Superlinear Convergence of Primal-Dual Interior Point Algorithms for Nonlinear Programming

Nicholas I.M. Gould   Dominique Orban   Annick Sartenaer   Philippe L. Toint

April 26, 2000

Abstract

The local convergence properties of a class of primal-dual interior point methods are analyzed. These methods are designed to minimize a nonlinear, nonconvex, objective function subject to linear equality constraints and general inequalities. They involve an inner iteration in which the log-barrier merit function is approximately minimized subject to satisfying the linear equality constraints, and an outer iteration that specifies both the decrease in the barrier parameter and the level of accuracy for the inner minimization. It is shown that, asymptotically, for each value of the barrier parameter, solving a single primal-dual linear system is enough to produce an iterate that already matches the barrier subproblem accuracy requirements. The asymptotic rate of convergence of the resulting algorithm is Q-superlinear and may be chosen arbitrarily close to quadratic. Furthermore, this rate applies componentwise. These results hold in particular for the method described in [2], and indicate that the details of its inner minimization are irrelevant in the asymptotics, except for its accuracy requirements.

Keywords : primal-dual interior point method, componentwise Q-superlinear convergence.

1 Introduction

In this paper, we aim to provide insight on the local behavior of a class of primal-dual interior point algorithms. The class of algorithms is intended to solve nonconvex nonlinear programs that involve linear equality and nonlinear inequality constraints using a barrier-type method. More specifically, we consider the following problem:

$$\text{NLP} \equiv \begin{cases} \min & f(x) \\ \text{s.t.} & Ax = b \\ & c(x) \geq 0, \end{cases} \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^p$ are assumed to be twice continuously differentiable, the matrix $A \in \mathbb{R}^{m \times n}$ $(m \leq n)$ has full rank and $b \in \mathbb{R}^m$. The method starts with a strictly feasible initial point, and, rather than solving NLP directly, instead approximately solves a sequence of barrier subproblems of the form

$$\text{BS}(\mu) \equiv \begin{cases} \min & \phi(x, \mu) \\ \text{s.t.} & Ax = b, \end{cases} \tag{1.2}$$

for a decreasing sequence $\{\mu_k\}$ of positive barrier parameters. Here the barrier function is

$$\phi(x,\mu) \stackrel{\text{def}}{=} f(x) - \mu \sum_{i=1}^{p} \log c_i(x), \tag{1.3}$$

where the functions $c_i(\cdot)$ are the components of the vector function $c(\cdot)$. Once an (approximate) solution x_{k+1} of BS(μ_k) is found, the parameter μ_k is updated and attention turns to the next barrier subproblem (see for instance [6, 16, 18] for a general survey and [19] for the linear case). Under reasonable conditions [2, 6, 16], it can be shown that the sequence $\{x_k\}$ converges to a stationary point x^* of NLP. A typical stopping criterion for the solution of BS(μ) is

$$\|P_{\mathcal{N}(A)}(\nabla_x \phi(x,\mu))\| \leq \vartheta(\mu), \tag{1.4}$$

where $P_{\mathcal{N}(A)}$ is the orthogonal projection onto the nullspace of A, $\|\cdot\|$ is some norm defined on $\mathbb{R}^n$ and the continuous function $\vartheta : \mathbb{R}_+ \to \mathbb{R}_+$ is a *forcing function*, that is $\vartheta(\mu) = 0$ if and only if $\mu = 0$.

The most intensive part of the solution procedure is in the approximate solution of successive barrier subproblems BS(μ) (the *inner minimizations*), whose difficulty depends on the chosen starting point. An obvious idea is to start the solution of BS(μ_{k+1}) from x_{k+1}. However, in the primal case, for both linear and nonlinear programming, it has been shown that the unit Newton step for a barrier subproblem is likely not to be accepted as a first step if the minimization process is started from x_{k+1}, even if x_{k+1} is close to a solution of NLP [17]. The determination of better initial points in interior methods has recently been examined by several authors, both for primal [3, 12] and primal-dual barrier methods [1, 22, 23], as well as for exterior penalty methods (see [10], whose results were the inspiration for [5]). In particular, Dussault [5] expands the primal central path about the current iterate. In this paper, we apply a similar analysis in the more general primal-dual framework. This framework offers the advantage of keeping the radius of the "sphere of convergence" of Newton's method away from zero (under nondegeneracy assumptions), whereas this radius is proportional to the barrier parameter in a purely primal scheme, as is shown in [14] for the case of linear programming and in [15] for nonlinear programming.

In this paper, we intend to determine conditions under which, asymptotically, a single Newton step is strictly feasible (in contrast with the purely primal case), and results in a point that satisfies suitable barrier subproblem termination rules, after every reduction of the barrier parameter (see for instance [1, 3, 4] for previous work on the subject). This is shown to imply a componentwise Q-superlinear rate of convergence, a stronger result than simply Q-superlinear convergence of the vector of variables and Lagrange multipliers. Furthermore, this rate of convergence may be made arbitrarily close to quadratic. The results we present hold independently of the particular algorithm used for the inner minimization. They may thus be interpreted as giving conditions on the inner minimization stopping criterion to ensure fast convergence, in a manner similar to that studied by [13] for linear complementarity problems.

The motivation for the results presented in this paper is that they cover the general algorithm of [2], as will be discussed below. This algorithm has been implemented as `HSL_VE12` in the Harwell Subroutine Library for the special case of quadratic programming problems. We refer the reader to [2] for further motivation and details, along with the results of tests performed on a number of large convex and nonconvex quadratic examples.

The paper is organized as follows. Section 2 describes the notation and assumptions used throughout the paper and §3 provides useful preliminary results. In §4, we state the class of algorithms that will be analyzed; in §5 we describe an extrapolation of the central path that provides a point which turns out to be a very good estimate of the solution of the next subproblem. We subsequently present the local convergence properties of the algorithm when using the aforementioned extrapolation in §6. In §7, we review the link between our class of algorithms and the method of [2]; we also briefly discuss the connections with other proposals. We conclude and give some comments in §8.

2 Notation and assumptions

In this section, we present our notation and outline the required assumptions for the algorithm to converge superlinearly.

2.1 Notation

The following notation will be used throughout the paper. For related positive quantities α and β, we write $\alpha = \mathcal{O}(\beta)$ if there is a constant $\kappa > 0$ such that $\alpha \le \kappa\beta$ for all β sufficiently small. We write $\alpha = o(\beta)$ if $\alpha/\beta \to 0$ as $\beta \to 0$. We also write $\alpha = \Omega(\beta)$ if $\beta = \mathcal{O}(\alpha)$, and write $\alpha = \Theta(\beta)$ if $\alpha = \mathcal{O}(\beta)$ and $\beta = \mathcal{O}(\alpha)$.

If x is any vector in $\mathbb{R}^n$, the corresponding capital letter X will denote the diagonal matrix $\mathrm{diag}(x)$. The components of x will be denoted by $[x]_1, \dots, [x]_n$. We shall sometimes define a vector $w \in \mathbb{R}^{n+m+p}$ from $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ and $z \in \mathbb{R}^p$ by $w = (x, y, z)$. For such a vector, we shall use the notation $[w]_x$, $[w]_y$ and $[w]_z$ to refer respectively to its x, y and z components.

In the remainder of the paper, the statement "μ small enough" is to be understood as "μ is positive and close enough to zero", and the notation $a \searrow b$ as "a decreases monotonically and converges to b". By "global convergence" we shall mean "convergence to a *local* solution, whatever the point from which the process was started". By "local convergence" we mean "convergence to a local solution when the process is started in the vicinity of that solution".

2.1.1 Optimality conditions

The Lagrangian function for NLP is

$$\mathcal{L}(w) = \mathcal{L}(x, y, z) = f(x) + (Ax - b)^T y - c^T(x) z \tag{2.1}$$

where the Lagrange multipliers $y \in \mathbb{R}^m$ correspond to the equality constraints and $z \in \mathbb{R}^p_+$ to inequalities. We shall conveniently express the optimality conditions and the local analysis developments of §5 and §6 in terms of the following family of functions, parameterized by a scalar $\mu \ge 0$:

$$\Psi(w; \mu) \stackrel{\text{def}}{=} \begin{bmatrix} \nabla_x \mathcal{L}(w) \\ Ax - b \\ C(x) z - \mu e \end{bmatrix}. \tag{2.2}$$

Here $\nabla_x \mathcal{L}(w) = \nabla_x f(x) + A^T y - J^T(x) z$, the matrix $J(x)$ denotes the Jacobian matrix of c at x, that is the p by n matrix whose i-th row is $(\nabla_x c_i(x))^T$, and e is the vector of all ones. If $w^* \stackrel{\text{def}}{=} (x^*, y^*, z^*)$ is a first-order critical point for NLP, it must satisfy the first-order Karush-Kuhn-Tucker (KKT) conditions, which are that

$$\Psi(w^*; 0) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{2.3}$$

and

$$(c(x^*), z^*) \ge 0. \tag{2.4}$$

When solving problem BS(μ), we seek $x(\mu)$ and $y(\mu)$ such that

$$\begin{aligned}
\nabla_x f(x(\mu)) + A^T y(\mu) - \mu J^T(x(\mu)) C^{-1}(x(\mu)) e &= 0, && (2.5)\\
A x(\mu) &= b, && (2.6)\\
c(x(\mu)) &> 0. && (2.7)
\end{aligned}$$

Solving this system corresponds to using a primal approach. Let $x(\mu)$ and $y(\mu)$ be solutions of (2.5)–(2.7). In this case, the set

$$\{x(\mu) \mid \mu > 0\} \tag{2.8}$$

is said to define a (local) primal central path. Under second-order sufficiency assumptions, a linear independence constraint qualification and a strict complementary slackness condition, the primal central path leads to a solution x^* of NLP and $y(\mu)$ converges to a corresponding vector of Lagrange multipliers y^* (see [6]) as μ decreases to zero. Crucially, (2.5)–(2.7) is equivalent to the system

$$\Psi(w(\mu);\mu) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \qquad (c(x(\mu)), z(\mu)) > 0, \tag{2.9}$$

in the sense that if $(x(\mu), y(\mu))$ solves (2.5)–(2.7), the vector $(x(\mu), y(\mu), z(\mu))$ with $z(\mu) = \mu C^{-1}(x(\mu))e$ solves (2.9), while if $(x(\mu), y(\mu), z(\mu))$ solves (2.9), $(x(\mu), y(\mu))$ solves (2.5)–(2.7) and we have $z(\mu) = \mu C^{-1}(x(\mu))e$. Treating $z(\mu)$ as an independent variable when iteratively solving (2.9) is a primal-dual approach, and is the one we adopt in this paper. As $\mu \searrow 0$, the solution $w(\mu)$ of (2.9) then converges to a solution w^* of (2.3) under the aforementioned conditions.

In accordance with the primal-dual theory for linear and convex programming, we use here the following terminology. The term *primal variables* refers to the x variables, *Lagrange multipliers* to y and *dual variables* to z, although the variables z are Lagrange multipliers too. When solving a primal-dual system, the set

$$\mathcal{C} \stackrel{\text{def}}{=} \{w(\mu) = (x(\mu), y(\mu), z(\mu)) \mid \mu > 0\} \tag{2.10}$$

is said to define a (local) primal-dual central path.

Note that for any $\mu \geq 0$, $\Psi(w;\mu)$ and $\Psi(w;0)$ satisfy the fundamental relationship

$$\Psi(w;\mu) = \Psi(w;0) - \begin{bmatrix} 0 \\ 0 \\ \mu e \end{bmatrix}.$$

This implies that the Jacobian matrices $\nabla_w \Psi(w;\mu)$ and $\nabla_w \Psi(w;0)$ are equal for every w in the domain of interest, and satisfy

$$\nabla_w \Psi(w;\mu) = \nabla_w \Psi(w;0) = \begin{bmatrix} \nabla_{xx}\mathcal{L}(w) & A^T & -J^T(x) \\ A & 0 & 0 \\ ZJ(x) & 0 & C(x) \end{bmatrix}. \tag{2.11}$$

Moreover, $\nabla_\mu \Psi(w;\mu) = [0 \; 0 \; -e^T]^T$.

