A Globally Convergent Interior Point Algorithm for General Non-Linear Programming Problems

Ioannis Akrotirianakis[1] and Berc Rustem[2]

Abstract

This paper presents a primal-dual interior point algorithm for solving general constrained non-linear programming problems. The initial problem is transformed to an equivalent equality constrained problem, with inequality constraints incorporated into the objective function by means of a logarithmic barrier function. Satisfaction of the equality constraints is enforced through the incorporation of an adaptive quadratic penalty function into the objective. The penalty parameter is determined using a strategy that ensures a descent property for a merit function. It is shown that the adaptive penalty does not grow indefinitely. The algorithm applies Newton's method to solve the first order optimality conditions of the equivalent equality problem. Global convergence of the algorithm is achieved through the monotonic decrease of a merit function. Locally the algorithm is shown to be quadratically convergent.

Key Words: Non-linear Programming, primal-dual interior point methods, adaptive penalty parameter, augmented Lagrangian, convergence analysis.

1991 Mathematics Subject Classification: Primary 90C30, Secondary 49M29, 65K05

November 1997,
Technical Report 97/14,
Department of Computing,
Imperial College of Science, Technology and Medicine,
180 Queen's Gate, London SW7 2BZ, U.K.

1. e-mail: ia4@doc.ic.ac.uk
2. e-mail: br@doc.ic.ac.uk

1 Introduction

Since Karmarkar's seminal work [14], there has been substantial interest in interior point algorithms for linear programming (LP). These algorithms consider LP as a special case of non-linear programming (NLP). Among different interior point approaches, primal-dual algorithms have attracted most interest. Computational experiments (eg, [16], [18]) and theoretical developments (eg, [1], [24]) have shown that they perform much better than other interior point algorithms and outperform the simplex method in many large-scale LP problems. Primal-dual methods basically apply Newton's algorithm directly to the primal-dual system of equations for both feasibility and approximate (or perturbed) complementarity conditions. A rigorous treatment of primal-dual methods in LP can be found in Wright [12].

Motivated by the computational success of primal-dual methods in LP, the investigation has focussed on possible extensions to NLP. The bulk of the effort has concentrated on convex quadratic (eg, [10], [19]) and convex NLP problems (eg, [17], [13], [11]), showing that primal-dual methods provide an efficient solution framework. However, only recently general (non-convex) NLP problems have been the subject of research in this area. El-Bakry *et al.* [22], McCormick and Falk [20], and Yamashita [7] have developed globally convergent primal-dual algorithms for that class of problems. Also Lasdon *et al.* [15] have considered various primal-dual formulations of those problems and presented their computational experience.

In this paper, we discuss a primal-dual interior point algorithm for general NLP problems. Our approach basically derives from the premise that the solution of the first order optimality conditions of any NLP problem, which exists in the core of interior point algorithms, is not sufficient to guarantee the convergence to an optimum solution, unless the problem is convex. In other words, the algorithm, applied for example on a minimization problem, may converge to a local maximum or even worse to a saddle point, since the first order optimality conditions are also satisfied in those points. To avoid such cases a merit function is incorporated within the primal-dual interior point algorithm. This is achieved by using an Armijo rule to determine the step-size, which guarantees the monotonic decrease of our merit function.

The algorithm is partially motivated by two different approaches. The first is the augmented Lagrangian sequential quadratic programming (SQP) framework for general constrained optimization problems, discussed in Rustem [2]. The SQP algorithms possess good theoretical and practical properties and are very efficient for solving general NLP problems [4]. The second approach is the primal-dual interior point method, where a barrier function and a damped Newton framework are used in order to solve NLP problems. This is closely related to the SQP framework, since after the initial incorporation of the inequality constraints into the objective function an equivalent equality constrained problem is obtained. The latter is solved by applying Newton's method to the first order optimality conditions. Although our algorithm is related to the approaches proposed by El-Bakry *et at.* [22] and Yamashita [7], it differs in significant aspects, such as the choice of the merit function, the adaptive penalty selection rule and the step-size rules. Recently, it has come to our attention that Vanderbei and

Shanno [9] and Gajulapalli and Lasdon [21] report very encouraging numerical results with algorithms which also use an adaptive penalty. The present paper, however, provides the full analysis of an algorithm which substantially differs from [9] and [21], in the adaptive penalty term, discussed in Rustem [2], the barrier parameter, the merit function and the step-size rules.

In section 2 we introduce the basic features of the augmented Lagrangian methods, used in this paper. In section 3 we present the basic algorithmic framework of primal-dual methods for NLP problems. Section 4 describes the primal-dual interior point algorithm. In section 5 we establish the global convergence of the algorithm. In section 6 we examine the local behaviour of the algorithm and show that it converges quadratically to the optimum solution, provided that the standard conditions associated with the Newton method hold.

2 Augmented Lagrangian Methods

Penalty methods are mainly used for equality constrained optimization problems. The aim is to eliminate the constraints and augment the cost function with a penalty term that associates a high cost to infeasible points. The severity of the penalty is determined by a parameter, denoted by c. As c takes higher values feasibility is increasingly ensured.

Consider the equality constrained problem

$$\begin{aligned} &min \quad f(x) \\ &ST \quad\; g(x) = 0, \end{aligned} \tag{1}$$

where $f : \Re^n \to \Re$, and $g(x) : \Re^n \to \Re^q$ are given functions. The Lagrangian function of this problem is $L(x, y) = f(x) - y^T g(x)$. Augmenting $L(x, y)$ with a quadratic penalty term yields the augmented Lagrangian function given by

$$L_c(x, y) = f(x) - y^T g(x) + \frac{c}{2} \parallel g(x) \parallel^2,$$

which can be considered the Lagrangian function of

$$\begin{aligned} &min \quad f(x) + \tfrac{c}{2} \parallel g(x) \parallel^2 \\ &ST \quad\; g(x) = 0 \end{aligned} \tag{2}$$

Problem (2) has the same local minima as problem (1). The gradient and the Hessian of L_c with respect to x are

$$\nabla_x L_c(x, y) = \nabla f(x) + \nabla g(x)^T (cg(x) - y),$$

$$\nabla^2_{xx} L_c(x,y) = \nabla^2 f(x) + \sum_{i=1}^{m} \nabla^2 g_i(x)(cg_i(x) - y_i) + c\nabla g(x) \nabla g(x)^T. \quad (3)$$

In particular, if x_* and y_* satisfy the first order optimality conditions, then $\nabla_x L_c(x_*, y_*) = \nabla L(x_*, y_*) = 0$ and $\nabla^2_{xx} L_c(x_*, y_*) = \nabla^2_{xx} L(x_*, y_*) + c\nabla g(x_*) \nabla g(x_*)^T$. For a detailed treatment of penalty and augmented Lagrangian methods we refer to [3] and [4].

3 Basic Iteration in Primal-Dual Methods

Consider the following constrained problem

$$\begin{array}{ll} min & f(x) \\ ST & g(x) = 0, \quad x \geq 0, \end{array} \quad (4)$$

where $x \in \Re^n$, $f : \Re^n \rightarrow \Re$ and $g(x) : \Re^n \rightarrow \Re^q$.

In barrier methods, (4) is approximated by augmenting the objective with the logarithmic barrier function $B(x;\mu) : \Re^n \rightarrow \Re$, $B(x;\mu) = -\mu \sum_{i=1}^n log(x^i)$. Thus, the initial problem is approximated by

$$\begin{array}{ll} min & f(x) - \mu \sum_{i=1}^n log(x^i) \\ ST & g(x) = 0, \end{array} \quad (5)$$

where $x > 0$ and the barrier parameter μ is a given sufficiently small and strictly positive constant [23], [4]. The optimality conditions of (5) are

$$\begin{array}{rcl} \nabla f(x) - \nabla g(x)^T y - \mu X^{-1} e & = & 0 \\ g(x) & = & 0, \end{array} \quad (6)$$

where X is the diagonal matrix given by $X = diag(x^1, ..., x^n)$. Also $e \in \Re^n$ is the vector of all ones. Introducing of the non-linear transformation $z = \mu X^{-1} e$, (6) becomes

$$\begin{array}{rcl} \nabla f(x) - \nabla g(x)^T y - z & = & 0 \\ g(x) & = & 0 \\ XZe & = & \mu e, \end{array} \quad (7)$$

where $x, z > 0$ and $Z = diag(z^1, ..., z^n)$. The introduction of z is essential to the numerical success of the barrier methods (see for example [22]).

Consider the Lagrangian function of the equality and inequality constrained problem (4)

$$L(x,y,z) = f(x) - y^T g(x) - z^T x, \tag{8}$$

where $y \in \Re^q$ and $z \in \Re^n_+ \equiv \{v \in \Re^n : v \geq 0\}$ are the Lagrange multiplier vectors of the equality constraints $g(x) = 0$ and non-negativity constraints $x \geq 0$, respectively. The KKT conditions of (4) are given by the nonlinear system of equations

$$F(x,y,z) = \begin{pmatrix} \nabla_x L(x,y,z) \\ \\ g(x) \\ \\ XZe \end{pmatrix} = 0, \tag{9}$$

where $x, z \geq 0$ and the gradient of the Lagrangian with respect to x is

$$\nabla_x L(x,y,z) \quad = \quad \nabla f(x) - \nabla g(x)^T y - z \quad = \quad \nabla f(x) - \sum_{i=1}^{m} \nabla g_i(x) y_i - z.$$

The perturbed KKT conditions are taken by introducing a positive perturbation to the third equation of (9), namely to the complimentarity equation. Hence, for $x, z \geq 0$ the perturbed KKT conditions are

$$F(x,y,z;\mu) = \begin{pmatrix} \nabla_x L(x,y,z) \\ \\ g(x) \\ \\ XZe - \mu e \end{pmatrix} = 0. \tag{10}$$

A point $(x(\mu), y(\mu), z(\mu))$ is said to belong to the central path C, if it is the solution of the perturbed KKT conditions (10), for a fixed value of μ. Conditions (10) approximate the KKT conditions (9) increasingly accurately as $\mu \to 0$. Hence, as $\mu \to 0$, the sequence $\{(x(\mu), y(\mu), z(\mu))\}$ of converges to the solution of the KKT conditions (9), of the initial constrained problem (4).

The perturbed KKT conditions of the initial problem (4), given by (10), are equivalent to the KKT conditions of the logarithmic barrier function problem (5), given by (6). El-Bakry et al., in [22], have proved that the perturbed KKT conditions (10) are not the KKT conditions (6) of the logarithmic barrier function problem. Furthermore, the iterates of the Newton method applied to the perturbed KKT conditions (10) are not the same as the iterates of the Newton method applied to the KKT conditions (6) of the logarithmic barrier function problem. In other words systems (10) and (6) have the same solutions (i.e., they are equivalent) but they are not Newton algorithmically equivalent.

