

# 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<sup>†</sup>  
Department of Electrical and Computer Engineering  
and Institute for Systems Research  
University of Maryland  
College Park, MD 20742  
[andre@umd.edu](mailto:andre@umd.edu)

## 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) multiplier estimates are components of the Newton (or quasi-Newton)

---

\*The work of the first author was supported in part by the School of Computational Science of the 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.

<sup>†</sup>Corresponding author.

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, to appear], 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 mild assumptions for both algorithms. Numerical results on randomly generated problems suggest that the proposed algorithms hold promise, even when the nondegeneracy assumptions are violated.

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

## 1 Introduction

Consider the quadratic programming problem

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

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, a primal-dual algorithm was first proposed and analyzed in [MA89]. Interior-point methods have also been proposed for the computation of local solutions to general, nonlinear programming problems (e.g., [Yam98, ETTZ96, GOW98, FG98, BGN00, VS99, PTH88, TWB<sup>+</sup>03]), 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 ([Ye92, Tse04, TY02]) on Dikin’s algorithm [Dik67], and that of Bonnans and Bouhtou [BB95] and of Coleman and Liu [CL99], all of which consider trust-region-based algorithms.

In this paper, we propose and analyze two interior-point methods, one of the affine scaling type, one of the barrier type, for the solution of problem  $(P)$ . Like the algorithms of [Ye92, Tse04, TY02, BB95, CL99], 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 enhanced, affine-scaling version of the barrier-based general nonlinear programming algorithm of [PTH88] and [TWB<sup>+</sup>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<sup>+</sup>03], in early iterations, the Newton-KKT direction is replaced by a quasi-Newton direction obtained by replacing  $H$  (or the Hessian of the Lagrangian in the general case of [TWB<sup>+</sup>03]) by 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<sup>+</sup>03] is adopted for the KKT multiplier estimates. Global convergence as well as local quadratic convergence of the constructed sequence to KKT points is proved under mild 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<sup>+</sup>03] and indeed, reduces to Algorithm A1 if the rule assigning a value to the barrier-parameter  $\mu^k$  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<sup>+</sup>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<sup>+</sup>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<sup>+</sup>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 quadratic convergence of the constructed sequence to KKT points is proved under mild assumptions.

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.

## 2 Problem Definition and Algorithm Statement

Let  $I = \{1, \dots, m\}$ , where  $m$  is the number of rows of  $A$ , and, for  $i \in I$ , let  $a_i$  be the *transpose* of the  $i$ th row of  $A$ , let  $b_i$  be the  $i$ th 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  $\Delta x$ ,

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

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

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

and the strictly feasible  $\mathcal{F}^0$  set by

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

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

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

(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  $(\Delta x, \Delta z)$

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

which we refer to as the Newton-KKT equations. It will be shown that, under mild assumptions, and after possible adjustment of  $H$ , if  $x \in \mathcal{F}^0$  (*strict primal feasibility*) and  $z_i > 0$  for all  $i \in I$  (*strict “dual feasibility”*),

then the solution  $\Delta x$  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 solution 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 satisfy, for some  $C > 0$ ,

$$\begin{aligned} \|x^+ - (x + \Delta x)\| &\leq C(\|\Delta x\| + \|\Delta z\|)^2 \\ z_i^+ &= \max\{z_i + \Delta z_i, \|\Delta x\|^2 + \|\Delta z\|^2\}, \quad i \in I. \end{aligned} \tag{4}$$

The quadratic convergence of Newton's method is then preserved because each component of the next iterate is "close enough" either to the corresponding component of the solution of the nonlinear system of equations being solved, or to the corresponding component of the next iterate given by the pure Newton iteration. Updates closely related to these are used in Algorithm A1, which we now state. Note that the algorithm statement implicitly requires that  $\mathcal{F}^0$  be nonempty.

**Algorithm A1.**

*Parameters.*  $\beta \in (0, 1)$ ,  $\underline{z} > 0$ ,  $\bar{z} > \underline{z}$ ,  $\sigma > 0$ ,  $\gamma > 1$ .

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

*Step 0. Initialization.* Set  $k := 0$ . Set  $\bar{I} := \emptyset$ . Set  $\bar{\alpha}_i := 0$ ,  $i = 1, \dots, m$ . Set  $\bar{E} := I$ .<sup>1</sup>

*Step 1. Computation of modified Hessian.* If  $H \succeq \sigma I$ , set  $E^k := 0$  and  $W^k := H$ . Else, set  $W^k := H + E^k$ , with  $E^k \succeq 0$  "small" such that  $W^k + \sum_{i=1}^m \frac{z_i^k}{|g_i(x^k)|} a_i a_i^T \succeq \sigma I$ ; specifically:

- If  $\frac{z_i^k}{|g_i(x^k)|} \leq \bar{\alpha}_i$  for some  $i \in \bar{I}$  or ( $\bar{E} \neq 0$  and  $\bar{I} = \emptyset$ ) or ( $\bar{E} \neq 0$  and  $\frac{z_i^k}{|g_i(x^k)|} \geq \gamma^2 \bar{\alpha}_i$  for some  $i \in \bar{I}$ ) then

---

<sup>1</sup>The initial values assigned to the components of  $\bar{\alpha}$  and to  $\bar{E}$  are immaterial as long as  $\bar{E} \neq 0$ .

- Set  $\bar{I} := \{i : \frac{z_i^k}{|g_i(x^k)|} \geq 1\}$ ; set  $\bar{\alpha}_i := \frac{1}{\gamma} \frac{z_i^k}{|g_i(x^k)|}$ ,  $i \in \bar{I}$ ; if  $H + \sum_{\bar{I}} \bar{\alpha}_i a_i a_i^T \succeq \sigma I$ , set  $\bar{E} := 0$ , else set  $\bar{E} \succeq 0$  “small” such that  $H + \sum_{\bar{I}} \bar{\alpha}_i a_i a_i^T + \bar{E} \succeq \sigma I$ .

- Set  $E^k := \bar{E}$ .

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

$$W^k \Delta x + A^T \zeta = -\nabla f(x^k) \quad (5a)$$

$$z_i^k \langle a_i, \Delta x \rangle + g_i(x^k) \zeta_i = 0 \quad \forall i \in I. \quad (5b)$$

Set  $\Delta z^k := \zeta^k - z^k$ . If  $\Delta x^k = 0$ , stop.

*Step 3. Updates.*

(i) Set

$$\bar{t}^k := \begin{cases} \infty & \text{if } \langle a_i, \Delta x^k \rangle \leq 0 \quad \forall i \in I, \\ \min \left\{ \frac{|g_i(x^k)|}{\langle a_i, \Delta x^k \rangle} : \langle a_i, \Delta x^k \rangle > 0, i \in I \right\} & \text{otherwise.} \end{cases} \quad (6)$$

Set

$$t^k := \min \left\{ \max\{\beta \bar{t}^k, \bar{t}^k - \|\Delta x^k\|\}, 1 \right\}. \quad (7)$$

Set  $x^{k+1} := x^k + t^k \Delta x^k$ .

(ii) Set

$$z_i^{k+1} := \min \left\{ \max \left\{ \min \{ \|\Delta x^k\|^2 + \|\Delta z^k\|^2, \underline{z} \}, \zeta_i^k \right\}, \bar{z} \right\}, \quad \forall i \in I. \quad (8)$$

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

□

**Remark.** In Step 3(i),  $\bar{t}^k$  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. In Step 3(ii),  $\bar{z}$  is introduced to ensure boundedness of  $\{z^k\}$ . (This is used in the proof of Lemma 3.2; it may not be important in practice.) Notice that, in contrast to usual practice,  $z_i^{k+1}$  is not obtained by a step in the direction of  $\Delta z^k$ . Update rule (8) is inspired from [PTH88]. Numerical experiments (see Section 6) suggest that, in the context of indefinite quadratic programming, this update rule often performs better than an update along  $\Delta z^k$ . Finally, it follows from Step 3

that any  $x^k$  and  $z^k$  computed by Algorithm A1 are primal and dual strictly feasible, namely

$$g_i(x^k) < 0, \quad \forall i \in I, \quad (9)$$

$$z_i^k > 0, \quad \forall i \in I. \quad (10)$$

◇

For later reference, let

$$M(x, z, W) := \begin{bmatrix} W & A^T \\ \text{diag}(z_i)A & \text{diag}(g_i(x)) \end{bmatrix} \quad (11)$$

denote the full Newton-KKT matrix, so that the linear system (5a)-(5b) appearing in Step 2 of Algorithm A1 reads

$$M(x^k, z^k, W^k) \begin{bmatrix} \Delta x \\ \zeta \end{bmatrix} = \begin{bmatrix} -\nabla f(x^k) \\ 0 \end{bmatrix}. \quad (12)$$

Below we will also make use of a “general” instance of this system, viz.,

$$M(x, z, W) \begin{bmatrix} \Delta x \\ \zeta \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}. \quad (13)$$

Also define the condensed Newton-KKT matrix

$$S^k := W^k - \sum_{i=1}^m \frac{z_i^k}{g_i(x^k)} a_i a_i^T, \quad (14)$$

which is the Schur complement of  $\text{diag}(g_i(x^k))$  in  $M(x^k, z^k, W^k)$ .