2.1.2 Norms

We use the symbol $\|\cdot\|$ to represent the Euclidean ℓ_2-norm, unless otherwise specified. We thus have

$$\|X\| = \|x\|_\infty \leq \|x\| \tag{2.12}$$

for any vector x. If S is a symmetric positive definite matrix, the S-norm of x, $\|x\|_S$, is defined as usual by $\|x\|_S^2 \stackrel{\text{def}}{=} x^T S x$.

We let the columns of the n by $n-m$ matrix N be an orthonormal basis for the nullspace of A. We denote the smallest and largest eigenvalues of any $n \times n$ symmetric matrix M by $\lambda^{\min}[M]$ and $\lambda^{\max}[M]$. Such a matrix is said to be *second-order sufficient* (with respect to A) if and only if the reduced matrix

$$R[M] = N^T M N$$

is positive definite (see, for instance, [9]).

In the context of our algorithm, we shall choose to measure gradients and related quantities in a seminorm induced by a second-order sufficient iteration-dependent preconditioner M_k, where k is the index of the current iteration[1]. We define the *k-seminorm* of a vector g, $\|g\|_{[k]}$, by

$$\|g\|_{[k]}^2 \stackrel{\text{def}}{=} q^T g, \tag{2.13}$$

where q solves the system

$$\begin{bmatrix} M_k & A^T \\ A & 0 \end{bmatrix} \begin{pmatrix} q \\ r \end{pmatrix} = \begin{pmatrix} g \\ 0 \end{pmatrix}.$$

This is actually a norm if g lies in the nullspace of A. In particular,

$$\|g\|_{[k]} = 0 \quad \text{if and only if} \quad \|N^T g\| = 0.$$

A simple calculation (see, for example, [8], Section 5.4.1) reveals that (2.13) may be expressed as

$$\|g\|_{[k]}^2 = g^T N R^{-1}[M_k] N^T g = \|N^T g\|_{R^{-1}[M_k]}^2 = \|R^{-\frac{1}{2}}[M_k] N^T g\|^2. \tag{2.14}$$

The simplest choice $M_k = I$, for which $R[I] = I$, simply measures the size of the projection of g into the nullspace of A. Note that the k-seminorm is invariant for displacements in the range space of A^T, *i.e.*:

$$\|g + A^T \overline{g}\|_{[k]} = \|g\|_{[k]} \tag{2.15}$$

for any $g \in \mathbb{R}^n$ and any $\overline{g} \in \mathbb{R}^m$.

In addition, because gradients can be interpreted as linear forms on the space of the problem variables, it is natural to measure quantities directly involving these variables, such as the distance between iterates, in a seminorm corresponding to the dual of $\|\cdot\|_{[k]}$. It is easy to verify that such a seminorm is given by

$$\|s\|_k \stackrel{\text{def}}{=} \|N^T s\|_{R[M_k]}, \tag{2.16}$$

and is, in fact, a norm in the nullspace of A. As a consequence, for all $v, s \in \mathbb{R}^n$ such that $As = 0$, *i.e.* such that $s = NN^T s$, we have that

$$|v^T s| = |v^T N (N^T M_k N)^{-\frac{1}{2}} (N^T M_k N)^{\frac{1}{2}} N^T s| \leq \|v\|_{[k]} \|s\|_k, \tag{2.17}$$

because of the Cauchy-Schwarz inequality. We stress that there is no need for M_k itself to be positive definite, merely that $N^T M_k N$ has to be.

[1] Strictly, this seminorm also depends on A, but we hide this dependence since A is fixed throughout this paper.

If U is any symmetric matrix, we also define the reduced matrix

$$R[U, M_k] \stackrel{\text{def}}{=} (N^T M_k N)^{-\frac{1}{2}} N^T U N (N^T M_k N)^{-\frac{1}{2}}, \tag{2.18}$$

we denote its smallest and largest eigenvalues by $\lambda_{M_k}^{\min}[U] = \lambda^{\min}[R[U, M_k]]$ and $\lambda_{M_k}^{\max}[U] = \lambda^{\max}[R[U, M_k]]$. We also note that the inertia of $R[U, M_k]$ and $R[U, I] \equiv N^T U N$ are the same. In particular, we have that

$$\lambda_{M_k}^{\min}[U] \geq 0 \quad \text{is equivalent to} \quad \lambda_I^{\min}[U] \geq 0. \tag{2.19}$$

We write $\|v\|_\diamond \stackrel{\text{def}}{=} \|N^T v\| = \|N N^T v\|$, the Euclidean norm of the projection of v onto the nullspace of A, and observe that $\|\cdot\|_\diamond$ is a self-dual norm in this nullspace, and that the k-seminorm and $\|\cdot\|_\diamond$ are equivalent if $M_k = I$. Moreover, we have

$$\|g + A^T \overline{g}\|_\diamond = \|g\|_\diamond \tag{2.20}$$

for any $g \in \mathbb{R}^n$ and any $\overline{g} \in \mathbb{R}^m$, which parallels (2.15).

We also notice the following equivalence

$$P_{\mathcal{N}(A)}(\nabla_x \phi(x, \mu)) = 0 \Longleftrightarrow N^T \nabla_x \phi(x, \mu) = 0, \tag{2.21}$$

which in turn is equivalent to

$$\|\nabla_x \phi(x, \mu)\|_\diamond = 0 \quad \text{and to} \quad \|\nabla_x \phi(x, \mu)\|_{[k]} = 0, \tag{2.22}$$

for any second-order sufficient matrix M_k. For future reference, we state the expressions of the first and second derivatives of the barrier function $\phi(x, \mu)$ with respect to x:

$$\nabla_x \phi(x, \mu) = \nabla_x f(x) - \sum_{i=1}^{p} \frac{\mu}{c_i(x)} \nabla_x c_i(x), \tag{2.23}$$

$$\nabla_{xx} \phi(x, \mu) = \nabla_{xx} f(x) + \sum_{i=1}^{p} \frac{\mu}{c_i^2(x)} \nabla_x c_i(x) (\nabla_x c_i(x))^T - \sum_{i=1}^{p} \frac{\mu}{c_i(x)} \nabla_{xx} c_i(x). \tag{2.24}$$

2.2 Assumptions

Let $\mathcal{I} = \{x \mid c(x) \geq 0\}$ be the set of points satisfying the inequalities, $\mathcal{E} = \{x \mid Ax = b\}$ be the set of points satisfying the equality constraints, and the intersection $\mathcal{F} \stackrel{\text{def}}{=} \mathcal{I} \cap \mathcal{E}$ be the set of feasible points for NLP. We assume that

AS1 there exists x_0 such that $Ax_0 = b$ and $c(x_0) > 0$,

AS2 the functions $f(\cdot)$ and $c_i(\cdot)$ are twice continuously differentiable over an open set containing $\mathcal{F}$.

Furthermore, if w^* is a solution of (2.3)–(2.4), if we let $\mathcal{A} \stackrel{\text{def}}{=} \{i \mid c_i(x^*) = 0\}$ be the set of active inequality constraints at x^* and if $a^{(j)}$ denotes the j-th column of A^T, we assume that

AS3 the vectors $\{\nabla_x c_i(x^*)\}_{i \in \mathcal{A}}$ and $\{a^{(j)}\}_{j=1}^m$ form a linearly independent set in $\mathbb{R}^n$,

AS4 the strong second-order sufficiency condition is satisfied at w^*: $v^T \nabla_{xx} \mathcal{L}(w^*) v > 0$ for any vector $v \neq 0, v \in \mathcal{N}(A)$ such that $\nabla_x c_i(x^*)^T v = 0$ for every $i \in \mathcal{A}$,

AS5 strict complementary slackness holds, that is $[z^*]_i + c_i(x^*) > 0$ for all $i = 1, \ldots, p$.

Note that **AS3** implies that the Lagrange multipliers y^* and the dual variables z^* are unique, and that the matrix A has full rank, which is not restrictive since it can always be satisfied by preprocessing the linear system $Ax = b$. Under **AS3**, **AS4** and **AS5**, x^* is an isolated (and thus strict) local solution of NLP. Throughout the paper, the dependence of $\mathcal{A}$ on x^* will remain hidden as only one local solution of NLP is considered.

3 Preliminary results

In this section, we state some results about central paths which will be useful later on.

It is shown in [6, 16] that under **AS3**, **AS4** and **AS5**, $\nabla_w \Psi(w^*;0)$ is nonsingular, and a continuity argument yields that it remains nonsingular in a small neighborhood of w^*. In the following technical lemma, which is a simple extension of that proved in [20] to the case of linear equality constraints, we now verify that the central path is well-defined in the intersection of this neighborhood and $\mathcal{E}$, and show that it has useful continuity properties.

Lemma 1 *Under* **AS2**–**AS5**, *let the vector*

$$w(l,r,\zeta) = (x(l,r,\zeta), y(l,r,\zeta), z(l,r,\zeta))$$

be defined implicitly as the solution of the following nonlinear system

$$\Psi(w;0) = \begin{bmatrix} NN^T l \\ r \\ \zeta \end{bmatrix}, \tag{3.1}$$

for given $l \in I\!R^n, r \in I\!R^m$ and $\zeta \in I\!R^p$, and with Ψ defined as in (2.2). Then there exist constants $\varepsilon > 0$ and $\kappa > 0$ for which the following statements hold.

(i) $w(l,r,\zeta)$ is a continuously differentiable function of (l,r,ζ) in the neighborhood

$$\mathcal{N}(\varepsilon) \stackrel{\rm def}{=} \{(l,r,\zeta) \mid \|l\|_\diamond + \|r\| + \|\zeta\| \leq \varepsilon\}.$$

(ii) For $\zeta > 0$ and $(l,r,\zeta) \in \mathcal{N}(\varepsilon)$, we have $[z(l,r,\zeta)]_i > 0$ and $c_i(x(l,r,\zeta)) > 0$ for all $i = 1,\ldots,p$.

(iii) Let $(l_1,r_1,\zeta_1), (l_2,r_2,\zeta_2) \in \mathcal{N}(\varepsilon)$ and w_1 and w_2 be the corresponding solutions of (3.1). We then have

$$w_2 - w_1 = (\nabla_w \Psi(w_1;0))^{-1} \begin{bmatrix} NN^T(l_1 - l_2) \\ r_1 - r_2 \\ \zeta_1 - \zeta_2 \end{bmatrix} + \rho \tag{3.2}$$

where $\rho \in I\!R^{n+m+p}$ and $\|\rho\| \leq \kappa\left(\|l_1 - l_2\|_\diamond + \|r_1 - r_2\| + \|\zeta_1 - \zeta_2\|\right)^2$, i.e. $\rho = \mathcal{O}(\varepsilon^2)$.

Proof. To prove *(i)*, we note that since $\nabla_w \Psi(w^*;0)$ is nonsingular, $\Psi(w^*;0) = 0$ and $\Psi(\cdot;0)$ is continuously differentiable in a neighborhood of w^*, the implicit function theorem implies that there exists $\varepsilon > 0$ such that the implicitly-defined function $w(l,r,\zeta)$ is continuously differentiable in (l,r,ζ) over the neighborhood $\mathcal{N}(\varepsilon)$.