Furthermore, primal-dual methods solve approximately the perturbed KKT conditions (10), for a fixed value of μ. Therefore, the first order change of the above system needs to be found. The k-th Newton iteration for solving (10) can be written as

$$\begin{pmatrix} \nabla^2_{xx}L(x_k, y_k, z_k) & -\nabla g(x_k)^T & -I \\ \nabla g(x_k) & 0 & 0 \\ Z_k & 0 & X_k \end{pmatrix} \begin{pmatrix} \Delta x_k \\ \Delta y_k \\ \Delta z_k \end{pmatrix} = - \begin{pmatrix} \nabla_x L(x_k, y_k, z_k) \\ g(x_k) \\ X_k Z_k e - \mu_k e \end{pmatrix}$$

or in matrix-vector form

$$J(w_k)\Delta w_k = -r(w_k), \tag{11}$$

where $w_k = (x_k, y_k, z_k)^T$, and $\Delta w_k = (\Delta x_k, \Delta y_k, \Delta z_k)^T$. The solution of (11) gives a direction vector Δw_k which is used to find the next approximation of the solution of (10). That is, the next iterate is $w_{k+1} = w_k + \mathrm{A}_k \Delta w_k$, where A_k is the diagonal matrix $\mathrm{A}_k = diag(\alpha_{xk} I_n, \alpha_{yk} I_q, \alpha_{zk} I_n)$ and I_n, I_q are the n-th and q-th order identity matrices respectively. The step-lengths α_{xk}, α_{yk}, and α_{zk} belong to the interval $(0, 1]$ and may all be equal to or different from each other.

A unit step along the Newton direction is often not allowed because it violates the non-negativity constraints on x and z in (10). To avoid this violation, the step-sizes α_{xk} and α_{zk} are selected such that the new iterates x_{k+1} and z_{k+1} are strictly positive for all k. When an approximation of the central point corresponding to the value where μ is fixed is found, the barrier parameter μ is fixed onto a strictly smaller value and the iterations proceed until μ becomes zero.

4 Description of the Algorithm

The algorithm discussed below solves problem (4) and is based on a sequence of optimization problems characterized by a penalty $c \geq 0$ and a barrier $\mu \geq 0$ parameter. The following assumptions are used throughout the paper.

Assumptions:

A1: The second order derivatives of the objective function f and the constraints g are Lipschitz continuous at the optimum x_*.

A2: The columns of the matrix $[\nabla g(x), e_i : i \in I_x^0\}$ are linear independent, where $I_x^0 = \{i : \liminf_{k\to\infty} x_k^i = 0,\ i = 1, 2, ..., n\}$ and e_i represents the i-th column of the $n \times n$ identity matrix. Also the sequence $\{x_k\}$ is bounded.

A3: Strict complementarity of the solution $w_* = (x_*, y_*, z_*)$ is satisfied, that is if $z_*^i > 0$ then $x_*^i = 0$, for $i = 1, 2, ..., n$ and vice versa.

A4: The second order sufficiency condition for optimality is satisfied at the solution point, i.e., if for all vectors $0 \neq v \in \Re^n$ such that $\nabla g^i(x_*)^T v = 0$, $i = 1, 2, ..., q$, and $e_i^T v = 0$, for $i \in I_x^0$, then $v^T \nabla_{xx} L(x, y, z) v > 0$. Also, the matrix $\Omega_k = H_k + X_k^{-1} Z_k$ is invertible, with $H_k = \nabla^2_{xx} L_c$ defined in (3).

The original equality and inequality constrained optimization problem (4) is approximated by

$$\begin{array}{ll} min & f(x) + \frac{c}{2} \parallel g(x) \parallel_2^2 - \mu \sum_{i=1}^{n} log(x^i) \\ & \\ ST & g(x) = 0, \end{array} \tag{12}$$

for $c, \mu \geq 0$. The objective in (4) is augmented by the penalty and the logarithmic barrier functions. The penalty is used to enforce satisfaction of the equality constraints by adding a high cost to the objective function for infeasible points. The barrier is needed to introduce an interior point method to solve the initial problem (4), since it creates a positive singularity at the boundary of the feasible region. Thus, strict feasibility is enforced, while approaching the optimum solution. Both the penalty and the barrier functions are continuous and differentiable, since f and g are assumed to possess these properties.

Penalty-barrier methods involve outer and inner iterations [5]. Outer iterations are associated with decreasing the barrier parameter μ, such that μ approaches zero. Inner iterations determine the penalty parameter c and then solve the optimization problem (12) for the corresponding values of μ and c.

The Lagrangian associated with the optimization problem (12) is given by

$$L(x, y; c, \mu) = f(x) + \frac{c}{2} \parallel g(x) \parallel_2^2 - \mu \sum_{i=1}^{n} log(x^i) - g(x)^T y,$$

while the first order optimality conditions are the system of nonlinear equations

$$\begin{pmatrix} \nabla f(x) - \mu X^{-1} e + c \nabla g(x)^T g(x) - \nabla g(x)^T y \\ \\ g(x) \end{pmatrix} = 0,$$

for $x > 0$. By invoking the nonlinear transformation $z = \mu X^{-1} e$ the above conditions become

$$F(x, y, z; c, \mu) = \begin{pmatrix} \nabla f(x) - z + c \nabla g(x)^T g(x) - \nabla g(x)^T y \\ \\ g(x) \\ \\ XZe - \mu e \end{pmatrix} = 0, \tag{13}$$

with $x, z > 0$. For μ fixed, system (13) is solved by using the Newton method. At the k-th iteration, the Newton system is

$$\begin{pmatrix} H_k & -\nabla g_k^T & -I \\ \\ \nabla g_k & 0 & 0 \\ \\ Z_k & 0 & X_k \end{pmatrix} \begin{pmatrix} \Delta x_k \\ \\ \Delta y_k \\ \\ \Delta z_k \end{pmatrix} = - \begin{pmatrix} \nabla f_k - z_k + c_k \nabla g_k^T g_k - \nabla g_k^T y_k \\ \\ g_k \\ \\ X_k Z_k e - \mu e, \end{pmatrix} \tag{14}$$

where

$$H_k = \nabla^2 f_k + \sum_{i=1}^{m} \nabla^2 g_k^i (c_k g_k^i - y_k^i) + c_k \sum_{i=1}^{m} \nabla g_k^i (\nabla g_k^i)^T \equiv \nabla_{xx}^2 L_c(x_k, y_k)$$

is the Hessian of the augmented Lagrangian defined in (3). In matrix-vector form (14) can be written as

$$J(w_k; c_k)\Delta w_k = -F(w_k; c_k, \mu_k), \tag{15}$$

where $w_k = (x_k, y_k, z_k)^T$, $\Delta w_k = (\Delta x_k, \Delta y_k, \Delta z_k)^T$, and $J(w_k; c_k)$ is the Jacobian matrix of the vector function $F(w_k; c_k, \mu_k)$. Equation (15) is different from the corresponding equation (11) due to the introduction of the penalty term in the objective function.

The solution of (14) is given by

$$\begin{aligned}
\Delta x_k &= \Omega_k^{-1} \nabla g_k^T \Delta y_k - \Omega_k^{-1} (\nabla f_k - z_k + c \nabla g_k^T g_k - \nabla g_k^T y_k) \\
\Delta y_k &= -[\nabla g_k \Omega_k^{-1} \nabla g_k^T]^{-1} (g_k - \nabla g_k \Omega_k^{-1} (\nabla f_k - z_k + c \nabla g_k^T g_k - \nabla g_k^T y_k)) \\
\Delta z_k &= -z_k + \mu X_k^{-1} e - X_k^{-1} Z_k \Delta x_k,
\end{aligned}$$

where $\Omega_k = H_k + X_k^{-1} Z_k$ is assumed to be invertible. The algorithm uses different step-sizes for the primal and dual variables. Hence, the next iterate $w_{k+1} = (x_{k+1}, y_{k+1}, z_{k+1})$ is defined as

$$x_{k+1} = x_k + \alpha_{xk} \Delta x_k, \quad y_{k+1} = y_k + \alpha_{zk} \Delta y_k, \quad z_{k+1} = z_k + \alpha_{zk} \Delta z_k,$$

where α_{xk} and α_{zk} are the step-lengths for the primal variables x and the pair of dual variables y and z, respectively.

To initiate the algorithm, a strictly interior starting point is needed, that is a point $w^0 = (x^0, y^0, z^0)$, with x^0, $z^0 > 0$. By controlling the step lengths α_{xk} and α_{zk}, the algorithm ensures that the generated iterates remain strictly in the interior of the feasible region. Moreover, the algorithm moves from one inner iteration to another inner iteration (i.e., with μ fixed) by seeking to minimize the merit function

$$\Phi(x; c, \mu) = f(x) + \frac{c}{2} \parallel g(x) \parallel_2^2 - \mu \sum_{i=1}^{n} log(x^i), \tag{16}$$

which is basically the objective function of problem (12). This is achieved by properly selecting the values of the penalty parameter c at each inner iteration. As shown later, the monotonic decrease of (16) and the rules for determining the primal and dual step-sizes, guarantee that the inner iterates converge to the solution of (12), for μ fixed. Subsequently, by reducing μ, such that $\{\mu\} \to 0$, the optimum of the initial problem (4) is reached.

The primal-dual interior point algorithm is given in Figure 1. Throughout the algorithm, subscript k indicates variables changed in inner iterations (i.e., while μ is fixed) and superscript l indicates variables changed in outer iterations (i.e., when μ decreases). Superscript i denotes elements of vectors.

Algorithm 1

STEP 0: Initialization: Given $\tilde{x}^0, \tilde{z}^0 \in \Re^n$ and $\tilde{y}^0 \in \Re^q$, such that $\tilde{x}^0,\ \tilde{z}^0\ >\ 0$, penalty and barrier parameters $c_0 > 0$, $\mu^0 > 0$ and parameters: $\gamma, \eta, \rho \in (0,1)$, $\delta > 0$.

For $l = 0, 1, 2, \ldots$ do the following steps:

STEP 1: Test for convergence of outer iterations:

If $\parallel F(\tilde{x}^l, \tilde{y}^l, \tilde{z}^l; c_*, \mu^l) \parallel \leq \epsilon_0$, then Stop.

STEP 2: Start of inner iterations: (μ is fixed to μ^l throughout)

Set $(x_0, y_0, z_0) = (\tilde{x}^l, \tilde{y}^l, \tilde{z}^l)$

For k=0,1,2,... do the following steps:

Step 2.1: Test for convergence of inner iterations:

If $\parallel F(x_k, y_k, z_k; c_k, \mu^l) \parallel \leq \eta\mu^l$ then

Set $(\tilde{x}^l, \tilde{y}^l, \tilde{z}^l) = (x_k, y_k, z_k)$ and GoTo step 3

Step 2.2: Solve Newton system (14) to obtain $(\Delta x_k, \Delta y_k, \Delta z_k)$

Step 2.3: Penalty parameter selection:

If $\Delta x_k^T \nabla f_k - c_k \parallel g_k \parallel_2^2 - \mu^l \Delta x_k^T X_k^{-1} e + \parallel \Delta x_k \parallel_{H_k}^2 \leq 0$ then $c_{k+1} = c_k$.

Else set

$c_{k+1} = max\{\frac{\Delta x_k^T \nabla f_k - \mu^l \Delta x_k^T X_k^{-1} e + \|\Delta x_k\|_{H_k}^2}{\|g_k\|_2^2},\ c_k + \delta\}$

Step 2.4: Step-length selection rules:

Set $\alpha_{xk}^{max} = \min_{1 \leq i \leq n} \{\frac{-x_k^i}{\Delta x_k^i}\ :\ \Delta x_k^i < 0\}$ and $\hat{\alpha}_{xk} = min\{\gamma \alpha_{xk}^{max}, 1\}$.

