordingly, $\underline{z}$ should be set to a "small" value.)

The z update adopted in (9) has a significant difference from that used in [PTH88, TZ94, TWB$^+$03] however, in that, in the latter, the lower bound involves only $\|\Delta x^k\|^2$. In itself such lower bound does not prevent convergence to non-KKT stationary points—where $\|\Delta x^k\|$ goes to zero but ζ^k has some significantly negative components, so that ζ_-^k does not go to zero. Accordingly, additional safeguards (less "natural", we feel) were then incorporated in [PTH88, TZ94, TWB$^+$03].

Finally, one may wonder why the arguably even more "natural" lower bound $\|\Delta x^k\|^2 + \|\Delta z^k\|^2$ is not used in (9). Indeed, just like ζ_-^k, $\|\Delta z^k\|$ should go to zero when KKT points are approached and (since z^k is forced to remain in the positive orthant throughout) it is bounded away from zero in the neighborhood of non-KKT stationary points. The reason we chose to use $\|\zeta_-^k\|^2$ is that the proofs are then somewhat simpler, in particular that of part (iii) of Proposition 3.11,[3] and that the numerical results we obtained were essentially identical under both schemes.

It is readily checked that the unique solution of (6a)-(6b) is given by

$$\Delta x^k = -(S^k)^{-1}\nabla f(x^k) \tag{11a}$$

$$\zeta^k = -\text{diag}\left(\frac{z_i^k}{g_i(x^k)}\right) A\Delta x^k. \tag{11b}$$

Expression (11a) shows that Algorithm A1 belongs to the affine scaling family. Since $S^k \succeq \sigma \mathrm{I}$, it also follows from (11a) that

$$\langle \nabla f(x^k), \Delta x^k\rangle \leq -\sigma\|\Delta x^k\|^2 \tag{12}$$

which shows that Δx^k is a direction of descent for f at x^k. Since $z_i^k/g_i(x^k) < 0$

[3] In fact, when $\|\Delta z^k\|^2$ is used, Proposition 3.11(iii) no longer holds as stated but, under the condition that $\underline{z} < 1/m$, z^k still converges to z^* when all components of z^* are less than z_{u}.

for all i and k, it follows from (11b) that

$$\langle \zeta^k, A\Delta x^k \rangle \geq 0 \tag{13}$$

for all k.

Next we establish that Algorithm A1 is well defined and only stops at unconstrained KKT points.

Proposition 2.1 *Algorithm A1 constructs an infinite sequence* $\{x^k\}$ *unless the stopping criterion in Step 2 (*$\Delta x^k = 0$*) is satisfied at some iteration* k_0*. In the latter case,* $\nabla f(x^{k_0}) = 0$*, hence* x^{k_0} *is a KKT point.*

Proof. The computations in Step 3(i) of Algorithm A1 ensure that every constructed x^{k+1} belongs to $\mathcal{F}^o$, so that $\text{diag}(g_i(x^k))$ is nonsingular for all k such that Step 3 is executed at iteration $k-1$ (as well as for $k=0$). Since its Schur complement S^k also is nonsingular for all such k, it follows that $M(x^k, z^k, W^k)$ is nonsingular for all such k, thus that (6) has a unique solution whenever Step 2 of Algorithm A1 is attempted. Since it is clear that all other operations performed are well defined, the entire algorithm is well defined and can only terminate when the stopping criterion in Step 2 is satisfied, say at x_{k_0}. In such case, since $g_i(x^{k_0}) < 0$, substitution of $\Delta x^{k_0} = 0$ in (6) yields $\zeta^{k_0} = 0$ and $\nabla f(x^{k_0}) = 0$, i.e., x^{k_0} is a KKT point (indeed, an unconstrained KKT point). □

The next section is devoted to analyzing the sequences constructed by Algorithm A1 in the case where the stopping criterion is never satisfied. Before proceeding with this analysis, we conclude this section with three lemmas which will be of repeated use. The first lemma further characterizes the relationship between Δx vanishing and x being stationary.

Lemma 2.2 *Let* $(\Delta x, \zeta)$ *satisfy (4) for some* x*,* z*, and with* H *replaced by* W*, for some symmetric* W*. Then (i) if* $x \in \mathcal{F}$ *and* $\Delta x = 0$*, then* x *is stationary for* (P) *and* ζ *is an associated multiplier vector; and (ii) if* $x \in \mathcal{F}$ *is stationary for* (P) *and* $M(x, z, W)$ *is nonsingular, then* $\Delta x = 0$ *and* ζ *is the unique multiplier vector associated with* x*.*

Proof. To prove the first claim, simply substitute $\Delta x = 0$ into (4). Concerning the second claim, let η be a multiplier vector associated with stationary point x. It is readily verified that $(0, \eta)$ satisfies (4). The claim then follows from nonsingularity of $M(x, z, W)$. □

Conditions guaranteeing the nonsingularity required in (ii) above are established in the next lemma, borrowed from [TWB$^+$03, Lemma PTH-3.1*].

Lemma 2.3 [4] *Let $x \in \mathcal{F}$ be such that $\{a_i : i \in I(x)\}$ is a linearly independent set and let $z \geq 0$, with $z_i > 0$ for all $i \in I(x)$. Suppose W satisfies*

$$\left\langle v, \left(W + \sum_{i \notin I(x)} \frac{z_i}{|g_i(x)|} a_i a_i^{\mathrm{T}} \right) v \right\rangle > 0 \ \forall v \in \mathcal{T}(x) \setminus \{0\},$$

where

$$\mathcal{T}(x) := \{v : \langle a_i, v \rangle = 0 \ \forall i \in I(x)\}.$$

Then $M(x, z, W)$ is nonsingular.

The final lemma builds on Lemma 2.3 to show that $M(x, z, W)$ is nonsingular at all accumulation points of certain sequences. It is a simplified version of the first portion of Lemma PTH-3.5* in [TWB$^+$03], and is reproduced here with proof for ease of reference. It relies on a linear independence assumption.

Assumption 1 (linear independence constraint qualification) *For all $x \in \mathcal{F}$, $\{a_i : i \in I(x)\}$ is a linearly independent set.*

Lemma 2.4 [5] *Suppose Assumption 1 holds. Let $\{x^k\}$, $\{z^k\}$, and $\{W^k\}$ be arbitrary infinite sequences such that $\{x^k\}$ converges to x^*, $\{z^k\}$ to z^*, and $\{W^k\}$ to W^*, for some x^*, z^*, and W^*. Suppose that $g(x^k) < 0$ for all k, that $z^k > 0$ for all k, that S^k defined by* (10) *satisfies $S^k \succeq \sigma \mathrm{I}$ for all k and that $z_j^* > 0$ for all $j \in I(x^*)$. Then $M(x^*, z^*, W^*)$ is nonsingular.*

Proof. It suffices to show that (x^*, z^*, W^*) satisfies the assumptions of Lemma 2.3. Thus let $v \neq 0$ be such that

$$\langle a_i, v \rangle = 0 \quad \forall i \in I(x^*). \tag{14}$$

It then follows from (10) and positive semidefiniteness of $S^k - \sigma \mathrm{I}$, by adding terms that vanish in view of (14), that for all k

$$\left\langle v, \left(W^k + \sum_{i \notin I(x^*)} \frac{z_i^k}{|g_i(x_k)|} a_i a_i^{\mathrm{T}} \right) v \right\rangle \geq \sigma \|v\|^2.$$

[4] It is readily checked that the result still holds if $z \geq 0$ is omitted and $z_i > 0$ is replaced by $z_i \neq 0$ in the statement. Only the case $z_i > 0$ is needed in this paper, though.

[5] Much as Lemma 2.3, Lemma 2.4 holds under weaker hypotheses, but the stated version is adequate for the purpose of this paper.

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

$$\left\langle v, \left(W^* + \sum_{i \notin I(x^*)} \frac{z_i^*}{|g_i(x^*)|} a_i a_i^{\mathrm{T}} \right) v \right\rangle \geq \sigma \|v\|^2 > 0.$$

The proof is complete.

3 Convergence Analysis

3.1 Global Convergence

We first show (Proposition 3.5) that, under Assumption 1, the accumulation points of $\{x^k\}$ are stationary for (P). Then, under the additional assumption that stationary points are isolated, we show (Theorem 3.9) that these accumulation points are KKT for (P).

First of all, at every iteration, the values of the objective function and of all constraint functions with negative multiplier estimates decrease. This is established in our first proposition.

Proposition 3.1 *Let $\{x^k\}$, $\{\Delta x^k\}$, and $\{\zeta^k\}$ be as constructed by Algorithm A1. Suppose $\Delta x^k \neq 0$.*
(i) If $\langle \Delta x^k, H \Delta x^k \rangle \leq 0$, then

$$f(x^k + t\Delta x^k) < f(x^k) \quad \forall t > 0, \tag{15}$$

(ii) If $\langle \Delta x^k, H \Delta x^k \rangle > 0$, then

$$f(x^k + t\Delta x^k) \leq f(x^k) + \frac{t}{2} \langle \nabla f(x^k), \Delta x^k \rangle < f(x^k) \quad \forall t \in (0, 1], \tag{16}$$

(iii)

$$g_i(x^k + t\Delta x^k) = g_i(x^k) + t\langle a_i, \Delta x^k \rangle < g_i(x^k) \quad \forall t > 0, \quad \forall i \text{ s.t. } \zeta_i^k < 0. \tag{17}$$

Proof. Claim (i) follows directly from (12). To prove claim (ii) observe that (1) yields

$$f(x^k + t\Delta x^k) = f(x^k) + t \left(\langle \nabla f(x^k), \Delta x^k \rangle + \frac{t}{2} \langle \Delta x^k, H \Delta x^k \rangle \right), \tag{18}$$

and that (6a) and (13) yield

$$\langle \Delta x^k, W^k \Delta x^k \rangle \leq -\langle \nabla f(x^k), \Delta x^k \rangle$$

which, since $H \preceq W^k$, yields

$$\frac{t}{2}\langle \Delta x^k, H\Delta x^k \rangle \leq \frac{1}{2}\langle \Delta x^k, H\Delta x^k \rangle \leq -\frac{1}{2}\langle \nabla f(x^k), \Delta x^k \rangle \quad \forall t \in [0,1].$$

Substituting into (18) yields the left-hand inequality in (16). Since $\Delta x^k \neq 0$, the right-hand inequality in (16) follows as a direct consequence of (12), completing the proof of claim (ii). Finally, since g is linear,

$$g_i(x^k + t^k \Delta x^k) = g_i(x^k) + t^k \langle a_i, \Delta x^k \rangle, \quad i = 1, \ldots, m.$$

Since $z_i^k > 0$ and $g_i(x^k) < 0$ for all $i \in I$, it follows from (6b) that $\langle a_i, \Delta x^k \rangle < 0$ whenever $\zeta_i^k < 0$, proving claim (iii). □

With these results in hand, we now proceed to show that the accumulation points of $\{x^k\}$ are stationary points for (P). The argument is a simplified version of that used in [PTH88]. It is given here for ease of reference. The rationale is as follows: Given an infinite index set K such that $\{x_k\}_{k \in K} \to x^*$ for some x^*, either (i) $\{\Delta x^k\}_{k \in K} \to 0$, or (ii) there exists an infinite index set $K' \subseteq K$ such that $\{\Delta x^{k-1}\}_{k \in K'} \to 0$ (Lemma 3.2). In case (i), it follows from Lemma 3.3 below that x^* is stationary. In case (ii), $\{x^{k-1}\}_{k \in K'} \to x^*$ (Lemma 3.4) and it again follows from Lemma 3.3 that x^* is stationary. Based on these results, stationarity of x^* is then established in Proposition 3.5. The lemmas prove somewhat more than is needed here, to be used in the proof of convergence to KKT points and in the local convergence analysis of the next subsection.

Lemma 3.2 *Let $\{x^k\}$, $\{\Delta x^k\}$, and $\{\zeta_-^k\}$ be as constructed by Algorithm A1. Suppose Assumption 1 holds. Let K be an infinite index set such that $\{x^k\}_{k \in K} \to x^*$ for some x^* and*

$$\inf\{\|\Delta x^{k-1}\|^2 + \|\zeta_-^{k-1}\|^2 : \; k \in K\} > 0. \tag{19}$$

Then $\Delta x^k \to 0$ as $k \to \infty$, $k \in K$.