For *(ii)*, let $(l,r,\zeta) \in \mathcal{N}(\varepsilon)$ with $\zeta > 0$. Taking a smaller ε if necessary, note that $c_i(x(l,r,\zeta)) > 0$ for all $i \notin \mathcal{A}$ and that **AS5** yields $[z(l,r,\zeta)]_i > 0$ for all $i \in \mathcal{A}$. On the other hand, since $[z(l,r,\zeta)]_i c_i(x(l,r,\zeta)) = \zeta_i$ for all $i = 1,\ldots,p$, we also obtain $[z(l,r,\zeta)]_i > 0$ for all $i \notin \mathcal{A}$ and $c_i(x(l,r,\zeta)) > 0$ for all $i \in \mathcal{A}$.

Finally, *(iii)* is shown noting that, by the implicit function theorem, the Jacobian of the function found in *(i)* is $-(\nabla_w \Psi(w;0))^{-1}$, from which the result follows, using a first-order Taylor expansion. □

We may now apply Lemma 1 to obtain a fundamental result relating the optimal w^* to $w(\mu)$. We first observe that, because of Lemma 1 *(i)*, when μ is sufficiently small, $w(0,0,\mu e)$ is unique and is therefore equal to $w(\mu)$, since $w(\mu)$ solves this system by definition. Moreover, because of **AS2**, $\nabla_w \Psi(\cdot; 0)$ is uniformly nonsingular in a neighborhood of w^*, and the left-hand side of (3.2) is dominated by the first term in its right-hand side for small enough ε. Consequently,

$$\|w_1 - w_2\| = \Theta\left(\|l_1 - l_2\|_\diamond + \|r_1 - r_2\| + \|\zeta_1 - \zeta_2\|\right). \tag{3.3}$$

Substituting $(l_1, r_1, \zeta_1) = (0, 0, \mu e)$ and $(l_2, r_2, \zeta_2) = (0, 0, 0)$ in (3.3) and using the equivalence of $w(0, 0, \mu e)$ and $w(\mu)$, we obtain the important result

$$\|w(\mu) - w^*\| = \Theta(\mu), \tag{3.4}$$

for all sufficiently small μ.

We now show a property of the behavior of the path $x(\mu)$ when it approaches its limit x^*. We know from (2.2) and (2.9) that

$$\Psi(w(\mu); \mu) = \begin{bmatrix} \nabla_x \mathcal{L}(w(\mu)) \\ Ax(\mu) - b \\ C(x(\mu))z(\mu) - \mu e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{3.5}$$

Differentiating system (3.5) with respect to μ and rearranging gives

$$\nabla_w \Psi(w(\mu); \mu) \begin{bmatrix} \dot{x}(\mu) \\ \dot{y}(\mu) \\ \dot{z}(\mu) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ e \end{bmatrix} \tag{3.6}$$

for any nonnegative μ. At $\mu = 0$, taking only active constraints into account, the third equation of (3.6) together with (2.11) shows that $\dot{x}(0) \neq 0$ and

$$(\nabla_x c_i(x^*))^T \dot{x}(0) = \frac{1}{[z^*]_i} \quad \forall i \in \mathcal{A}. \tag{3.7}$$

Equation (3.7) means that the trajectory $x(\mu)$ does not skirt the active constraints to reach x^*, that is, its approach is nontangential. Note that this is a consequence of the strict complementarity assumption **AS5** (see [21] for details).

Remark 3.1. Similarly to (3.7), if we write out the third equation of (3.6) for inactive constraints, we obtain

$$[\dot{z}(0)]_i = \frac{1}{c_i(x^*)} \quad \forall i \notin \mathcal{A}, \tag{3.8}$$

and this shows that the dual trajectory $z(\mu)$ does not approach its bounds tangentially. Identities (3.7) and (3.8) share a dual-like relationship which is, once again, due to strict complementarity. ■

4 The algorithm

We now state our class of algorithms. Let us first define, for every strictly feasible (x, z),

$$B(x, z) \stackrel{\text{def}}{=} J^T(x) C^{-1}(x) Z J(x), \tag{4.1}$$

the Lagrangian

$$L(x, z) \stackrel{\text{def}}{=} f(x) - c^T(x) z, \tag{4.2}$$

and finally

$$V(x,z) \stackrel{\text{def}}{=} \nabla_{xx} L(x,z) + B(x,z). \tag{4.3}$$

Note that for any strictly feasible (x,z) and for any $y \in \mathbb{R}^m$,

$$\nabla_{xx} L(x,z) = \nabla_{xx} \mathcal{L}(w) = \nabla_{xx} f(x) - \sum_{i=1}^{p} [z]_i \nabla_{xx} c_i(x). \tag{4.4}$$

Note also that $V(x,z)$ is the same as the Hessian of the barrier function (2.24) in the special case where the dual variables $z = \mu C^{-1}(x)e$.

To distinguish the overall algorithm from the inner minimization, that is the approximate solution of the barrier subproblem, we call the former the "outer minimization". Our outer minimization may be formally stated as Algorithm 4.1.

Algorithm 4.1: [Outer Minimization]

Initialization. An initial barrier parameter $\mu_0 > 0$ and the forcing functions $\epsilon^{\mathrm{C}}(\mu)$ and $\epsilon^{\mathrm{D}}(\mu)$ are given. Set $k = 0$.

Inner Minimization. Approximately minimize the log-barrier function $\phi(x, \mu_k)$. Stop this inner algorithm as soon as an inner iterate (x_{k+1}, z_{k+1}) is found such that

$$\begin{aligned} Ax_{k+1} &= b & (4.5)\\ (c(x_{k+1}), z_{k+1}) &> 0 & (4.6)\\ \|C(x_{k+1})z_{k+1} - \mu_k e\| &\le \epsilon^{\mathrm{C}}(\mu_k) \quad \text{and} & (4.7)\\ \|\nabla_x f(x_{k+1}) - J^T(x_{k+1})z_{k+1}\|_{[k+1]} &\le \epsilon^{\mathrm{D}}(\mu_k), & (4.8) \end{aligned}$$

where the norm $\|\cdot\|_{[k+1]}$ is defined with respect to some second-order sufficient preconditioning matrix M_{k+1}. Choose $\mu_{k+1} < \mu_k$, increment k by one, and perform next inner minimization.

A crucial feature of Algorithm 4.1 is that at every stage, it generates iterates lying in the constraint manifold $Ax = b$. This allows us to concentrate on the natural curvature of the problem. It also has the important consequence that the Lagrange multipliers y neither appear nor are used anywhere in the algorithm. However, for the needs of the local analysis of §6, let us define the Lagrange multipliers y_{k+1} by

$$\begin{aligned} A^T y_{k+1} &= NN^T \left[\nabla_x f(x_{k+1}) - J^T(x_{k+1})z_{k+1}\right] \\ &\quad - \left(\nabla_x f(x_{k+1}) - J^T(x_{k+1})z_{k+1}\right), \end{aligned} \tag{4.9}$$

with x_{k+1} and z_{k+1} as given by Algorithm 4.1. Note that the system (4.9) uniquely determines y_{k+1} because the matrix A has full rank. It is easy to check that under our assumptions, as $x_{k+1} \to x^*$ and $z_{k+1} \to z^*$, we have $y_{k+1} \to y^*$. Moreover, the definition (4.9) implies

$$\begin{aligned} &\nabla_x f(x_{k+1}) + A^T y_{k+1} - J^T(x_{k+1})z_{k+1} \\ &= NN^T \left[\nabla_x f(x_{k+1}) - J^T(x_{k+1})z_{k+1}\right] \\ &= NN^T \left[\nabla_x f(x_{k+1}) + A^T y_{k+1} - J^T(x_{k+1})z_{k+1}\right], \end{aligned} \tag{4.10}$$

or in other words, if we let $w_{k+1} = (x_{k+1}, y_{k+1}, z_{k+1})$, we have

$$\nabla_x \mathcal{L}(w_{k+1}) = NN^T \nabla_x \mathcal{L}(w_{k+1}). \tag{4.11}$$

Note that conditions (4.7)–(4.8) are relaxations of a part of the optimality system (2.9), and that the stopping condition (4.8) is equivalent to

$$\|\nabla_x \mathcal{L}(w_{k+1})\|_{[k+1]} \leq \epsilon^{\mathrm{D}}(\mu_k),$$

using the identity (2.15).

We do not describe an inner minimization algorithm here (we will return to this question in §7), but focus on the choice of the preconditioning matrices M_k. Since this preconditioning aims to locally represent the geometry of the log-barrier function, it is natural to assume that M_k is chosen as

$$M_k = W_k + B(x_k, z_k), \tag{4.12}$$

where the matrix W_k is chosen so that M_k is second-order sufficient, and might, for example, be a suitable approximation of the natural choice $\nabla_{xx} L(x_k, z_k)$. A discussion of the possible practical choices for W_k is out of the scope of this work and we refer the interested reader to [2] for theoretical arguments and results of a practical implementation. Our assumption on M_k is the following.

AS6 There exist $\epsilon_M \in (0,1)$ and $\kappa_{\mathrm{W}} > 0$ such that, for all k, the preconditioner $M_k = W_k + B(x_k, z_k)$ and its component W_k satisfy

$$\lambda^{\min}[N^T M_k N] \geq \epsilon_M \tag{4.13}$$

and

$$\|N^T W_k N\| \leq \kappa_{\mathrm{W}}. \tag{4.14}$$

AS6 allows us to analyze the interrelationship of the preconditioners, in that we can deduce an important relation between the norm $\|\cdot\|_\diamond$ and the seminorms (2.13)–(2.14). It is worth emphasising that this relation does not enforce uniform equivalence between those norms, as it only gives one of the two inequalities required for such an equivalence.

Lemma 2 *Suppose* **AS6** *is satisfied and that there exists a constant $\kappa_J > 0$ such that, for all k,*

$$\|J(x_k)\| \leq \kappa_J. \tag{4.15}$$

Then, for any vector $v \in I\!R^n$ and for all k,

$$\|v\|_k \geq \epsilon_M^{1/2} \|v\|_\diamond, \tag{4.16}$$

and

$$\|v\|_{[k]} \leq \epsilon_M^{-1/2} \|v\|_\diamond. \tag{4.17}$$

Moreover if there exists $\kappa^{(c)} > 0$ such that

$$\lim_{\mu \to 0} \frac{\epsilon^C(\mu)}{\mu} \leq \kappa^{(c)}, \tag{4.18}$$

then there exists $\kappa_\diamond > 0$ such that for all k,

$$\|v\|_{[k]} \geq \kappa_\diamond \min\left[\frac{\min_i c_i(x_k)}{\sqrt{\mu_{k-1}}}, 1\right] \|v\|_\diamond. \tag{4.19}$$

Proof. The following proof is inspired by [2, Lemma 4.2 and Theorem 4.12].

Inequalities (4.16) and (4.17) clearly hold for any v orthogonal to $\mathcal{N}(A)$. Using the identity (2.16), the positive definiteness of $N^T M_k N$ and **AS6**, we have for all v such that $N^T v \neq 0$,

$$\begin{aligned}\frac{\|v\|_\diamond^2}{\|v\|_k^2} = \frac{\|N^T v\|^2}{\|N^T v\|^2_{R[M_k]}} &= \frac{\|(N^T M_k N)^{-\frac{1}{2}}(N^T M_k N)^{\frac{1}{2}} N^T v\|^2}{\|(N^T M_k N)^{\frac{1}{2}} N^T v\|^2} \\ &\leq \|(N^T M_k N)^{-1}\| \\ &\leq \epsilon_M^{-1},\end{aligned}$$

which proves (4.16). The proof of (4.17) is similar.