Let $\alpha_{xk} = \beta^\theta \hat{\alpha}_{xk}$, where θ is the smallest non-negative integer such that

$\Phi(x_{k+1}; c_{k+1}, \mu^l) - \Phi(x_k; c_{k+1}, \mu^l) \leq \rho\, \alpha_{xk} \nabla \Phi(x_k; c_{k+1}, \mu^l)^T \Delta x_k$,

with $x_{k+1} = x_k + \alpha_{xk} \Delta x_k$.

Set $LB_k^i = \min\{\frac{1}{2} m\mu, x_{k+1}^i z_k^i\}$ and $UB_k^i = \max\{2M\mu, x_{k+1}^i z_k^i\}$, $m, M > 0$

For $i = 1, 2, \ldots, n$ find: $\alpha_{zk}^i = \max\{\alpha_i : LB_k^i \leq x_{k+1}^i (z_k^i + \alpha_i \Delta z_k^i) \leq UB_k^i\}$

Set $\alpha_{zk} = \min\{1, \min_{1 \leq i \leq n} \{\alpha_{zk}^i\}\}$

Set $y_{k+1} = y_k + \alpha_{zk} \Delta y_k$ and $z_{k+1} = z_k + \alpha_{zk} \Delta z_k$.

Step 2.5: Set $k = k + 1$ and GoTo Step 2.1

STEP 3: Reduction of barrier parameter: Set $\mu^{l+1} = (1 - \nu)\mu^l$, where $0 < \nu < 1$.

STEP 4: Set $l = l + 1$ and GoTo Step 1.

Figure 1

4.1 Penalty parameter selection rule

The penalty parameter c plays an important role in the algorithm. At each iteration, its value is determined such that a descent property is ensured for the merit function $\Phi(x; c, \mu)$. For μ fixed, the gradient of Φ at the k-th iteration is

$$\nabla \Phi(x_k; c_k, \mu) = \nabla f_k + c_k \nabla g_k^T g_k - \mu X_k^{-1} e. \tag{17}$$

The direction Δx_k is a descent direction for Φ, at the current point x_k, if

$$\Delta x_k^T \nabla \Phi(x_k; c_k, \mu) \leq 0. \tag{18}$$

By considering the second equation of the Newton system (14), the directional derivative $\Delta x_k^T \nabla \Phi(x_k; c_k, \mu)$ can be written as

$$\Delta x_k^T \nabla \Phi(x_k; c_k, \mu) = \Delta x_k^T \nabla f(x_k) - c_k \parallel g_k \parallel^2 - \mu \Delta x_k^T X_k^{-1} e, \tag{19}$$

where c_k is the value of the penalty parameter at the beginning of the k-th iteration. Since the barrier parameter μ is fixed throughout the inner iterations, we can deduce from (19) that the sign of $\Delta x_k^T \nabla \Phi(x_k; c_k, \mu)$ depends on the value of the penalty parameter. If c_k is not large enough then the descent property (18) may not be satisfied. Thus, a new value $c_{k+1} > c_k$ must be determined to guarantee the satisfaction of the descent property. The next lemma shows that Algorithm 1 chooses the value of the penalty parameter in such a way that (18) holds.

Lemma 1 *Let f and g be differentiable functions and let $g_k \neq 0$. If Δx_k is calculated by solving the Newton system (14) and c_{k+1} is chosen as in step 2.3 of Algorithm 1 then Δx_k is a descent direction for the merit function Φ at the current point x_k. Furthermore*

$$\Delta x_k^T \nabla \Phi(x_k; c_{k+1}, \mu) \leq - \parallel \Delta x_k \parallel_{H_k}^2 \leq 0. \tag{20}$$

Proof In step 2.3, Algorithm 1 initially checks the inequality

$$\Delta x_k^T \nabla f_k - c_k \parallel g_k \parallel_2^2 - \mu \Delta x_k^T X_k^{-1} e + \parallel \Delta x_k \parallel_{H_k}^2 \leq 0. \tag{21}$$

If (21) is satisfied then by setting $c_{k+1} = c_k$ and re-arranging (21) we obtain (20). On the other hand if (21) is not satisfied, by setting

$$c_{k+1} \geq \frac{\Delta x_k^T \nabla f_k - \mu \Delta x_k^T X_k^{-1} e + \parallel \Delta x_k \parallel_{H_k}^2}{\parallel g_k \parallel_2^2},$$

and substituting it into (19) it can be verified that (20) also holds. •

In the previous lemma it is assumed that $g_k \neq 0$. The next lemma demonstrates that Δx_k remains a descent direction for the merit function Φ when $g_k = 0$, i.e., when feasibility of the equality constraints has been achieved.

Lemma 2 *Let f and g be differentiable functions and let $\Delta w_k = (\Delta x_k, \Delta y_k, \Delta z_k)$ be the Newton direction taken by solving system (14). If for some or all iterations k, $g_k = 0$, then the descent property (20) is satisfied for any choice of the penalty parameter $c_k \in [0, \infty)$.*

Proof If $g_k = 0$ then (19) yields

$$\Delta x_k^T \nabla \Phi(x_k; c_k, \mu) = \Delta x_k^T \nabla f_k - \mu \Delta x_k^T X_k^{-1} e \tag{22}$$

and the second equation of the Newton system (14) becomes

$$\nabla g_k \Delta x_k = 0. \tag{23}$$

Furthermore, solving the first equation of (14) for Δz_k we have

$$\Delta z_k = -X_k^{-1} Z_k \Delta x_k - z_k + \mu X_k^{-1} e.$$

Substituting Δz_k into the first equation of (14) yields

$$\nabla f_k - c_k \nabla g_k^T g_k + \mu X_k^{-1} e = -(H_k + X_k^{-1} Z_k) \Delta x_k + \nabla g_k^T (y_k + \Delta y_k). \tag{24}$$

Pre-multiplying (24) by Δx_k^T yields

$$\Delta x_k^T \nabla f_k - c_k \Delta x_k^T \nabla g_k^T g_k + \mu \Delta x_k^T X_k^{-1} e = -\Delta x_k^T (H_k + X_k^{-1} Z_k) \Delta x_k + \Delta x_k^T \nabla g_k^T (y_k + \Delta y_k).$$

Using (23), the above equation becomes

$$\Delta x_k^T \nabla f_k + \mu \Delta x_k^T X_k^{-1} e = -\Delta x_k^T (H_k + X_k^{-1} Z_k) \Delta x_k, \tag{25}$$

and from (22), equation (25) yields

$$\Delta x_k^T \nabla \Phi(x_k; c_k, \mu) = -\Delta x_k^T (H_k + X_k^{-1} Z_k) \Delta x_k. \tag{26}$$

From the fact that x_k and z_k are strictly positive and Assumption (A4), we have

$$-\Delta x_k^T (H_k + X_k^{-1} Z_k) \Delta x_k < -\Delta x_k^T H_k \Delta x_k. \tag{27}$$

From (26) and (27) it is derived that (20) holds for every $c_k \in [0, \infty)$. •

Lemma 3 *Let the assumptions of the previous lemma hold and let $g_k = 0$, for some k. Then the algorithm chooses $c_{k+1} = c_k$, in step 2.3. Also, Δx_k is still a descent direction for the merit function Φ at x_k.*

Proof In the previous lemma it was proved that the descent property (20) is satisfied for $g_k = 0$. This basically means that the condition in step 2.3 of Algorithm 1 is always satisfied. Consequently, the algorithm does not need to increase the value of the penalty parameter and simply sets $c_{k+1} = c_k$. For this choice of the penalty parameter it can be verified that the descent property (20) still holds •

Lemma 4 *Let f and g be continuously differentiable functions. Then for μ fixed and for all iterations $k \geq 0$,*

(i) there always exists a constant $c_{k+1} \geq 0$, satisfying step 2.3 of Algorithm 1.

(ii) assuming that the sequence $\{x_k\}$ is bounded, there exists an iteration $k_* \geq 0$, such that if, for all iterations $k \geq k_*$, the values of the penalty parameters $c_k = c_* \in [0, \infty)$, then the descent property (20) is always satisfied, where c_* is the value of the penalty parameter corresponding to iteration k_*.

Proof Part (i) is a direct consequence of Lemmas 1 and 2, since a value $c_{k+1} \geq 0$ is always generated, in step 2.3. To show part (ii), we note that at each iteration, step 2.3 generates a value $c_{k+1} \geq 0$, such that the descent property (20) holds. If c_k is increased infinitely often, then $\{c_k\} \to \infty$. As there exists $c_{k+1} \geq 0$ that Algorithm 1 can choose at every iteration, then there exists $c_* \geq 0$ such that step 2.3 is always satisfied if $c_k = c_*$, for all iterations $k \geq k_*$. •

4.2 Primal step-size rule

In step 2.4 of the algorithm we adopt Armijo's rule to determine the new iterate x_{k+1}. The maximum allowable step-size is determined by the boundary of the feasible region and is given by

$$\alpha_{xk}^{max} = \min_{1 \leq i \leq n} \{ \frac{x_k^i}{-\Delta x_k^i} \; : \; \Delta x_k^i < 0 \}.$$

This is indeed the maximum allowed step, since α_{xk}^{max} gives an infinitely large value to at least one term of the logarithmic barrier function $\sum_{i=1}^{n} log(x_{k+1}^i)$. However, if the step-size is in the interval $[0, \alpha_{xk}^{max})$ then the next primal iterate x_{k+1} is strictly feasible and none of the logarithmic terms becomes infinitely large.

We take as initial step $\hat{\alpha}_{xk}$ a number very close to α_{xk}^{max} and we ensure that it is never greater than one, i.e., $\hat{\alpha}_{xk} = min\{\gamma_k \alpha_{xk}^{max}, 1\}$, with $\gamma_k \in (0, 1)$. The final step is $\alpha_{xk} = \beta^{\theta} \hat{\alpha}_{xk}$, where θ is the first non-negative integer for which Armijo's rule is satisfied and the factor β is usually chosen in the interval $[0.1, 0.5]$, depending on the confidence we have on the initial step $\hat{\alpha}_{xk}$. The value of the parameter ρ is chosen in the interval $[10^{-5}, 10^{-1}]$.

4.3 Dual step-size rule

In this section we discuss the determination of the step-size of the dual variables z. The strategy uses the information provided by the new primal iterate x_{k+1}, in order to find the new iterate z_{k+1}. It is a modification of the strategy suggested by Yamashita [7] and Yamashita and Yabe [8].

While the barrier parameter μ is fixed, we determine a step α_{zk}^i along the direction Δz_k^i, for each dual variable z_k^i, $i = 1, 2, ...n$, such that the box constraints are satisfied

$$\alpha_{zk}^i = \max\{\alpha > 0 : \; LB_k^i \leq (x_k^i + \alpha_{xk}\Delta x_k^i)(z_k^i + \alpha \Delta z_k^i) \leq UB_k^i\}. \tag{28}$$

The lower bounds LB_k^i and upper bounds UB_k^i, $i = 1, 2, ..., n$ are defined as

$$LB_k^i = \min\{\frac{1}{2}m\mu, (x_k^i + \alpha_{xk}\Delta x_k^i)z_k^i\} \text{ and } UB_k^i = \max\{2M\mu, (x_k^i + \alpha_{xk}\Delta x_k^i)z_k^i\}, \tag{29}$$

where the parameters m and M are chosen such that

$$0 < m \leq \min\{1, \frac{(1-\gamma_k)(1-\frac{\gamma_k}{2^\mu})\min_i\{x_k^i z_k^i\}}{\mu}\}, \tag{30}$$

and

$$M \geq \max\{1, \frac{\max_i\{x_k^i z_k^i\}}{\mu}\} > 0, \tag{31}$$

with $\gamma_k \in (0, 1)$. These two parameters are always fixed to constants, while μ is fixed. These constants satisfy inequalities (30) and (31) for $k = 0$. The values of m and M change when the barrier parameter μ is decreased.