Proof. Proceeding by contradiction, assume that $\{x^k\}_{k \in K} \to x^*$, that (19) holds, and that, for some infinite index set $K' \subseteq K$, $\inf\limits_{k \in K'} ||\Delta x^k|| > 0$. Since

$\{z^k\}$ and $\{W^k\}$ are bounded, we may assume, without loss of generality, that for some z^* and W^*,

$$\begin{aligned}\{z^k\} &\to z^* \quad \text{as} \quad k \to \infty, \quad k \in K',\\ \{W^k\} &\to W^* \quad \text{as} \quad k \to \infty, \quad k \in K'.\end{aligned}$$

In view of (19) and (9), it follows that $z^* > 0$. Since in view of Lemma 2.4 $M(x^*, z^*, W^*)$ is nonsingular, it follows that $\{\Delta x^k\}_{k\in K'}$ and $\{\zeta^k\}_{k\in K'}$ are bounded and that, for some Δx^* with $\Delta x^* \neq 0$ (since $\inf_{k\in K'} \|\Delta x^k\| > 0$),

$$\{\Delta x^k\} \to \Delta x^* \quad \text{as} \quad k \to \infty, \quad k \in K'.$$

On the other hand, it follows from (6b) and (7) that

$$\bar{t}^k = \begin{cases} \infty & \text{if } \zeta_i^k \leq 0 \ \ \forall i \in I, \\ \min\{(z_i^k/\zeta_i^k) : \zeta_i^k > 0, \ i \in I\} & \text{otherwise.} \end{cases}$$

Since, on K', $\{\zeta^k\}$ is bounded and, for each $i \in I$, $\{z_i^k\}$ is bounded away from zero, it follows that $\{\bar{t}^k\}$ is bounded away from zero on K', and that so is $\{t^k\}$ (Step 3 (i) in Algorithm A1); say, $t^k > \underline{t}$ for all $k \in K'$, with $\underline{t} \in (0, 1)$.

To complete the proof by contradiction, we show that the above implies that $f(x^k) \to -\infty$ as $k \to \infty$, contradicting convergence of $\{x^k\}_{k\in K}$. Since $\{f(x^k)\}$ is nonincreasing (Proposition 3.1) it suffices to show that for some $\delta > 0$,

$$f(x^{k+1}) \leq f(x^k) - \delta \tag{20}$$

infinitely many times. We show that (20) holds for all $k \in K'$. We consider two cases. First, if $\langle \Delta x^*, H\Delta x^*\rangle > 0$, then since $t^k \in (\underline{t}, 1]$ for all k, the claim follows from Proposition 3.1(ii) and (12). Suppose now that $\langle \Delta x^*, H\Delta x^*\rangle \leq 0$. Then $\langle \Delta x^k, H\Delta x^k\rangle \leq \frac{1}{2}\sigma\|\Delta x^*\|^2$ for $k \in K'$, k large enough. Also, in view of (12), $\langle \nabla f(x^k), \Delta x^k\rangle \leq -\frac{1}{2}\sigma\|\Delta x^*\|^2$ for $k \in K'$, k large enough. Since $t_k \in (\underline{t}, 1]$ for $k \in K'$, it follows from (1) that (20) again holds on K', with $\delta = \frac{1}{4}\underline{t}\sigma\|\Delta x^*\|^2$. □

Lemma 3.3 *Let $\{x^k\}$, $\{\Delta x^k\}$, and $\{\zeta^k\}$ be as constructed by Algorithm A1. Suppose Assumption 1 holds. Let x^* be such that, for some infinite index set K, $\{x^k\}$ converges to x^* on K. If $\{\Delta x^k\}$ converges to zero on K, then x^* is stationary and $\{\zeta^k\}$ converges to z^* on K, where z^* is the unique multiplier vector associated with x^*.*

Proof. Suppose $\Delta x^k \to 0$ as $k \to \infty,\ k \in K$. Since $\{z^k\}$ is bounded, it follows from (6b) that $\zeta_i^k \to 0$ as $k \to \infty$, $k \in K$, for all $i \notin I(x^*)$. Since in view of (6a) and boundedness of W^k, $\{A^T\zeta^k\}$ converges on K, it follows from Assumption 1 that $\{\zeta^k\}$ converges on K, say to z^*. Taking limits in (6a)–(6b) as $k \to \infty$, $k \in K$, and using the fact that $\{z^k\}$ is bounded on K yields

$$\nabla f(x^*) + A^{\mathrm{T}} z^* = 0$$

$$z_i^* g_i(x^*) = 0,\ i = 1, \ldots, m,$$

implying that x^* is stationary, with multiplier vector z^*. The multiplier is unique because of Assumption 1. □

The next lemma is a direct consequence of the fact that, by construction, $x^{k+1} = x^k + t^k \Delta x^k$, with $t^k \in (0, 1]$.

Lemma 3.4 *Let $\{x^k\}$ and $\{\Delta x^k\}$ be as constructed by Algorithm A1. Let K be an infinite index set such that $\{x^k\}_{k\in K} \to x^*$ for some x^*. If $\{\Delta x^{k-1}\}_{k\in K} \to 0$ then $\{x^{k-1}\}_{k\in K} \to x^*$.*

In view of the rationale given above, the following holds.

Proposition 3.5 *Under Assumption 1, every accumulation point of $\{x^k\}$ constructed by Algorithm A1 is a stationary point for (P).*

In the remainder of this section we show that, if the stationary points are isolated, then the accumulation points of $\{x^k\}$ are KKT points for (P).

Lemma 3.6 *Let $\{x^k\}$, $\{\Delta x^k\}$, and $\{\zeta_-^k\}$ be as constructed by Algorithm A1. Suppose Assumption 1 holds. Suppose that K, an infinite index set, is such that, for some x^*, $\{x^k\}_{k\in K}$ tends to x^*, and $\{\Delta x^{k-1}\}_{k\in K}$ and $\{\zeta_-^{k-1}\}_{k\in K}$ tend to zero. Then x^* is a KKT point.*

Proof. In view of Lemma 3.4, $\{x^{k-1}\}_{k\in K} \to x^*$, and, in view of Lemma 3.3, $\{\zeta^{k-1}\}_{k\in K}$ converges to z^*, the multiplier vector associated with stationary point x^*. Since $\{\zeta_-^{k-1}\}_{k\in K}$ tends to zero, it follows that $z^* \geq 0$, thus that x^* is a KKT point. □

The remainder of the proof is essentially identical to the proof of Theorem 3.11 in [PTH88]. It is reproduced here for the reader's ease of reference. For clarity, part of the proof is given as Lemmas 3.7 and 3.8.

Lemma 3.7 *Let $\{x^k\}$ and $\{\Delta x^k\}$ be as constructed by Algorithm A1. Suppose Assumption 1 holds. Suppose that K, an infinite index set, is such that, for some x^*, $\{x^k\}_{k\in K} \to x^*$, and x^* is not KKT. Then $\{\Delta x^k\}_{k\in K} \to 0$.*

Proof. Proceeding by contradiction, suppose $\{\Delta x^k\}_{k\in K}$ does not tend to zero. It then follows from Lemma 3.2 that

$$\inf\{\|\Delta x^{k-1}\|^2 + \|\zeta_-^{k-1}\|^2 : \ k \in K\} = 0.$$

Lemma 3.6 then implies that x^* is KKT, a contradiction. $\square$

Assumption 2 *All stationary points are isolated.*

Given $r > 0$ and x in some Euclidean space, let $\overline{B}(x^*, r)$ denote the closed ball $\{x : \|x - x^*\| \leq r\}$.

Lemma 3.8 *Let $\{x^k\}$ be as constructed by Algorithm A1. Suppose Assumptions 1 and 2 hold. Suppose $\{x^k\}$ has x^* as an accumulation point, and x^* is not KKT. Then the entire sequence $\{x^k\}$ converges to x^*.*

Proof. In view of Proposition 3.5, x^* is stationary. In view of Assumption 2, there exists $\epsilon > 0$ such that the closed ball $\overline{B}(x^*, 2\epsilon)$ does not contain any other stationary point than x^* and hence, in view of Proposition 3.5, does not contain any other accumulation point than x^*. We show that $x^k \in \overline{B}(x^*, \epsilon)$ for all k large enough, i.e., that $\{x^k\}$ converges to x^*. Proceeding by contradiction, suppose that, for some infinite index set K, $x^k \in \overline{B}(x^*, \epsilon)$ but $x^{k+1} \notin \overline{B}(x^*, \epsilon)$. Since $\{x^k\}_{k\in K}$ converges to x^*, it follows from Lemma 3.7 that $\{\Delta x^k\}_{k\in K}$ tends to zero. Since, in view of (8) and the update formula for x^k in Step3(i) of Algorithm A1, $\|x^{k+1} - x^k\| \leq \|\Delta x^k\|$ for all k, it follows that $x^{k+1} \in \overline{B}(x^*, 2\epsilon) \setminus \overline{B}(x^*, \epsilon)$ for all $k \in K$, k large enough, contradicting the fact that x^* is the only accumulation point in $\overline{B}(x^*, 2\epsilon)$. $\square$

Theorem 3.9 *Under Assumptions 1 and 2, every accumulation point of $\{x^k\}$ constructed by Algorithm A1 is a KKT point.*

Proof. Let x^* be an accumulation point of $\{x^k\}$. Proceeding by contradiction, suppose that x^* is not KKT. In view of Lemma 3.8, the entire sequence $\{x^k\}$ converges to x^*. In view of Lemma 3.7, it follows that $\{\Delta x^k\}$ tends to zero, and from Lemma 3.3 that $\{\zeta^k\}$ converges to z^*, the multiplier vector associated with stationary point x^*. Further, since x^* is not KKT, there exists $i_0 \in I(x^*)$ such that $z^*_{i_0} < 0$. In view of (17), it follows that, for k_0 large enough,

$$g_{i_0}(x^k) \leq g_{i_0}(x^{k-1} \leq \cdots \leq g_{i_0}(x^{k_0+1}) \leq g_{i_0}(x^{k_0}) < 0,$$

contradicting the fact that $g_{i_0}(x^*) = 0$ (since $i_0 \in I(x^*)$). $\square$

3.2 Local Rate of Convergence

We will now assume that some accumulation point of $\{x^k\}$ enjoys certain additional properties. First we will assume that strict complementarity holds at one such point, which will imply that the entire sequence $\{x^k\}$ converges to that point and that sequences $\{\Delta x^k\}$, $\{\zeta^k\}$, and $\{z^k\}$ converge appropriately as well. Then we will show that if, in addition, the second order sufficiency condition of optimality holds at that point, then convergence is q-quadratic.

Assumption 3 (strict complementarity) *The sequence $\{x^k\}$ generated by Algorithm A1 has an accumulation point x^* with associated multiplier vector z^* satisfying $z_i^* > 0$ for all $i \in I(x^*)$.*

We first show that the entire sequence $\{x^k\}$ converges to x^*. The proof makes use of the following lemma, adapted from [PTH88, Lemma 4.1] and [BT03, Lemma 9].

Lemma 3.10 *Let $\{x^k\}$ and $\{\Delta x^k\}$ be as constructed by Algorithm A1. Suppose Assumptions 1 and 3 hold. Let K be an infinite index set such that $\{x^k\}_{k\in K} \to x^*$. 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 3.2, there exists an infinite index set $K'' \subseteq K'$ such that $\{\Delta x^{k-1}\}_{k\in K''}$ and $\{\zeta_-^{k-1}\}_{k\in K''}$ go to zero. It follows from Lemma 3.4 that $\{x^{k-1}\}_{k\in K''} \to x^*$. In view of Lemma 3.3 it follows that $\{\zeta^{k-1}\}_{k\in K''} \to z^*$ where z^* is the multiplier vector associated to x^*. In view of Assumption 3, $z^* \geq 0$. It then follows from (9) that, for all j, $\{z_j^k\}_{k\in K''} \to \hat{z}_j := \min\{z_j^*, z_{\rm u}\}$ which, in view of Assumption 3 is positive for all $j \in I(x^*)$. Also, since $\{W^k\}$ is bounded there is no loss of generality in assuming that, for some W^*, $\{W^k\}_{k\in K''}$ converges to W^*. In view of Lemma 2.4 it then follows that $M(x^*, \hat{z}, W^*)$ is nonsingular. Since x^* is stationary (indeed, KKT) it follows from Lemma 2.2 that $\{\Delta x^k\}_{k\in K''}$ goes to zero, a contradiction. □

Convergence of the entire sequence $\{x^k\}$ to x^* can now be proved. The following proposition is adapted from [PTH88, Proposition 4.2].

Proposition 3.11 *Let $\{x^k\}$, $\{\Delta x^k\}$, $\{z^k\}$, and $\{\zeta^k\}$ be as constructed by Algorithm A1. Suppose Assumptions 1, 2 and 3 hold. Then the entire sequence $\{x^k\}$ converges to x^*. Moreover, (i) $\{\Delta x^k\} \to 0$, (ii) $\{\zeta^k\} \to z^*$, and (iii) $\{z_j^k\} \to \min\{z_j^*, z_{\rm u}\}$ for all j.*

Proof. Consider a closed ball of radius $\epsilon > 0$ about x^* where there is no KKT point other than x^* (in view of Assumption 2 such ϵ exists). Proceeding by contradiction to prove the first claim, suppose without loss of generality that the sequence $\{x^k\}$ leaves the ball infinitely many times. Consider the infinite subsequence $\{x^k\}_{k\in K}$ of points such that x^k is in the ball and x^{k+1} is out of the ball. Then $\{x^k\}_{k\in K} \to x^*$, otherwise the closed ϵ-ball would contain an accumulation point other than x^* and this point would be a KKT point by Theorem 3.9. In particular, $\|x^k - x^*\| < \epsilon/4$ for all $k \in K$ large enough. On the other hand, it follows from Lemma 3.10 that $\|\Delta x^k\| < \epsilon/4$ for all $k \in K$ large enough. Consequently, for all $k \in K$ large enough, we have

$$\|x^{k+1} - x^*\| \leq \|x^{k+1} - x^k\| + \|x^k - x^*\| \leq \|\Delta x^k\| + \|x^k - x^*\| \leq \epsilon/2.$$

That is, x^{k+1} is in the ϵ-ball for $k \in K$ large enough, a contradiction. Thus, the first claim holds. Claim (i) then follows from Lemma 3.10 and claim (ii) follows from Lemma 3.3. Finally claim (iii) follows from a careful inspection of (9). □

In order for quadratic convergence to set in, it is desirable that $\{W^k\}$ converge to H, at least in the tangent plane of the active constraints at x^*. For this to be possible with $M(x^k, z^k, W^k)$ still remaining bounded away from singularity, we must assume that x^* is in fact a minimizer.