To prove (4.19), first observe that **AS6**, (2.12) and (4.15) imply that for all k,

$$\|N^T M_k N\| \leq \|N^T W_k N\| + \|N^T B(x_k, z_k) N\| \leq \kappa_{\mathrm{W}} + \kappa_J^2 \max_i \frac{[z_k]_i}{c_i(x_k)}. \tag{4.20}$$

If r denotes the vector $N^T v$, we have $\|r\| = \|v\|_\diamond$. Consider first the case where there exists $\kappa_\infty > 0$ such that for all $i = 1, \ldots, p$,

$$\limsup_{k \to \infty} \frac{[z_k]_i}{c_i(x_k)} \leq \kappa_\infty < +\infty.$$

Inequality (4.20) then becomes

$$\|N^T M_k N\| \leq \kappa_{\mathrm{W}} + \kappa_J^2 \kappa_\infty,$$

and we have from the positive definiteness of $N^T M_k N$:

$$\|v\|_{[k]} = \|(N^T M_k N)^{-\frac{1}{2}} r\| \geq (\kappa_{\mathrm{W}} + \kappa_J^2 \kappa_\infty)^{-1/2} \|r\|. \tag{4.21}$$

Now consider the other possibility, namely that

$$\limsup_{k \to \infty} \frac{[z_k]_{i_0}}{c_{i_0}(x_k)} = +\infty,$$

for some index i_0. Then, using (4.7) and (4.18), we have for all i and for sufficiently large k,

$$\frac{[z_k]_i}{c_i(x_k)} \leq \frac{\mu_{k-1}}{c_i(x_k)^2} + \frac{|c_i(x_k)[z_k]_i - \mu_{k-1}|}{c_i(x_k)^2} \leq \frac{\mu_{k-1}}{c_i(x_k)^2} + \frac{\epsilon^{\mathrm{C}}(\mu_{k-1})}{c_i(x_k)^2} \leq (1 + 2\kappa^{(\mathrm{c})}) \frac{\mu_{k-1}}{c_i(x_k)^2}.$$

We thus obtain from (4.20) that for large enough k,

$$\|N^T M_k N\| \leq \kappa_{\mathrm{W}} + \kappa_J^2 \max_i \frac{[z_k]_i}{c_i(x_k)} \leq 2\kappa_J^2 (1 + 2\kappa^{(\mathrm{c})}) \frac{\mu_{k-1}}{\min_i c_i(x_k)^2},$$

so that

$$\|v\|_{[k]} = \|(N^T M_k N)^{-\frac{1}{2}} r\| \geq (2\kappa_J^2(1 + 2\kappa^{(\mathrm{c})}))^{-\frac{1}{2}} \frac{\min_i c_i(x_k)}{\sqrt{\mu_{k-1}}} \|r\|. \tag{4.22}$$

Putting (4.21) and (4.22) together yields (4.19) with

$$\kappa_\diamond \stackrel{\mathrm{def}}{=} \min\left[(2\kappa_J^2(1 + 2\kappa^{(\mathrm{c})}))^{-\frac{1}{2}}, (\kappa_{\mathrm{W}} + \kappa_J^2 \kappa_\infty)^{-1/2}\right].$$

□

5 Choosing the starting point for the inner minimization

As stated in the introduction, a computationally critical part of the algorithm is the choice of the starting point for the inner minimization and we already indicated that choosing x_{k+1} to start the solution of $BS(\mu_{k+1})$ is likely to be inefficient. The purpose of this section is to examine alternative choices, from the point of view of improving the local convergence rate. Of course, this rate of convergence depends on the particular choice of the functions $\epsilon^{\mathrm{C}}(\mu)$ and $\epsilon^{\mathrm{D}}(\mu)$. We therefore start by considering appropriate choices for these functions.

5.1 Stopping tolerances

Formally, we shall suppose that the inner iteration, that is the approximate minimization of $\mathrm{BS}(\mu_k)$, starts from the (as yet, undefined) primal-dual point $(x_{k,0}, z_{k,0})$, generates a sequence of iterates $\{(x_{k,j}, z_{k,j})\}_{j\geq 0}$, and terminates at the point $(x_{k,j_k}, z_{k,j_k}) \equiv (x_{k+1}, z_{k+1})$ at which (4.5)–(4.8) are satisfied for some appropriate second-order sufficient matrix M_{k+1}. Let us define the Lagrange multipliers y_{k+1} according to (4.9) and let $w_{k+1} = (x_{k+1}, y_{k+1}, z_{k+1})$. We assume from now on that the tolerances $\epsilon^{\mathrm{C}}(\mu_k)$ and $\epsilon^{\mathrm{D}}(\mu_k)$ asymptotically have the particular form given in Fig.1. Observe that (4.7) and (5.1) imply that $c_i(x_{k+1})[z_{k+1}]_i - \mu_k = \mathcal{O}(\mu_k)$ for all $i = 1, \ldots, p$.

Stopping Tolerances:

We assume that there exist constants $0 < \underline{\kappa}_\mu^{\mathrm{C}} \leq \overline{\kappa}_\mu^{\mathrm{C}} < 1$ and $0 < \underline{\kappa}_\mu^{\mathrm{D}} \leq \overline{\kappa}_\mu^{\mathrm{D}}$ such that, for sufficiently large values of k,

$$\underline{\kappa}_\mu^{\mathrm{C}} \mu_k \leq \epsilon^{\mathrm{C}}(\mu_k) \leq \overline{\kappa}_\mu^{\mathrm{C}} \mu_k \tag{5.1}$$

and

$$\underline{\kappa}_\mu^{\mathrm{D}} \mu_k^{\gamma_k+1} \leq \epsilon^{\mathrm{D}}(\mu_k) \leq \overline{\kappa}_\mu^{\mathrm{D}} \mu_k^{\gamma_k+1}, \tag{5.2}$$

where

$$0 < \gamma_k < 1. \tag{5.3}$$

Figure 1: Stopping tolerances for Algorithm 4.1

In the context of our local analysis, we now assume that the vector w^*, a solution of (2.3)–(2.4), is a limit point of the sequence $\{w_{k+1}\}$, satisfying (4.5)–(4.8), with y_{k+1} defined by (4.9), as $\mu_k \searrow 0$. More specifically, we assume that there exists an infinite index set $\mathcal{K}$ such that $w_{k+1} \to w^*$ as $k \to \infty$, $k \in \mathcal{K}$. The subsequence of $\{w_{k+1}\}$ indexed by $\mathcal{K}$ is denoted by $\{w_{k+1}\}_{\mathcal{K}}$ and we write $\{w_{k+1}\}_{\mathcal{K}} \to w^*$. In what follows, we consider only $k \in \mathcal{K}$. In addition, **AS2** implies that, for all $i = 1, \ldots, p$ and all x sufficiently close to x^*,

$$\|\nabla_x f(x)\| \leq \kappa_{\mathrm{g}}, \|\nabla_{xx} f(x)\| \leq \kappa_{\mathrm{H}}, \|\nabla_x c_i(x)\| \leq \kappa_\gamma \text{ and } \|\nabla_{xx} c_i(x)\| \leq \kappa_\Gamma. \tag{5.4}$$

for some $\kappa_{\mathrm{g}}, \kappa_{\mathrm{H}}, \kappa_\gamma, \kappa_\Gamma > 0$. The third of these bounds and the fact that we restrict our attention to $\mathcal{K}$ imply that (4.15) holds for k sufficiently large, and Lemma 2 can thus be applied asymptotically within $\mathcal{K}$.

The next theorem provides bounds on the active and inactive quantities involved in Algorithm 4.1 that result from our choice of stopping tolerances.

Theorem 3 *Assume w^* is a solution of NLP and $\{w_{k+1}\}_{\mathcal{K}} \to w^*$, where $\{w_{k+1}\}$ is a sequence of iterates generated by Algorithm 4.1 with y_{k+1} defined by (4.9). Under* **AS1**–**AS6** *and (5.1), we have that for sufficiently large $k \in \mathcal{K}$,* *(i) for all $i \in \mathcal{A}$, there exist $\kappa^z \geq \kappa_z > 0$ such that*

$$\frac{1}{\kappa^z}(1 - \overline{\kappa}_\mu^C)\mu_k \leq c_i(x_{k+1}) \leq \frac{1}{\kappa_z}(1 + \overline{\kappa}_\mu^C)\mu_k, \tag{5.5}$$

$$\kappa_z \leq [z_{k+1}]_i \leq \kappa^z, \tag{5.6}$$

(ii) for all $i \notin \mathcal{A}$, there exist $\kappa^c \geq \kappa_c > 0$ such that

$$\frac{1}{\kappa^c}(1 - \overline{\kappa}_\mu^C)\mu_k \leq [z_{k+1}]_i \leq \frac{1}{\kappa_c}(1 + \overline{\kappa}_\mu^C)\mu_k, \tag{5.7}$$

$$\kappa_c \leq c_i(x_{k+1}) \leq \kappa^c, \tag{5.8}$$

where $\overline{\kappa}_\mu^C$ is defined in (5.1).

Proof. The strict complementarity assumption **AS5** implies that for $i \in \mathcal{A}$, we have for sufficiently large $k \in \mathcal{K}$

$$0 < \tfrac{1}{2}[z^*]_i \leq [z_{k+1}]_i \leq 2[z^*]_i. \tag{5.9}$$

Relation (5.6) then follows with $\kappa_z = \frac{1}{2}\min_{i \in \mathcal{A}}[z^*]_i$ and $\kappa^z = 2\max_{i \in \mathcal{A}}[z^*]_i$.

On the other hand, stopping condition (4.7) yields that for all $i = 1, \ldots, p$,

$$-\epsilon^{\mathrm{C}}(\mu_k) \leq c_i(x_{k+1})[z_{k+1}]_i - \mu_k \leq \epsilon^{\mathrm{C}}(\mu_k). \tag{5.10}$$

Using the rightmost inequality in (5.1), (5.6) and (5.10) yields (5.5).

Observe that if $i \notin \mathcal{A}$, then for sufficiently large $k \in \mathcal{K}$,

$$0 < \tfrac{1}{2}c_i(x^*) \leq c_i(x_{k+1}) \leq 2c_i(x^*). \tag{5.11}$$

Relation (5.8) then follows with $\kappa_c = \frac{1}{2}\min_{i \notin \mathcal{A}} c_i(x^*)$ and $\kappa^z = 2\max_{i \notin \mathcal{A}} c_i(x^*)$. The proof of (5.7) is similar to that of (5.5) using (5.8) and (5.10). □

We now show that a termination criterion such as (4.7)–(4.8) coupled with (5.1) and (5.2) guarantees that w_{k+1} lies within a constant factor of $(\mu_k^{\gamma_k+\frac{1}{2}} + \mu_k)$ from an exact solution $w(\mu_k)$ of $\mathrm{BS}(\mu_k)$ and from w^* in the usual, Euclidean norm.