**Remark.** Step 1 in Algorithm A1 ensures that  $S^k \succeq \sigma I$ , as we now explain. The case when  $H \succeq \sigma I$  is clear. Assume that  $H \not\succeq \sigma I$ . Then, after completion of Step 1, the relation  $W^k + \sum_{\bar{I}} \bar{\alpha}_i a_i a_i^T \succeq \sigma I$  is satisfied, and  $\frac{z_i^k}{|g_i(x^k)|} \geq \bar{\alpha}_i$  for all  $i \in \bar{I}$ . Consequently,  $S^k \succeq W^k + \sum_{\bar{I}} \bar{\alpha}_i a_i a_i^T \succeq \sigma I$ , hence the desired conclusion. Next, under appropriate assumptions,  $\{E^k\}$  converges to zero (see Lemma 3.10 below) and Newton-related superlinear convergence can set in. Step 1 also ensures that  $\{W^k\}$  is bounded, for any reasonable interpretation of “small” in Step 1 of Algorithm A1 (see, e.g., (46) below). Finally, an interesting feature of Step 1 is that  $\bar{E}$  need not be updated at each step: update occurs in particular if one of the ratios  $\frac{z_i^k}{|g_i(x^k)|}$ ,  $i \in \bar{I}$ , leaves

the interval  $(\bar{\alpha}_i, \gamma^2 \bar{\alpha}_i)$  through its lower end, or if  $\bar{E} \neq 0$  and one of these ratios leave the corresponding interval through its upper end.  $\diamond$

It is readily checked that the unique solution of (5a)-(5b) is given by

$$\Delta x^k = -(S^k)^{-1} \nabla f(x^k) \quad (15a)$$

$$\zeta^k = -\text{diag} \left( \frac{z_i^k}{g_i(x^k)} \right) A \Delta x^k. \quad (15b)$$

Expression (15a) shows that Algorithm A1 belongs to the affine scaling family. Since  $S^k \succeq \sigma I$ , it also follows from (15a) that

$$\langle \nabla f(x^k), \Delta x^k \rangle \leq -\sigma \|\Delta x^k\|^2 \quad (16)$$

which shows that  $\Delta x^k$  is a direction of descent for  $f$  at  $x^k$ . Since  $z_i^k/g_i(x^k) < 0$  for all  $i$  and  $k$ , it follows from (15b) that

$$\langle \zeta^k, A \Delta x^k \rangle \geq 0 \quad (17)$$

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  $\mathcal{F}^0$ , 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 (5) 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 at the stopping criterion in Step 2, say at  $x_{k_0}$ . In such case, since  $g_i(x^{k_0}) < 0$ , substitution of  $\Delta x^{k_0} = 0$  in (5) 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).  $\square$

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 to this analysis, we conclude this section with three lemmas which will be of repeated use. The first lemma further characterizes the relationship between  $\Delta x$  vanishing and  $x$  being stationary.

**Lemma 2.2** *Let  $(\Delta x, \zeta)$  satisfy (13) for some  $x, z$ , and  $W$ . Then (i) if  $x \in \mathcal{F}$  and  $\Delta x = 0$ , then  $x$  is stationary for (P) and  $\zeta$  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  $\zeta$  is the unique multiplier vector associated with  $x$ .*

*Proof.* To prove the first claim, simply substitute  $\Delta x = 0$  into (13). Concerning the second claim, let  $\eta$  be a multiplier vector associated with stationary point  $x$ . It is readily verified that  $(0, \eta)$  satisfies (13). The claim then follows from nonsingularity of  $M(x, z, W)$ .  $\square$

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

**Lemma 2.3**<sup>2</sup> *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*

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

where

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

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<sup>+</sup>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**<sup>3</sup> *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 (14) satisfies  $S^k \succeq \sigma I$  for all  $k$  and that  $z_j^* > 0$  for all  $j \in I(x^*)$ . Then  $M(x^*, z^*, W^*)$  is nonsingular.*

---

<sup>2</sup>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.

<sup>3</sup>Much as Lemma 2.3, Lemma 2.4 holds under weaker hypotheses, but the stated version is adequate for the purpose of this paper.

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

$$\langle a_i, v \rangle = 0 \quad \forall i \in I(x^*). \quad (18)$$

It then follows from (14) and positive semidefiniteness of  $S^k - \sigma I$ , by adding terms that vanish in view of (18), that for all  $k$

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

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

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

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 3.7) 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  $\{z^k\}$  be as constructed by Algorithm A1. If  $\Delta x^k \neq 0$ , then, for all  $t \in (0, 2)$ ,*

$$f(x^k + t\Delta x^k) \leq f(x^k) - t(1 - t/2)\langle \Delta x^k, H\Delta x^k \rangle - t\langle \zeta^k, A\Delta x^k \rangle < f(x^k) \quad (19)$$

and, for all  $t > 0$ ,

$$g_i(x^k + t\Delta x^k) = g_i(x^k) + t\langle a_i, \Delta x^k \rangle < g_i(x^k) \quad \forall i \text{ s.t. } \zeta_i^k < 0. \quad (20)$$

*Proof.* First, in view of (1), it follows from (5a) that, for any  $t \geq 0$ ,

$$\begin{aligned}
f(x^k + t\Delta x^k) &= f(x^k) + t\langle \nabla f(x^k), \Delta x^k \rangle + \frac{1}{2}t^2\langle \Delta x^k, H\Delta x^k \rangle & (21) \\
&= f(x^k) - t(\langle \Delta x^k, W^k \Delta x^k \rangle + \langle \Delta x^k, A^T \zeta^k \rangle) + \frac{1}{2}t^2\langle \Delta x^k, H\Delta x^k \rangle \\
&= f(x^k) - t(\langle \Delta x^k, H\Delta x^k \rangle + \langle \zeta^k, A\Delta x^k \rangle) + \frac{1}{2}t^2\langle \Delta x^k, H\Delta x^k \rangle \\
&\quad - t\langle \Delta x^k, (W^k - H)\Delta x^k \rangle \\
&\leq f(x^k) - t(1 - t/2)\langle \Delta x^k, H\Delta x^k \rangle - t\langle \zeta^k, A\Delta x^k \rangle,
\end{aligned}$$

where we have used the fact that, by construction,  $W^k - H$  is positive semidefinite for all  $k$ . The last inequality yields the first inequality in (19). Next, if  $\langle \Delta x^k, H\Delta x^k \rangle > 0$ , it follows from (17) that, for any  $t \in (0, 2)$ ,

$$f(x^k) - t(1 - t/2)\langle \Delta x^k, H\Delta x^k \rangle - t\langle \zeta^k, A\Delta x^k \rangle < f(x^k),$$

and the second inequality in (19) then follows. On the other hand, if  $\langle \Delta x^k, H\Delta x^k \rangle \leq 0$ , then, in view of (15a) and positive definiteness of  $S^k$ , it follows from (21) that  $f(x^k + t\Delta x^k) < f(x^k)$ , for all  $t > 0$ , and the second inequality in (19) follows in that case as well. Finally, since  $g$  is linear,

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

Since  $z_i^k > 0$  and  $g_i(x^k) < 0$  for all  $i \in I$ , it follows from (5b) that  $\langle a_i, \Delta x^k \rangle < 0$  whenever  $\zeta_i^k < 0$ , proving (20).  $\square$

With these results in hand, we now proceed to show that the accumulation points of  $\{x^k\}$  are stationary points of  $(P)$ . 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  $\{x_k\}_{k \in K} \rightarrow x^*$  for some  $x^*$ , one has either (i)  $\{\Delta x^k\}_{k \in K} \rightarrow 0$ , or (ii) there exists an infinite index set  $K' \subseteq K$  such that  $\{\Delta x^{k-1}\}_{k \in K'} \rightarrow 0$  (Lemma 3.2). In case (i), it follows from Lemma 3.3 below that  $x^*$  is stationary. In case (ii),  $\{x^{k-1}\}_{k \in K'} \rightarrow x^*$  (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 local convergence analysis of the next subsection.