Assumption 4 *x^* is a local (or global) minimizer.*

Since $f(x^k)$ is reduced at each iteration, this assumption is rather mild: points x^* that are not local minimizers are unstable under perturbations.

Assumptions 2 and 4 imply that a second order sufficiency condition of optimality holds, specifically (under Assumption 3) that

$$\langle \Delta x, H\Delta x\rangle > 0 \text{ for all } \Delta x \neq 0 \text{ such that } \langle a_i, \Delta x\rangle = 0\ \forall i \in I(x^*). \tag{21}$$

In turn, under Assumption 3, (21) implies Assumption 4.

Lemma 3.12 *Let $\{W^k\}$ be as constructed by Algorithm A1. Under Assumptions 1–4, if σ in Algorithm A1 is such that $\sigma < \min\{\langle v, Hv\rangle : \langle a_i, v\rangle = 0\ \forall i \in I(x^*),\ \|v\| = 1\}$, then $W^k = H$ for all k sufficiently large.*

Proof. It follows from Assumption 3 and Proposition 3.11(iii) that $\{z_j^k/|g_j(x^k)|\}$ goes to ∞ for all $j \in I(x^*)$. It then follows from (21) and Step 1 in Algorithm A1 that $E^k = 0$ for all k large enough. □

To prove q-quadratic convergence of $\{(x^k, z^k)\}$, the following property of Newton's method, taken from [TZ94], will be used.

Proposition 3.13 *Let $F : \mathcal{R}^n \to \mathcal{R}^n$ be twice continuously differentiable and let $w^* \in \mathcal{R}^n$. Let $\rho > 0$ be such that $F(w^*) = 0$ and $\frac{\partial F}{\partial w}(w)$ is invertible whenever $w \in \overline{B}(w^*, \rho)$. Let $\delta^{\mathrm{N}} : \overline{B}(w^*, \rho) \to \mathcal{R}^n$ be the Newton increment $\delta^{\mathrm{N}}(w) := -\left(\frac{\partial F}{\partial w}(w)\right)^{-1} F(w)$. Then given any $c_1 > 0$ there exists $c_2 > 0$ such that the following statement holds:*

> *For every $w \in \overline{B}(w^*, \rho)$ and every $w^+ \in \mathcal{R}^n$ for which, for each $i \in \{1, \ldots, n\}$, either*
>
> *(i)* $|w_i^+ - w_i^*| \leq c_1 \|\delta^{\mathrm{N}}(w)\|^2$
>
> *or*
>
> *(ii)* $|w_i^+ - (w_i + \delta_i^{\mathrm{N}}(w))| \leq c_1 \|\delta^{\mathrm{N}}(w)\|^2,$
>
> *it holds that*
> $$\|w^+ - w^*\| \leq c_2 \|w - w^*\|^2. \tag{22}$$

Q-quadratic convergence of the sequence $\{(x^k, z^k)\}$ follows. The proof is essentially identical to that of [TZ94, Theorem 3.11] and is reproduced here for ease of reference.

Theorem 3.14 *Let $\{x^k\}$ and $\{z^k\}$ be as constructed by Algorithm A1. Suppose Assumptions 1–4 hold. Then, if $z_i^* < z_{\mathrm{u}}\ \forall i \in I$ and $\sigma < \min\{\langle v, Hv\rangle : \langle a_i, v\rangle = 0\ \forall i \in I(x^*),\ \|v\| = 1\}$, $\{(x^k, z^k)\}$ converges to (x^*, z^*) q-quadratically.*

Proof. We establish that the conditions in Proposition 3.13 hold with $F : R^n \times R^m \to R^n \times R^m$ given by

$$F(x, z) := \begin{bmatrix} \nabla f(x) + A^{\mathrm{T}} z \\ z_1 g_1(x) \\ \vdots \\ z_m g_m(x) \end{bmatrix}.$$

Clearly, the Jacobian of F at (x, z) is $M(x, z, H)$ and $(\Delta x, \Delta z)$, with $(\Delta x, z + \Delta z)$ solution of (4), is the Newton direction for the solution of $F(x, z) = 0$. With reference to Proposition 3.13, let $\rho > 0$ be such that $M(x, z, H)$ is nonsingular for all $(x, z) \in \overline{B}((x^*, z^*), \rho)$. (In view of (21), Assumption 3, and Lemma 2.3, such ρ exists.) Since, in view of the stated assumption on z^* and of Proposition 3.11, $\{(x^k, z^k)\} \to (x^*, z^*)$, it follows from Lemma 3.12 that there exists k_0 such that, for all $k \geq k_0$, $(x^k, z^k) \in \overline{B}((x^*, z^*), \rho)$ and $W^k = H$.

Now, with the aim of verifying conditions (i)/(ii) of Proposition 3.13 along $\{(x^k, z^k)\}$ with F as specified above, let us first consider $\{z^k\}$, updated in Step 3(ii) of Algorithm A1. Since $z_i^* < z_{\rm u}$, if follows that, for $i \in I(x^*)$ (so $z_i^* > 0$), z_i^{k+1} is equal to ζ_i^k for k large enough, so that condition (ii) in Proposition 3.13 holds (with any $c_1 > 0$) for k large enough. For $i \notin I(x^*)$ (so $z_i^* = 0$), for each k large enough either again $z_i^{k+1} = \zeta_i^k$ or $z_i^{k+1} = \|\Delta x^k\|^2 + \|\zeta_-^k\|^2 \leq \|\Delta x^k\|^2 + \|\Delta z^k\|^2$, where $\Delta z^k := \zeta^k - z^k$. In the latter case, since $z_i^* = 0$, condition (i) in Proposition 3.13 holds with any $c_1 \geq 1$. Next, consider $\{x^k\}$. For $i \notin I(x^*)$, in view of Proposition 3.11, we have

$$\frac{|g(x_i^k)|}{|\langle a_i, \Delta x^k\rangle|} \to \infty \quad \text{as} \quad k \to \infty.$$

Thus, if $I(x^*) = \emptyset$, then, in view of Step 3 (i) in Algorithm A1, $t^k = 1$ for k large enough and otherwise, since in view of (6b) $\text{sign}(\langle a_i, \Delta x^k\rangle) = \text{sign}(\zeta_i^k) = 1$, for all k large enough and $i \in I(x^*)$,

$$\bar{t}^k = \min\left\{\frac{z_i^k}{\zeta_i^k} : \ i \in I(x^*)\right\}$$

and

$$t^k = \min\left\{1, \frac{z_{i_k}^k}{\zeta_{i_k}^k} - \|\Delta x^k\|\right\} \tag{23}$$

for k large enough, for some $i_k \in I(x^*)$. (In particular, in view of Proposition 3.11, $\{t^k\}$ converges to 1.) Thus, for k large enough,

$$\begin{aligned}
\|x^{k+1} - (x^k + \Delta x^k)\| &= \|x^k + t^k\Delta x^k - (x^k + \Delta x^k)\| \\
&= |t^k - 1|\|\Delta x^k\| \\
&\leq \left|\|\Delta x^k\| + \frac{\zeta_{i_k}^k - z_{i_k}^k}{\zeta_{i_k}^k}\right| \ \|\Delta x^k\| \\
&\leq \left(\|\Delta x^k\| + (\zeta_{i_k}^k)^{-1}\|\Delta z^k\|\right) \ \|\Delta x^k\|.
\end{aligned}$$

Since $z_i^* > 0$ for all $i \in I(x^*)$, it follows that for some $C > 0$ and all k large enough

$$\begin{aligned}
\|x^{k+1} - (x^k + \Delta x^k)\| &\leq (\|\Delta x^k\| + C\|\Delta z^k\|)\|\Delta x^k\| \\
&\leq (1 + C)(\|\Delta x^k\| + \|\Delta z^k\|)^2.
\end{aligned}$$

Thus condition (ii) of Proposition 3.13 holds (with $c_1 = 1 + C$). The claim then follows from Lemma 3.9 and Proposition 3.13. □

4 Refinement: A Barrier-Based Algorithm

While the Newton-KKT affine scaling algorithm we have considered so far has the advantage of simplicity, it has been observed in various contexts that improved behavior is likely achieved if instead of linearizing the KKT equations (2) (yielding (3)) one linearizes a perturbed version of (2), with the right-hand side in the second equation set to be a certain positive number μ rather than 0. It is well known that the resulting iteration can be viewed as a "primal-dual" *barrier* iteration. Typically, μ is progressively reduced and made to tend to 0 as a solution of the problem is approached. Variants of such methods have been proved to be very effective in the solution of linear programming, and more generally convex programming problems. Various extensions have been proposed for general nonlinear programming (see Section 1).

In this section, we propose and analyze a "barrier-based" algorithm which is specially tailored to quadratic programming, and is closely related to the affine scaling algorithm of Section 2 (Algorithm A1). Like Algorithm A1, it is strongly inspired from the algorithm of [PTH88, TWB$^+$03], as well as from the related algorithm of [TZ94] for linear programming and convex quadratic programming. At iteration k, the value μ^k of the barrier parameter is determined via computation of the affine scaling direction Δx^k (used in Algorithm A1), in such a way that (i) the resulting $\Delta x^{\mu,k}$ is still a good descent direction for f at x^k; and (ii) μ^k goes to zero fast enough, as Δx^k goes to zero, that quadratic convergence can be maintained. Specifically μ^k is assigned the value $\varphi^k \|\Delta x^k\|^\nu z^k_{\min}$, where $\nu > 2$ is prescribed, $z^k_{\min} := \min_i z^k_i > 0$, and φ^k is the largest scalar in $(0, \overline{\varphi}]$, $\overline{\varphi} > 0$ prescribed, such that

$$\langle \nabla f(x^k), \Delta x^{\mu,k} \rangle \leq \theta \langle \Delta x^k, \nabla f(x^k) \rangle$$

where $\theta \in (0,1)$ is prescribed; it then follows from (12) that

$$\langle \nabla f(x^k), \Delta x^{\mu,k} \rangle \leq -\theta\sigma \|\Delta x^k\|^2. \tag{24}$$

As noted in [BT03], such φ^k can be expressed as

$$\varphi^k = \begin{cases} \overline{\varphi} & \text{if } \sum_j \frac{\zeta^k_j}{z^k_j} \leq 0 \\ \min\left\{ \frac{(1-\theta)|\langle \nabla f(x^k), \Delta x^k \rangle|}{\|\Delta x^k\|^\nu z^k_{\min} \sum_j \frac{\zeta^k_j}{z^k_j}}, \overline{\varphi} \right\} & \text{otherwise.} \end{cases} \tag{25}$$

With μ^k thus computed, a second linear system is solved, with the same coefficient matrix and modified right-hand side. The solutions $\Delta x^{\mu,k}$ and $\zeta^{\mu,k}$ of this modified system are then substituted for Δx^k and ζ^k in the remainder of the iteration. The only other difference between Algorithm A1 and Algorithm A2 is linked to the fact that, while with the affine-scaling direction $f(x^k + t\Delta x^k) < f(x^k)$ for all $t \in (0, 1]$, this is no longer guaranteed with the modified direction $\Delta x^{\mu,k}$ when $\langle \Delta x^{\mu,k}, H\Delta x^{\mu,k}\rangle > 0$. It is readily verified however that in that case it holds that, for any descent direction Δx,

$$f(x^k + t\Delta x) < f(x^k) \;\forall t \in \left(0, 2\frac{|\langle \nabla f(x^k), \Delta x\rangle|}{\langle \Delta x, H\Delta x\rangle}\right).$$

Step 3(i) of the algorithm is modified accordingly.

The proposed algorithm is as follows. Note that, again, the algorithm statement implicitly assumes that $\mathcal{F}^o$ is nonempty.

Algorithm A2.

Parameters. $\beta \in (0, 1)$, $\underline{z} \in (0, 1)$, $z_{\mathrm{u}} > 0$, $\sigma > 0$, $\gamma > 1$, $\theta \in (0, 1)$, $\overline{\varphi} > 0$, $\nu > 2$, $\psi \in (1, 2)$.

Data. $x^0 \in \mathcal{F}^o$, $z_i^0 > 0 \; \forall i \in I$.

Step 0. Initialization. Set $k := 0$. Set $\overline{I} := \emptyset$. Set $\overline{\alpha}_i := 0$, $i = 1, \ldots, m$. Set $\overline{E} := \mathrm{I}$.[6]

Step 1. Computation of modified Hessian. Set $W^k := H + E^k$, where $E^k \succeq 0$ is computed as follows.