Theorem 4 *Suppose that* **AS1**–**AS6**, *(5.1) and (5.2) are satisfied. Assume furthermore that w^* is a solution of NLP and that $\{w_{k+1}\}_\mathcal{K} \to w^*$, where $\{w_{k+1}\}$ is a sequence of iterates generated by Algorithm 4.1 with y_{k+1} defined by (4.9). Then, we have that, for sufficiently large $k \in \mathcal{K}$, there exist constants $\kappa_{dst}, \kappa_{dst}^* > 0$ such that*

$$\|w_{k+1} - w(\mu_k)\| \leq \kappa_{dst}\,(\mu_k^{\gamma_k+\frac{1}{2}} + \mu_k), \tag{5.12}$$

and

$$\|w_{k+1} - w^*\| \leq \kappa_{dst}^*\,(\mu_k^{\gamma_k+\frac{1}{2}} + \mu_k). \tag{5.13}$$

Proof. Observe that, under the stated assumptions, we may apply (4.19) which, together with the relations (2.20), (4.8) and (5.2), yields

$$\begin{aligned} \|\nabla_x \mathcal{L}(w_{k+1})\|_\diamond &\leq \kappa_\diamond^{-1}\max\left[\frac{\epsilon^{\mathrm{D}}(\mu_k)\sqrt{\mu_k}}{\min_i c_i(x_{k+1})}, \epsilon^{\mathrm{D}}(\mu_k)\right] && (5.14)\\ &\leq \kappa_\diamond^{-1}\overline{\kappa}_\mu^{\mathrm{D}}\max\left[\frac{\mu_k^{\gamma_k+3/2}}{\min_i c_i(x_{k+1})}, \mu_k^{\gamma_k+1}\right]. && (5.15)\end{aligned}$$

First consider the case where the active set $\mathcal{A}$ is nonempty. In view of (5.5) and (5.8), the index i that realizes the minimum in (5.15) certainly asymptotically satisfies (5.5), which implies

$$\min_i c_i(x_{k+1}) = \min_{i\in\mathcal{A}} c_i(x_{k+1}) \geq \frac{1}{\kappa^z}(1-\overline{\kappa}_\mu^{\mathrm{C}})\mu_k. \tag{5.16}$$

Combining (5.15) and (5.16), we obtain

$$\|\nabla_x \mathcal{L}(w_{k+1})\|_\diamond \leq \kappa_{\mathcal{L}} \max\left[\frac{\kappa^z}{1-\overline{\kappa}_\mu^{\mathrm{C}}}\mu_k^{\gamma_k+\frac{1}{2}}, \mu_k^{\gamma_k+1}\right] = \kappa_{\mathcal{L}}\frac{\kappa^z}{1-\overline{\kappa}_\mu^{\mathrm{C}}}\mu_k^{\gamma_k+\frac{1}{2}}, \tag{5.17}$$

for sufficiently large $k\in\mathcal{K}$, where we have set $\kappa_{\mathcal{L}} = \kappa_\diamond^{-1}\overline{\kappa}_\mu^{\mathrm{D}}$.

Consider now the case where there are no active constraints. This time, the index i that realizes the minimum in (5.15) satisfies (5.8) and we have $\min_i c_i(x_{k+1}) \geq \kappa_c$. Thus, (5.15) gives that for sufficiently large $k\in\mathcal{K}$,

$$\|\nabla_x \mathcal{L}(w_{k+1})\|_\diamond \leq \kappa_\diamond^{-1}\overline{\kappa}_\mu^{\mathrm{D}} \max\left[\kappa_c^{-1}\mu_k^{\gamma_k+3/2}, \mu_k^{\gamma_k+1}\right] = \kappa_{\mathcal{L}}\mu_k^{\gamma_k+1}. \tag{5.18}$$

From the definition (4.9) of y_{k+1}, (4.11) together with (4.5) guarantees that we have $w_{k+1} = w(\nabla_x\mathcal{L}(w_{k+1}), 0, C(x_{k+1})z_{k+1})$ in Lemma 1. Moreover, (4.7), (5.1), (5.17) and (5.18) guarantee that $\|\nabla_x\mathcal{L}(w_{k+1})\|_\diamond + \|C(x_{k+1})z_{k+1}\|$ is smaller than the threshold ε defined in Lemma 1 for sufficiently large $k\in\mathcal{K}$. Invoking (3.3) with the parameters $(\nabla_x\mathcal{L}(w_{k+1}), 0, C(x_{k+1})z_{k+1})$ and $(0,0,\mu_k e)$ thus gives

$$\begin{aligned}
\|w_{k+1} - w(\mu_k)\| &= \Theta(\|\nabla_x\mathcal{L}(w_{k+1})\|_\diamond + \|C(x_{k+1})z_{k+1} - \mu_k e\|) &(5.19)\\
&= \mathcal{O}(\|\nabla_x\mathcal{L}(w_{k+1})\|_\diamond + \epsilon^{\mathrm{C}}(\mu_k)), &(5.20)\\
&= \mathcal{O}(\|\nabla_x\mathcal{L}(w_{k+1})\|_\diamond + \mu_k), &(5.21)
\end{aligned}$$

where we have used (4.7) and (5.1).

When $\mathcal{A}$ is nonempty, we obtain from (5.17) and (5.21) that, for sufficiently large $k\in\mathcal{K}$,

$$\|w_{k+1} - w(\mu_k)\| = \mathcal{O}(\mu_k^{\gamma_k+\frac{1}{2}} + \mu_k), \tag{5.22}$$

while if $\mathcal{A}$ is empty, (5.18) and (5.21) yield

$$\begin{aligned}
\|w_{k+1} - w(\mu_k)\| &= \mathcal{O}(\mu_k^{\gamma_k+1} + \mu_k) &(5.23)\\
&= \mathcal{O}(\mu_k^{\gamma_k+\frac{1}{2}} + \mu_k), &(5.24)
\end{aligned}$$

because of (5.3). Putting (5.22) and (5.24) together proves (5.12).

Using the triangle inequality, (3.4), (5.3) and (5.12), there exists a constant $\kappa^* > 0$ such that

$$\begin{aligned}
\|w_{k+1} - w^*\| &\leq \|w_{k+1} - w(\mu_k)\| + \|w(\mu_k) - w^*\| &(5.25)\\
&\leq \kappa_{\mathrm{dst}}(\mu_k^{\gamma_k+\frac{1}{2}} + \mu_k) + \kappa^*\mu_k, &(5.26)\\
&\leq (\kappa_{\mathrm{dst}} + \kappa^*)(\mu_k^{\gamma_k+\frac{1}{2}} + \mu_k), &(5.27)
\end{aligned}$$

which proves (5.13) with $\kappa_{\mathrm{dst}}^* = \kappa_{\mathrm{dst}} + \kappa^*$. □

Remark 5.1. In condition (5.14), the minimum is certainly asymptotically attained for an active index, if any. In that case, if **AS5** is satisfied, this minimum is of the order of μ_k (and of the order of $\mu_k^{1/2}$ if **AS5** fails to be satisfied)—see for instance [11, 21]. In our case, in order for the sequence

$\{\|\nabla_x \mathcal{L}(w_{k+1})\|_\diamond\}$ to converge to zero, we thus require that $\epsilon^{\mathrm{D}}(\mu_k)$ converges to zero faster than $\mu_k^{1/2}$ (as guaranteed by (5.2)), which is usually sufficient in practice to ensure convergence of the outer minimization. ■

Examining (4.19), we see that in the nondegenerate case, and when there are active constraints, we asymptotically have $\|v\|_{[k]} \geq \kappa_\diamond^{\mathrm{nd}}\sqrt{\mu_k}\|v\|_\diamond$ for some constant $\kappa_\diamond^{\mathrm{nd}} > 0$. Hence, conditions (4.8) and (5.2) amount to $\|\nabla_x \mathcal{L}(w_{k+1})\|_\diamond = \mathcal{O}(\mu_k^{\gamma_k+\frac{1}{2}})$ which may be weaker than the usual stopping criterion $\|\nabla_x \mathcal{L}(w_{k+1})\|_\diamond = \mathcal{O}(\mu_k)$ whenever $\gamma_k < \frac{1}{2}$. In this case, the right-hand sides of (5.12) and (5.13) are $\mathcal{O}(\mu_k^{\gamma_k+\frac{1}{2}})$, which is weaker than the usual bound $\mathcal{O}(\mu_k)$. If $\gamma_k > \frac{1}{2}$, the stopping criterion is tightened but the right-hand sides of (5.12) and (5.13) are required to take the traditional form $\mathcal{O}(\mu_k)$. Finally, if $\gamma_k = \frac{1}{2}$, the stopping criterion (4.8) and the bounds (5.12) and (5.13) coincide with the traditional ones.

In the degenerate case, (4.19) becomes $\|v\|_{[k]} \geq \kappa_\diamond^{\mathrm{d}}\|v\|_\diamond$ for some constant $\kappa_\diamond^{\mathrm{d}} > 0$, and shows, together with (4.17), that the norms $\|\cdot\|_{[k]}$ and $\|\cdot\|_\diamond$ are equivalent in the nullspace of A. Using (5.2), condition (4.8) then amounts to $\|\nabla_x \mathcal{L}(w_{k+1})\|_\diamond = \mathcal{O}(\mu_k^{\gamma_k+1})$ and shows that we are more restrictive in this case for any $\gamma_k > 0$.

5.2 Simple choices for the starting point

In the unpreconditioned primal case, it has previously been suggested [3] that the tolerance $\epsilon^{\mathrm{D}}(\mu)$ be set to $\mathcal{O}(\mu)$. In that case, inequality (4.8) implies that

$$\|\nabla_x \mathcal{L}(w_{k+1}^{\mathrm{P}})\| = \mathcal{O}(\mu_k) = \mathcal{O}\left(\frac{\mu_k}{\mu_{k+1}}\right) \tag{5.28}$$

for the primal choice $w_{k+1}^{\mathrm{P}} = (x_{k+1}, y_{k+1}, \mu_{k+1}C^{-1}(x_{k+1})e)$. This observation parallels [5, Lemma 1] and reinforces the results in [17] by suggesting that if the parameter μ_k is reduced too fast (*i.e.* $\mu_{k+1} \ll \mu_k$), it is unlikely that w_{k+1}^{P} will be an accurate estimate of the solution $w(\mu_{k+1})$ for the forthcoming outer iteration.

By contrast, [3] suggests that, letting $x_{k+1}^{\mathrm{O}} = x_{k+1}$ and $z_{k+1}^{\mathrm{O}} = \mu_k C^{-1}(x_{k+1})e$ (as opposed to the value $\mu_{k+1}C^{-1}(x_{k+1})e$ which would have been used in a purely primal context without extrapolation), a good initial point for $\mathrm{BS}(\mu_{k+1})$ might be $w_{k+1,0} = w_{k+1}^{\mathrm{O}} + d_{k+1}^{\mathrm{N}} = (x_{k+1,0}, y_{k+1,0}, z_{k+1,0})$, where d_{k+1}^{N} is the full Newton step taken from $w_{k+1}^{\mathrm{O}} = (x_{k+1}^{\mathrm{O}}, y_{k+1}^{\mathrm{O}}, z_{k+1}^{\mathrm{O}})$ (for some y_{k+1}^{O}). Simply restating a few vital steps from [3], it can be shown that d_{k+1}^{N} is asymptotically feasible and

$$\begin{aligned}
\|\nabla_x \mathcal{L}(w_{k+1,0})\| &= \mathcal{O}(\mu_k^2), \\
\|C(x_{k+1,0})z_{k+1,0} - \mu_{k+1}e\| &= \mathcal{O}(\mu_k^2),
\end{aligned}$$

which would be accepted by any primal-dual stopping rule for which $\epsilon^{\mathrm{C}}(\mu) = \epsilon^{\mathrm{D}}(\mu) = \mathcal{O}(\mu)$ provided that $\mu_{k+1} = \Omega(\mu_k^2)$. Continuing in this vein, we would subsequently have

$$z_{k+1,0} = \mu_{k+1}C^{-1}(x_{k+1,0})e + \mathcal{O}\left(\frac{\mu_k^2}{\mu_{k+1}}\right),$$

which then provides the bound

$$\left\|\nabla_x f(x_{k+1,0}) - \mu_{k+1}J^T(x_{k+1,0})C^{-1}(x_{k+1,0})e\right\| = \mathcal{O}\left(\frac{\mu_k^2}{\mu_{k+1}}\right).$$

Hence, if the inner minimization corresponding to μ_{k+1} is started with $x_{k+1,0} = x^{\mathrm{O}}_{k+1} + [d^{\mathrm{N}}_{k+1}]_x$ and $z_{k+1,0} = z^{\mathrm{O}}_{k+1} + [d^{\mathrm{N}}_{k+1}]_z$, and assuming a subsequent Newton step is acceptable to the inner minimization method (this can be shown to be the case), the size of $\nabla_x \mathcal{L}(w)$ at the resulting iterate will be $\mathcal{O}(\mu_k^4/\mu_{k+1}^2)$. Consequently, if we wish the resulting iterate to satisfy a primal stopping rule of the form $\epsilon^{\mathrm{D}}(\mu) = \mathcal{O}(\mu)$, this requires that

$$\mu_{k+1} = \Omega(\mu_k^{4/3}), \tag{5.29}$$

which suggests a superlinear rate of convergence in μ is possible.