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

$$\inf\{\|\Delta x^{k-1}\|^2 + \|\Delta z^{k-1}\|^2 : k \in K\} > 0. \quad (22)$$

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

*Proof.* Proceeding by contradiction, assume that  $\{x^k\}_{k \in K} \rightarrow x^*$ , that (22) holds, and that, for some infinite index set  $K' \subseteq K$ ,  $\inf_{k \in K'} \|\Delta x^k\| > 0$ . Since  $\{z^k\}$  and  $\{W^k\}$  are bounded, we may assume, without loss of generality, that for some  $z^*$  and  $W^*$ ,

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

In view of (22) and (8), it follows that  $z^* > 0$ . Since in view of Lemma 2.4  $M(x^*, z^*, W^*)$  is nonsingular, it follows that, for some  $\Delta x^*$  and  $z^*$ , with  $\Delta x^* \neq 0$  (since  $\inf_{k \in K'} \|\Delta x^k\| > 0$ ),

$$\begin{aligned} \{\Delta x^k\} &\rightarrow \Delta x^* \quad \text{as } k \rightarrow \infty, \quad k \in K', \\ \{\zeta^k\} &\rightarrow z^* \quad \text{as } k \rightarrow \infty, \quad k \in K'. \end{aligned}$$

On the other hand, it follows from (5b) and (6) that

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

Since, on  $K'$ ,  $\{\zeta^k\}$  is bounded (because it converges) 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  $\underline{t} > 0$ .

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

$$f(x^{k+1}) \leq f(x^k) - \delta \tag{23}$$

infinitely many times. We show that it 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  $k$ , it follows from Proposition 3.1 that, for  $k \in K'$ ,  $k$  large enough,

$$f(x^k + t^k \Delta x^k) < f(x^k) - (\underline{t} - \frac{1}{2}\underline{t}^2)\langle \Delta x^k, H \Delta x^k \rangle - t^k \langle \zeta^k, A \Delta x^k \rangle. \tag{24}$$

The claim then directly follows from (17). Suppose now that  $\langle \Delta x^*, H \Delta x^* \rangle \leq 0$ . Then  $\langle \Delta x^k, H \Delta x^k \rangle \leq \frac{1}{2}\underline{t}\sigma\|\Delta x^*\|^2$  for  $k \in K'$ ,  $k$  large enough. Also, in view of (16),  $\langle \nabla f(x^k), \Delta x^k \rangle \leq -\frac{1}{2}\sigma\|\Delta x^*\|^2$  for  $k \in K'$ ,  $k$  large enough. Since  $t_k \in (\underline{t}, 1]$  for  $k \in K'$ , it follows from (1) that (23) again holds on  $K'$ .  $\square$

**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  $\Delta x^k \rightarrow 0$  as  $k \rightarrow \infty$ ,  $k \in K$ . Since  $\{z^k\}$  is bounded, it follows from (5b) that  $\zeta_i^k \rightarrow 0$  as  $k \rightarrow \infty$ ,  $k \in K$ , for all  $i \notin I(x^*)$ . Since in view of (5a),  $\{A^T \zeta^k\}$  converges on  $K$ , it follows from Assumption 1 that  $\{\zeta^k\}$  converges on  $K$ , say to  $z^*$ . Taking limits in (5a)–(5b) as  $k \rightarrow \infty$ ,  $k \in K$ , and using the fact that  $\{z^k\}$  is bounded on  $K$  yields

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

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

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

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  $\{x^k\}_{k \in K} \rightarrow x^*$  for some  $x^*$ . If  $\{\Delta x^{k-1}\}_{k \in K} \rightarrow 0$  then  $\{x^{k-1}\}_{k \in K} \rightarrow x^*$ .*

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  $\{\Delta z^k\}$  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  $\{\Delta z^{k-1}\}_{k \in K}$  tend to zero. Then  $x^*$  is a KKT point.*

*Proof.* In view of Lemma 3.4,  $\{x^{k-1}\}_{k \in K} \rightarrow x^*$ , 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  $\{\Delta z^{k-1}\}_{k \in K}$  tends to zero and  $z^{k-1} > 0$  for all  $k$ , it follows that  $z^* \geq 0$ , thus that  $x^*$  is a KKT point.  $\square$

**Assumption 2** *All stationary points are isolated.*

**Theorem 3.7** *Under Assumptions 1 and 2, every accumulation point of  $\{x^k\}$  constructed by Algorithm A1 is a KKT point.<sup>4</sup>*

The proof is identical to the proof of Theorem 3.11 in [PTH88], with references to Lemmas 3.7, 3.8 and 3.9 replaced by references to Lemmas 3.3, 3.6 and 3.2 above, references to Assumptions A5 and A7 in [PTH88] replaced by references to Assumptions 1 and 2 above, and  $\tilde{\Delta}x_k$  set to zero (and subscripts  $k$  are changed to superscripts). Also, at the end of the proof, “ $i_0 \in J_k$ ” should be replaced by “ $\zeta_{i_0}^k < 0$ ”, and the reference to Equation (2.11) of [PTH88] should be replaced by one to (20).

## 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 3.8** *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  $\{x^k\}_{k \in K} \rightarrow x^*$ . Then  $\{\Delta x^k\}_{k \in K} \rightarrow 0$ .*

*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  $\{\Delta z^{k-1}\}_{k \in K''}$  go to

---

<sup>4</sup>Strictly speaking, for the purpose of this theorem, Assumption 2 can be replaced by the milder assumption that *non-KKT* stationary points are isolated.

zero. It follows from Lemma 3.4 that  $\{x^{k-1}\}_{k \in K''} \rightarrow x^*$ . In view of Lemma 3.3 it follows that  $\{\zeta^{k-1}\}_{k \in K''} \rightarrow z^*$  where  $z^*$  is the multiplier vector associated to  $x^*$ . It then follows from (8) that, for all  $j$ ,  $\{z_j^k\}_{k \in K''} \rightarrow \hat{z}_j := \min\{z_j^*, \bar{z}\}$  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.  $\square$

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

**Proposition 3.9** *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)  $\{\Delta x^k\} \rightarrow 0$ , (ii)  $\{\zeta^k\} \rightarrow z^*$ , and (iii)  $\{z_j^k\} \rightarrow \min\{z_j^*, \bar{z}\}$  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  $\epsilon$  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  $\{x^k\}_{k \in K} \rightarrow x^*$ , otherwise the closed  $\epsilon$ -ball would contain an accumulation point other than  $x^*$  and this point would be a KKT point by Theorem 3.7. In particular,  $\|x^k - x^*\| < \epsilon/4$  for all  $k \in K$  large enough. On the other hand, it follows from Lemma 3.8 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  $\epsilon$ -ball for  $k \in K$  large enough, a contradiction. Thus, the first claim holds. Claim (i) then follows from Lemma 3.8 and claim (ii) follows from Lemma 3.3. Finally claim (iii) follows from a careful inspection of (8).  $\square$

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

$$\langle \Delta x, H \Delta x \rangle > 0 \text{ for all } \Delta x \text{ such that } \langle a_i, \Delta x \rangle = 0 \ \forall i \in I(x^*). \quad (25)$$

In turn, under Assumption 3, (25) implies Assumption 4.

**Lemma 3.10** *Let  $\{W^k\}$  be as constructed by Algorithm A1. Under Assumptions 1–4, if  $\sigma$  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 3.9(iii) that  $\{z_j^k/|g_j(x^k)|\}$  goes to  $\infty$  for all  $j \in I(x^*)$ . It then follows from (25) and Step 1 in Algorithm A1 that  $E^k = 0$  for all  $k$  large enough.  $\square$

To prove Q-quadratic convergence of  $\{(x^k, z^k)\}$ , the following property of Newton's method, taken from [TZ94], will be used. Here and in the proof of Theorem 3.12, given  $r > 0$  and  $s^*$  in some Euclidean space,  $B(s^*, r)$  denotes the closed ball  $\{s : \|s - s^*\| \leq r\}$ .

**Proposition 3.11** *Let  $F : \mathcal{R}^n \rightarrow \mathcal{R}^n$  be twice continuously differentiable and let  $w^* \in \mathcal{R}^n$  and  $\rho > 0$  be such that  $F(w^*) = 0$  and  $\frac{\partial F}{\partial w}(w)$  is invertible whenever  $w \in B(w^*, \rho)$ . Let  $\delta^N : B(w^*, \rho) \rightarrow \mathcal{R}^n$  be the Newton increment  $\delta^N(w) = -(\frac{\partial F}{\partial w}(w))^{-1} F(w)$ . Then given any  $c_1 > 0$  there exists  $c_2 > 0$  such that the following statement holds:*

*For every  $w \in B(w^*, \rho)$  and every  $w^+ \in \mathcal{R}^n$  for which, for each  $i \in \{1, \dots, n\}$ , either*

$$(i) \ |w_i^+ - w_i^*| \leq c_1 \|\delta^N(w)\|^2 \quad \forall w \in B(w^*, \rho)$$

*or*

$$(ii) \ |w_i^+ - (w_i + \delta_i^N(w))| \leq c_1 \|\delta^N(w)\|^2 \quad \forall w \in B(w^*, \rho),$$

*it holds that*

$$\|w^+ - w^*\| \leq c_2 \|w - w^*\|^2 \quad \forall w \in B(w^*, \rho). \quad (26)$$