The common dual step length α_{zk} is the minimum of all individual step lengths α_{zk}^i with the restriction of being always not more that one, namely

$$\alpha_{zk} = min\{1, \; \min_{1 \leq i \leq n}\{\alpha_{zk}^i\}\}.$$

The step-size for the dual variables y can be either $\alpha_{yk} = 1$ or $\alpha_{yk} = \alpha_{zk}$. The convergence results mentioned in the following sections hold for both choices.

The dual step-size rule is similar to the one that Yamashita and Yabe [8] proposed. The dual step-size rule of Algorithm 1, however, uses different lower and upper bounds in the box constraints defined in (28). It is shown in section 6 that both the lower and upper bounds of Algorithm 1 are smaller than the ones used in [8]. As a result, the dual steps of Algorithm 1 are greater than those determined in [8].

5 Global Convergence

In this section we show that the algorithm is globally convergent, because it always guarantees progress towards a solution from any starting point. El-Bakry *et al.* [22] and Yamashita [7] have also shown global convergence for their primal-dual algorithms. In [22] the global convergence is achieved by determining a single step-size for all the variables. The Armijo rule is used to guarantee that the Euclidean norm of the KKT conditions, which plays the role of a merit function, is reduced at each iteration.

Algorithm 1 is closely related to that in [7], since both algorithms use the same strategies to determine different step-sizes for the primal and dual variables. They use, however, different merit functions to guarantee global convergence. In [7], the non-differentiable merit function

$$\phi(x;\hat{\rho},\mu) = f(x) + \hat{\rho}\sum_{i=1}^{q} |g^i(x)| - \mu \sum_{i=1}^{n} \log(x^i)$$

is used, while we have chosen the merit function defined by (16), which has the useful property to be continuously differentiable. An advantage of the differentiability of the merit function Φ is that the adaptive strategy, developed by Rustem in [2], is used to determine the penalty parameter c, to ensure a descent property for Φ, leading to the global convergence of the algorithm.

We show that, while the barrier parameter is fixed to a value μ^l, the algorithm produces iterates $w_k(\mu^l) = (x_k(\mu^l), y_k(\mu^l), z_k(\mu^l))$, for $k = 1, 2, \ldots$, which are bounded and converge to a $w_*(\mu^l) = (x_*(\mu^l), y_*(\mu^l), z_*(\mu^l))$ such that

$$\| F(x_*(\mu^l), y_*(\mu^l), z_*(\mu^l); c_*, \mu^l) \| = 0,$$

where $F(x, y, z; c, \mu)$ is the vector of the perturbed KKT conditions, defined in (13). In other words, we show that the inner iterations (2.1) - (2.5) of Algorithm 1, converge to a solution of the perturbed KKT conditions. For simplicity we suppress the index l, and we use w_k instead of $w_k(\mu^l)$ to denote the iterates produced while $\mu = \mu^l$.

The basic result of Lemmas 1 to 4 is that the direction Δx_k, taken from the solution of the Newton system (14), is a descent direction for the merit function Φ at the current point x_k, that is inequality (20) holds. In the next theorem we show that the sequence $\{\Phi(x_k; c_*, \mu)\}$ is monotonically decreasing if the barrier parameter μ is fixed. We also show that the step α_{xk}, chosen in step 2.4 is always positive.

Theorem 1 *Assume that the following conditions hold*

i. the objective function f and the constraints g are twice continuously differentiable,

ii. the Hessian matrix $H_k = \nabla^2 f_k + \sum_{i=1}^{m} \nabla^2 g_k^i (c_k g_k^i - y_k^i) + c_k \nabla g_k^T \nabla g_k$, is such that, for every iteration k and for every vector $v \in \Re^n$ there exist constants $M' > m' > 0$, such that

$$m' \| v \|_2^2 \leq v^T H_k v \leq M' \| v \|_2^2,$$

iii. for each iteration k, there exists a triple $(\Delta x_k, \Delta y_k, \Delta z_k)$, as a solution to the Newton system (14),

iv. there exists an iteration k_* and a scalar $c_* \geq 0$, such that the penalty parameter restriction in step 2.3

$$\Delta x_k^T \nabla f_k - c_k \| g_k \|_2^2 - \mu \Delta x_k^T X_k^{-1} e + \| \Delta x_k \|_{H_k}^2 \leq 0$$

is satisfied for all $k \geq k_*$ with $c_{k+1} = c_k = c_*$.

Then the step-size computed in step 2.4 is such that $\alpha_{xk} \in (0, 1]$ and hence the sequence $\{\Phi(x_k; c_*, \mu)\}$ is monotonically decreasing, for $k \geq k_*$ and μ fixed.

Proof Consider the first order approximation with remainder of the function $\Phi(x; c_*, \mu)$ around the point $x_{k+1} = x_k + \alpha_{xk}\Delta x_k$

$$\Phi(x_{k+1}; c_*, \mu) - \Phi(x_k; c_*, \mu) =$$

$$\alpha_{xk}\Delta x_k^T \nabla\Phi(x_k; c_*, \mu) + \alpha_{xk}^2 \int_0^1 (1-t)\Delta x_k^T \nabla_x^2 \Phi(x_k + t\alpha_{xk}\Delta x_k; c_*, \mu)\Delta x_k \, dt.$$

The above equation, after adding and subtracting the Hessian matrix H_k in the remainder, yields

$$\Phi(x_{k+1}; c_*, \mu) - \Phi(x_k; c_*, \mu) \ \leq \ \alpha_{xk}\Delta x_k^T \nabla\Phi(x_k; c_*, \mu) \ + \ \tfrac{1}{2}\alpha_{xk}^2 \Delta x_k^T H_k \Delta x_k$$

$$+\alpha_{xk}^2 \int_0^1 (1-t) \mid \Delta x_k^T \ (\ \nabla_x^2 \Phi(x_k + t\alpha_{xk}\Delta x_k; c_*, \mu) - H_k \ ) \ \Delta x_k \mid dt. \tag{32}$$

We note that for every symmetric matrix A (eg, [6])

$$\parallel A \parallel_2 \quad = \quad \text{max eigenvalue of } A.$$

Hence inequality (32) can take the form

$$\Phi(x_{k+1}; c_*, \mu) \ - \ \Phi(x_k; c_*, \mu) \ \leq$$

$$\alpha_{xk}\Delta x_k^T \nabla\Phi(x_k; c_*, \mu) \ + \ \frac{\alpha_{xk}^2}{2}\Delta x_k^T H_k \Delta x_k + \alpha_{xk}^2 \psi_k \parallel \Delta x_k \parallel_2^2, \tag{33}$$

where

$$\psi_k = \int_0^1 (1-t) \ \parallel \nabla_x^2 \Phi(x_k + t\alpha_{xk}\Delta x_k; c_*, \mu) - H_k \ \parallel_2 \ dt.$$

Furthermore, from assumption (ii) we have

$$\parallel \Delta x_k \parallel_2^2 \leq \frac{1}{m'}\Delta x_k^T H_k \Delta x_k, \tag{34}$$

and from Lemmas 1 and 2

$$\Delta x_k^T H_k \Delta x_k \leq -\Delta x_k^T \nabla\Phi(x_k; c_*, \mu). \tag{35}$$

Substituting (34) and (35) into (33) yields

$$\Phi(x_{k+1}; c_*, \mu) - \Phi(x_k; c_*, \mu) \ \leq \alpha_{xk}\Delta x_k^T \nabla\Phi(x_k; c_*, \mu)(1 - \alpha_{xk}(\frac{1}{2} + \frac{\psi_k}{m'})). \tag{36}$$

The scalar ρ in Armijo's rule in step 2.4 determines a step-length α_{xk} such that

$$\rho \leq 1 - \alpha_{xk}(\frac{1}{2} + \frac{\psi_k}{m'}) \leq \frac{1}{2}.$$

Since from Lemmas 1 to 4 we always have $\Delta x_k^T \nabla\Phi(x_k; c_*, \mu) \leq 0$, there must exist $\alpha_{xk} \in (0,1]$, to ensure (36) and Armijo's rule in step 2.4. Assume that α^0 is the largest step in the interval $(0,1]$ satisfying both (36) and Armijo's rule. Consequently for every $\alpha \leq \alpha^0$, inequality (36) and Armijo's rule are also satisfied. Hence the strategy in step 2.4 always selects a step-length $\alpha_{xk} \in [\beta\alpha^0, \alpha^0]$, where $0 < \beta \leq 1$. From the above analysis, it follows that the sequence $\{\Phi(x_k; c_*, \mu)\}$ is monotonically decreasing. •

Remark 1 *The results of the above theorem can be proved to hold before the penalty parameter c_k achieves a constant value c_*. This can be done by considering the difference $\Phi(x_{k+1}; c_{k+1}, \mu) - \Phi(x_k; c_{k+1}, \mu)$ and the Taylor expansion of the function $\Phi(x; c_{k+1}, \mu)$ instead of $\Phi(x; c_*, \mu)$. In the above theorem we chose to prove the case where $c_k = c_*$ has been achieved, in order to show that asymptotically, Φ is monotonically decreasing and the strategy in step 2.4 selects a step-length $\alpha_{xk} \in (0,1]$.*

An immediate consequence of the above theorem is that the sequence $\{x_k\}$ is bounded away from zero. This is established in the following corollary.

Corollary 1 *The sequence $\{x_k\}$ of primal variables generated by Algorithm 1, with μ fixed, is bounded away from zero.*

Proof Assume to the contrary that the sequence $\{x_k\} \to 0$. Then the sequence $\{-\sum_{i=1}^n \log(x_k^i)\} \to \infty$. From the assumption that the feasible region is bounded we conclude that the sequences $\{f(x_k)\}$ and $\{\| g(x_k) \|\}$ are also bounded. Hence $\{\Phi(x_k; c_*, \mu)\} \to \infty$ which contradicts the monotonic decrease of Φ. •

The following lemma, proved by Yamashita in [7], shows that the dual step-size rule, used by the Algorithm 1, generates iterates z_k which are also bounded above and away from zero.

Lemma 5 *While μ is fixed, the lower bounds LB_k^i and the upper bounds UB_k^i, $i = 1, 2, ..., n$, of the box constraints in the dual step-size rule, are bounded away from zero and bounded from above respectively, if the corresponding components x_k^i, of the iterates x_k are also bounded above and away from zero.*

Proof The proof can be found in [7]. •

Having established that the sequences of iterates $\{x_k\}$ and $\{z_k\}$ are bounded above and away from zero, we show that the iterates $\{y_k\}$, $k \geq 0$ are also bounded. In particular Lemma 7 shows that if at each iteration of the algorithm we take a unit step along the direction Δy_k, then the resulting sequence $\{y_k + \Delta y_k\}$ is bounded. In addition to this, Lemma 7 also shows that the Newton direction $\Delta w_k = (\Delta x_k, \Delta y_k, \Delta z_k)$ is bounded, while μ is fixed. We first establish the following technical result.