5.3 An alternative choice based on extrapolating the central path

We now intend to parallel the approach of Section 5.2 in the primal-dual case, with the hope of improving the bound (5.28). Assume we update the barrier parameter from μ_k to μ_{k+1}. In order to solve problem BS(μ_{k+1}) efficiently, it is natural to aim to choose a starting point which is as close as possible to a stationary point $w(\mu_{k+1})$ of this problem. We thus wish to (approximately) solve the system

$$\Psi(w;\mu_{k+1}) = 0. \tag{5.30}$$

An attractive possibility is therefore to choose the starting point for the inner minimization as the result of a single Newton iteration for this system. This point, which we denote w^{PD}_{k+1}, is obtained from the solution of the linearized version of (5.30), that is

$$\nabla_w \Psi(w_{k+1};\mu_{k+1})(w^{\mathrm{PD}}_{k+1} - w_{k+1}) = -\begin{bmatrix} \nabla_x \mathcal{L}(w_{k+1}) \\ 0 \\ C(x_{k+1})z_{k+1} - \mu_{k+1} e \end{bmatrix}, \tag{5.31}$$

where $\nabla_w \Psi(w;\mu)$ is given by (2.11). If we let

$$d_{k+1} = ([d_{k+1}]_x, [d_{k+1}]_y, [d_{k+1}]_z) = w^{\mathrm{PD}}_{k+1} - w_{k+1}, \tag{5.32}$$

we may eliminate $[d_{k+1}]_z$, use the first identity of (4.4) together with (4.3), the fact that $\nabla_x \mathcal{L}(w_{k+1}) = \nabla_x f(x_{k+1}) + A^T y_{k+1} - J^T(x_{k+1}) z_{k+1}$, rearrange and obtain the *reduced* system

$$\begin{bmatrix} V(x_{k+1}, z_{k+1}) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} [d_{k+1}]_x \\ [w^{\mathrm{PD}}_{k+1}]_y \end{bmatrix} = -\begin{bmatrix} \nabla_x \phi(x_{k+1}, \mu_{k+1}) \\ 0 \end{bmatrix} \tag{5.33}$$

from which we may recover

$$[d_{k+1}]_z = -z_{k+1} + \mu_{k+1} C^{-1}(x_{k+1}) e - C^{-1}(x_{k+1}) Z_{k+1} J(x_{k+1}) [d_{k+1}]_x.$$

Note that the right-hand side of (5.33) is independent of z_{k+1}. This system entirely determines $[d_{k+1}]_x$ and $[w^{\mathrm{PD}}_{k+1}]_y$ provided that the matrix $V(x_{k+1}, z_{k+1})$ is positive definite on the nullspace of A and A has full rank. This is equivalent to requiring that the matrix $V(x_{k+1}, z_{k+1})$ is second-order sufficient (see §2.1.2).

Before studying the implications of this step in terms of local convergence rate, we first give an alternative geometrical interpretation of the step d_{k+1}, in the spirit of [5], arising from the implicit function theorem.

From the point w_{k+1} and with the barrier parameter μ_k, assume we wish to find an estimate w^{EX}_{k+1} to an exact solution w_+ of

$$\Psi(w;\mu_{k+1}) = \Psi(w_{k+1};\mu_k), \tag{5.34}$$

since we know from (4.7) and (4.8) that the right-hand side of (5.34) is "small". Notice that if we compute w_+ as the solution of (5.34), the equality constraints remain satisfied.

Let us consider the following system

$$\Psi(w;\mu) = \Psi(w_{k+1};\mu_k) \tag{5.35}$$

and think of its solution in the variable w as a function of the barrier parameter μ, say $w = \varphi(\mu)$, where the (implicit) function $\varphi(\mu)$ is defined in a small neighborhood of μ_k, and where $w_{k+1} = \varphi(\mu_k)$. As already mentioned in §3, for small enough values of μ_k, the Jacobian $\nabla_w \Psi(w_{k+1};\mu_k) = \nabla_w \Psi(w_{k+1};0)$ is nonsingular and therefore the implicit function theorem yields that, in a small vicinity of μ_k, the function φ is well-defined and differentiable. In particular, if we let $w'_{k+1} = \varphi'(\mu_k)$, we have

$$w'_{k+1} = -\left(\nabla_w \Psi(w_{k+1};\mu_k)\right)^{-1} \nabla_\mu \Psi(w_{k+1};\mu_k) = \left(\nabla_w \Psi(w_{k+1};\mu_k)\right)^{-1} \begin{bmatrix} 0 \\ 0 \\ e \end{bmatrix}. \tag{5.36}$$

The estimate w^{EX}_{k+1} is computed as the following first-order Taylor expansion of $\varphi(\mu)$ about μ_k, which represents an extrapolation from the parameters $(\mu_k, \Psi(w_{k+1};\mu_k))$ to $(\mu_{k+1}, \Psi(w_{k+1};\mu_k))$ and defines the step

$$w^{\text{EX}}_{k+1} = \varphi(\mu_k) + \varphi'(\mu_k)(\mu_{k+1} - \mu_k) = w_{k+1} + w'_{k+1}(\mu_{k+1} - \mu_k). \tag{5.37}$$

Assuming all the functions of interest are three times continuously differentiable, so that φ is twice continuously differentiable, we have by Taylor's theorem that the point w^{EX}_{k+1} is within $\mathcal{O}((\mu_{k+1} - \mu_k)^2)$ of $w_+ = \varphi(\mu_{k+1})$, see [5].

Since the matrix $\nabla_w \Psi(w_{k+1};\mu_k)$ used in (5.36) is *the same* as the one needed to compute a Newton step from w_{k+1} (see (2.11)), it is now possible to take, from w^{EX}_{k+1}, the Newton step we would have taken had we stayed at w_{k+1}, which then defines the step

$$d^{\text{NW}}_{k+1} = -\left(\nabla_w \Psi(w_{k+1};\mu_k)\right)^{-1} \Psi(w_{k+1};\mu_k),$$

and thereby define our composite extrapolation step

$$\begin{aligned}
w^{\text{PD}}_{k+1} &\stackrel{\text{def}}{=} w^{\text{EX}}_{k+1} + d^{\text{NW}}_{k+1} \\
&= w_{k+1} + \left(\nabla_w \Psi(w_{k+1};\mu_k)\right)^{-1} \left(\begin{bmatrix} 0 \\ 0 \\ e \end{bmatrix} (\mu_{k+1} - \mu_k) - \Psi(w_{k+1};\mu_k) \right) \\
&= w_{k+1} - \left(\nabla_w \Psi(w_{k+1};\mu_k)\right)^{-1} \begin{bmatrix} \nabla_x \mathcal{L}(w_{k+1}) \\ 0 \\ C(x_{k+1})z_{k+1} - \mu_{k+1} e \end{bmatrix},
\end{aligned} \tag{5.38}$$

where we have used (2.2), (5.36), (5.37) and the definition of d^{NW}_{k+1}. The step (5.38) amounts to an extrapolation from the parameters $(\mu_k, \Psi(w_{k+1};\mu_k))$ to $(\mu_{k+1}, 0)$, in the spirit of the predictor-corrector approach used in linear programming. Notice that, since the Jacobian matrix $\nabla_w \Psi(w;\mu)$ is independent of μ, the steps (5.31) and (5.38) are identical. This is not true in the purely primal case, as the Jacobian matrix $\nabla_w \Psi(w;\mu)$ is no longer independent of μ. An illustration of the decomposition of the step (5.38) appears in Fig. 2.

As $\mu \searrow 0$, the trajectory given by (5.35) and represented by $\varphi(\mu)$ obviously gets closer and closer to the primal-dual central path $\mathcal{C}$ represented by $w(\mu)$, until both coincide at w^*. Moreover, in this case, its derivative

$$\varphi'(\mu) = -\left(\nabla_w \Psi(w_{k+1};\mu)\right)^{-1} \nabla_\mu \Psi(w_{k+1};\mu)$$

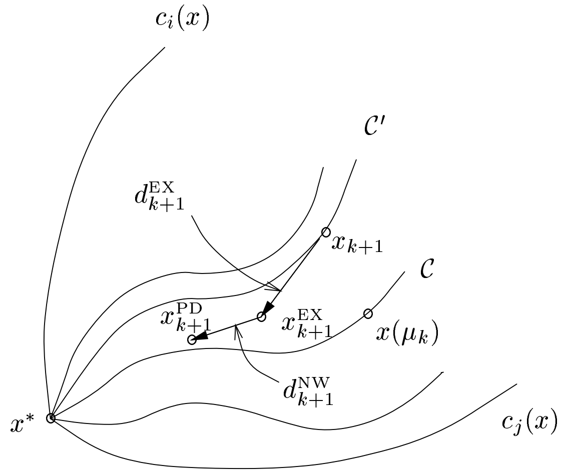

Figure 2: The above picture simplifies the situation as the x-space only is considered. The infeasible region is the "outer" part of the picture. Constraints i and j are active at the local constrained minimizer x^*. The exact solution of the current barrier subproblem, $x(\mu_k)$, lies exactly on the primal-dual central path $\mathcal{C}$. The second path $\mathcal{C}'$ that is pictured is defined implicitely by $\Psi(w;\mu) = \Psi(w_{k+1};\mu_k)$ as in equation (5.35). The two outermost paths represent the neighborhood of $\mathcal{C}$ defined by the forcing functions $\epsilon^{\text{C}}(\mu)$ and $\epsilon^{\text{D}}(\mu)$. The extrapolated step d_{k+1}^{EX} taken from w_{k+1} leaves the path $\mathcal{C}'$ tangentially and leads to w_{k+1}^{EX}. To that step is added the Newton step d_{k+1}^{NW} that would have been taken from w_{k+1}, as represented on the picture. The step $d_{k+1}^{\text{EX}} + d_{k+1}^{\text{NW}}$ is the step leading to w_{k+1}^{PD} as given by (5.31).

converges to

$$-\left(\nabla_w \Psi(w^*;0)\right)^{-1} \begin{bmatrix} 0 \\ 0 \\ -e \end{bmatrix}$$

because of (2.11), which, in view of (3.6), equals $\dot{w}(0)$. The two paths thus coincide up to first order at w^*. This intuitively guarantees that expanding the function φ asymptotically gives an accurate approximation of the primal-dual central path.

6 Local convergence analysis

Having proposed the point w_{k+1}^{PD} as a possible starting point for the inner minimization, we now wish to analyze its properties from the point of view of improving the local convergence rate of the algorithm. We will prove that this particular choice is not only strictly feasible—in contrast with the purely primal case [17]—but also dramatically improves on the bound (5.28). In fact we shall show that this point asymptotically satisfies the stopping criterion (4.5)–(4.8), which means that the inner minimization algorithm is ultimately not needed. In particular, this means that, asymptotically, only one linear system (5.31) need be solved per update of the barrier parameter μ.

We start by verifying that this system is asymptotically well posed, at least along the converging subsequence.

Lemma 5 *Under the assumptions of Theorem 3, there exists a closed and bounded neighborhood $\mathcal{V}$ of w^* such that the matrix $V(x,z)$ defined by (4.3) is positive definite over the nullspace of A*

for all $w \in \mathcal{V}$.

Proof. The result follows from the application of [16, Theorem 8, (iii)] to the reduced Hessian matrix $N^T V(x,z)N$. $\square$

This result indicates that the systems (5.31) and (5.33) are asymptotically well posed for $k \in \mathcal{K}$ sufficiently large.