- If $H \succeq \sigma \mathrm{I}$, then set $E^k := 0$.
- Else
  - If $\frac{z_i^k}{|g_i(x^k)|} \leq \overline{\alpha}_i$ for some $i \in \overline{I}$ or ($\overline{E} \neq 0$ and $\overline{I} = \emptyset$) or ($\overline{E} \neq 0$ and $\frac{z_i^k}{|g_i(x^k)|} \geq \gamma^2 \overline{\alpha}_i$ for some $i \in \overline{I}$) then (i) set $\overline{I} := \{i : \frac{z_i^k}{|g_i(x^k)|} \geq 1\}$ and $\overline{\alpha}_i := \frac{1}{\gamma}\frac{z_i^k}{|g_i(x^k)|}$, $i \in \overline{I}$; and (ii)
    - ∗ if $H + \sum_{\overline{I}} \overline{\alpha}_i a_i a_i^{\mathrm{T}} \succeq \sigma \mathrm{I}$, set $\overline{E} := 0$;
    - ∗ else pick $\overline{E} \succeq 0$, with $\overline{E} \preceq (\|H\|_{\mathrm{F}} + \sigma)\mathrm{I}$, such that $H + \sum_{\overline{I}} \overline{\alpha}_i a_i a_i^{\mathrm{T}} + \overline{E} \succeq \sigma \mathrm{I}$. (Here, as in Algorithm A1, $\|H\|_{\mathrm{F}}$ is the Frobenius norm of H.)

[6] Again, the initial values assigned to the components of $\overline{\alpha}$ and to $\overline{E}$ are immaterial, as long as $\overline{E} \neq 0$.

– Set $E^k := \overline{E}$.

Step 2. Computation of a search direction.
(i) Let $(\Delta x^k, \zeta^k)$ solve the linear system in $(\Delta x, \zeta)$

$$W^k \Delta x + A^{\mathrm{T}} \zeta = -\nabla f(x^k) \tag{26a}$$

$$z_i^k \langle a_i, \Delta x \rangle + g_i(x^k) \zeta_i = 0 \quad \forall i \in I. \tag{26b}$$

If $\Delta x^k = 0$, stop.
(ii) Set $\mu^k := \varphi^k \|\Delta x^k\|^\nu z_{\min}^k$, where $z_{\min}^k := \min_i z_i^k$ and φ^k is given by

$$\varphi^k = \begin{cases} \overline{\varphi} & \text{if } \sum_j \frac{\zeta_j^k}{z_j^k} \le 0 \\ \min\left\{ \frac{(1-\theta)|\langle \nabla f(x^k), \Delta x^k \rangle|}{\|\Delta x^k\|^\nu z_{\min}^k \sum_j \frac{\zeta_j^k}{z_j^k}}, \overline{\varphi} \right\} & \text{otherwise.} \end{cases} \tag{27}$$

Let $(\Delta x^{\mu,k}, \zeta^{\mu,k})$ solve the linear system in $(\Delta x^\mu, \zeta^\mu)$

$$W^k \Delta x^\mu + A^{\mathrm{T}} \zeta^\mu = -\nabla f(x^k) \tag{28a}$$

$$z_i^k \langle a_i, \Delta x^\mu \rangle + g_i(x^k) \zeta_i^\mu = -\mu^k \quad \forall i \in I. \tag{28b}$$

Step 3. Updates.
(i) Set

$$\overline{t}^k := \begin{cases} \infty & \text{if } \langle a_i, \Delta x^{\mu,k} \rangle \le 0 \ \ \forall i \in I, \\ \min\left\{ \frac{|g_i(x^k)|}{\langle a_i, \Delta x^k \rangle} : \langle a_i, \Delta x^{\mu,k} \rangle > 0, i \in I \right\} & \text{otherwise.} \end{cases} \tag{29}$$

Set

$$t^k := \begin{cases} \min\left\{ \max\{\beta \overline{t}^k, \overline{t}^k - \|\Delta x^{\mu,k}\|\},\ 1 \right\} & \text{if } \langle \Delta x^{\mu,k}, H \Delta x^{\mu,k} \rangle \le 0, \\ \min\left\{ \max\{\beta \overline{t}^k, \overline{t}^k - \|\Delta x^{\mu,k}\|\},\ 1, \psi \frac{|\langle \nabla f(x^k), \Delta x^{\mu,k} \rangle|}{\langle \Delta x^{\mu,k}, H \Delta x^{\mu,k} \rangle} \right\} & \text{otherwise.} \end{cases} \tag{30}$$

Set $x^{k+1} := x^k + t^k \Delta x^{\mu,k}$.
(ii) Set $\left(\zeta_-^{\mu,k}\right)_i := \min\{\zeta_i^{\mu,k}, 0\}$, $\forall i \in I$. Set

$$z_i^{k+1} := \min\{\max\{\min\{\|\Delta x^{\mu,k}\|^2 + \|\zeta_-^{\mu,k}\|^2, \underline{z}\}, \zeta_i^{\mu,k}\},\ z_{\mathrm{u}}\},\ \forall i \in I. \tag{31}$$

(iii) Set $k := k + 1$. Go to Step 1.

$\square$

We now proceed to establish global and local convergence properties for Algorithm A2. The results are the same as for Algorithm A1, with the same assumptions. Many of the steps in the analysis are analogous to those of Section 3. The numbering of the first 12 lemmas, propositions and theorems is parallel to that used in Section 3. We start with a modified (and expanded) version of Proposition 3.1.

Proposition 4.1 *Let $\{x^k\}$, $\{\Delta x^{\mu,k}\}$, and $\{\zeta^{\mu,k}\}$ be as constructed by Algorithm A2, and let $\psi \in (0,2)$. Suppose $\Delta x^k \neq 0$.*
(i) If $\langle \Delta x^{\mu,k}, H\Delta x^{\mu,k}\rangle \leq 0$, then

$$f(x^k + t\Delta x^{\mu,k}) < f(x^k) \quad \forall t > 0. \tag{32}$$

(ii) If $\langle \Delta x^{\mu,k}, H\Delta x^{\mu,k}\rangle > 0$, then

$$f(x^k + t\Delta x^{\mu,k}) \leq f(x^k) + t\left(1 - \frac{\psi}{2}\right)\langle \nabla f(x^k), \Delta x^{\mu,k}\rangle < f(x^k)$$
$$\forall t \in \left[0, \psi\frac{|\langle \nabla f(x^k), \Delta x^{\mu,k}\rangle|}{\langle \Delta x^{\mu,k}, H\Delta x^{\mu,k}\rangle}\right] \tag{33}$$

and furthermore

$$f(x^k + t\Delta x^{\mu,k}) < f(x^k) \quad \text{ iff } t \in \left(0, 2\frac{|\langle \nabla f(x^k), \Delta x^{\mu,k}\rangle|}{\langle \Delta x^{\mu,k}, H\Delta x^{\mu,k}\rangle}\right]. \tag{34}$$

(iii)

$$g_i(x^k + t\Delta x^{\mu,k}) = g_i(x^k) + t\langle a_i, \Delta x^{\mu,k}\rangle < g_i(x^k) \quad \forall t > 0, \forall i \text{ s.t. } \zeta_i^{\mu,k} < 0. \tag{35}$$

Proof. Claim (i) follows directly from (24) and the second part of (ii) follows from (24) via routine manipulations on the quadratic function

$$f(x^k + t\Delta x^{\mu,k}) = f(x^k) + t\langle \nabla f(x^k), \Delta x^{\mu,k}\rangle + \frac{1}{2}t^2\langle \Delta x^{\mu,k}, H\Delta x^{\mu,k}\rangle. \tag{36}$$

To prove the first part of (ii), observe that (1) yields

$$f(x^k + t\Delta x^{\mu,k}) = f(x^k) + t\left(\langle \nabla f(x^k), \Delta x^{\mu,k}\rangle + \frac{1}{2}t\langle \Delta x^{\mu,k}, H\Delta x^{\mu,k}\rangle\right) \tag{37}$$

and that $t\langle \Delta x^{\mu,k}, H\Delta x^{\mu,k}\rangle \le -\psi\langle \nabla f(x^k), \Delta x^{\mu,k}\rangle$ for all $t \in \left[0, \psi\frac{-\langle \nabla f(x^k), \Delta x^{\mu,k}\rangle}{\langle \Delta x^{\mu,k}, H\Delta x^{\mu,k}\rangle}\right]$. In view of (24), substituting into (37) yields the left-hand inequality in (33). Since $\Delta x^k \neq 0$ and $\psi < 2$, the right-hand inequality in (33) follows as a direct consequence of (24), completing the proof of claim (ii). Finally, since g is linear,

$$g_i(x^k + t^k\Delta x^{\mu,k}) = g_i(x^k) + t^k\langle a_i, \Delta x^{\mu,k}\rangle \quad i = 1, \ldots,\ m.$$

Since $z_i^k > 0$ for all $i \in I$, it follows from (28b) that $\langle a_i, \Delta x^{\mu,k}\rangle < 0$ whenever $\zeta_i^{\mu,k} < 0$, proving claim (iii). □

The following is the critical step in the global convergence analysis.

Lemma 4.2 *Let $\{x^k\}$, $\{\Delta x^{\mu,k}\}$, and $\{\zeta_-^{\mu,k}\}$ be as constructed by Algorithm A2. Suppose Assumption 1 holds. Let K be an infinite index set such that $\{x^k\}_{k\in K} \to x^*$ for some x^* and*

$$\inf\{\|\Delta x^{\mu,k-1}\|^2 + \|\zeta_-^{\mu,k-1}\|^2 :\ k \in K\} > 0. \tag{38}$$

Then $\{\Delta x^k\}_{k\in K} \to 0$ and $\{\Delta x^k\}_{k\in K} \to 0$.

Proof. First we prove that $\{\Delta x^k\}_{k\in K} \to 0$. As in the proof of Lemma 3.2, we proceed by contradiction by assuming that, for some infinite index set $K' \subseteq K$, $\inf_{k\in K'} ||\Delta x^k\| > 0$. The same argument that was used in that proof shows that, on K', the components of $\{z^k\}$ are bounded away from zero and, without loss of generality, $\{z^k\}_{k\in K'} \to z^*$ for some $z^* > 0$, $\{W^k\}_{k\in K'} \to W^*$ for some W^*, $M(x^*, z^*, W^*)$ is nonsingular, and

$$\{\Delta x^k\}_{k\in K'} \to \Delta x^*$$

for some Δx^*. Then it follows from Step 2(ii) of Algorithm A2 that $\{\mu^k\}_{k\in K'}$ is bounded. It follows that $\{\Delta x^{\mu,k}\}_{k\in K'}$ and $\{\zeta^{\mu,k}\}_{k\in K'}$ are bounded and that, without loss of generality, for some $\Delta x^{\mu,*}$,

$$\{\Delta x^{\mu,k}\}_{k\in K'} \to \Delta x^{\mu,*}.$$

(Boundedness of $\{\Delta x^{\mu,k}\}_{k\in K'}$ and $\{\zeta^{\mu,k}\}_{k\in K'}$ follows from the fact that they solve the linear system (28), whose right-hand side is bounded and whose system matrix $M(x^k, z^k, W^k)$, nonsingular for all k, converges to the nonsingular matrix $M(x^*, z^*, W^*)$.) On the other hand, it follows from (28b) that, for all i such that $\langle a_i, \Delta x^{\mu,k}\rangle \neq 0$,

$$-\frac{g_i(x^k)}{\langle a_i, \Delta x^{\mu,k}\rangle}\zeta_i^{\mu,k} = z_i^k + \frac{\mu^k}{\langle a_i, \Delta x^{\mu,k}\rangle}.$$

For all i, the right-hand side is positive whenever $\langle a_i, \Delta x^{\mu,k} \rangle > 0$ and under the same condition is bounded away from zero on K' since $\{z_i^k\}$ is. Since $\zeta_i^{\mu,k}$ is bounded on K', it follows that $\{-g_i(x^k)/\langle a_i, \Delta x^{\mu,k}\rangle\}$ is (positive and) bounded away from zero on K' when $\langle a_i, \Delta x^{\mu,k}\rangle > 0$. In view of Step 3(i) of Algorithm A2, it follows that $\{\bar{t}^k\}$ is bounded away from zero on K'.

To proceed with the proof, we show that t^k is bounded away from zero on some infinite index set $K'' \subseteq K'$. If $\langle \Delta x^{\mu,k}, H\Delta x^{\mu,k}\rangle \leq 0$ holds infinitely many times on K', then in view of (30), the case is clear. Otherwise, there exists an infinite index set $K'' \subseteq K'$ such that $\langle \Delta x^{\mu,k}, H\Delta x^{\mu,k}\rangle > 0$ for all $k \in K''$. In view of (24), for all such k,

$$\frac{|\langle \Delta x^{\mu,k}, \nabla f(x^k)\rangle|}{\langle \Delta x^{\mu,k}, H\Delta x^{\mu,k}\rangle} > \frac{\theta\sigma \|\Delta x^k\|^2}{\langle \Delta x^{\mu,k}, H\Delta x^{\mu,k}\rangle}.$$

Since $\{\|\Delta x^k\|\}_{k\in K'}$ is bounded away from zero and $K'' \subseteq K'$, if follows from (30) that $\{t^k\}_{k\in K''}$ is bounded away from zero, say, $t^k > \underline{t}$ for all $k \in K'$, with $\underline{t} \in (0,1)$.