Lemma 6 *Let w_k is a sequence of vectors generated by Algorithm 1 for μ fixed. Then the matrix sequence $\{\Theta_k^{-1}\}$ is bounded, where*

$$\Theta_k = \begin{pmatrix} 0 & \nabla g_k \\ -\nabla g_k^T & H_k + X_k^{-1} Z_k \end{pmatrix}.$$

Proof The inverse of the partitioned matrix Θ_k is

$$\Theta_k^{-1} = \begin{pmatrix} [\nabla g_k \Omega_k \nabla g_k^T]^{-1} & -[\nabla g_k \Omega_k \nabla g_k^T]^{-1} \nabla g_k \Omega_k \\ \Omega_k \nabla g_k^T [\nabla g_k \Omega_k \nabla g_k^T]^{-1} & \Omega_k - \Omega_k \nabla g_k^T [\nabla g_k \Omega_k \nabla g_k^T]^{-1} \nabla g_k \Omega_k \end{pmatrix},$$

where $\Omega_k = (H_k + X_k^{-1} Z_k)^{-1}$. According to Assumption (A4), Corollary 1 and Lemma 5, the matrices Ω_k and $[\nabla g_k \Omega_k \nabla g_k^T]^{-1}$ exist and are bounded. Hence the matrix Θ_k^{-1} is bounded, since all matrices involved in it are bounded. •

Lemma 7 *Let w_k is a sequence of vectors generated by Algorithm 1 for μ fixed. Then the sequence of vectors $\{(\Delta x_k, y_k + \Delta y_k, \Delta z_k)\}$ is bounded.*

Proof Solving the third equation of the Newton system (14) for Δz_k yields

$$\Delta z_k = -z_k + \mu X_k^{-1} e - X_k^{-1} Z_k \Delta x_k. \tag{37}$$

Substituting (37) into the first equation of (14) and re-arranging the first two equations, yields the following reduced system

$$\begin{pmatrix} 0 & \nabla g_k \\ -\nabla g_k^T & H_k + X_k^{-1} Z_k \end{pmatrix} \begin{pmatrix} y_k' \\ \Delta x_k \end{pmatrix} = - \begin{pmatrix} g_k \\ \nabla f_k + c_k \nabla g_k^T g_k - \mu X_k^{-1} e \end{pmatrix} \tag{38}$$

where $y_k' = y_k + \Delta y_k$. From the previous lemma we have that the inverse of the matrix in the left side of (38) exists and is bounded. Hence the sequences $\{\Delta x_k\}$ and $\{y_k'\}$ are also bounded. Considering now (37), we can easily deduce that the sequence $\{\Delta z_k\}$ is bounded, since it is a sum of bounded sequences. •

Lemmas 8 and 9 provide the necessary results needed by Theorem 2, which shows that the sequence of $\{w_k\}$ converges to a point $w_* = (x_*, y_*, z_*)$, satisfying the KKT conditions of problem (12).

Lemma 8 *Let the assumptions of Theorem 1 be satisfied and the barrier parameter μ is fixed. Also let for some iteration $k_0 \geq 0$, the level set*

$$S_1 = \{x \in \Re_+^n : \Phi(x; c_*, \mu) \leq \Phi(x_{k_0}; c_*, \mu)\} \tag{39}$$

is compact. Then for all $k \geq k_0$ we have

$$\lim_{k\to\infty} \Delta x_k^T \nabla\Phi(x_k; c_*, \mu) = 0. \tag{40}$$

Proof The scalar $\rho \in (0, 1/2)$ in the step-size strategy at step 2.4, determines α_{xk} such that

$$\rho \leq 1 - \alpha_{xk}(\frac{1}{2} + \frac{\psi_k}{m'}) \leq \frac{1}{2},$$

and by solving for α_{xk} we obtain

$$\frac{1/2}{1/2 + \psi_k/m'} \leq \alpha_{xk} \leq \frac{1-\rho}{1/2 + \psi_k/m'}.$$

Hence the largest value that the step-length α_{xk} can take and still satisfy Armijo's rule in step 2.4 is

$$\alpha_{xk}^0 = \min\{1, \frac{1-\rho}{1/2 + \psi_k/m'}\}.$$

Recall that the step-length α_{xk} is chosen by reducing the maximum allowable step-length $\hat{\alpha}_{xk}$ until Armijo's rule is satisfied. Therefore $\alpha_{xk} \in [\beta\alpha_{xk}^0, \alpha_{xk}^0]$ and thereby also satisfies Armijo's rule.

As the merit function Φ is twice continuously differentiable and the level set S_1 is bounded, there exists a scalar $\bar{M} < \infty$ such that

$$\psi_k = \int_0^1 (1-t) \parallel \nabla_x \Phi(x_k + t\alpha_{xk}\Delta x_k; c_*, \mu) - H_k \parallel_2 dt \leq \bar{M} < \infty.$$

Thus we always have $\alpha_{xk} \geq \bar{\alpha}_{xk} > 0$, where

$$\bar{\alpha}_{xk} = min\{1, \frac{1-\rho}{1/2 + \bar{M}/m'}\}.$$

Hence the step-size α_{xk} is always bounded away from zero.

Furthermore, from Armijo's rule and lemmas 1 and 2 we have

$$\Phi(x_{k+1}; c_*, \mu) - \Phi(x_k; c_*, \mu) \leq \rho\alpha_{xk}\nabla\Phi(x_k; c_*, \mu)^T \Delta x_k < 0. \tag{41}$$

From our assumption that the level set S_1 is bounded, it can be deduced that

$$\lim_{k\to\infty} |\Phi(x_{k+1}; c_*, \mu) - \Phi(x_k; c_*, \mu)| = 0.$$

Consequently, from (41)

$$\lim_{k\to\infty} (\rho\alpha_{xk}\nabla\Phi(x_k; c_*, \mu)^T \Delta x_k) = 0.$$

Finally, since ρ, $\alpha_{xk} > 0$ it can be deduced that (40) holds. •

Lemma 9 *Let the assumptions of the previous lemma hold. Then*

$$\lim_{k\to\infty} \parallel \Delta x_k \parallel_{H_k}^2 = 0. \tag{42}$$

Proof From (20) we have

$$-\nabla\Phi(x_k; c_*, \mu)^T \Delta x_k \geq \| \Delta x_k \|_{H_k}^2 . \tag{43}$$

Hence from (43) and (40) we have that (42) holds. •

Theorem 2 *Let the assumptions of the previous lemma hold. Then the algorithm terminates at a KKT point, satisfying the first order necessary conditions of problem (12), and at that point the perturbed KKT conditions (13) are satisfied, for μ fixed.*

Proof Let $x_*(\mu)$, $z_*(\mu) \in \Re^n$ and $y_*(\mu) \in \Re^q$ be such that $\lim_{k\to\infty} x_k = x_*(\mu)$, $\lim_{k\to\infty} z_k = z_*(\mu)$, and $\lim_{k\to\infty} y_k = y_*(\mu)$, $\forall k \geq k_*$, $k \in K \subseteq \{1, 2, ...\}$. The existence of such points is ensured since by Assumption A2 and Lemmas 5 and 7, the sequence $\{(x_k(\mu), y_k(\mu), z_k(\mu))\}$ is bounded for μ fixed, and by Theorem 1 the algorithm always decreases the merit function Φ sufficiently at each iteration, thereby ensuring $x_k \in S_1$, with S_1 compact.

Solving the third equation of the Newton system (14) for Δz_k yields

$$\Delta z_k = -X_k^{-1} Z_k \Delta x_k - z_k + \mu X_k^{-1} e.$$

Taking limits and using Lemma 7, the above equation yields

$$\lim_{k\to\infty} (z_k + \Delta z_k) = \lim_{k\to\infty} \mu X_k^{-1} e. \tag{44}$$

By defining $z'_k = z_k + \Delta z_k$, from Lemmas 5 and 7, it can be derived that the sequence $\{z'_k\}$ is bounded, since it is a sum of bounded sequences. Hence, there exits a vector z'_* such that

$$\lim_{k\to\infty} z'_k = z'_* = \mu X_*^{-1} e. \tag{45}$$

Furthermore, writing $y'_k = y_k + \Delta y_k$, the first equation of the Newton system (14) takes the form

$$H_k \Delta x_k - \nabla g_k^T y'_k = -\nabla f_k + z'_k - c_k \nabla g_k^T g_k.$$

Letting $k \to \infty$, and applying Lemma 9 and (45) the above equation becomes

$$\nabla f_* - \mu X_*^{-1} e + c_* \nabla g_*^T g_* - \nabla g_*^T y'_* = 0, \tag{46}$$

where we have set $y'_* = y'_*(\mu) = \lim_{k\to\infty}(y_k + \Delta y_k)$. Similarly, letting $k \to \infty$ the second equation of the Newton system (14) yields

$$\lim_{k\to\infty} (\nabla g_k \Delta x_k) = \lim_{k\to\infty} (-g_k).$$

Applying Lemma 9 the above equation yields

$$g_* = 0. \tag{47}$$

From (45), (46) and (47) we can conclude that the vector $x_*(\mu)$ is the optimum solution of problem (12), while the triple $(x_*(\mu), y'_*(\mu), z'_*(\mu))$ solves the perturbed KKT conditions (13), for μ fixed. •

An immediate consequence of Theorem 2 is that, for any convergent subsequence, produced by the algorithm, for $\mu = \mu^l$, there is an iteration $\tilde{k}$, such that

$$\| F(x_{\tilde{k}}, y_{\tilde{k}}, z_{\tilde{k}}; c_*, \mu) \| \leq \eta\mu, \tag{48}$$

for all $k \geq \tilde{k}$, where $\eta \in (0,1)$ and $F(x, y, z; c, \mu)$ is given by (13). At this point, we record the value of the current iterate

$$(\tilde{x}^l, \tilde{y}^l, \tilde{z}^l) = (x_{\tilde{k}}, y_{\tilde{k}}, z_{\tilde{k}}),$$

and set μ to a smaller value $\mu^{l+1} < \mu^l$. Therefore a sequence of approximate central points $\{(\tilde{x}^l, \tilde{y}^l, \tilde{z}^l)\}$ is generated.

In the remaining part of this section, we show that the sequence of approximate central points converges to a KKT point $(\tilde{x}^*, \tilde{y}^*, \tilde{z}^*)$ of the initial constrained optimization problem (4).

For a given $\epsilon \geq 0$, consider the set of all the approximate central points, generated by Algorithm 1

$$S_2(\epsilon) = \{(\tilde{x}^l, \tilde{y}^l, \tilde{z}^l) : \epsilon \leq \| F(\tilde{x}^l, \tilde{y}^l, \tilde{z}^l; c_*, \mu^l) \| \leq \| F(\tilde{x}^0, \tilde{y}^0, \tilde{z}^0; c_*, \mu^0) \|, \ \forall \mu^l < \mu^0\}.$$

If $\epsilon > 0$ then the step-size rules, described in section 4 guarantee that $\tilde{x}^l, \tilde{z}^l \in S_2(\epsilon)$ are bounded away from zero, for $l \geq 0$. Consequently $(\tilde{x}^l)^T \tilde{z}^l$ is also bounded away from zero in $S_2(\epsilon)$. The following lemma shows that the sequence $\{\tilde{y}^l\}$ is bounded if the sequence $\{\tilde{z}^l\}$ is also bounded.