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 3.12** *Let  $\{x^k\}$  and  $\{z^k\}$  be as constructed by Algorithm A1. Suppose Assumptions 1–4 hold. Then, if  $z_i^* < \bar{z} \forall i \in I$  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 3.11 hold with  $F : R^n \times R^m \rightarrow R^n \times R^m$  given by

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

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 (13), is the Newton direction for the solution of  $F(x, z) = 0$ . With reference to Proposition 3.11, let  $\rho > 0$  be such that  $M(x, z, H)$  is nonsingular for all  $(x, z) \in B((x^*, z^*), \rho)$ . (In view of (25), Assumption 3, and Lemma 2.3, such  $\rho$  exists.) Since, in view of the stated assumption on  $z^*$  and of Proposition 3.9,  $\{(x^k, z^k)\} \rightarrow (x^*, z^*)$ , it follows from Lemma 3.10 that there exists  $k_0$  such that, for all  $k \geq k_0$ ,  $(x^k, z^k) \in B((x^*, z^*), \rho)$  and  $W^k = H$ . Now, with the aim of verifying conditions (i)/(ii) of Proposition 3.11 along  $\{(x^k, z^k)\}$  with  $F$  as specified above, let us first consider  $\{z^k\}$ , updated in Step 3(ii) of Step 1 of Algorithm A1. Since  $z_i^* < \bar{z}$ , it follows that, for  $i \in I(x^*)$  (so  $z_i^* > 0$ ),  $z_i^{k+1}$  is equal to  $\zeta_i^k$  for  $k$  large enough, so that condition (ii) in Proposition 3.11 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  $z_i^{k+1} = \|\Delta x^k\|^2 + \|\Delta z^k\|^2$ . In the latter case, since  $z_i^* = 0$ , condition (i) in Proposition 3.11 holds with any  $c_1 \geq 1$ . Next, consider  $\{x^k\}$ . For  $i \notin I(x^*)$ , in view of Proposition 3.9, we have

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

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 (5b)  $\text{sign}(\langle a_i, \Delta x^k \rangle) = \text{sign}(\zeta_i^k) = 1$ , 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

$$t^k = \min \left\{ 1, \frac{z_{i_k}^k}{\zeta_{i_k}^k} - \|\Delta x^k\| \right\} \quad (27)$$

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

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

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 3.11 holds (with  $c_1 = 1 + C$ ). The claim then follows from Lemma 3.9 and Proposition 3.11.  $\square$

## 4 Refinement: A Barrier-Based Algorithm

While the 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  $\mu$  rather than 0. It is well known that the resulting iteration can be viewed as a “primal-dual” *barrier* iteration. Typically,  $\mu$  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 (e.g., [Yam98, ETTZ96, GOW98, FG98, BGN00, VS99, PTH88, TWB<sup>+</sup>03]).

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<sup>+</sup>03], as well as from the related algorithm of [TZ94] for linear programming and convex quadratic programming. At iteration  $k$ , the value  $\mu^k$  of the barrier parameter

is determined via computation of the affine scaling direction  $\Delta x^k$  (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)  $\mu^k$  goes to zero fast enough, as  $\Delta x^k$  goes to zero, that quadratic convergence can be maintained. Specifically  $\mu^k$  is assigned the value  $\varphi^k \|\Delta x^k\|^\nu z_{\min}^k$ , where  $\nu > 2$  is prescribed,  $z_{\min}^k > 0 := \min_i z_i^k$ , and  $\varphi^k$  is the largest scalar in  $(0, \bar{\varphi}]$ ,  $\bar{\varphi} > 0$  prescribed, such that

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

where  $\theta \in (0, 1)$  is prescribed; it then follows from (16) that

$$\langle \Delta x^{\mu,k}, \nabla f(x^k) \rangle \leq -\theta \sigma \|\Delta x^k\|^2. \quad (28)$$

As noted in [BT03], such  $\varphi^k$  can be expressed as

$$\varphi^k = \begin{cases} \bar{\varphi} & \text{if } \sum_j \frac{\zeta_j^k}{z_j^k} \leq 0 \\ \min \left\{ \frac{(1-\theta) |\langle \Delta x^k, \nabla f(x^k) \rangle|}{\|\Delta x^k\|^\nu z_{\min}^k \sum_j \frac{\zeta_j^k}{z_j^k}}, \bar{\varphi} \right\} & \text{otherwise.} \end{cases} \quad (29)$$

With  $\mu^k$  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  $\Delta x^k$  and  $\zeta^k$  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  $t \in (0, 2)$ , 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  $\Delta x$ ,

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

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

The proposed algorithm is as follows. Note that, again, the algorithm statement implicitly assumes that  $\mathcal{F}^0$  is nonempty.

### Algorithm A2.

*Parameters.*  $\beta \in (0, 1)$ ,  $\underline{z} \in (0, 1)$ ,  $\bar{z} > \underline{z}$ ,  $\sigma > 0$ ,  $\gamma > 1$ ,  $\theta \in (0, 1)$ ,  $\bar{\varphi} > 0$ ,  $\nu > 2$ ,  $\psi \in (1, 2)$ .

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

*Step 0. Initialization.* Set  $k := 0$ . Set  $\bar{I} := \emptyset$ . Set  $\bar{\alpha}_i := 0$ ,  $i = 1, \dots, m$ . Set  $\bar{E} := \mathbf{I}$ .<sup>5</sup>

*Step 1. Computation of modified Hessian.* If  $H \succeq \sigma \mathbf{I}$ , set  $E^k := 0$  and  $W^k := H$ . Else, set  $W^k := H + E^k$ , with  $E^k \succeq 0$  “small” such that  $W^k + \sum_{i=1}^m \frac{z_i^k}{|g_i(x^k)|} a_i a_i^T \succeq \sigma \mathbf{I}$ ; specifically:

- If  $\frac{z_i^k}{|g_i(x^k)|} \leq \bar{\alpha}_i$  for some  $i \in \bar{I}$  or ( $\bar{E} \neq 0$  and  $\bar{I} = \emptyset$ ) or ( $\bar{E} \neq 0$  and  $\frac{z_i^k}{|g_i(x^k)|} \geq \gamma^2 \bar{\alpha}_i$  for some  $i \in \bar{I}$ ) then
  - Set  $\bar{I} := \{i : \frac{z_i^k}{|g_i(x^k)|} \geq 1\}$ ; set  $\bar{\alpha}_i := \frac{1}{\gamma} \frac{z_i^k}{|g_i(x^k)|}$ ,  $i \in \bar{I}$ ; if  $H + \sum_{\bar{I}} \bar{\alpha}_i a_i a_i^T \succeq \sigma \mathbf{I}$ , set  $\bar{E} := 0$ , else set  $\bar{E} \succeq 0$  “small” such that  $H + \sum_{\bar{I}} \bar{\alpha}_i a_i a_i^T + \bar{E} \succeq \sigma \mathbf{I}$ ;
- Set  $E^k := \bar{E}$ .

*Step 2. Computation of a search direction.*

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

$$W^k \Delta x + A^T \zeta = -\nabla f(x^k) \quad (30a)$$

$$z_i^k \langle a_i, \Delta x \rangle + g_i(x^k) \zeta_i = 0 \quad \forall i \in I. \quad (30b)$$

Set  $\Delta z^k := \zeta^k - z^k$ . 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  $\varphi^k$  is given by

$$\varphi^k = \begin{cases} \bar{\varphi} & \text{if } \sum_j \frac{\zeta_j^k}{z_j^k} \leq 0 \\ \min \left\{ \frac{(1-\theta) |\langle \Delta x^k, \nabla f(x^k) \rangle|}{\|\Delta x^k\|^\nu z_{\min}^k \sum_j \frac{\zeta_j^k}{z_j^k}}, \bar{\varphi} \right\} & \text{otherwise.} \end{cases} \quad (31)$$

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

$$W^k \Delta x^\mu + A^T \zeta^\mu = -\nabla f(x^k) \quad (32a)$$

$$z_i^k \langle a_i, \Delta x^\mu \rangle + g_i(x^k) \zeta_i^\mu = -\mu^k \quad \forall i \in I. \quad (32b)$$

---

<sup>5</sup>Again, the initial values assigned to the components of  $\bar{\alpha}$  and to  $\bar{E}$  are immaterial, as long as  $\bar{E} \neq 0$ .

Set  $\Delta z^{\mu,k} := \zeta^{\mu,k} - z^k$ .