Proposition 4.1 and Step 3(i) in Algorithm A2 imply that $\{f(x^k)\}$ is monotonic decreasing. Thus, as in the proof of Lemma 3.2, to complete the contradiction argument, it then suffices to show that for some $\delta > 0$,

$$f(x^{k+1}) \leq f(x^k) - \delta \tag{39}$$

infinitely many times. We show that it holds for all $k \in K''$. When $\langle \Delta x^{\mu,*}, H\Delta x^{\mu,*}\rangle > 0$, the result follows from Proposition 4.1(ii) and (24). When $\langle \Delta x^{\mu,*}, H\Delta x^{\mu,*}\rangle \leq 0$, essentially the same argument as that used in the proof of Lemma 3.2 applies. Namely, $\langle \Delta x^{\mu,k}, H\Delta x^{\mu,k}\rangle \leq \frac{1}{2}\sigma\|\Delta x^*\|^2$ for $k \in K'$, k large enough; in view of (24), $\langle \nabla f(x^k), \Delta x^{\mu,k}\rangle \leq -\frac{1}{2}\sigma\|\Delta x^*\|^2$ for $k \in K'$, k large enough; since $t_k \in (\underline{t}, 1]$ for $k \in K'$, it follows from (1) that (39) again holds on K', with $\delta = \frac{1}{4}\underline{t}\sigma\|\Delta x^*\|^2$. Hence (39) holds on K'', and the proof that $\{\Delta x^k\}_{k\in K} \to 0$ is complete.

It remains to prove that $\{\Delta x^{\mu,k}\}_{k\in K} \to 0$. Suppose not. Then there is an infinite index set $K' \subseteq K$ such that $\{\Delta x^{\mu,k}\}_{k\in K'}$ is bounded away from zero. Since the sequences $\{z^k\}$ and $\{W^k\}$ are bounded, we assume without loss of generality that $\{z^k\}_{k\in K'} \to z^*$ for some $z^* > 0$ and $\{W^k\}_{k\in K'} \to W^*$ for some W^*. It then follows from Lemma 2.4 that $M(x^*, z^*, W^*)$ is nonsingular. Since $\{\Delta x^k\}_{k\in K} \to 0$, it follows from the choice of μ^k in Algorithm A2 that $\{\mu^k\}_{k\in K} \to 0$. Therefore, as k tends to infinity in $K' \subseteq K$, the right-hand side of (26) tends to the right-hand side of (28). Since $M(x^*, z^*, W^*)$ is nonsingular, it follows that $\lim_{k\to\infty, k\in K'} \Delta x^{\mu,k} = \lim_{k\to\infty, k\in K'} \Delta x^k = 0$, a contradiction with the supposition that $\{\Delta x^{\mu,k}\}_{k\in K'}$ is bounded away from zero. □

Lemma 4.3 *Let $\{x^k\}$, $\{\Delta x^{\mu,k}\}$, $\{\zeta^k\}$, and $\{\zeta^{\mu,k}\}$ be as constructed by Algorithm A2. Suppose Assumption 1 holds. Let x^* be such that, for some infinite index set K, $\{x^k\}_{k\in K}$ converges to x^*. If $\{\Delta x^{\mu,k}\}_{k\in K}$ converges to zero, then x^* is stationary and $\{\zeta^k\}_{k\in K}$ and $\{\zeta^{\mu,k}\}_{k\in K}$ both converge to z^*, where z^* is the unique multiplier vector associated with x^*.*

Proof. First, since $\{\Delta x^{\mu,k}\}_{k\in K} \to 0$, it follows from (24) that $\{\Delta x^k\}_{k\in K} \to 0$. The claims are then proved using the same argument as in the proof of Lemma 3.3, first starting from (26a)–(26b) to show that $\{\zeta^k\}_{k\in K}$ converges to z^*, then starting from (28a)–(28b) and using the fact that, due to the boundedness of $\{z^k\}$, $\{\Delta x^k\}_{k\in K} \to 0$ implies that $\{\mu^k\}_{k\in K} \to 0$, to show that $\{\zeta^{\mu,k}\}_{k\in K}$ converges to z^*. □

The proofs of the next six results are direct extensions of those of the corresponding results in Section 3 and are omitted.

Lemma 4.4 *Let $\{x^k\}$ and $\{\Delta x^{\mu,k}\}$ be as constructed by Algorithm A2. Let K be an infinite index set such that $\{x^k\}_{k\in K} \to x^*$ for some x^*. If $\{\Delta x^{\mu,k-1}\}_{k\in K} \to 0$ then $\{x^{k-1}\}_{k\in K} \to x^*$.*

Proposition 4.5 *Under Assumption 1, every accumulation point of $\{x^k\}$ constructed by Algorithm A2 is a stationary point for (P).*

Lemma 4.6 *Let $\{x^k\}$, $\{\Delta x^{\mu,k}\}$, and $\{\zeta_-^{\mu,k}\}$ be as constructed by Algorithm A2. Suppose Assumption 1 holds. Suppose that K, an infinite index set, is such that, for some x^*, $\{x^k\}_{k\in K}$ tends to x^*, and $\{\Delta x^{\mu,k-1}\}_{k\in K}$ and $\{\zeta_-^{\mu,k-1}\}_{k\in K}$ tend to zero. Then x^* is a KKT point.*

Lemma 4.7 *Let $\{x^k\}$ and $\{\Delta x^{\mu,k}\}$ be as constructed by Algorithm A2. Suppose Assumption 1 holds. Suppose that K, an infinite index set, is such that, for some x^*, $\{x^k\}_{k\in K} \to x^*$, and x^* is not KKT. Then $\{\Delta x^{\mu,k}\}_{k\in K} \to 0$.*

Lemma 4.8 *Let $\{x^k\}$ be as constructed by Algorithm A2. Suppose Assumptions 1 and 2 hold. Suppose $\{x^k\}$ has x^* as an accumulation point, and x^* is not KKT. Then the entire sequence $\{x^k\}$ converges to x^*.*

Theorem 4.9 *Under Assumptions 1 and 2, every accumulation point of $\{x^k\}$ constructed by Algorithm A2 is a KKT point.*

Lemma 4.10 *Let* $\{x^k\}$ *and* $\{\Delta x^{\mu,k}\}$ *be as constructed by Algorithm A2. Suppose Assumptions 1 and 3 hold. Let* K *be an infinite index set such that* $\{x^k\}_{k\in K} \to x^*$*. Then* $\{\Delta x^{\mu,k}\}_{k\in K} \to 0$.

Proposition 4.11 *Let* $\{x^k\}$, $\{\Delta x^{\mu,k}\}$, $\{z^k\}$, *and* $\{\zeta^{\mu,k}\}$ *be as constructed by Algorithm A2. Suppose Assumptions 1, 2 and 3 hold. Then the entire sequence* $\{x^k\}$ *converges to* x^**. Moreover, (i)* $\{\Delta x^{\mu,k}\} \to 0$*, (ii)* $\{\zeta^{\mu,k}\} \to z^*$*, and (iii)* $\{z_j^k\} \to \min\{z_j^*, z_{\mathrm{u}}\}$ *for all* j.

Lemma 4.12 *Let* $\{W^k\}$ *be as constructed by Algorithm A2. Under Assumptions 1–4, if* σ *in Algorithm A2 is such that* $\sigma < \min\{\langle v, Hv\rangle : \langle a_i, v\rangle = 0\ \forall i \in I(x^*),\ \|v\| = 1\}$*, then* $W^k = H$ *for all* k *sufficiently large.*

The remainder of the analysis departs somewhat from that of Section 3. We use the following additional lemma, which hinges on μ^k going to zero at least as fast as the smallest component of z^k.

Lemma 4.13 *Let* $\{\Delta x^k\}$, $\{\Delta x^{\mu,k}\}$, $\{\zeta^k\}$ *and* $\{\zeta^{\mu,k}\}$ *be as constructed by Algorithm A2. Suppose Assumptions 1–4 hold. For* k *large enough,*

$$\|\Delta x^{\mu,k} - \Delta x^k\| + \|\zeta^{\mu,k} - \zeta^k\| = O(\|\Delta x^k\|^\nu). \tag{40}$$

Furthermore,

$$\max\{0, -\langle \zeta^{\mu,k}, A\Delta x^{\mu,k}\rangle\} = O(\|\Delta x^{\mu,k}\|^\nu).$$

Proof. The first claim is a direct consequence of nonsingularity of $M(x^*, z^*, H)$ (which, like in the proof of Theorem 3.14, follows from Lemma 2.3), the fact that $\mu^k \leq \overline{\varphi}\|\Delta x^k\|^\nu z_{\mathrm{u}}$. Next, since $g_i(x^k) < 0$ for all i and all k, (28b) implies that, for all $i \in I$,

$$-\zeta_i^{\mu,k}\langle a_i, \Delta x^{\mu,k}\rangle = \frac{-(\zeta_i^{\mu,k})^2}{z_i^k}|g_i(x^k)| + \frac{\mu^k \zeta_i^{\mu,k}}{z_i^k} \leq \frac{\mu^k \zeta_i^{\mu,k}}{z_i^k}.$$

The second claim then follows from positiveness of z_i^k, boundedness of $\{\zeta^k\}$ (since it converges) and the fact that $\mu^k \leq \overline{\varphi}\|\Delta x^k\|^\nu z_{\min}^k$ in Step 2(ii) of Algorithm A2. □

Theorem 4.14 *Let* $\{x^k\}$, $\{z^k\}$ *be as constructed by Algorithm A2. Suppose Assumptions 1–4 hold. Then, if* $z_i^* < z_{\mathrm{u}}\ \forall i \in I$ *and* $\sigma < \min\{\langle v, Hv\rangle : \langle a_i, v\rangle = 0\ \forall i \in I(x^*),\ \|v\| = 1\}$, $\{(x^k, z^k)\}$ *converges to* (x^*, z^*) *q-quadratically.*

Proof. (Only the differences with the proof of Theorem 3.14 are pointed out.) First consider $\{z^k\}$. For $i \in I(x^*)$, $z_i^{k+1} = \zeta_i^{\mu,k}$ for k large enough. In view of Lemma 4.13 and since $\nu > 2$, it follows that condition (ii) in Proposition 3.13 holds for k large enough. For $i \notin I(x^*)$, for k large enough, either again $z_i^{k+1} = \zeta_i^{\mu,k}$ or $z_i^{k+1} = \|\Delta x^{\mu,k}\|^2 + \|\zeta_-^{\mu,k}\|^2$. In the latter case, we note that

$$\begin{aligned}|z_i^{k+1} - 0| &\leq \|\Delta x^{\mu,k}\|^2 + \|\zeta^{\mu,k}\|^2 \\ &\leq \|\Delta x^{\mu,k} - \Delta x^k\|^2 + \|\Delta x^k\|^2 + \|\zeta^{\mu,k} - \zeta_k\|^2 + \|\zeta^k\|^2.\end{aligned}$$

For k large enough, this yields

$$|z_i^{k+1} - 0| \leq \|\Delta x^{\mu,k} - \Delta x^k\| + \|\zeta^{\mu,k} - \zeta_k\| + \|\Delta x^k\|^2 + \|\zeta^k\|^2.$$

Again using Lemma 4.13 and the fact that $\nu > 2$, we conclude that condition (i) in Proposition 3.13 holds.

To conclude the proof, we first show that, when $\langle \Delta x^{\mu,k}, H\Delta x^{\mu,k}\rangle > 0$,

$$\psi \frac{|\langle \Delta x^{\mu,k}, \nabla f(x^k)\rangle|}{\langle \Delta x^{\mu,k}, H\Delta x^{\mu,k}\rangle} > 1 \tag{41}$$

for k large enough, implying that, for all k large enough, t^k is given by (see Step 3(i) of Algorithm A2)

$$t^k = \min\left\{\max\{\bar{t}^k - \|\Delta x^{\mu,k}\|, \beta\bar{t}^k\},\ 1\right\}. \tag{42}$$

Taking the inner product of both sides of equation (28a) by $\Delta x^{\mu,k}$ and using the fact that $W^k = H$ for all k large enough (Lemma 4.12), we get

$$\langle \nabla f(x^k), \Delta x^{\mu,k}\rangle = -\langle \Delta x^{\mu,k}, H\Delta x^{\mu,k}\rangle - \langle \zeta^{\mu,k}, A\Delta x^{\mu,k}\rangle$$

Now let $\tau = 2/\psi$. Then $\tau \in (1,2)$. Since f is quadratic, it follows that

$$\begin{aligned}f(x^k + \tau\Delta x^{\mu,k}) &= f(x^k) + \tau\langle \nabla f(x^k), \Delta x^{\mu,k}\rangle + \frac{\tau^2}{2}\langle \Delta x^{\mu,k}, H\Delta x^{\mu,k}\rangle \\ &= f(x^k) + (\tau - \frac{\tau^2}{2})\langle \nabla f(x^k), \Delta x^{\mu,k}\rangle - \frac{\tau^2}{2}\langle \zeta^{\mu,k}, A\Delta x^{\mu,k}\rangle. \\ &\leq f(x^k) - \tau(1 - \frac{\tau}{2})\theta\sigma\|\Delta x^k\|^2 + O(\|\Delta x^{\mu,k}\|^\nu).\end{aligned}$$

where we have used Lemma 4.13 and equation (24). Since $\nu > 2$, it follows that $f(x^k + \tau \Delta x^{\mu,k}) < f(x^k)$ for k large enough, and Proposition 4.1 ("only if" portion of (34)) then implies that for k large enough,

$$\tau < 2\frac{|\langle \Delta x^{\mu,k}, \nabla f(x^k)\rangle|}{\langle \Delta x^{\mu,k}, H\Delta x^{\mu,k}\rangle},$$

i.e., (41) holds for k large enough. Hence (42) holds for all k large enough.