Lemma 10 *Assuming that the columns of $\nabla g(\tilde{x}^l)$ are linearly independent and the iterates $\tilde{x}^l$ are in a compact set for $l \geq 0$, then there exists a constant $M_1 > 0$ such that*

$$\| \tilde{y}^l \| \leq M_1(1 + \| \tilde{z}^l \|).$$

Proof By defining $r^l = \nabla f(\tilde{x}^l) - \tilde{z}^l + c\nabla g(\tilde{x}^l)^T g(\tilde{x}^l) - \nabla g(\tilde{x}^l)^T \tilde{y}^l$ and solving for $\nabla g(\tilde{x}^l)^T \tilde{y}^l$ we obtain

$$\nabla g(\tilde{x}^l)^T \tilde{y}^l = \nabla f(\tilde{x}^l) - \tilde{z}^l + c\nabla g(\tilde{x}^l)^T g(\tilde{x}^l) - r^l.$$

From our assumptions the above equation can be written as

$$\begin{aligned} \tilde{y}^l &= [\nabla g(\tilde{x}^l) \nabla g(\tilde{x}^l)^T]^{-1} \nabla g(\tilde{x}^l) \, (\nabla f(\tilde{x}^l) + c\nabla g(\tilde{x}^l)^T g(\tilde{x}^l) - r^l) \\ &\quad - [\nabla g(\tilde{x}^l) \nabla g(\tilde{x}^l)^T]^{-1} \nabla g(\tilde{x}^l) \, \tilde{z}^l. \end{aligned}$$

Taking norms in both sides of the above equation yield

$$\begin{aligned} \| \tilde{y}^l \| &\leq \| [\nabla g(\tilde{x}^l) \nabla g(\tilde{x}^l)^T]^{-1} \nabla g(\tilde{x}^l) \| \, \| \nabla f(\tilde{x}^l) + c\nabla g(\tilde{x}^l)^T g(\tilde{x}^l) - r^l \| \\ &\quad + \| [\nabla g(\tilde{x}^l) \nabla g(\tilde{x}^l)^T]^{-1} \nabla g(\tilde{x}^l) \| \, \| \tilde{z}^l \| \\ &\leq M_1(1 + \| \tilde{z}^l \|). \end{aligned}$$

where the constant $M_1 < \infty$ is defined as

$$\begin{aligned} M_1 \;\geq\; & \max\{ \parallel [\nabla g(\tilde{x}^l)\nabla g(\tilde{x}^l)^T]^{-1}\nabla g(\tilde{x}^l) \parallel \parallel \nabla f(\tilde{x}^l) + c\nabla g(\tilde{x}^l)^T g(\tilde{x}^l) - r^l \parallel, \\ & \parallel [\nabla g(\tilde{x}^l)\nabla g(\tilde{x}^l)^T]^{-1}\nabla g(\tilde{x}^l) \parallel \}. \end{aligned}$$

•

Lemma 11 *If $(\tilde{x}^l, \tilde{y}^l, \tilde{z}^l) \in S_2(\epsilon)$ for all $l \geq 0$, then the sequence $\{(\tilde{x}^l, \tilde{y}^l, \tilde{z}^l)\}$ is bounded above.*

Proof From Lemma 10, it suffices to prove that the sequences $\{\tilde{x}^l\}$ and $\{\tilde{z}^l\}$ are bounded from above. By assumption (A2), the sequence $\{\tilde{x}^l\}$ is bounded. Assume that there exists a non-empty set I_z^∞, which contains the indices i of those elements, $(\tilde{z}^l)^i$, of the vector $\tilde{z}^l$, for which $\lim_{l\to\infty}(\tilde{z}^l)^i = \infty$. From the boundedness of the sequences $\{(\tilde{x}^l)^i(\tilde{z}^l)^i\}$, $i = 1, 2, ..., n$, we obtain $\liminf_{l\to\infty}(\tilde{x}^l)^i = 0$, for those indices $i \in I_z^\infty$. Furthermore from the definition of the set I_x^0, in Assumption (A4), it is evident that $I_z^\infty \subseteq I_x^0$.

From (48) and the fact that $\{\mu^l\} \to 0$ we have that the sequence

$$\{\parallel \nabla f(\tilde{x}^l) - \tilde{z}^l + c_* \nabla g(\tilde{x}^l)^T g(\tilde{x}^l) - \nabla g(\tilde{x}^l)^T \tilde{y}^l \parallel\}$$

is bounded. Using this and the fact that the sequences $\{\parallel \nabla f(\tilde{x}^l) \parallel\}$ and $\{\parallel c\nabla g(\tilde{x}^l)^T g(\tilde{x}^l) \parallel\}$ are bounded, we conclude that $\{\parallel -\tilde{z}^l - \nabla g(\tilde{x}^l)^T \tilde{y}^l \parallel\}$ is also bounded. Hence, we have

$$\frac{\parallel \tilde{z}^l + \nabla g(\tilde{x}^l)^T \tilde{y}^l \parallel}{\parallel (\tilde{y}^l, \tilde{z}^l) \parallel} \to 0 \tag{49}$$

By setting $\tilde{u}^l = \frac{(\tilde{y}^l, \tilde{z}^l)}{\|(\tilde{y}^l, \tilde{z}^l)\|}$, we have $\{\tilde{u}^l\}$ bounded and $\{\tilde{u}^l\} \to \tilde{u}^*$. It is clear that $\parallel \tilde{u}^* \parallel = 1$ and the components of $\tilde{u}^*$, corresponding to those indices $i \notin I_z^\infty$, i.e., $\{(\tilde{z}^l)^i\} < \infty$, are zero. If $\hat{u}^*$ is the vector consisting of the components of $\tilde{u}^*$ which correspond to the indices $i \in I_z^\infty$, then $\parallel \hat{u}^* \parallel = \parallel \tilde{u}^* \parallel = 1$. Furthermore, from (49) we have

$$\frac{\nabla g(\tilde{x}^l)^T \tilde{y}^l + \tilde{z}^l}{\parallel (\tilde{y}^l, \tilde{z}^l) \parallel} = \frac{[\nabla g(\tilde{x}^l)^T,\ I_n]\,(\tilde{y}^l, \tilde{z}^l)}{\parallel (\tilde{y}^l, \tilde{z}^l) \parallel} = [\nabla g(\tilde{x}^l)^T,\ e_i : i \in I_x^0]\,\hat{u}^* \to 0.$$

However, this result contradicts Assumption (A2). Hence, the set I_z^∞ is empty, or for all indices $i = 1, 2, ..., n$, the sequences $\{(\tilde{z}^l)^i\}$ are bounded. Consequently, $\{\tilde{z}^l\}$ is also bounded. •

The following theorem shows that the sequence $\{(\tilde{x}^l, \tilde{y}^l, \tilde{z}^l)\}$ converges to $(\tilde{x}^*, \tilde{y}^*, \tilde{z}^*)$ which is a KKT point of the initial constrained optimization problem (4).

Theorem 3 *Let $\{\mu^l\}$ is a positive monotonically decreasing sequence of barrier parameters with $\{\mu^l\} \to 0$, and let $\{(\tilde{x}^l, \tilde{y}^l, \tilde{z}^l)\}$ be a sequence of approximate central points satisfying (48) for $\mu = \mu^l$, $l \geq 0$. Then the sequence $\{(\tilde{x}^l, \tilde{y}^l, \tilde{z}^l)\}$ is bounded and its limit point $(\tilde{x}^*, \tilde{y}^*, \tilde{z}^*)$ satisfies the KKT conditions of problem (4).*

Proof From Lemma 10 the sequence $\{(\tilde{x}^l, \tilde{y}^l, \tilde{z}^l)\}$ is bounded. Then it is convergent and let $(\tilde{x}^*, \tilde{y}^*, \tilde{z}^*)$ be its limit point. From (48) and the fact that $\mu^l \to 0$ we easily obtain that $\lim_{l\to\infty} \| F(\tilde{x}^l, \tilde{y}^l, \tilde{z}^l) \| = 0$. Therefore,

$$
\begin{aligned}
\nabla f(\tilde{x}^*) - \tilde{z}^* - \nabla g(\tilde{x}^*)^T \tilde{y}^* &= 0 \\
g(\tilde{x}^*) &= 0 \\
\tilde{X}^* \tilde{Z}^* e &= 0.
\end{aligned}
$$

Clearly from the above equations we may derive that $(\tilde{x}^*, \tilde{y}^*, \tilde{z}^*)$ is a KKT point of the initial constrained optimization problem (4). •

6 Local Convergence

In this section the local convergence of the algorithm is discussed. Without loss of generality, we assume that the value of the barrier parameter changes at each iteration and Armijo's rule does not backtrack and hence the primal step-size α_{xk} determined in step 2.4 is equal to the maximum allowable step-size $\hat{\alpha}_{xk}$. We show that Algorithm 1 converges quadratically to the optimum solution. This is essentially a preservation of the property of the basic Newton algorithm.

Algorithm 1 uses a technique similar to Yamashita and Yabe [8] to determine different step-sizes. The difference in the present algorithm is in the rule used for the step-size α_{zk} of the dual variables y_k and z_k. In particular, the term $1 - \gamma_k/2^{\mu_k} \in (0, 1)$ in the definition of the parameter m, given by (30), results in the lower bound LB_k^i of the box constraints (28), being smaller than the corresponding bound, defined in [8]. As $\{\mu_k\} \to 0$, the term $1 - \gamma_k/2^{\mu_k}$ approaches $1 - \gamma_k$. By noting

$$z_k^i + \alpha_{zk} \Delta z_k^i \geq (1 - \gamma_k) z_k^i > 0, \forall i = 1, 2, ..., n,$$

it can be shown that the inclusion of $1 - \gamma_k/2^{\mu_k}$ in the lower bounds LB_k^i, $i = 1, 2, ..., n$, allows a dual step-size α_{zk}, greater than that in [8]. Also, as $\{\mu_k\} \to 0$, α_{zk} approaches the maximum allowable step-size for the variables z, i.e., $\alpha_{zk} \to \hat{\alpha}_{zk}$, with

$$\hat{\alpha}_{zk} = \max\{1, \gamma_k \max_{1 \leq i \leq n} \{\frac{z_k^i}{-\Delta z_k^i} : \Delta z_k^i < 0\}\}.$$

The next lemma provides useful bounds on the Newton directions near the optimum solution. It is based on the complementarity condition

$$Z_k \Delta x_k + X_k \Delta z_k = -X_k Z_k e + \mu_k e \tag{50}$$

of the perturbed Newton system (14). It has been proved by Yamashita and Yabe in [8] and has also been used by El-Bakry *et al* in [22]. We include this lemma for convenience, since its results are used frequently in the sequel.

Lemma 12 *Let strict complementarity at the optimum solution w_* holds and there exists a constant $\epsilon > 0$ such that if $\| w_k - w_* \| < \epsilon$ then for all i such that $x_*^i = 0$, we have*

$$\frac{\Delta x_k^i}{x_k^i} = -1 + \frac{\mu_k}{x_k^i z_k^i} + \frac{-\Delta z_k^i}{z_k^i} \tag{51}$$

$$\frac{|\Delta z_k^i|}{z_k^i} \leq \kappa \| \Delta w_k \|, \tag{52}$$

while for all i such that $x_*^i > 0$, we have

$$\frac{\Delta z_k^i}{z_k^i} = -1 + \frac{\mu_k}{x_k^i z_k^i} + \frac{-\Delta x_k^i}{x_k^i} \tag{53}$$

$$\frac{|\Delta x_k^i|}{x_k^i} \leq \kappa \| \Delta w_k \|, \tag{54}$$

where κ is defined as follows

$$\kappa = 2 \max\{ \{\frac{1}{x_*^i} : x_*^i > 0, i = 1, 2, ..., n\}, \{\frac{1}{z_*^i} : z_*^i > 0, i = 1, 2, ..., n\} \}.$$

Proof The proof can be found in [8]. •

The following two lemmas show that the step-sizes α_{xk} and α_{zk} are always strictly positive. They also show that both step-sizes approach unity, when we are in a close neighbourhood of w_*.