*Step 3. Updates.*

(i) Set

$$\bar{t}^k := \begin{cases} \infty & \text{if } \langle a_i, \Delta x^{\mu,k} \rangle \leq 0 \quad \forall i \in I, \\ \min \left\{ \frac{|g_i(x^k)|}{\langle a_i, \Delta x^{\mu,k} \rangle} : \langle a_i, \Delta x^{\mu,k} \rangle > 0, i \in I \right\} & \text{otherwise.} \end{cases} \quad (33)$$

Set

$$t^k := \begin{cases} \min \left\{ \max\{\beta \bar{t}^k, \bar{t}^k - \|\Delta x^{\mu,k}\|\}, 1 \right\} & \text{if } \langle \Delta x^{\mu,k}, H \Delta x^{\mu,k} \rangle \leq 0, \\ \min \left\{ \max\{\beta \bar{t}^k, \bar{t}^k - \|\Delta x^{\mu,k}\|\}, 1, \psi \frac{|\langle \Delta x^{\mu,k}, \nabla f(x^k) \rangle|}{\langle \Delta x^{\mu,k}, H \Delta x^{\mu,k} \rangle} \right\} & \text{otherwise.} \end{cases} \quad (34)$$

Set  $x^{k+1} := x^k + t^k \Delta x^{\mu,k}$ .

(ii) Set

$$z_i^{k+1} := \min\{\max\{\min\{\|\Delta x^{\mu,k}\|^2 + \|\Delta z^{\mu,k}\|^2, \underline{z}\}, \zeta_i^{\mu,k}\}, \bar{z}\}, \quad \forall i \in I. \quad (35)$$

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

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 10 lemmas, propositions and theorems is parallel to that used in Section 3. We start with a modified (and refined) version of Proposition 3.1.

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

$$f(x^k + t \Delta x^{\mu,k}) < f(x^k) \quad \forall t > 0, \quad (36)$$

and otherwise

$$f(x^k + t \Delta x^{\mu,k}) < f(x^k) \quad \text{iff } t \in \left( 0, 2 \frac{|\langle \Delta x^{\mu,k}, \nabla f(x^k) \rangle|}{\langle \Delta x^{\mu,k}, H \Delta x^{\mu,k} \rangle} \right); \quad (37)$$

and (ii)

$$g_i(x^k + t \Delta x^{\mu,k}) = g_i(x^k) + t \langle a_i, \Delta x^{\mu,k} \rangle < g_i(x^k) \quad \forall t > 0, \forall i \text{ s.t. } \zeta_i^{\mu,k} < 0. \quad (38)$$

*Proof.* Claim (i) follows from (28) via routine manipulations on the quadratic function

$$f(x^k + t\Delta x^{\mu,k}) = f(x^k) + t\langle \nabla f(x^k), \Delta x^{\mu,k} \rangle + \frac{1}{2}t^2\langle \Delta x^{\mu,k}, H\Delta x^{\mu,k} \rangle. \quad (39)$$

Further, since  $g$  is linear,

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

Since  $z_i^k > 0$  for all  $i \in I$ , it follows from (32b) that  $\langle a_i, \Delta x^{\mu,k} \rangle < 0$  whenever  $\zeta_i^{\mu,k} < 0$ , proving claim (ii).  $\square$

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

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

$$\inf\{\|\Delta x^{\mu,k-1}\|^2 + \|\Delta z^{\mu,k-1}\|^2 : k \in K\} > 0. \quad (40)$$

*Then  $\{\Delta x^{\mu,k}\}_{k \in K} \rightarrow 0$ .*

*Proof.* As in the proof of Lemma 3.2, proceed by contradiction by assuming that, for some infinite index set  $K' \subseteq K$ ,  $\inf_{k \in K'} \|\Delta x^{\mu,k}\| > 0$ . 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,  $\{z^k\}_{k \in K'} \rightarrow z^*$  for some  $z^* > 0$ ,  $\{W^k\}_{k \in K'} \rightarrow W^*$  for some  $W^*$ ,  $M(x^*, z^*, W^*)$  is nonsingular, and  $\{\Delta x^k\}_{k \in K'}$  is bounded. 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. On the other hand, it follows from (32b) 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 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 often on  $K'$ , then in view of (34), 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 (28), for all such  $k$ ,

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

It remains to show that  $\{\|\Delta x^k\|\}_{k \in K''}$  is bounded away from zero to conclude from (34) that  $\{t^k\}_{k \in K''}$  is bounded away from zero. To this end, proceeding by contradiction, suppose that  $\inf_{k \in K''} \|\Delta x^k\| = 0$ , i.e.,  $\{\Delta x^k\}_{k \in K''} \rightarrow 0$  for some  $K''' \subseteq K''$ . Since  $\{z^k\}_{k \in K'}$  is bounded, it follows from Step 2(ii) of Algorithm A2 that  $\{\mu^k\}_{k \in K''} \rightarrow 0$ . Also, from (30) and (32), we have

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

Nonsingularity of  $M(x^k, z^k, W^k)$  then implies that  $\{\Delta x^k - \Delta x^{\mu,k}\}_{k \in K''} \rightarrow 0$ , and therefore  $\{\Delta x^{\mu,k}\}_{k \in K''} \rightarrow 0$ , a contradiction since  $K''' \subseteq K'$ . We thus have shown that  $\{t^k\}_{k \in K''}$  is bounded away from zero.

As in the proof of Lemma 3.2, in view of Proposition 4.1, to complete the proof, it then suffices to show that for some  $\delta > 0$ ,

$$f(x^{k+1}) \leq f(x^k) - \delta \tag{41}$$

infinitely many times. We show that it holds for all  $k \in K''$ . For the case where  $\langle \Delta x^*, H \Delta x^* \rangle \leq 0$ , essentially the same argument as that used in the proof of Lemma 3.2 applies, with the reference to (16) replaced by a reference to (28). Suppose then that  $\langle \Delta x^*, H \Delta x^* \rangle > 0$ . Then, for  $k \in K''$ ,  $k$  large enough,  $\langle \Delta x^{\mu,k}, H \Delta x^{\mu,k} \rangle > 0$ . To conclude the proof, observe that

$$f(x^{k+1}) - f(x^k) = t^k \left( \langle \nabla f(x^k), \Delta x^{\mu,k} \rangle + \frac{t^k}{2} \langle \Delta x^{\mu,k}, H \Delta x^{\mu,k} \rangle \right). \tag{42}$$

The first factor ( $t^k$ ) is positive and bounded away from zero on  $K''$ . In view of the update rule for  $t^k$  in Step 3 of Algorithm A2, the second factor is smaller than  $\langle \nabla f(x^k), \Delta x^{\mu,k} \rangle + \frac{\psi}{2} |\langle \nabla f(x^k), \Delta x^{\mu,k} \rangle|$  for  $k \in K''$ ,  $k$  large enough. In view of (28) and since  $\|\Delta x^{k,\mu}\|$  is bounded away from zero on  $K''$ , it follows that  $\langle \nabla f(x^k), \Delta x^{\mu,k} \rangle$  is negative and bounded away from zero on  $K''$ . Since moreover  $\psi < 2$ , it follows that the second factor in (42) is negative and bounded away from zero on  $K''$ . Hence (41) holds on  $K''$ , and the proof is complete.  $\square$

**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  $\{\Delta x^{\mu,k}\}_{k \in K} \rightarrow 0$ , it follows from (28) that  $\{\Delta x^k\}_{k \in K} \rightarrow 0$ . The claims are then proved using the same argument as in the proof of Lemma 3.3, first starting from (30a)–(30b) to show that  $\{\zeta^k\}_{k \in K}$  converges to  $z^*$ , then starting from (32a)–(32b) and using the fact that, due to the boundedness of  $\{z^k\}$ ,  $\{\Delta x^k\}_{k \in K} \rightarrow 0$  implies that  $\{\mu^k\}_{k \in K} \rightarrow 0$ , to show that  $\{\zeta^{\mu,k}\}_{k \in K}$  converges to  $z^*$ .  $\square$

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  $\{x^k\}_{k \in K} \rightarrow x^*$  for some  $x^*$ . If  $\{\Delta x^{\mu,k-1}\}_{k \in K} \rightarrow 0$  then  $\{x^{k-1}\}_{k \in K} \rightarrow x^*$ .*

**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  $\{\Delta z^{\mu,k}\}$  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  $\{\Delta z^{\mu,k-1}\}_{k \in K}$  tend to zero. Then  $x^*$  is a KKT point.*

**Theorem 4.7** *Under Assumptions 1 and 2, every accumulation point of  $\{x^k\}$  constructed by Algorithm A2 is a KKT point.*

**Lemma 4.8** *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  $\{x^k\}_{k \in K} \rightarrow x^*$ . Then  $\{\Delta x^{\mu,k}\}_{k \in K} \rightarrow 0$ .*

**Proposition 4.9** *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)  $\{\Delta x^{\mu,k}\} \rightarrow 0$ , (ii)  $\{\zeta^{\mu,k}\} \rightarrow z^*$ , and (iii)  $\{z_j^k\} \rightarrow \min\{z_j^*, \bar{z}\}$  for all  $j$ .*

**Lemma 4.10** *Let  $\{W^k\}$  be as constructed by Algorithm A2. Under Assumptions 1–4, if  $\sigma$  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  $\mu^k$  going to zero at least as fast as the smallest component of  $z^k$ .