Now, for $i \notin I(x^*)$,

$$\frac{|g(x_i^k)|}{|\langle a_i, \Delta x^{\mu,k}\rangle|} \to \infty \quad \text{as} \quad k \to \infty.$$

Thus, if $I(x^*) = \emptyset$, then, in view of Step 3(i) in Algorithm A2, $t^k = 1$ for k large enough. Further, since, whenever $\langle a_i, \Delta x^{\mu,k}\rangle > 0$ and $\zeta_i^{\mu,k} > 0$,

$$\frac{z_i^k}{\zeta_i^{\mu,k}} = \frac{|g_i(x^k)|}{\langle a_i, \Delta x^{\mu,k}\rangle} - \frac{\mu_i^k \zeta_i^{\mu,k}}{\langle a_i, \Delta x^{\mu,k}\rangle} \leq \frac{|g_i(x^k)|}{\langle a_i, \Delta x^{\mu,k}\rangle},$$

it follows that, when $I(x^*)$ is not empty, then

$$\bar{t}^k = \min\left\{\frac{z_i^k}{\zeta_i^{\mu,k}} : \ i \in I(x^*)\right\}.$$

and

$$t^k = \min\left\{1, \frac{z_{i_k}^k}{\zeta_{i_k}^{\mu,k}} - \|\Delta x^{\mu,k}\|\right\} \tag{43}$$

for k large enough, for some $i_k \in I(x^*)$. (In particular, in view of Proposition 4.11, $\{t^k\}$ converges to 1.) Thus,

$$\begin{aligned}\|x^{k+1} - (x^k + \Delta x^k)\| &= \|x^k + t^k \Delta x^{\mu,k} - (x^k + \Delta x^k)\| \\ &= |t^k - 1|\|\Delta x^k\| + t^k\|\Delta x^{\mu,k} - \Delta x^k\|\end{aligned}$$

In view of (43), we get

$$\begin{aligned}\|x^{k+1} - (x^k + \Delta x^k)\| &= \left|\|\Delta x^{\mu,k}\| + \frac{\zeta_{i_k}^{\mu,k} - z_{i_k}^k}{\zeta_{i_k}^{\mu,k}}\right| \ \|\Delta x^k\| + t^k\|\Delta x^{\mu,k} - \Delta x^k\| \\ &\leq \left|\|\Delta x^k\| + \|\Delta x^{\mu,k} - \Delta x^k\| + \frac{\zeta_{i_k}^{\mu,k} - z_{i_k}^k}{\zeta_{i_k}^{\mu,k}}\right| \ \|\Delta x^k\| + t^k\|\Delta x^{\mu,k} - \Delta x^k\|.\end{aligned}$$

In view of Lemma 4.13 and since $\nu > 2$, we obtain, for k large enough,

$$\|x^{k+1} - (x^k + \Delta x^k)\| \le 2\left(\|\Delta x^k\| + (\zeta_{i_k}^{\mu,k})^{-1}\|\Delta z^k\|\right)\ \|\Delta x^k\|.$$

The remainder of the proof is as in the proof of Theorem 3.14. □

5 Implementation issues

We still have to define explicitly a way of choosing $\overline{E} \succeq 0$ "small" such that $H + \sum_{\overline{I}} \overline{\alpha}_i a_i a_i^{\mathrm{T}} + \overline{E} \succeq \sigma \mathrm{I}$ (Step 1 of Algorithms A1 and A2). We use the following method, borrowed from [TWB+03, p. 192]: $\overline{E} = h\mathrm{I}$, with

$$h = \begin{cases} 0 & \text{if } \lambda_{\min} \geq \sigma, \\ -\lambda_{\min} + \sigma & \text{if } |\lambda_{\min}| < \sigma, \\ 2|\lambda_{\min}| & \text{otherwise,} \end{cases} \tag{44}$$

where $\lambda_{\min}$ denotes the leftmost eigenvalue of $H + \sum_{\overline{I}} \overline{\alpha}_i a_i a_i^{\mathrm{T}}$. [7] (This idea of choosing $\overline{E}$ as a nonnegative multiple of the identity has been used in interior-point algorithms for nonlinear programming, e.g., in [VS99].)

Another implementation issue concerns the feasibility of the iterates. The presence of the expression $\overline{t}^k - \|\Delta x^k\|$ in the definition (8) of the primal stepsize t^k is key to obtaining quadratic convergence. However, when $\|\Delta x^k\|$ becomes small, t^k may become so close to the maximal feasible stepsize $\overline{t}^k$ that, in finite precision arithmetic, $g_i(x^{k+1})$ returns a nonnegative number, due to an accumulation of numerical errors in the evaluation of $\overline{t}^k$, t, x^{k+1} and finally $g_i(x^{k+1})$. This was indeed observed in our numerical experiments, where the evaluation of a constraint function sometimes returned a nonnegative number—but never larger than a small multiple of the machine relative accuracy $\mathbf{u}$.[8] To circumvent this difficulty, we allow the constraints to recede slightly. At each iteration, before Step 1, we define $\tilde{g}_i^k = \min\{g_i(x^k), -\epsilon\}$, where ϵ is a small positive multiple of $\mathbf{u}$, and we replace $g_i(x^k)$ by $\tilde{g}_i^k$ throughout the iteration. If, as a result, some $g_i(x^k)$ eventually becomes too positive, then it is possible to apply a small perturbation to the current iterate along the corresponding a_i's in order to make it numerically feasible

[7] Note that the use of eigenvalues to construct the matrix $\overline{E}$ could potentially be avoided by using a modified Cholesky factorization instead. In some cases, such as large-scale problems, computing a (modified) Cholesky factorization is impractical, and iterative methods are preferred.

[8] The floating point relative accuracy is the distance from 1.0 to the next (larger) floating point number.

again. In our numerical experiments, we never had to apply this latter procedure since the $g_i(x^k)$'s always remained below $10^3\mathbf{u}$, which should be typical under our linear independence assumption (Assumption 1): indeed, since $(W^k + \sum_{i=1}^m \frac{z_i^k}{|g_i(x^k)|} a_i a_i^{\mathrm{T}})\Delta x^k = \nabla f(x^k)$ in view of (11a), it then follows that when $g_i(x^k)$ is close to zero and $\|\nabla f(x^k)\|$ is not very large, the component of Δx^k along a_i is very small.

The major computational tasks in Algorithms A1 and A2 are the following. First, in Step 1 (when the $\overline{\alpha}_i$'s are recomputed), compute a sufficiently good approximation $\tilde{\lambda}_{\min}$ of the leftmost eigenvalue of $H + \sum_{\overline{I}} \overline{\alpha}_i a_i a_i^{\mathrm{T}}$, to be used in (44); many possibilities are available for computing $\lambda_{\min}$, see e.g. [BDD$^+$00], [ABG04] and references therein. Second, in Step 2, solve system (6)—in Algorithm A1—or systems (26) and (28)—in Algorithm A2. The other operations require minimal computation (note that $A\Delta x^k$ computed in (6b) can be reused in the computation of $\bar{t}^k$).

Structured linear systems like (6), or (26) and (28), are ubiquitous in primal-dual interior-point methods; see e.g. [Wri98] or [FG98]. Since these three systems have the same matrix $M(x^k, z^k, W^k)$, we focus on one of them, say (6). One option is to solve (6) explicitly using a classical technique. However, if the number m of constraints is very large, then the cost of a "black-box" method may be prohibitive. An alternative is to eliminate ζ from the second line (6b) and substitute into the first line (6a), leading to the condensed primal-dual system (11a). This expresses Δx^k as the solution of an $n \times n$ linear system. The multiplier estimate ζ is then obtained at low cost from (11b). When the matrices are dense, solving a linear system of the form (6) via the Schur complement (11) costs $O(mn^2)$ flops to form S^k, $O(n^3)$ to solve (11a) and $O(mn)$ to solve (11b).

Solving the condensed primal-dual system (11a) for the update Δx^k may seem inappropriate because the condition number of the condensed primal-dual matrix S^k grows unbounded if some constraints are active at x^*. However, as shown by M. Wright [Wri98], this ill-conditioning is benign: the singular values of S^k split into two well-behaved groups, one very large and the other of order 1 (this is responsible for the large condition number of S^k), and the expected inaccuracies preserve this structure. (In her paper, M. Wright shows this for our $S^k - E^k$. Because our E^k is a bounded multiple of the identity, it applies to S^k as well.) It follows that the absolute error on Δx^k computed with a backward-stable method is comparable to $\mathbf{u}$, and so is the error on the multipliers computed via (11b). Moreover, the absolute error in the computed solution of the full, well-conditioned primal-dual system (6) by any backward-stable method is also comparable to $\mathbf{u}$. We refer to [Wri98] for details. In our implementations (see Section 6), we solve (6), (26) and (28) via the condensed approach.

Finally, we point out that Algorithms A1 and A2 can be adapted to deal with equality constraints by means of elimination techniques; see e.g. [NW99, Section 15.2] for details.

6 Numerical experiments and comparisons

In order to assess the practical value of Algorithms A1 and A2, we performed preliminary comparative tests using Matlab implementations of Algorithm A1 (affine scaling, $\mu = 0$), Algorithm A2 (barrier function, $\mu > 0$) and two interior ellipsoid trust-region (TR) algorithms based on [Ye89, Ye92, Tse04].

We considered indefinite quadratic programs in *standard inequality form*

$$\text{minimize } \frac{1}{2}\langle x, Hx\rangle + \langle c, x\rangle \ \text{ s.t. } \ Cx \leq d, \quad x \geq 0, \quad x \in \mathcal{R}^n,$$

because they fit within the framework of (P) and, by means of slack variables, the constraints are readily transformed into the form $Ax = b$, $x \geq 0$ used in [Ye89, Ye92, Tse04]. We chose the entries of C independently from a uniform distribution on the interval $(10^{-6}, 1 + 10^{-6})$. This ensures that the feasible set $\{Cx \leq d, x \geq 0\}$ has a nonempty interior and does not feature exceedingly acute angles that would compromise the behavior of a strictly feasible method. The number $m - n$ of rows in C was itself chosen randomly in $[1, 2n]$ with uniform distribution. The algorithms were initialized with $x^0 = e$, the vector of all ones. The vector d was selected as $d = Cx^0 + e$. The matrix H, with condition number 10^{ncond} and number of negative eigenvalues approximately *negeig*, was generated as described by Moré and Toraldo [MT89, p. 392]. Finally, the vector c was defined as $c = -Hx^*$ (so $\nabla f(x^*) = 0$) where x^* was chosen from the normal distribution $\mathcal{N}(0, 1)$. The algorithms were tested on a common set of sample problems, with $n = 100$ and varying values of *ncond* and *negeig*. Ten problems were generated in each category, for a total of 250 test problems, and the algorithms were compared on these problems with regard to the number of iterations, the number of times the leftmost eigenvalue of a large $(n \times n)$ matrix has to be computed, the number of times a large $(n \times n)$ linear system has to be solved, and the final value obtained for the cost function.

In order to assess the usefulness of the procedure defined in Step 1, we also tested a simpler version of Algorithms A1 and A2 where the leftmost eigenvalue of $H+\sum_{i=1}^{m}\frac{z_i^k}{|g_i(x^k)|}a_ia_i^{\mathrm{T}}$ is evaluated at *every* iteration in order to compute E^k according to (44). We ran comparative tests with Y. Ye's `spsolqp` algorithm, or more precisely with a modification thereof, provided to us by Prof. Ye, tuned to address the standard inequality form described above; we dub this algorithm `spsolqp-std-ineq`. The `spsolqp` code, initially written in 1988, is based on the Interior Ellipsoid TR method described and analyzed in [Ye89, Ye92]. The TR subproblems are tackled using an inner iteration based on [Ye89, Procedure 3.1]; the inner iteration terminates as soon as a feasible point is found that produces a decrease of the cost function (note that the decrease guaranteed by [Ye89, Lemma 3.5] assumes that H is positive semidefinite). Accordingly, this algorithm does not attempt to solve the TR subproblems accurately.