Lemma 13 *Let the assumptions of lemma (12) are satisfied. If*

$$\kappa \| \Delta w_k \| \leq \frac{1}{2}\gamma_k, \quad \textit{with } \gamma_k \in (0, 1),$$

then we have,

$$1 \geq \alpha_{xk} \geq \gamma_k - \kappa \| \Delta w_k \| . \tag{55}$$

Proof The proof can be found in [8] •

Lemma 14 *Let the assumptions of the previous lemma hold. If*

$$\kappa \parallel \Delta w_k \parallel \leq \frac{1}{2}\gamma_k, \quad with\ \gamma_k \in (0,1),$$

then the step-size α_{zk} is determined by the formula:

$$\alpha_{zk} = mim\{1, \min\{\alpha > 0 : \ z_k^i + \alpha \Delta z_k^i = \frac{\frac{1}{2} m \mu_k}{x_k^i + \alpha_{xk} \Delta x_k^i}, \ \Delta z_k^i < 0\}. \tag{56}$$

Also the values that the step-size α_{zk} take are in the following interval:

$$1 \geq \alpha_{zk} \geq 1 - \kappa \parallel \Delta w_k \parallel -(1-\gamma_k)(1 - \frac{\gamma_k}{2^{\mu_k}}). \tag{57}$$

Proof Equality (56) basically implies that only those indices i for which $\Delta z_k^i < 0$, contribute to the determination of α_{zk}. First we prove that the upper bound constraint in (28) is always not active. This is obvious for all those indices i for which $\Delta z_k^i < 0$. Hence, we examine whether the upper bound constraint is also not active when $\Delta z_k^i \geq 0$. To prove this, it suffices to show that if $\alpha_{zk}^i = 1$ (i.e., the maximum allowed step we can take when $\Delta z_k^i \geq 0$) then we always have

$$(x_k^i + \alpha_{xk} \Delta x_k^i)(z_k^i + \Delta z_k^i) \leq 2M\mu_k, \tag{58}$$

where M is defined in (31). We can distinguish the following four cases:

CASE A: If $x_*^i = 0$ and $\Delta x_k^i \geq 0$ then by using (51) and (52) we have

$$\begin{aligned}
(x_k^i + \alpha_{xk} \Delta x_k^i)(z_k^i + \Delta z_k^i) &\leq x_k^i z_k^i (1 + \frac{\Delta x_k^i}{x_k^i})(1 + \frac{\Delta z_k^i}{z_k^i}) \\
&= x_k^i z_k^i (1 - 1 + \frac{\mu_k}{x_k^i z_k^i} - \frac{\Delta z_k^i}{z_k^i})(1 + \frac{\Delta z_k^i}{z_k^i}) \\
&= x_k^i z_k^i (1 + \frac{\mu_k}{x_k^i z_k^i})(1 + \frac{\Delta z_k^i}{z_k^i}) - x_k^i z_k^i (1 + \frac{\Delta z_k^i}{z_k^i})^2 \\
&\leq x_k^i z_k^i (1 + \frac{\mu_k}{x_k^i z_k^i})\frac{3}{2} - x_k^i z_k^i \\
&\leq \frac{1}{2} \max_i \{x_k^i z_k^i\} + \frac{3}{2}\mu_k \\
&\leq \frac{1}{2} \max\{1, \frac{\max_i\{x_k^i z_k^i\}}{\mu_k}\}\mu_k + \frac{3}{2} \max\{1, \frac{\max_i\{x_k^i z_k^i\}}{\mu_k}\}\mu_k \\
&= 2M\mu_k.
\end{aligned}$$

CASE B: If $x_*^i = 0$ and $\Delta x_k^i < 0$ then from (52) we have

$$\begin{aligned}
(x_k^i + \alpha_{xk} \Delta x_k^i)(z_k^i + \Delta z_k^i) &= x_k^i z_k^i (1 + \alpha_{xk} \frac{\Delta x_k^i}{x_k^i})(1 + \frac{\Delta z_k^i}{z_k^i}) \\
&< x_k^i z_k^i (1 + \frac{\Delta z_k^i}{z_k^i})
\end{aligned}$$

$$
\begin{aligned}
&\le x_k^i z_k^i (1 + \frac{1}{2}) \\
&\le \frac{3}{2} \max_i \{x_k^i z_k^i\} \\
&\le \frac{3}{2} \max\{1, \frac{\max_i\{x_k^i z_k^i\}}{\mu_k}\}\mu_k \\
&< 2M\mu_k.
\end{aligned}
$$

CASE C: If $x_*^i > 0$ and $\Delta x_k^i \ge 0$ then, working as in case (A), and using (53) and (54) we obtain

$$
(x_k^i + \alpha_{xk}\Delta x_k^i)(z_k^i + \Delta z_k^i) \quad \le \quad 2M\mu_k.
$$

CASE D: If $x_*^i > 0$ and $\Delta x_k^i < 0$ then, working as in case (B), and using (53) we obtain

$$
(x_k^i + \alpha_{xk}\Delta x_k^i)(z_k^i + \Delta z_k^i) \quad \le \tfrac{3}{2} M\mu_k \ < 2M\mu_k.
$$

Hence the step-size α_{zk}^i is determined from the lower bound constraint, which becomes active when $\Delta z_k^i < 0$. We first show that, once the new primal iterate $x_{k+1}^i = x_k^i + \alpha_{xk}\Delta x_k^i$ is known, the corresponding product $x_{k+1}^i z_k^i$ is always strictly less than the lower bound constraint. Indeed, by observing that for all $i = 1, 2, ..., n$, $x_k^i + \alpha_{xk}\Delta x_k^i \ge (1-\gamma_k)x_k^i$, we have

$$
\begin{aligned}
\frac{\frac{1}{2}m\mu_k}{x_k^i + \alpha_{xk}\Delta x_k^i} &\le \frac{\frac{1}{2}m\mu_k}{(1-\gamma_k)x_k^i} \\
&< \frac{\frac{1}{2}m\mu_k}{(1-\gamma_k)(1-\frac{\gamma_k}{2^{\mu_k}})x_k^i z_k^i} z_k^i \\
&\le \frac{\frac{1}{2}m\mu_k}{(1-\gamma_k)(1-\frac{\gamma_k}{2^{\mu_k}})\min_i\{x_k^i z_k^i\}} z_k^i \\
&\le \frac{\frac{1}{2}m\mu_k}{\min\{\ 1,\ \frac{(1-\gamma_k)(1-\frac{\gamma_k}{2^{\mu_k}})\min_i\{x_k^i z_k^i\}}{\mu_k}\}\mu_k} z_k^i \\
&< z_k^i.
\end{aligned}
$$

Therefore from the above analysis we deduce that the step-size α_{zk} is determined by (56).

Finally we show that (57) holds. If $\Delta z_k^i \ge 0$ for all i, then from the previous discussion can be shown that $\alpha_{zk}^i = 1$ and thus (57) holds. Therefore assume that there exists at least one index i such that $\Delta z_k^i < 0$. Working similarly as in cases (A)-(D), it can be shown that if $x_*^i = 0$ then the lower bound constraint in (28) is not violated by taking a unit step-size along the direction $\Delta z_k^i < 0$. Indeed, we have

$$
(x_k^i + \alpha_{xk}\Delta x_k^i)(z_k^i + \Delta z_k^i) \quad \ge \quad (1-\gamma_k)x_k^i z_k^i(1 + \frac{\Delta z_k^i}{z_k^i})
$$

$$
\begin{aligned}
&\geq (1-\gamma_k) x_k^i z_k^i (1-\frac{1}{2}) \\
&> \frac{1}{2}(1-\gamma_k)(1-\frac{\gamma_k}{2^{\mu_k}}) \min_i \{x_k^i z_k^i\} \\
&\geq \frac{1}{2} \min\{\, 1,\ \frac{(1-\gamma_k)(1-\frac{\gamma_k}{2^{\mu_k}}) \min_i\{x_k^i z_k^i\}}{\mu_k} \} \mu_k \\
&= \frac{1}{2} m \mu_k .
\end{aligned}
$$

Hence (57) is again satisfied if $\Delta z_k^i < 0$ and $x_*^i = 0$.

Consider the case we have not yet investigated that is, when $\Delta z_k^i < 0$ and $x_*^i > 0$. It is not certain whether unit step-sizes are allowed, as this might violate the lower bound constraint in (28) or the feasible region (i.e., $z_k^i + \Delta z_k^i < 0$). Therefore, the step-size α_{zk}^i for the i-th dual variable is determined by

$$
z_k^i + \alpha_{zk}^i \Delta z_k^i = \frac{\frac{1}{2} m \mu_k}{x_k^i + \alpha_{xk} \Delta x_k^i} .
$$

Noting that $x_k^i + \alpha_{xk} \Delta x_k^i \geq x_k^i (1 - \kappa \parallel \Delta w_k \parallel)$, the above equation becomes

$$
z_k^i + \alpha_{zk}^i \Delta z_k^i \leq \frac{\frac{1}{2} m \mu_k}{x_k^i (1 - \kappa \parallel \Delta w_k \parallel)},
$$

and solving for α_{zk}^i yields

$$
\alpha_{zk}^i \geq -\frac{z_k^i}{\Delta z_k^i} (1 - \frac{\frac{1}{2} m \mu_k}{x_k^i z_k^i (1 - \kappa \parallel \Delta w_k \parallel)}). \tag{59}
$$

Since $x_*^i > 0$, from (53) we have

$$
\frac{\Delta z_k^i}{z_k^i} = -1 + \frac{\mu_k}{x_k^i z_k^i} - \frac{\Delta x_k^i}{x_k^i} > -1 - \frac{\Delta x_k^i}{x_k^i} \geq -1 - \kappa \parallel \Delta w_k \parallel .
$$

Substituting the above inequality into (59) we obtain

$$
\alpha_{zk}^i \geq \frac{1}{1 + \kappa \parallel \Delta w_k \parallel} (1 - \frac{\frac{1}{2} m \mu_k}{x_k^i z_k^i (1 - \kappa \parallel \Delta w_k \parallel)}). \tag{60}
$$

Furthermore, since it is assumed that $\kappa \parallel \Delta w_k \parallel \leq \frac{1}{2}$ we have $1 - (\kappa \parallel \Delta w_k \parallel)^2 \leq 1$, and therefore

$$
\frac{1}{1 + \kappa \parallel \Delta w_k \parallel} > 1 - \kappa \parallel \Delta w_k \parallel .
$$

Substituting the above inequality into (60) yields

$$
\begin{aligned}
\alpha_{zk}^i &\geq 1 - \kappa \parallel \Delta w_k \parallel - \frac{1}{2} \frac{m \mu_k}{x_k^i z_k^i} \\
&\geq 1 - \kappa \parallel \Delta w_k \parallel - \frac{1}{2} \frac{m \mu_k}{(1-\gamma_k)(1-\frac{\gamma_k}{2^{\mu_k}}) \min_i\{x_k^i z_k^i\}} (1-\gamma_k)(1-\frac{\gamma_k}{2^{\mu_k}}) \\
&\geq 1 - \kappa \parallel \Delta w_k \parallel - (1-\gamma_k)(1-\frac{\gamma_k}{2^{\mu_k}}).
\end{aligned}
$$

Since the common dual step-size is defined as $\alpha_{zk} = \min\{1, \min_{1\le i\le n}\{\alpha_{zk}^i : \Delta z_k^i < 0\}\}$ the previous inequality establishes (57). •

The result of the following lemma is used in Theorem 4, in which Q-quadratic convergence of the algorithm is established. Recall that the next iterate is given by $w_{k+1} = w_k + \mathrm{A}_k \Delta w_k$, where the matrix A_k is defined as $\mathrm{A}_k = diag\{\alpha_{xk} I_n, \alpha_{yk} I_q, \alpha_{zk} I_n\}$.