**Lemma 4.11** *Let  $\{\Delta x^k\}$ ,  $\{\Delta x^{\mu,k}\}$ ,  $\{\Delta z^k\}$ ,  $\{\Delta z^{\mu,k}\}$ , and  $\{\zeta^{\mu,k}\}$  be as constructed by Algorithm A2. Suppose Assumptions 1–4 hold. For  $k$  large enough,*

$$\|\Delta x^{\mu,k} - \Delta x^k\|^2 + \|\Delta z^{\mu,k} - \Delta z^k\|^2 = O(\|\Delta x^k\|^\nu). \quad (43)$$

Furthermore,

$$\max\{0, -\langle \zeta^{\mu,k}, A\Delta x^{\mu,k} \rangle\} = o(\|\Delta x^{\mu,k}\|^2).$$

*Proof.* The first claim is a direct consequence of nonsingularity of  $M(x^*, z^*, H)$  (which, like in the proof of Theorem 3.12, follows from Lemma 2.3) and boundedness of  $\{z^k\}$ . Next, since  $g_i(x^k) < 0$  for all  $i$  and all  $k$ , (32b) 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 definition of  $\mu^k$  in Step 2(ii) of Algorithm A2.  $\square$

**Theorem 4.12** *Let  $\{x^k\}$ ,  $\{z^k\}$  be as constructed by Algorithm A2. Suppose Assumptions 1–4 hold. Then, if  $z_i^* < \bar{z} \forall i \in I$  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 3.12 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 4.11 and since  $\nu > 2$ , it follows that condition (ii) in Proposition 3.11 holds for  $k$  large enough. For  $i \notin I(x^*)$ , for  $k$  large enough, either again  $z_i^{k+1} = \zeta_i^k$  or  $z_i^{k+1} = \|\Delta x^{\mu,k}\|^2 + \|\Delta z^{\mu,k}\|^2$ . In the latter case, again using Lemma 4.11 and the fact that  $\nu > 2$ , we conclude that condition (i) in Proposition 3.11 holds.

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

$$\psi \frac{|\langle \Delta x^{\mu,k}, \nabla f(x^k) \rangle|}{\langle \Delta x^{\mu,k}, H\Delta x^{\mu,k} \rangle} > 1 \quad (44)$$

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

$$t^k = \min \left\{ \max\{\bar{t}^k - \|\Delta x^{\mu,k}\|, \beta \bar{t}^k\}, 1 \right\}. \quad (45)$$

Taking the inner product of both sides of equation (32a) by  $\Delta x^{\mu,k}$  and using the fact that  $W^k = H$  for all  $k$  large enough (Lemma 4.10), 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

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

where we have used Lemma 4.11 and equation (28). 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 (37)) 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., (44) holds for  $k$  large enough. Hence (45) holds for all  $k$  large enough.

For  $i \notin I(x^*)$ ,

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

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

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

The remainder of the proof is as in the proof of Theorem 3.12.  $\square$

## 5 Implementation issues

We still have to define explicitly a way of choosing  $\bar{E} \succeq 0$  “small” such that  $\bar{S} + \bar{E} \succeq \sigma I$ , where  $\bar{S} := H + \sum_{\bar{I}} \bar{\alpha}_i a_i a_i^T$  (Step 1 of Algorithms A1 and A2). We use the following method, borrowed from [TWB<sup>+</sup>03, p. 192]:  $\bar{E} = hI$ , with

$$h = \begin{cases} 0 & \text{if } \lambda_{\min} > \sigma, \\ -\lambda_{\min} + \sigma & \text{if } |\lambda_{\min}| \leq \sigma, \\ 2|\lambda_{\min}| & \text{otherwise,} \end{cases} \quad (46)$$

where  $\lambda_{\min}$  denotes the leftmost eigenvalue of  $\bar{S}$ .<sup>6</sup>

Another implementation issue concerns the feasibility of the iterates. It may happen in the course of the iterations that some  $|g_i(x^k)|$  becomes close to the machine relative accuracy  $\mathbf{u}$ ,<sup>7</sup> whereas the chosen stopping criterion is not yet satisfied. For example, this is especially likely to happen with Algorithm A1 when the Hessian  $H$  is ill-conditioned and a constraint boundary is almost parallel to a level curve of the cost function  $f$  over a long distance. Then, two phenomena may concur to produce undesirable consequences: (i) due to numerical errors,  $g_i(x^k + \Delta x^k)$  may be positive and close to  $\mathbf{u}$ ; (ii) the evaluation of  $g_i(x^k)$  itself has a relative error of order 1, which may result in a large relative error on  $\bar{t}^k$ . Consequently, an unnecessarily small step may be taken and, in spite of this,  $g(x^{k+1})$  may well be evaluated to be positive (infeasible point).

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  $\epsilon$  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 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  $(W^k - \sum_{i=1}^m \frac{z_i^k}{g_i(x^k)} a_i a_i^T) \Delta x^k = \nabla f(x^k)$  in view of (15a), it then follows that

---

<sup>6</sup>Since  $\bar{S} \succeq H$ , this guarantees that  $\{W^k\}$  is bounded (in particular,  $W^k \preceq H + 2 \max\{\sigma, -\underline{\lambda}\}I$ , where  $\underline{\lambda}$  is the leftmost eigenvalue of  $H$ ), which was taken for granted in the analysis.

<sup>7</sup>The floating point relative accuracy is the distance from 1.0 to the next (larger) floating point number.

when  $g_i(x^k)$  is close to zero and  $\|\nabla f(x^k)\|$  is not very large, the component of  $\Delta x^k$  along  $a_i$  is very small.

The major computational tasks in Algorithms A1 and A2 are the following. First, in Step 1 (when the  $\bar{\alpha}_i$ 's are recomputed), compute a sufficiently good approximation  $\lambda_{\min}$  of the leftmost eigenvalue of  $\bar{S}$ , to be used in (46); many possibilities are available for computing  $\lambda_{\min}$ , see e.g. [BDD<sup>+</sup>00], [ABG04] and references therein. Second, in Step 2, solve system (5)—in Algorithm A1—or systems (30) and (32)—in Algorithm A2. The other operations require minimal computation (note that  $A\Delta x^k$  computed in (5b) can be reused in the computation of  $\bar{t}^k$ ).

Structured linear systems like (5), or (30) and (32), 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 (5). One option is to solve (5) 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  $\zeta$  from the second line (5b) and substitute into the first line (5a), leading to the condensed primal-dual system (15a). This expresses  $\Delta x^k$  as the solution of an  $n \times n$  linear system. The multiplier estimate  $\zeta$  is then obtained at low cost from (15b). When the matrices are dense, solving a linear system of the form (5) via the Schur complement (15) costs  $O(mn^2)$  flops to form  $S^k$ ,  $O(n^3)$  to solve (15a) and  $O(mn)$  to solve (15b).

Solving the condensed primal-dual system (15a) for the update  $\Delta x^k$  may seem inappropriate because the condition number of the condensed primal-dual matrix  $S$  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  $S$  split into two well-behaved groups, one very large and the other of order 1 (this is responsible for the large condition number of  $S$ ), and the expected inaccuracies preserve this structure. It follows that the absolute error on  $\Delta x^k$  computed with a backward-stable method is comparable to  $\mathbf{u}$ , and so is the error on the multipliers computed via (15b). Moreover, the absolute error in the computed solution of the full, well-conditioned primal-dual system (5) 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 (5), (30) and (32) 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

To our knowledge, Algorithms A1 and A2 are the first to be tailored and analyzed specifically for the optimization of an indefinite quadratic function under general linear inequality constraints. Indefinite quadratic programs with constraints of the form  $Ax = b$ ,  $x \geq 0$ , have received more attention in the literature. In particular, an affine-scaling algorithm, related to Dikin’s algorithm [Dik67], was proposed by Ye [Ye92], which alternatively may be viewed as an interior trust-region method in which the Dikin ellipsoid is used as the trust region [CGT00, Section 13.11]. Variants of Ye’s algorithm have been proposed and analyzed by Bonnans and Bouhtou [BB95], Tseng and Ye [TY02], and Tseng [Tse04].

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 the interior trust-region (TR) algorithm analyzed and tested by Tseng [Tse04]. We considered indefinite quadratic programs of the form

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

because they fit within the framework of  $(P)$  and the constraints are readily transformed into the form  $Ax = b$ ,  $x \geq 0$ , using slack variables. 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 use 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  $x^0$  chosen from a uniform distribution in  $(0, 1)^n$ . The vector  $d$  was selected as  $d = Cx^0 + y$  where  $y$  was chosen from a uniform distribution in  $(0, 1)^{m-n}$ . The matrix  $H$ , with condition number<sup>8</sup>  $10^{ncond}$  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  $ncond$ <sup>9</sup> and  $negeig$ . Ten problems were generated in each category, for a total of 200 test

---

<sup>8</sup>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.)