In the next major stage of our analysis, we verify that the stopping conditions for the inner minimization are all satisfied at $w_{k+1}^{\rm PD}$, provided we impose further conditions on the barrier parameter updating rule. Furthermore, we also show that the bound (5.28) can be improved in this context.

Theorem 6 *Under* **AS1**–**AS6**, *assume* w^* *is a solution of NLP, that the sequence* $\{w_{k+1}\}_{\mathcal{K}} \to w^*$, *where* $\{w_{k+1}\}$ *is a sequence of iterates generated by Algorithm 4.1 with* y_{k+1} *defined by (4.9), and that the functions* f *and* c_i $(i = 1, \ldots, p)$ *are three times continuously differentiable over an open neighborhood of* x^*. *Assume furthermore that (5.1)–(5.3) are satisfied, that* $0 < \epsilon_\tau < 1/2$ *is a given constant, that* $0 < \gamma_k \le (1 - 2\epsilon_\tau)/(1 + 2\epsilon_\tau)$, *and that the barrier parameter updating rule satisfies*

$$\mu_{k+1} = \Omega(\mu_k^{\tau_k}) \qquad 1 + \epsilon_\tau \le \tau_k \le \frac{2}{1 + \gamma_{k+1}} - \epsilon_\tau. \tag{6.1}$$

Then, we have that, for $k \in \mathcal{K}$ *sufficiently large,*

$$\begin{aligned} A x_{k+1}^{PD} &= b & (6.2)\\ (c(x_{k+1}^{PD}), z_{k+1}^{PD}) &> 0 & (6.3)\\ \|C(x_{k+1}^{PD}) z_{k+1}^{PD} - \mu_{k+1} e\| &\le \epsilon^C(\mu_{k+1}) \quad \textit{and} & (6.4)\\ \|\nabla_x f(x_{k+1}^{PD}) - J^T(x_{k+1}^{PD}) z_{k+1}^{PD}\|_{[k+2]} &\le \epsilon^D(\mu_{k+1}), & (6.5) \end{aligned}$$

and

$$\|\Psi(w_{k+1}^{PD}; \mu_{k+1})\| = o(\mu_{k+1}). \tag{6.6}$$

Proof. Observe first that (5.3) and (6.1) imply that

$$\mu_k^2 = o(\mu_{k+1}). \tag{6.7}$$

We start by proving (6.2). From (5.31), the direction d_{k+1} satisfies the equations

$$\nabla_{xx} L(x_{k+1}, z_{k+1})[d_{k+1}]_x + A^T [d_{k+1}]_y - J^T(x_{k+1})[d_{k+1}]_z = -\nabla_x \mathcal{L}(w_{k+1}), \tag{6.8}$$

$$A[d_{k+1}]_x = 0, \tag{6.9}$$

and

$$Z_{k+1} J(x_{k+1})[d_{k+1}]_x + C(x_{k+1})[d_{k+1}]_z = \mu_{k+1} e - C(x_{k+1}) z_{k+1}, \tag{6.10}$$

where $\nabla_{xx} L(x,z)$ is defined in (4.4). It follows from (6.9) that $x_{k+1}^{\rm PD}$ satisfies the equality constraints, which implies that (6.2) holds for all $k \in \mathcal{K}$. Note that since the right-hand side of (6.8)–(6.10), which is $-((\mu_{k+1} - \mu_k)(0\ 0\ \ -e^T)^T + \Psi(w_{k+1}; \mu_k))$, is $\mathcal{O}(\mu_k)$ because of (4.7),(4.8) and (5.1)–(5.3), and as the Jacobian $\nabla_w \Psi(w_{k+1}; \mu_{k+1})$ remains uniformly nonsingular in the vicinity of w^*, we also obtain that

$$d_{k+1} = \mathcal{O}(\mu_k), \tag{6.11}$$

for all sufficiently large $k \in \mathcal{K}$. As a consequence, the sequence $\{w_{k+1}^{\rm PD}\}_{k \in \mathcal{K}}$ also converges to w^* since μ_k converges to zero.

We next show that $c(x_{k+1}^{\rm PD}) > 0$, which is part of (6.3). If constraint i is inactive, $\{c_i(x_{k+1})\}_{\mathcal{K}} \to c_i(x^*) > 0$ as $\mu_k \searrow 0$. Taylor's expansion of c_i around x_{k+1}, (5.4) and (6.11) give that

$$c_i(x_{k+1} + [d_{k+1}]_x) = c_i(x_{k+1}) + \mathcal{O}(\mu_k),$$

and thus, asymptotically,

$$0 < \tfrac{1}{2} c_i(x_{k+1}) \leq c_i(x_{k+1}^{\rm PD}) \leq 2 c_i(x_{k+1}), \tag{6.12}$$

which shows that $x_{k+1}^{\rm PD}$ is strictly feasible with respect to the inactive constraints. Now consider the active constraints, if any. Premultiplying (6.10) by Z_{k+1}^{-1} and rearranging, we obtain that for all $i = 1, \ldots, p$:

$$c_i(x_{k+1}) + \nabla_x c_i(x_{k+1})^T [d_{k+1}]_x = \mu_{k+1} [z_{k+1}]_i^{-1} - [z_{k+1}]_i^{-1} c_i(x_{k+1}) [[d_{k+1}]_z]_i. \tag{6.13}$$

For all active indices, (5.5), (5.6) and (6.11) show that the last term of the right-hand side of (6.13) is $\mathcal{O}(\mu_k^2)$ so that we obtain:

$$c_i(x_{k+1}) + \nabla_x c_i(x_{k+1})^T [d_{k+1}]_x = \mu_{k+1} [z_{k+1}]_i^{-1} + \mathcal{O}(\mu_k^2) \qquad (i \in \mathcal{A}). \tag{6.14}$$

Substituting this equation in the expansion

$$c_i(x_{k+1} + [d_{k+1}]_x) = c_i(x_{k+1}) + \nabla_x c_i(x_{k+1})^T [d_{k+1}]_x + \mathcal{O}(\|[d_{k+1}]_x\|^2), \tag{6.15}$$

where we have used (5.4), and using (6.11) gives, for all $i \in \mathcal{A}$:

$$c_i(x_{k+1} + [d_{k+1}]_x) = \mu_{k+1} [z_{k+1}]_i^{-1} + \mathcal{O}(\mu_k^2). \tag{6.16}$$

If constraint i is active, $\{c_i(x_{k+1})\}_{\mathcal{K}} \to c_i(x^*) = 0$ as $\mu_k \searrow 0$. Using now the bounds (5.6), (6.16) yields

$$\frac{\mu_{k+1}}{\kappa^z} + \mathcal{O}(\mu_k^2) \leq c_i(x_{k+1}^{\rm PD}) \leq \frac{\mu_{k+1}}{\kappa_z} + \mathcal{O}(\mu_k^2). \tag{6.17}$$

Combining (6.7) and (6.17), we obtain

$$0 < \frac{1}{2\kappa^z} \mu_{k+1} \leq c_i(x_{k+1}^{\rm PD}) \leq \frac{2}{\kappa_z} \mu_{k+1} \tag{6.18}$$

as soon as μ_k is sufficiently small. Relations (6.12) and (6.18) together show that $x_{k+1}^{\rm PD}$ is asymptotically strictly feasible.

In order to complete our proof of (6.3), we now consider the feasibility of $z_{k+1}^{\rm PD}$. Note that since the direction d_{k+1} is $\mathcal{O}(\mu_k)$ because of (6.11), the same holds for $[d_{k+1}]_z$. We then have that, for every $i \in \mathcal{A}$,

$$[z_{k+1}]_i + [[d_{k+1}]_z]_i = [z_{k+1}]_i + \mathcal{O}(\mu_k), \tag{6.19}$$

which implies, by **AS5** and for sufficiently large $k \in \mathcal{K}$, that

$$0 < \tfrac{1}{2} [z_{k+1}]_i \leq [z_{k+1}^{\rm PD}]_i \leq 2 [z_{k+1}]_i, \tag{6.20}$$

so that $[z_{k+1}^{\rm PD}]_i$ is asymptotically positive. For the inactive constraints, (6.11) indicates that there exists $\kappa^d > 0$ such that $\|[d_{k+1}]_x\| \leq \kappa^d \mu_k$. From (6.10) and the Cauchy-Schwarz inequality, we know that for all $i = 1, \ldots, p$,

$$\begin{aligned} [z_{k+1}]_i + [[d_{k+1}]_z]_i &= c_i^{-1}(x_{k+1}) \left(\mu_{k+1} - [z_{k+1}]_i (\nabla_x c_i(x_{k+1}))^T [d_{k+1}]_x \right) && (6.21) \\ &\geq c_i^{-1}(x_{k+1}) \left(\mu_{k+1} - [z_{k+1}]_i \|\nabla_x c_i(x_{k+1})\| \|[d_{k+1}]_x\| \right). && (6.22) \end{aligned}$$

Using (5.4), (5.7) and (6.22), we have that, for $i \notin \mathcal{A}$ and $k \in \mathcal{K}$ sufficiently large,

$$\begin{aligned}
[z^{\rm PD}_{k+1}]_i &\geq c_i^{-1}(x_{k+1}) \left(\mu_{k+1} - [z_{k+1}]_i \kappa_\gamma \kappa^d \mu_k\right) \\
&\geq c_i^{-1}(x_{k+1}) \left(\mu_{k+1} - \frac{(1+\overline{\kappa}^{\rm C}_\mu)}{\kappa_{\rm c}} \kappa_\gamma \kappa^d \mu_k^2\right) \\
&> 0
\end{aligned}$$

where the last inequality follows from (6.7). Thus (6.3) holds for sufficiently large $k \in \mathcal{K}$.

We next prove (6.6). In view of (2.11), our differentiability assumptions, (5.6), (5.7) and (6.11) imply that the partial derivatives with respect to x, y and z of each of the elements of $\nabla_w \Psi(w;\mu)$ clearly remain bounded in a neighbourhood of w^* as μ goes to zero by (5.4). Consequently, applying Taylor's theorem to Ψ, we have

$$\Psi(w^{\rm PD}_{k+1};\mu_{k+1}) = \Psi(w_{k+1};\mu_{k+1}) + \nabla_w \Psi(w_{k+1};\mu_{k+1}) d_{k+1} + O(\|d_{k+1}\|^2). \tag{6.23}$$

From the definition (5.31) of $w^{\rm PD}_{k+1}$, the first two terms of (6.23) vanish, and hence we deduce

$$\|\Psi(w^{\rm PD}_{k+1};\mu_{k+1})\| = \mathcal{O}(\mu_k^2) \tag{6.24}$$

from (6.11). We finally deduce (6.6) from this bound and (6.7).

That (6.4) holds now immediately follows from (6.6) and (5.1). We next prove (6.5). Using (4.17), we have that

$$\|\nabla_x \mathcal{L}(w^{\rm PD}_{k+1})\|_{[k+2]} \leq \epsilon_M^{-1/2} \|\nabla_x \mathcal{L}(w^{\rm PD}_{k+1})\|_\diamond \leq \epsilon_M^{-1/2} \|\nabla_x \mathcal{L}(w^{\rm PD}_{k+1})\| = \mathcal{O}(\mu_k^2), \tag{6.25}$$

where the last equation follows from (6.24). Now, using (6.1) and (5.2), we have that

$$\mu_k^2 = \mathcal{O}\left(\mu_{k+1}^{\frac{2}{\tau_k}}\right) = o\left(\mu_{k+1}^{1+\gamma_{k+1}}\right) = o(\epsilon^{\rm D}(\mu_{k+1})),$$

which, with (6.25), implies that (6.5) holds for $k \in \mathcal{K}$ sufficiently large. □

An important consequence of this result is that, once $w^{\rm PD}_{k+1}$ has been computed, the inner minimization is ultimately unnecessary, since this "starting point" already satisfies the stopping conditions for this minimization. Thus we choose, in what follows,

$$x_{k+2} \stackrel{\rm def}{=} x^{\rm PD}_{k+1}, \quad z_{k+2} \stackrel{\rm def}{=} z^{\rm PD}_{k+1}, \quad \text{and} \quad y_{k+2} \text{ according to (4.9)}. \tag{6.26}$$

Observe that this makes the complete algorithm asymptotically independent of the procedure chosen for the inner minimization, since this procedure is asymptotically never used.