Lemma 15 *Let the assumptions of the previous lemma hold. Then*

$$\| I - A_k \| \le n(1-\gamma_k) + (q+n)(1-\gamma_k)(1 - \frac{\gamma_k}{2^{\mu_k}}) + O(\| F(w_k) \|) + O(\mu_k). \quad (61)$$

Proof From (55) we have

$$\begin{aligned} 0 \le 1 - \alpha_{xk} &\le 1 - \gamma_k + \kappa \| \Delta w_k \| \\ &\le 1 - \gamma_k + \kappa \| \nabla F(w_k)^{-1} \| (\| F(w_k) \| + \mu_k \| \hat{e} \|). \end{aligned}$$

while from (57) we have

$$\begin{aligned} 0 \le 1 - \alpha_{zk} &\le (1-\gamma_k)(1 - \frac{\gamma_k}{2^{\mu_k}}) + \kappa \| \Delta w_k \| \\ &\le (1-\gamma_k)(1 - \frac{\gamma_k}{2^{\mu_k}}) + \kappa \| \nabla F(w_k)^{-1} \| (\| F(w_k) \| + \mu_k \| \hat{e} \|). \end{aligned}$$

Assuming that $\alpha_{yk} = \alpha_{zk}$ and using the following inequality

$$\frac{1}{\sqrt{n}} \| \Pi \|_F \le \| \Pi \| \le \| \Pi \|_F,$$

which relates the l_2 matrix norm of any $n \times n$ matrix Π to the Frobenius one, we have

$$\begin{aligned} \| I - \mathrm{A}_k \| &= \| diag\{(1-\alpha_{xk}) I_n, (1-\alpha_{yk}) I_q, (1-\alpha_{zk}) I_n\} \| \\ &\le \| diag\{(1-\alpha_{xk}) I_n, (1-\alpha_{yk}) I_q, (1-\alpha_{zk}) I_n\} \|_F \\ &\le n(1-\gamma_k) + (q+n)(1-\gamma_k)(1 - \frac{\gamma_k}{2^{\mu_k}}) + O(\| F(w_k) \|) + O(\mu_k). \end{aligned}$$

•

The next theorem shows that Algorithm 1 converges to the optimum solution Q-quadratically. We define by $\mathcal{N}(\bar{v}, r)$ the open neighbourhood of radius r around $\bar{v}$, namely $\mathcal{N}(\bar{v}, r) = \{v \in \Re^n : \| v - \bar{v} \| < r\}$.

Theorem 4 *Assume that the sequence $\{w_k\}$ generated by Algorithm 1 converges to a solution w_* and suppose that assumptions (A1)-(A4) hold at that solution. Assume also that the parameters μ_k and γ_k are selected such that*

$$\mu_k = O(\| F(w_k) \|^2) \quad and \quad 1 - \gamma_k = O(\| F(w_k) \|). \quad (62)$$

Then there exists an $\epsilon > 0$ such that for all $w_0 \in \mathcal{N}(w_*, \epsilon)$ the sequence $\{w_k\}$ is well defined and converges to w_* Q-quadratically.

Proof We use induction to show that there always exists a positive constant ξ such that

$$\| w_{k+1} - w_* \| \leq \xi \| w_k - w_* \|^2 . \tag{63}$$

Since $w_0 \in \mathcal{N}(w_*, \epsilon)$ we have $\| w_0 - w_* \| < \epsilon$. For $\| w_k - w_* \| < \epsilon$, we also have,

$$\begin{aligned}
w_{k+1} - w_* &= w_k - w_* - \mathrm{A}_k \Delta w_k \\
&= w_k - w_* - \mathrm{A}_k \nabla F(w_k)^{-1} [F(w_k) - \mu_k \hat{e}] \\
&= \nabla F(w_k)^{-1} [\mathrm{A}_k F(w_*) - \mathrm{A}_k F(w_k) + \nabla F(w_k)(w_k - w_*)] \\
&\quad + \mu_k \mathrm{A}_k \nabla F(w_k)^{-1} \hat{e} \\
&= \nabla F(w_k)^{-1} [\mathrm{A}_k F(w_*) - \mathrm{A}_k F(w_k) - \mathrm{A}_k \nabla F(w_k)(w_k - w_*)] \\
&\quad + \nabla F(w_k)^{-1} [\mathrm{A}_k F(w_k)(w_k - w_*) - F(w_k)(w_k - w_*)] \\
&\quad + \mu_k \mathrm{A}_k \nabla F(w_k)^{-1} \hat{e} \\
&= \mathrm{A}_k \nabla F(w_k)^{-1} [F(w_*) - F(w_k) - \nabla F(w_k)(w_k - w_*)] \\
&\quad + (\mathrm{A}_k - I)(w_k - w_*) + \mu_k \mathrm{A}_k \nabla F(w_k)^{-1} \hat{e}.
\end{aligned}$$

Taking norms and applying the results of Lemma 15 we obtain

$$\begin{aligned}
\| w_{k+1} - w_* \| &\leq \| \mathrm{A}_k \nabla F(w_k)^{-1} \| \| F(w_*) - F(w_k) - \nabla F(w_k)(w_k - w_*) \| \\
&\quad + \| \mathrm{A}_k - I \| \| w_k - w_* \| + \mu_k \| \mathrm{A}_k \nabla F(w_k)^{-1} \hat{e} \| \\
&\leq O(\| w_k - w_* \|^2) + \| I - \mathrm{A}_k \| \| w_k - w_* \| + O(\mu_k) \\
&\leq O(\| w_k - w_* \|^2) \\
&\quad + [\, n(1 - \gamma_k) + (n + m)(1 - \gamma_k)(1 - \tfrac{\gamma_k}{2^{\mu_k}}) + O(\| F(w_k) \|) \\
&\quad + O(\mu_k) \,] \| w_k - w_* \| + O(\mu_k).
\end{aligned} \tag{64}$$

From Assumption (A1) we have that $F(w)$ is Lipschitz continuous. Hence, there exists a constant $\zeta > 0$ such that

$$\| F(w_k) \| = \| F(w_k) - F(w_*) \| \leq \zeta \| w_k - w_* \|, \tag{65}$$

for all $w_k \in \mathcal{N}(w_*, \epsilon)$. Choosing the parameters μ_k and γ_k as in (62) and considering (65), inequality (64) guarantees that there exists a positive constant ξ such that (63) is satisfied. Also by the induction hypothesis we have $\| w_k - w_* \| \leq \epsilon$ and therefore, for any sufficiently small ϵ we have $\| w_{k+1} - w_* \| \leq \epsilon$. Hence the sequence $\{w_k\}$ converges Q-quadratically to w_*. •

References

[1] Potra F. A. An infeasible interior-point predictor-corrector algorithm for linear programming. Technical Report 26, Department of Mathematics, University of Iowa, Iowa City, IA, 1992.

[2] Rustem B. Equality and inequality constrained optimization algorithms with convergent stepsizes. *JOTA*, 76(3):429–453, 1993.

[3] Bertsekas D. *Constrained Optimization and Lagrange Multiplier Methods.* Academic Press, 1982.

[4] Bertsekas D. *Nonlinear Programming.* Athena Scientific Publishing, 1995.

[5] den Hertog D. *Interior point approach to Linear, Quadratic and Convex Programming.* Kluwer Academic Publishers, 1994.

[6] Golub G. H. and Van Loan C. F. *Matrix Computations.* The Johns Hopkins University Press, Baltimore, MD, 1989.

[7] Yamashita H. A globally convergent primal-dual interior point method for constrained optimization. Technical report, Mathematical Systems Institute Inc, 2-5-3 Shinjuku, Shinjuku-ku, Tokyo, Japan, May 1995.

[8] Yamashita H. and Yabe H. Superlinear and quadratic convergence of some primal-dual interior point methods for constrained optimization. *Mathematical Programming*, 75:377–397, 1996.

[9] Vanderbei R. J. and Shanno D. F. An interior point algorithm for nonconvex nonlinear programming. Technical Report SOR-97-21, CEOR, Princeton University, Princeton, NJ, 1997.

[10] Vanderbei R. J. and Carpenter T. J. Symetric indefinite systems for interior point methods. *Mathematical Programming*, (58):1–32, 1991.

[11] Wright S. J. An interior point algorithm for linear constrained optimization. *SIAM Journal on Optimization*, 2:450–473, 1992.

[12] Wright S. J. *Primal-Dual Interior-Point Methods.* SIAM, 1997.

[13] Vial J.-P. Computational experience with a primal-dual interior point method for smooth convex programming. *Optimization Methods and Software*, 3:285–310, 1994.

[14] Karmarkar N. K. A new polynomial time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[15] Plummer J. Lasdon L. S. and Gang Y. Primal-dual and primal interior point algorithms for general nonlinear programms. *ORSA J. on Computing*, 7(3):321–332, 1995.

[16] Marsten R. J. Lasting I. J. and Shanno D. F. Computational experience with a primal-dual interior point method for linear programming. *Linear Algebra and its Applications*, (152):191–222, 1991.

[17] Anstreicher K. M. and Vial J.-P. On the convergence of an infeasible primal-dual interior point method for convex programming. *Optimization Methods and Software*, 3:273–283, 1994.

[18] Monma C. L. McShane K. A. and Shanno D. F. An implementation of a primal-dual interior point method for linear programming. *ORSA Journal on Computing*, 1(2):70–83, 1989.

[19] R. Monteiro and I Adler. Interior path following primal-dual algorithms-part 2: Convex quadratic programming. *Mathematical Programming*, 44:43–66, 1989.

[20] McCormick G. P. and Falk J. E. The superlinear convergence of a nonlinear primal-dual algorithm. Technical Report GWU/OR/Serial T-550/91, Department of Operations Research, The Ceorge Washington University, Washington, DC, 1991.

[21] Gajulapalli R.S. and Lasdon L.S. Computational experience with a safeguarded barrier algorithm for sparse nonlinear programming. Technical report, University of Texas at Austin, Austin, TX, 1997.

[22] El Bakry A. S. Tapia R. A. Tsuchiya T. and Zhang Y. On the formulation and theory of the newton interior point method for nonlinear programming. *JOTA*, 89(3):507–541, 1996.

[23] Fiacco A. V. and McCormick G. P. *Nonlinear Programming: Sequential Unconstrained minimization Techniques*. John Wiley and Sons, New York, 1968.

[24] Y. Zhang and R. Tapia. A superlinearly convergent polynomial primal dual interior point algorithm for linear programming. *SIAM J. Optimization*, 3(1):118–133, February 1993.