<sup>9</sup>Tests on symmetric  $100 \times 100$  matrices with elements chosen from a uniform distri-

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  $H - \sum_{i=1}^m \frac{z_i^k}{g_i(x^k)} a_i a_i^T$  is evaluated at *every* iteration in order to compute  $E^k$  according to (46).

Our implementation of the TR algorithm is based on a Matlab script for box constrained problems that was communicated to us by P. Tseng. We modified the script in accordance with [Tse04, Section 2] to include equality constraints. The inclusion of equality constraints led us to discard the stopping criterion specified in [Tse04, Section 5] and to utilize instead the stopping criterion  $\|x^k - x^{k-1}\|_\infty < 10^{-7}$ . For Algorithms A1 and A2, the following parameter values were used:  $\beta = .9$ ,  $\underline{z} = 10^{-4}$ ,  $\bar{z} = 10^{15}$ ,  $\sigma = 10^{-5}$ ,  $\gamma = 10^3$ ,  $\theta = .8$ ,  $\bar{\varphi} = 10^6$ ,  $\nu = 3$ ,  $\psi = 1.5$ ,  $\epsilon = 10^{-14}$  ( $\epsilon$  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  $z^0 = -(A^T)^\dagger \nabla f(x^0)$  (with reference to formulation (P)) and the superscript  $\dagger$  denotes the Moore-Penrose pseudo-inverse. Algorithms A1 and A2 are terminated at the first iteration  $k$  at which  $\max\{\|\nabla f(x^k) + A^T \zeta^{k-1}\|_\infty, \|G(x^k) \zeta^{k-1}\|_\infty\} \leq 10^{-7}$  and  $\zeta_i^{k-1} > -10^{-9}$ ,  $i \in I$ . 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 7) is approximately  $2.2 \cdot 10^{-16}$ .

The numerical results are presented on Tables 1 to 3 and Figures 1 to 3. They show that Algorithms A1 and A2 clearly outperform the TR algorithm regarding the number of iterations, number of eigensolves and number of system solves. The performance gap between the TR algorithm and the Newton-KKT algorithms is particularly large in terms of system solves. The reason the TR method requires so many system solves is that, as suggested in [Tse04, Section 5], the trust-region subproblems are solved with high accuracy using the Moré-Sorensen method [CGT00, Section 7.3]. It is reasonable to expect that an approximate solver, such as a truncated conjugate-gradient method, would yield better practical performances;<sup>10</sup> however, the analysis

---

bution on  $(-1, 1)$  suggest that the condition number of such matrices typically ranges between 50 and 5000.

<sup>10</sup>Also, the number of system solves might decrease if an “industrial-strength” implementation of the Moré and Sorensen algorithm is used instead.

in [Tse04] assumes that an exact solution of the trust-region subproblems is computed. This question is left for future research.

Barrier-based Algorithm A2 tends to outperform affine-scaling Algorithm A1 in terms of number of iterations and eigensolves for nonconvex problems ( $negeig > 0$ ). When the problem is convex, Algorithm A1 has the edge in terms of number of iterations, and both algorithms only require one eigensolve per problem (enough to notice that the problem is convex). The situation is less clear for the number of system solves. In Table 3, two system solves—(30) and (32)—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.

Notice that for ill-conditioned problems, Algorithm A2 performs better on indefinite problems than on positive-definite problems. We observed that this tendency disappears if in Step 2(ii) of that algorithm we modify the expression for  $\mu^k$  to  $\mu^k := \varphi^k \min\{\|\Delta x^k\|^\nu, 1\} z_{\min}^k$ , but with such modification the performance on indefinite problems slightly deteriorates.

The advantage of using Step 1—instead of computing the smallest eigenvalue of the Schur complement  $S^k$  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  $\bar{\alpha}_i$ ’s are recomputed (via eigenvalue decomposition) in Step 1, a dedicated Schur complement  $H + \sum_{\bar{I}} \bar{\alpha}_i a_i a_i^T$  has to be constructed. 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 3 is to compare the methods in terms of quality of the solutions.<sup>11</sup> In each figure, the top histogram shows on a log scale the distribution of the differences unfavorable to Algorithm A2, and the bottom histogram shows the distribution of the differences favorable to that algorithm. It appears that our Newton-KKT algorithms tend to produce better solutions than the

---

<sup>11</sup>The apparent explanation for the presence of two modes on each histogram is that the left mode corresponds to convergence to the same point (with differences in accuracy) and the right one to convergence to different points.

Mean number of iterations					
---------------------------	--	--	--	--	--

Algorithm A1 with simplified Step 1					
	<i>negeig</i>				
<i>ncond</i>	0	10	50	90	100
0	23.1	65.9	89.4	90.8	96.5
1	24.1	61.4	76.6	78.1	82.7
3	26.9	75.3	82.5	86.0	86.4
5	32.0	92.2	94.0	98.8	100.8

Algorithm A2 with simplified Step 1					
	<i>negeig</i>				
<i>ncond</i>	0	10	50	90	100
0	24.6	33.2	39.3	42.5	37.5
1	32.6	34.6	29.2	30.0	30.6
3	79.2	29.0	24.0	27.8	28.0
5	53.5	26.2	25.9	25.5	26.5

Interior trust-region [Tse04]					
	<i>negeig</i>				
<i>ncond</i>	0	10	50	90	100
0	99.8	193.6	219.6	274.6	267.5
1	110.0	203.8	209.6	222.0	227.5
3	133.5	194.8	220.0	244.2	222.8
5	163.2	194.8	196.0	210.5	208.0

Algorithm A1					
	<i>negeig</i>				
<i>ncond</i>	0	10	50	90	100
0	23.1	67.2	92.2	94.2	92.5
1	24.1	74.9	80.5	81.2	82.8
3	26.9	93.5	91.1	91.6	92.6
5	32.0	111.9	106.8	111.6	112.6

Algorithm A2					
	<i>negeig</i>				
<i>ncond</i>	0	10	50	90	100
0	24.6	36.0	35.8	38.2	40.8
1	32.6	33.4	34.8	32.0	32.2
3	79.2	40.4	27.8	29.8	30.8
5	53.5	30.3	27.0	29.1	29.7

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					
	<i>negeig</i>				
<i>ncond</i>	0	10	50	90	100
0	1.0	66.9	90.4	91.8	97.5
1	1.0	62.4	77.6	79.1	83.7
3	1.0	76.3	83.5	87.0	87.4
5	1.0	93.2	95.0	99.8	101.8
Algorithm A2 with simplified Step 1					
	<i>negeig</i>				
<i>ncond</i>	0	10	50	90	100
0	1.0	34.2	40.3	43.5	38.5
1	1.0	35.6	30.2	31.0	31.6
3	1.0	30.0	25.0	28.8	29.0
5	1.0	27.2	26.9	26.5	27.5
Interior trust-region [Tse04]					
	<i>negeig</i>				
<i>ncond</i>	0	10	50	90	100
0	99.8	193.6	219.6	274.6	267.5
1	110.0	203.8	209.6	222.0	227.5
3	133.5	194.8	220.0	244.2	222.8
5	163.2	194.8	196.0	210.5	208.0
Algorithm A1					
	<i>negeig</i>				
<i>ncond</i>	0	10	50	90	100
0	1.0	19.3	29.4	33.8	34.2
1	1.0	16.0	25.2	27.6	27.4
3	1.0	34.3	38.0	40.6	39.7
5	1.0	45.9	52.7	56.9	55.5
Algorithm A2					
	<i>negeig</i>				
<i>ncond</i>	0	10	50	90	100
0	1.0	16.6	15.0	13.4	14.1
1	1.0	16.7	17.3	18.4	17.8
3	1.0	17.7	16.4	18.5	18.3
5	1.0	17.5	17.6	19.3	19.6

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 solved. 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					
	<i>negeig</i>				
<i>ncond</i>	0	10	50	90	100
0	23.1	65.9	89.4	90.8	96.5
1	24.1	61.4	76.6	78.1	82.7
3	26.9	75.3	82.5	86.0	86.4
5	32.0	92.2	94.0	98.8	100.8

Algorithm A2 with simplified Step 1					
	<i>negeig</i>				
<i>ncond</i>	0	10	50	90	100
0	49.2	66.4	78.6	85.0	75.0
1	65.2	69.2	58.4	60.0	61.2
3	158.4	58.0	48.0	55.6	56.0
5	107.0	52.4	51.8	51.0	53.0

Interior trust-region [Tse04]					
	<i>negeig</i>				
<i>ncond</i>	0	10	50	90	100
0	1140.9	1540.1	1632.9	1982.2	1933.1
1	1252.8	1589.6	1567.1	1646.0	1696.6
3	1543.0	1532.0	1610.8	1760.8	1621.6
5	1875.4	1454.8	1429.2	1518.6	1504.2