We also tested our algorithm against an interior-ellipsoid-based algorithm that computes nearly exact solutions to the TR subproblems; we dub the algorithm `exact-TR`. The motivation for considering such an algorithm is that most convergence analyses, including the ones in [Ye89, Ye92, Tse04], assume that the TR subproblems are solved accurately. The `exact-TR` algorithm is based on a Matlab script for box constrained problems that was communicated to us by P. Tseng and was used for the tests reported in [Tse04].[9] We modified the script in accordance with [Tse04, Section 2] to include equality constraints and made further adjustments to improve its performance on our battery of test problems.

We chose to use the stopping criterion of `spsolqp` in all the algorithms. Thus, execution terminates when $(f(x^k) - f(x^{k+1})/(1 + |f(x^k)|)$ is smaller than a prescribed tolerance, which we set to 10^{-8}. For Algorithms A1 and A2, the following parameter values were used: $\beta = .9$, $\underline{z} = 10^{-4}$, $z_{\mathrm{u}} = 10^{15}$, $\sigma = 10^{-5}$, $\gamma = 10^3$, $\theta = .8$, $\overline{\varphi} = 10^6$, $\nu = 3$, $\psi = 1.5$, $\epsilon = 10^{-14}$ (ϵ appears in the definition of $\tilde{g}$, see Section 5). The multiplier estimates are assigned the initial value $z_j^0 = \max\{.1, z_j'^0\}$ where $z'^0 = -(A^{\mathrm{T}})^{\dagger}\nabla f(x^0)$ (with reference to formulation (P)) and the superscript † denotes the Moore-Penrose pseudo-inverse. All tests were run on a PC with Intel Pentium 4 CPU 2.60 GHz with 512 KB cache, running Linux kernel 2.6.1 and Matlab 6.5 (R13). The floating point relative accuracy (see definition in footnote 8) is approximately $2.2 \cdot 10^{-16}$.

[9] Note that in the results reported in [Tse04], *ncond* is the *natural* logarithm of the condition number. (This was confirmed to us by P. Tseng.)

The numerical results are presented in Tables 1 to 3 and Figures 1 to 4. Tables 1 to 3 show that Algorithms A1 and A2 clearly outperform the interior ellipsoid TR codes in terms of number of iterations and number of system solves, and the `exact-TR` code in terms of eigensolves. (`spsolqp-std-ineq` does not involve eigensolves.) The particularly large performance gap in terms of system solves between `exact-TR` and the other codes is to be attributed to the fact that in the former, consistent with the assumption made in the analysis in [Ye89, Ye92, Tse04], the TR subproblems are solved with high accuracy (using the Moré-Sorensen method [CGT00, Section 7.3]).

Barrier-based Algorithm A2 tends to outperform affine-scaling Algorithm A1 in terms of number of iterations, eigensolves, and system solves for nonconvex problems (*negeig* > 0). When the problem is convex, there is no clear winner in terms of number of iterations and system solves, and both algorithms only require one eigensolve per problem (enough to notice that the problem is convex). Concerning the number of system solves, note that, in Table 3, two system solves—(26) and (28)—are counted per iteration for the barrier-based algorithm A2, which possibly does a disservice to that algorithm: since the two systems have the same matrix, decomposition-based solvers will be able to reuse information to dramatically speed-up the second solve.

The advantage of using Step 1—instead of computing the smallest eigenvalue of $H + \sum_{i=1}^{m} \frac{z_i^k}{|g_i(x^k)|} a_i a_i^{\mathrm{T}}$ at each iteration—is clearly seen in the numerical results ("simplified Step 1" versus original algorithms): the number of eigensolves is significantly reduced whereas the number of iterations (and, thus, of system solves) is hardly affected. Note however that, every time the $\overline{\alpha}_i$'s are recomputed (via eigenvalue decomposition) in Step 1, a dedicated Schur complement $H + \sum_{\overline{I}} \overline{\alpha}_i a_i a_i^{\mathrm{T}}$ has to be formed. Fortunately, if A is sparse (which is the case in many applications), the cost of constructing the Schur complement is comparatively low. Moreover, if an inverse-free eigensolver is used, it may even not be profitable to form the Schur complement.

Since none of the tested algorithms is a global method, it is natural that they sometimes converge to different local minima. The purpose of Figures 1 to 4 is to compare the methods in terms of quality of the solutions. In each figure, the top histogram shows on a log scale the distribution of the differences favorable to Algorithm A2, and the bottom histogram shows the distribution of the differences unfavorable to that algorithm. Figures 1 and 2 show a strong tendency of Algorithm A2 to produce lower-cost solutions than the interior TR methods. The reasons are unknown. Figures 3 and 4 suggest that the variants of the Newton-KKT algorithms (simple or elaborate Step 1,

Mean number of iterations

Algorithm A1 with simplified Step 1					
	negeig				
ncond	0	10	50	90	100
0	22.0	71.4	86.1	102.2	89.7
3	29.9	61.3	79.2	82.6	81.6
6	34.4	73.7	86.5	81.8	89.1
9	42.1	95.6	86.8	98.5	99.7
12	47.0	95.6	110.4	118.4	112.1

Algorithm A2 with simplified Step 1					
	negeig				
ncond	0	10	50	90	100
0	21.2	28.5	32.6	42.1	36.6
3	91.6	22.8	28.1	22.8	24.6
6	31.2	19.6	23.4	20.3	24.0
9	33.6	30.0	33.8	37.4	34.8
12	31.9	35.6	40.0	42.3	43.9

Algorithm A1					
	negeig				
ncond	0	10	50	90	100
0	22.0	74.4	89.8	96.9	89.7
3	29.9	72.3	81.9	85.3	81.1
6	34.4	91.0	92.1	90.1	98.3
9	42.1	117.2	104.7	107.3	107.5
12	47.0	111.1	130.5	138.6	120.4

Algorithm A2					
	negeig				
ncond	0	10	50	90	100
0	21.2	26.5	33.9	37.8	38.1
3	91.6	25.8	29.9	24.8	27.0
6	31.2	23.1	25.8	27.4	27.4
9	33.6	39.4	35.6	42.2	36.2
12	31.9	47.7	45.4	45.3	47.2

spsolqp-stp-ineq					
	negeig				
ncond	0	10	50	90	100
0	53.7	117.7	91.0	181.3	153.8
3	69.8	79.1	111.1	114.3	126.4
6	75.3	81.8	83.3	112.0	102.7
9	76.8	93.3	88.8	124.5	86.9
12	66.5	83.8	79.5	113.5	77.7

exact-TR					
	negeig				
ncond	0	10	50	90	100
0	52.5	113.0	113.1	184.8	136.6
3	88.1	80.5	123.1	116.6	147.6
6	92.4	80.8	83.9	109.7	110.3
9	64.4	92.4	82.7	127.9	86.8
12	42.9	75.1	77.7	112.2	82.6

Table 1: Comparison of the various algorithms in terms of the mean number of iterations over 10 problems randomly selected as explained in the text.

Number of eigensolves per problem

Algorithm A1 with simplified Step 1					
	negeig				
ncond	0	10	50	90	100
0	1.0	72.4	87.1	103.2	90.7
3	1.0	62.3	80.2	83.6	82.6
6	1.0	74.7	87.5	82.8	90.1
9	1.0	96.6	87.8	99.5	100.7
12	1.0	96.6	111.4	119.4	113.1

Algorithm A2 with simplified Step 1					
	negeig				
ncond	0	10	50	90	100
0	1.0	29.5	33.6	43.1	37.6
3	1.0	23.8	29.1	23.8	25.6
6	1.0	20.6	24.4	21.3	25.0
9	1.0	31.0	34.8	38.4	35.8
12	1.0	36.6	41.0	43.3	44.9

Algorithm A1					
	negeig				
ncond	0	10	50	90	100
0	1.0	18.6	20.9	29.0	28.9
3	1.0	30.5	38.4	35.3	39.5
6	1.0	42.1	42.9	48.2	50.1
9	1.0	58.3	59.6	63.1	61.8
12	1.0	61.3	80.4	89.6	73.9

Algorithm A2					
	negeig				
ncond	0	10	50	90	100
0	1.0	8.0	9.9	10.7	11.9
3	1.0	13.6	16.6	14.5	14.7
6	1.0	13.3	16.3	16.4	16.9
9	1.0	23.7	29.0	36.2	29.3
12	1.0	27.4	37.9	40.1	38.5

spsolqp-stp-ineq					
	negeig				
ncond	0	10	50	90	100
0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.0
12	0.0	0.0	0.0	0.0	0.0

exact-TR					
	negeig				
ncond	0	10	50	90	100
0	52.5	113.0	113.1	184.8	136.6
3	88.1	80.5	123.1	116.6	147.6
6	92.4	80.8	83.9	109.7	110.3
9	64.4	92.4	82.7	127.9	86.8
12	42.9	75.1	77.7	112.2	82.6

Table 2: Comparison of the various algorithms in terms of the mean number of times the leftmost eigenvalue of an $n \times n$ matrix had to be computed. The mean is computed over 10 problems randomly selected as explained in the text.

Number of system solves per problem

Algorithm A1 with simplified Step 1					
	negeig				
ncond	0	10	50	90	100
0	22.0	71.4	86.1	102.2	89.7
3	29.9	61.3	79.2	82.6	81.6
6	34.4	73.7	86.5	81.8	89.1
9	42.1	95.6	86.8	98.5	99.7
12	47.0	95.6	110.4	118.4	112.1

Algorithm A2 with simplified Step 1					
	negeig				
ncond	0	10	50	90	100
0	42.4	57.0	65.2	84.2	73.2
3	183.2	45.6	56.2	45.6	49.2
6	62.4	39.2	46.8	40.6	48.0
9	67.2	60.0	67.6	74.8	69.6
12	63.8	71.2	80.0	84.6	87.8

Algorithm A1					
	negeig				
ncond	0	10	50	90	100
0	22.0	74.4	89.8	96.9	89.7
3	29.9	72.3	81.9	85.3	81.1
6	34.4	91.0	92.1	90.1	98.3
9	42.1	117.2	104.7	107.3	107.5
12	47.0	111.1	130.5	138.6	120.4

Algorithm A2					
	negeig				
ncond	0	10	50	90	100
0	42.4	53.0	67.8	75.6	76.2
3	183.2	51.6	59.8	49.6	54.0
6	62.4	46.2	51.6	54.8	54.8
9	67.2	78.8	71.2	84.4	72.4
12	63.8	95.4	90.8	90.6	94.4

spsolqp-stp-ineq					
	negeig				
ncond	0	10	50	90	100
0	54.9	156.0	122.9	243.7	206.1
3	74.3	103.6	149.9	153.5	169.5
6	81.3	108.1	112.1	151.7	136.9
9	79.6	119.0	120.5	168.9	117.0
12	67.5	109.6	107.0	153.2	103.9

exact-TR					
	negeig				
ncond	0	10	50	90	100
0	568.6	881.9	825.9	1332.5	999.9
3	936.5	641.1	901.2	870.4	1111.0
6	1052.8	645.8	644.2	810.6	808.4
9	757.0	823.6	615.8	944.2	630.0
12	528.6	608.8	572.0	804.8	607.6

Table 3: Comparison of the various algorithms in terms of the mean number of times an $n \times n$ linear system had to be solved. The mean is computed over 10 problems randomly selected as explained in the text.

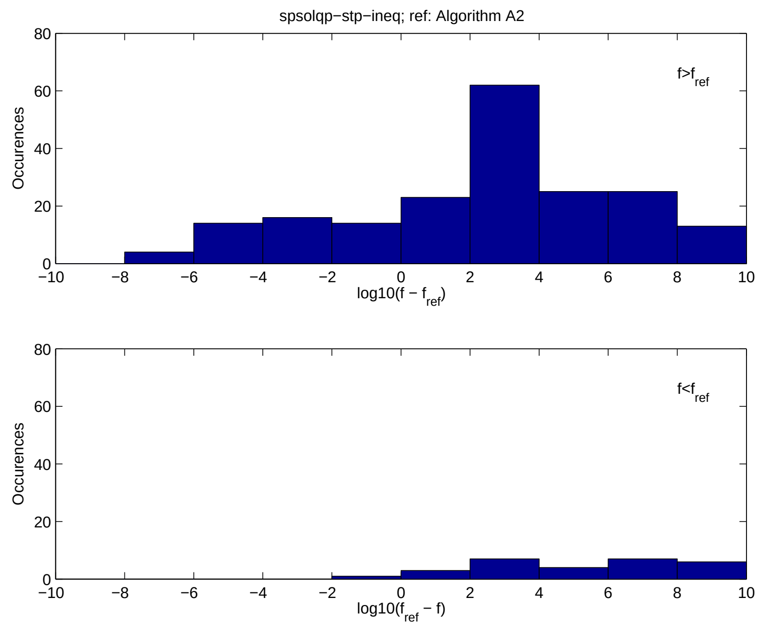

Figure 1: Comparison of the quality of solutions obtained by Algorithms A2 (f_{ref}) and `spsolqp-std-ineq` (f). The top histogram tallies instances where A2 reached a lower-cost solution than `spsolqp-std-ineq`, and vice-versa for the bottom histogram. The figure shows a strong tendency of Algorithm A2 to produce lower-cost solutions than `spsolqp-std-ineq`. The data comes from the 250 tests reported on in the tables.

barrier parameter or not) produce results of comparable quality.