Algorithm A1					
	<i>negeig</i>				
<i>ncond</i>	0	10	50	90	100
0	23.1	67.2	92.2	94.2	92.5
1	24.1	74.9	80.5	81.2	82.8
3	26.9	93.5	91.1	91.6	92.6
5	32.0	111.9	106.8	111.6	112.6

Algorithm A2					
	<i>negeig</i>				
<i>ncond</i>	0	10	50	90	100
0	49.2	72.0	71.6	76.4	81.6
1	65.2	66.8	69.6	64.0	64.4
3	158.4	80.8	55.6	59.6	61.6
5	107.0	60.6	54.0	58.2	59.4

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

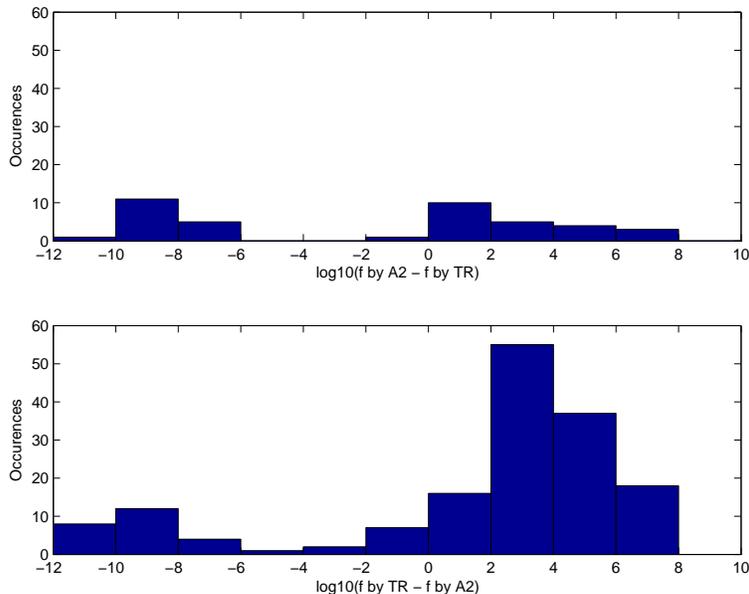


Figure 1: Comparison of the quality of the solutions obtained by Algorithm A2 and the TR algorithm. The horizontal axes give, on a logarithmic scale, the difference between the final values of the cost function  $f$  obtained by the two methods. The upper histogram tallies the cases where the TR algorithm reached a better quality solution than Algorithm A2, and vice versa for the lower histogram. The data comes from the 200 tests reported on in the tables.

interior trust-region method. The reasons are unknown. The variants of the Newton-KKT algorithms (simple or elaborate Step 1, barrier parameter or not) produce results of comparable quality.

We also tested variants of Algorithms A1 and A2 where the update of  $z^k$  was replaced by updates along  $\Delta z^k$  and  $\Delta z^{\mu,k}$ . For example, in the case of Algorithm A2,  $z^{k+1}$  was computed as  $z^{k+1} = z^k + t_d^k \Delta z^{\mu,k}$  where  $t_d^k$  was the largest stepsize such that

$$\min\{\|\Delta v^{\mu,k}\|^2 + \|\Delta z^{\mu,k}\|^2, \bar{z}\} \leq z^k + t_d^k \Delta z^{\mu,k} \leq \bar{z}, \forall i \in I.$$

With the modified Algorithm A1, the results were clearly not as good. With the modified Algorithm A2, results were alternatively slightly better and slightly worse.

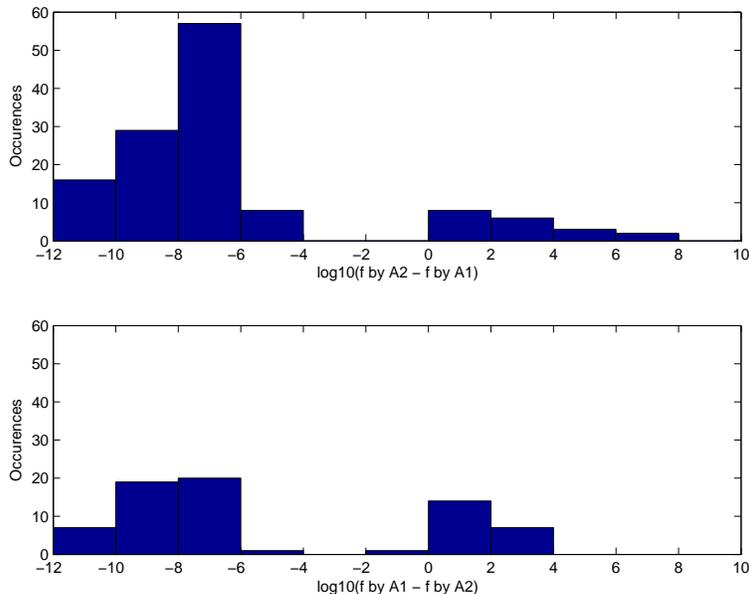


Figure 2: Comparison of the quality of solutions obtained by Algorithms A2 and A1.

Convergence properties of Algorithms A1 and A2 were proved in Sections 3 and 4 under certain nondegeneracy assumptions. To get a sense for how critical nondegeneracy may be to satisfactory behavior of the algorithms, as a final set of experiments, we tested Algorithms A1 and A2 on problems in which linear independence constraint qualification (Assumption 1) is violated. Specifically, we first randomly generated a number of (nondegenerate) problems as described above, for each value of the pair  $(ncond, negeig)$ . We then solved these problems with Algorithm A1 (resp. A2), yielding  $x^*$ . Ten new  $a_i$ 's,  $i = m + 1, \dots, m + 10$ , were then randomly chosen in the convex hull of  $\{a_i\}_{i \in I(x^*)}$  and appended to  $A$ , and  $b$  was appended with the values  $a_i^T x^*$ ,  $i = m + 1, \dots, m + 10$ . Consequently, linear independence constraint qualification did not hold at  $x^*$ , while the feasible set was unaltered. We observed that while, typically (but not always), more iterations were needed to solve the degenerate problems, in many cases the difference was no more than 10. Specifically, this happened with about two thirds of the problems in the case of Algorithm A1, and with over 90 percent of the problems in the case of Algorithm A2. Thus Algorithm A2 and to a lesser extent Algorithm A1 show promise even for degenerate problems.

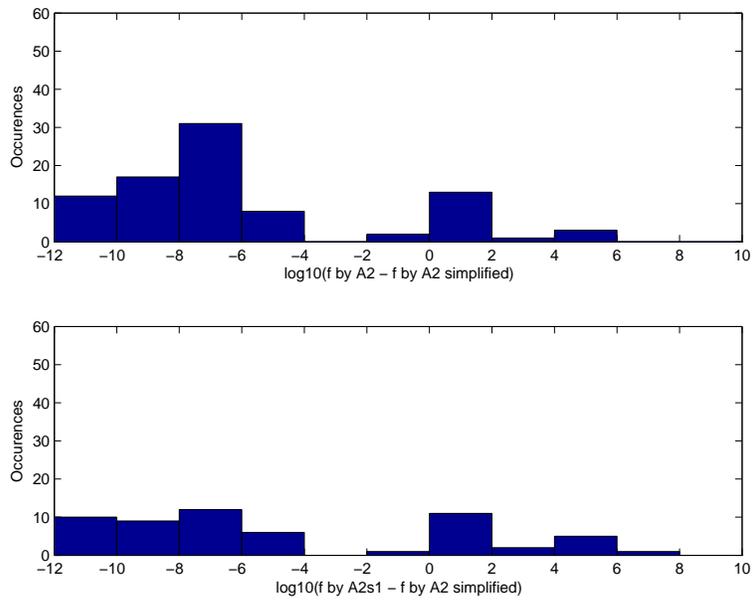


Figure 3: Comparison of the quality of solutions obtained by Algorithm A2 and Algorithm A2 with simplified Step 1.

## 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 for making available to them his Matlab code for the algorithm studied in [Tse04], and for valuable advice.

## References

- [ABG04] P.-A. Absil, C. G. Baker, and K. A. Gallivan, *A truncated-CG style method for symmetric generalized eigenvalue problems*, [http://www.csit.fsu.edu/~absil/Publi/SGEVP\\_tCG.htm](http://www.csit.fsu.edu/~absil/Publi/SGEVP_tCG.htm), submitted, 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<sup>+</sup>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.
- [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.
- [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.
- [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.
- [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. MR 90e:90091
- [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.
- [Tse04] P. Tseng, *Convergence properties of Dikin’s affine scaling algorithm for nonconvex quadratic minimization*, J. Global Optim. **30** (2004), no. 2–3, special issue on the International Conference on Optimization and Optimal Control, to appear.
- [TWB<sup>+</sup>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. MR 1 952 651
- [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.
- [Ye92] Y. Ye, *On affine scaling algorithms for nonconvex quadratic programming*, Mathematical Programming **56** (1992), 285–300.