Finally, we observed that the numerical behaviour of Algorithm A2 is further improved[10] when, in Step 2(ii), μ^k is assigned the value $\varphi^k(\|\Delta x^k\|^\nu + \|\zeta_-^k\|^\nu) z_{\min}^k$ instead of $\varphi^k\|\Delta x^k\|^\nu z_{\min}^k$ (with $(\|\Delta x^k\|^\nu + \|\zeta_-^k\|^\nu)$ also replacing $\|\Delta x^k\|^\nu$ in (27)). This is likely due to the modified μ^k being bounded away from zero near non-KKT stationary points, unlike the original μ^k. The theoretical properties of this modified algorithm are under investigation.

[10] In particular, the large number of iterations for *ncond*=3 and *negeig*=0 became much smaller.

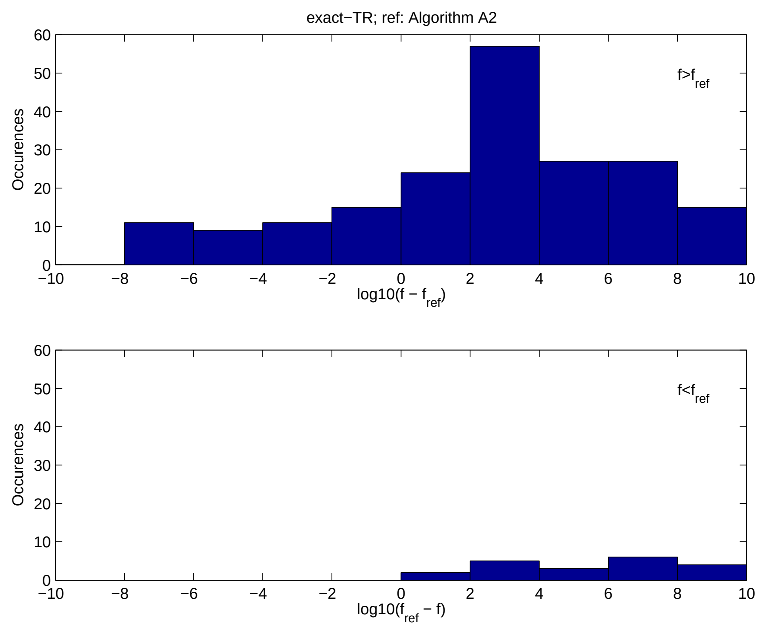

Figure 2: Comparison of the quality of solutions obtained by Algorithms A2 (f_{ref}) and `exact-TR` (f).

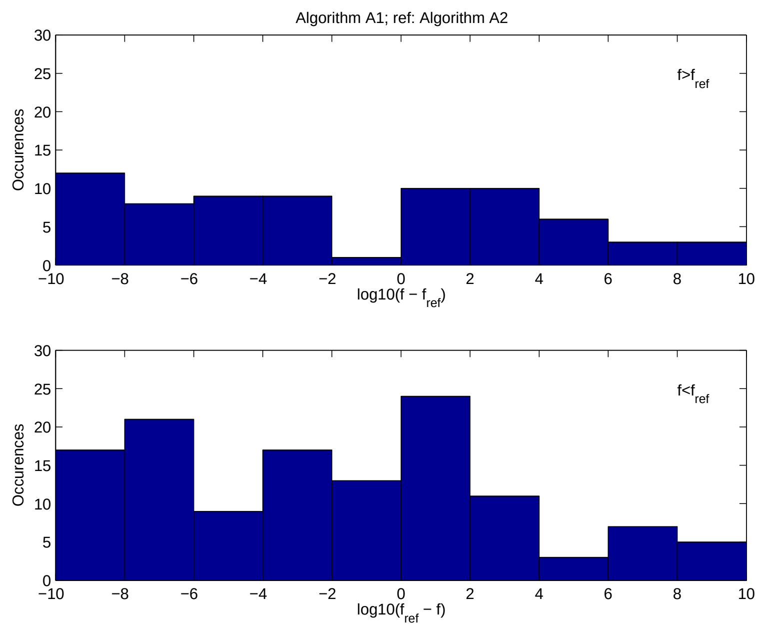

Figure 3: Comparison of the quality of solutions obtained by Algorithms A2 (f_{ref}) and A1 (f).

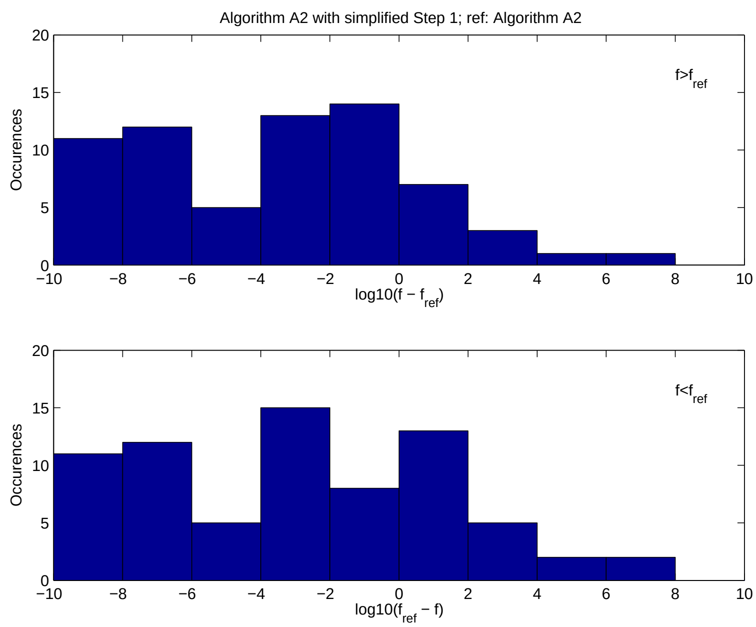

Figure 4: Comparison of the quality of solutions obtained by Algorithm A2 ($f_{\rm ref}$) and Algorithm A2 with simplified Step 1 (f).

7 Conclusion

Two "Newton-KKT" interior point algorithms for indefinite quadratic programming were proposed and analyzed, one of the affine-scaling type, the other barrier-based. Both were proved to converge globally and locally quadratically under nondegeneracy assumptions. Numerical results on randomly generated problems were reported that suggest that the proposed algorithms hold promise, even for degenerate problems.

Acknowledgements

The authors wish to thank Paul Tseng and Yinyu Ye for making available to them their Matlab codes, and for valuable advice. Further they wish to thank two anonymous referees and the anonymous associate editor for their careful reading of the manuscript and their many helpful comments; in particular, for uncovering a flow in the analysis in the original version of the paper and for making the authors aware of the `spsolqp` code.

References

[ABG04] P.-A. Absil, C. G. Baker, and K. A. Gallivan, *A truncated-CG style method for symmetric generalized eigenvalue problems*, accepted for publication in J. Comput. Appl. Math., special issue on ICCAM 2004 Conference, 2004.

[BB95] J.F. Bonnans and M. Bouhtou, *The trust region affine interior point algorithm for convex and nonconvex quadratic programming*, RAIRO Rech. Opér. **29** (1995), 195–217.

[BDD+00] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst (eds.), *Templates for the solution of algebraic eigenvalue problems: A practical guide*, Society for Industrial and Applied Mathematics, Philadelphia, 2000.

[BGN00] R.H. Byrd, J.C. Gilbert, and J. Nocedal, *A trust region method based on interior point techniques for nonlinear programming*, Mathematical Programming **89** (2000), 149–185.

[BP97] J. Frédéric Bonnans and Cecilia Pola, *A trust region interior point algorithm for linearly constrained optimization*, SIAM J. Optim. **7** (1997), no. 3, 717–731.

[BT03] S. Bakhtiari and A.L. Tits, *A simple primal-dual feasible interior-point method for nonlinear programming with monotone descent*, Comput. Optim. Appl. **25** (2003), 17–38.

[CGT00] A. R. Conn, N. I. M. Gould, and Ph. L. Toint, *Trust-region methods*, MPS/SIAM Series on Optimization, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, and Mathematical Programming Society (MPS), Philadelphia, PA, 2000.

[CL99] T.F. Coleman and J. Liu, *An interior Newton method for quadratic programming*, Math. Programming **85** (1999), 491–523.

[Dik67] I.I. Dikin, *Iterative solution of problems of linear and quadratic programming*, Soviet Math. Dokl. **8** (1967), 674–675.

[DX00] Chuangyin Dang and Lei Xu, *A barrier function method for the nonconvex quadratic programming problem with box constraints*, J. Global Optim. **18** (2000), no. 2, 165–188.

[ETTZ96] 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** (1996), 507–541.

[FG98] A. Forsgren and P.E. Gill, *Primal-dual interior methods for nonconvex nonlinear programming*, SIAM J. on Optimization **8** (1998), no. 4, 1132–1152.

[FGW02] A. Forsgren, P. E. Gill, and M. H. Wright, *Interior methods for nonlinear optimization*, SIAM Rev. **44** (2002), no. 4, 525–597.

[FM68] A.V. Fiacco and G.P. McCormick, *Nonlinear programming: Sequential unconstrained minimization techniques*, Wiley, New-York, 1968.

[GOW98] D. M. Gay, M. L. Overton, and M. H. Wright, *A primal-dual interior method for nonconvex nonlinear programming*, Advances in Nonlinear Programming (Y. Yuan, ed.), Kluwer Academic Publisher, 1998, pp. 31–56.

[GZ05] C. Grossmann and M. Zadlo, *A general class of penalty/barrier path-following Newton methods for nonlinear programming*, Optimization **54** (2005), no. 2, 161–190.

[Her82] J.N. Herskovits, *Développement d'une méthode numérique pour l'optimization non-linéaire*, Ph.D. thesis, Université Paris IX - Dauphine, Paris, France, January 1982.

[Her86] ______, *A two-stage feasible directions algorithm for nonlinear constrained optimization*, Math. Programming **36** (1986), no. 1, 19–38.

[MA89] R.D.C. Monteiro and I. Adler, *Interior path following primal-dual algorithms. Part ii: Convex quadratic programming*, Mathematical Programming **44** (1989), 43–66.

[MT89] Jorge J. Moré and Gerardo Toraldo, *Algorithms for bound constrained quadratic programming problems*, Numer. Math. **55** (1989), no. 4, 377–400.

[MT98] Renato D. C. Monteiro and Takashi Tsuchiya, *Global convergence of the affine scaling algorithm for convex quadratic programming*, SIAM J. Optim. **8** (1998), no. 1, 26–58 (electronic).

[NW99] J. Nocedal and S. Wright, *Numerical optimization*, Springer Series in Operations Research, Springer-Verlag, New York, 1999.

[PTH88] 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** (1988), no. 4, 788–811.

[QQ00] Hou-Duo Qi and Liqun Qi, *A new QP-free, globally convergent, locally superlinearly convergent algorithm for inequality constrained optimization*, SIAM J. Optim. **11** (2000), no. 1, 113–132 (electronic).

[Tse04] P. Tseng, *Convergence properties of Dikin's affine scaling algorithm for nonconvex quadratic minimization*, J. Global Optim. **30** (2004), no. 2, 285–300.

[TWB+03] 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 **14** (2003), no. 1, 173–199.

[TY02] Paul Tseng and Yinyu Ye, *On some interior-point algorithms for nonconvex quadratic optimization*, Math. Program. **93** (2002), no. 2, Ser. A, 217–225.

[TZ94] A.L. Tits and J.L. Zhou, *A simple, quadratically convergent algorithm for linear and convex quadratic programming*, Large Scale Optimization: State of the Art (W.W. Hager, D.W. Hearn, and P.M. Pardalos, eds.), Kluwer Academic Publishers, 1994, pp. 411–427.

[VS99] R.J. Vanderbei and D.F. Shanno, *An interior-point algorithm for nonconvex nonlinear programming*, Computational Optimization and Applications **13** (1999), 231–252.

[Wri98] M. H. Wright, *Ill-conditioning and computational error in interior methods for nonlinear programming*, SIAM J. Optim. **0** (1998), no. 1, 84–111.

[Yam98] H. Yamashita, *A globally convergent primal-dual interior point method for constrained optimization*, Optimization Methods and Software **10** (1998), 443–469.

[Ye87] Y. Ye, *Interior algorithms for linear, quadratic, and linearly constrained convex programming*, Ph.D. thesis, Stanford University, 1987.

[Ye89] Yinyu Ye, *An extension of Karmarkar's algorithm and the trust region method for quadratic programming*, Progress in mathematical programming (Pacific Grove, CA, 1987), Springer, New York, 1989, pp. 49–63.

[Ye92] Y. Ye, *On affine scaling algorithms for nonconvex quadratic programming*, Mathematical Programming **56** (1992), 285–300.

[Ye98] ———, *On the complexity of approximating a KKT point of quadratic programming*, Mathematical Programming **80** (1998), 195–211.

[YLQ03] Yu-Fei Yang, Dong-Hui Li, and Liqun Qi, *A feasible sequential linear equation method for inequality constrained optimization*, SIAM J. Optim. **13** (2003), no. 4, 1222–1244 (electronic).

[YT89] Yinyu Ye and Edison Tse, *An extension of Karmarkar's projective algorithm for convex quadratic programming*, Math. Programming **44** (1989), no. 2, (Ser. A), 157–179.