Observe also that, from (5.12) and (6.11),

$$\|w^{\rm PD}_{k+1} - w(\mu_k)\| = \mathcal{O}\left(\mu_k^{\gamma_k + \frac{1}{2}} + \mu_k\right),$$

and from (3.3), we have $\|w(\mu_k) - w(\mu_{k+1})\| = \mathcal{O}(\mu_k - \mu_{k+1}) = \mathcal{O}(\mu_k)$. Combining these two observations, we obtain

$$\|w^{\rm PD}_{k+1} - w(\mu_{k+1})\| = \mathcal{O}\left(\mu_k^{\gamma_k + \frac{1}{2}} + \mu_k\right). \tag{6.27}$$

In the primal-dual case, [1] and [15] show that the radius of the sphere of convergence of Newton's method for $\mathrm{BS}(\mu_{k+1})$ is both finite and bounded away from zero. Hence, equation (6.27) shows that $w^{\rm PD}_{k+1}$ asymptotically lies inside that sphere and thus that Newton's method started from

w_{k+1}^{PD} would generate points that converge quadratically to $w(\mu_{k+1})$, if an inner minimization were to be used. Also note that (6.1) indicates that the rate of decrease of the barrier parameter must not be too large.

Theorem 6 shows that, as soon as the barrier parameter is sufficiently small, the point (5.31) lies strictly inside the feasible region. In a practical implementation, it might be preferable to decide whether or not the algorithm is in a sufficiently advanced stage to use (5.31) by checking its feasibility in conjunction with a test of the form

$$\|\nabla_x \mathcal{L}(w_{k+1}^{\mathrm{PD}})\|_{[k+2]} \le \max(\eta, \|\nabla_x \mathcal{L}(w_{k+1})\|_{[k+2]}),$$

for some parameter $\eta > 0$, and to ignore the "improved" starting point if this test is violated.

If we now wish to pursue our rate of convergence analysis, we must be more specific about the rule used to update the barrier parameter. So far, we have assumed that $\mu_{k+1} \le \mu_k$ and (6.1); from now on, we will assume that

$$\mu_{k+1} = \Theta(\mu_k^{\tau_k}) \tag{6.28}$$

where τ_k remains within the bounds specified in (6.1).

The rate of convergence of $\{\mu_k\}$ implied by the updating rule (6.28) directly depends on the sequence $\{\gamma_k\}$ chosen in (5.2). The rule implies

$$\mu_{k+1} = \Omega\left(\mu_k^{\frac{2}{1+\gamma_{k+1}} - \epsilon_\tau}\right),$$

from which we may retrieve the rule (5.29) if we choose $\gamma_k \ge (2 - 3\epsilon_\tau)/(4 + 3\epsilon_\tau)$ for all k. If we choose ϵ_τ sufficiently small and impose $\lim_{k\to\infty} \gamma_k = 0$, then the rate at which the barrier parameter approaches zero can be made as close to quadratic as one desires.

It is important to make a distinction between a rate of convergence in μ and a rate of convergence in the variables w of the problem. In view of (3.4), if one is able to compute the *exact* solution of BS(μ) for every μ, $w(\mu)$ converges to w^* exactly as fast as μ decreases to zero. Intuitively, when $\gamma_k \ge \frac{1}{2}$, and because (5.12) and (5.13) are always satisfied, one can reasonably expect the same rate of convergence in the approximate solutions w_{k+1} as in μ_k, and we have observed this in practice. However, it is not immediately clear that this may be made rigorous, since the bound (5.13) is only one-sided. In the next results, we show that even in the case $\gamma_k < \frac{1}{2}$, not only can we show that $\{w_{k+1}\}_\mathcal{K}$ converges R-superlinearly to w^*, but we obtain Q-superlinear convergence of the whole sequence $\{w_{k+1}\}$ without restrictions on the sequence of scalars $\{\gamma_k\}$. The following lemma parallels [10, Lemma 5.13].

Lemma 7 *Under* **AS1**–**AS6**, *assume w^* is a solution of NLP, that $\{w_{k+1}\}_\mathcal{K} \to w^*$, where $\{w_{k+1}\}$ is a sequence of iterates generated by Algorithm 4.1 with y_{k+1} defined by (4.9), and that the functions f and c_i ($i = 1, \ldots, p$) are three times continuously differentiable over an open neighborhood of x^*. Assume furthermore that the barrier parameter μ_k is updated using (6.28) and that it is small enough to ensure that w_{k+1}^{PD} defined by (5.31) is strictly feasible. Then, we have the estimate*

$$w_{k+2} = w(0) + \mu_{k+1}\dot{w}(0) + o(\mu_{k+1}), \tag{6.29}$$

for all sufficiently large $k \in \mathcal{K}$, where w_{k+2} is defined by (6.26) and $\dot{w}(0) \ne 0$.

Proof. Proceeding as in the proof of Theorem 6, a second-order Taylor expansion of $\Psi(w;\mu)$ about $(w,\mu) = (w^*, 0)$ and the optimality conditions (2.3) yield

$$\begin{aligned}
\Psi(w_{k+2};\mu_{k+1}) &= \nabla_w \Psi(w^*;0)(w_{k+2} - w^*) + \nabla_\mu \Psi(w^*;0)\mu_{k+1} + r && (6.30)\\
&= \nabla_w \Psi(w^*;0)(w_{k+2} - w^*) + \begin{bmatrix} 0 \\ 0 \\ -\mu_{k+1} e \end{bmatrix} + r, && (6.31)
\end{aligned}$$

where

$$\|r\| = \mathcal{O}(\max(\|w_{k+2} - w^*\|^2, \mu_{k+1}^2)). \tag{6.32}$$

We may rewrite (6.31) as

$$\begin{bmatrix} \nabla_x \mathcal{L}(w_{k+2}) \\ 0 \\ C(x_{k+2}) z_{k+2} \end{bmatrix} = \nabla_w \Psi(w^*; 0)(w_{k+2} - w^*) + r. \tag{6.33}$$

Since $\nabla_x \mathcal{L}(w_{k+2})$ lies in the nullspace of A because of (4.11) and (6.26), we have,

$$\|\nabla_x \mathcal{L}(w_{k+2})\| = \|\nabla_x \mathcal{L}(w_{k+2})\|_\diamond = \|\nabla_x \mathcal{L}(w_{k+1}^{\mathrm{PD}})\|_\diamond = o(\mu_{k+1}) \tag{6.34}$$

and

$$C(x_{k+2}) z_{k+2} = \mu_{k+1} e + o(\mu_{k+1}), \tag{6.35}$$

using (6.6). Consequently, substituting into (6.33) and using the nonsingularity of $\nabla_w \Psi(w^*; 0)$, we obtain

$$w_{k+2} = w(0) + \mu_{k+1} \dot{w}(0) + o(\mu_{k+1}) + (\nabla_w \Psi(w^*; 0))^{-1} r, \tag{6.36}$$

where we observed that $w^* = w(0)$ and $(\nabla_w \Psi(w^*; 0))^{-1} [0\ 0\ e^T]^T = \dot{w}(0) \neq 0$ because of (3.6).

To complete the proof, it remains to show that $\|r\| = o(\mu_{k+1})$. From (6.34) and (6.35), for sufficiently large $k \in \mathcal{K}$, $\|\nabla_x \mathcal{L}(w_{k+2})\|_\diamond + \|C(x_{k+2}) z_{k+2}\|$ is smaller than the threshold ε given in Lemma 1, and thus, applying (3.3) with the parameters $(\nabla_x \mathcal{L}(w_{k+2}), 0, C(x_{k+2}) z_{k+2})$ and $(0, 0, 0)$ yields

$$\begin{aligned} \|w_{k+2} - w^*\| &= \Theta(\|\nabla_x \mathcal{L}(w_{k+2})\|_\diamond + \|C(x_{k+2}) z_{k+2}\|) && (6.37) \\ &= \mathcal{O}(\mu_{k+1}), && (6.38) \end{aligned}$$

where the last equality is due to (6.34) and (6.35). We thus obtain that $\|w_{k+2} - w^*\|^2 = \mathcal{O}(\mu_{k+1}^2) = o(\mu_{k+1})$, and thus (6.32) clearly implies that $\|r\| = o(\mu_{k+1})$ and proves (6.29). $\square$

The result contained in Lemma 7 parallels the well-known expansion

$$w(\mu_{k+1}) = w(0) + \mu_{k+1} \dot{w}(0) + o(\mu_{k+1}),$$

that holds for *exact* solutions of $\mathrm{BS}(\mu)$. It also confirms the suggestion in Fig. 2 that the trajectory $\mathcal{C}'$ is close to the primal-dual central path. Moreover, it reinforces the observation that the paths $\mathcal{C}$ and $\mathcal{C}'$ coincide up to first order at w^*.

Note that without loss of generality, we may assume that the components of the vector $\dot{w}(0)$ defined in (3.6) are all nonzero. A suitable change of coordinates can always transform this nonzero vector into the vector of all ones without modifying the nature of problem NLP. The identity (6.29) thus holds *componentwise*, so that we immediately have the following corollary.

Corollary 8 *Under the assumptions of Lemma 7, we have*

$$[w_{k+2}]_i = [w^*]_i + \mu_{k+1} [\dot{w}(0)]_i + o(\mu_{k+1}), \quad i = 1, \ldots, n + m + p, \tag{6.39}$$

for all sufficiently large $k \in \mathcal{K}$, *with* w_{k+2} *defined by (6.26) and* $[\dot{w}(0)]_i \neq 0$ *for all* $i = 1, \ldots, n + m + p$.

Consequently, there exists a constant $\kappa^w > 0$ such that, for all $k \in \mathcal{K}$ sufficiently large,

$$|[w_{k+2}]_i - [w^*]_i| \leq \kappa^w \mu_{k+1}, \quad i = 1, \ldots, n + m + p, \tag{6.40}$$

which, unlike (5.13), is independant of the sequence $\{\gamma_k\}$.

So far, we have simply assumed that w^* is a limit point of the sequence $\{w_{k+1}\}_{k\in\mathcal{K}}$. We are now in position to prove that the whole sequence of iterates w_{k+1} converges Q-superlinearly to w^*. Moreover, as in (6.39), this convergence occurs componentwise, showing that all errors $|[w_{k+2}]_i - [w^*]_i|$ are of comparable size, and thus that *all* variables converge to their limit at a comparable rate. The following result is inspired by [10, Theorem 5.14].

Theorem 9 *Under* **AS1**–**AS6**, *assume w^* is a solution of NLP, that the sequence $\{w_{k+1}\}_{\mathcal{K}} \to w^*$, where $\{w_{k+1}\}$ is a sequence of iterates generated by Algorithm 4.1 with y_{k+1} defined by (4.9), and that the functions f and c_i $(i = 1, \ldots, p)$ are three times continuously differentiable over an open set containing $\mathcal{F}$. Assume furthermore that (5.1)–(5.3) are satisfied and that the barrier parameter μ_k is updated using (6.28). Assume finally that (6.26) is used for all $k \in \mathcal{K}$ large enough to ensure that Theorem 6 holds. Then the complete sequence $\{w_{k+1}\}$ converges to w^* and*

$$\left|\frac{[w_{k+2}]_i - [w^*]_i}{[w_{k+1}]_i - [w^*]_i}\right| = \Theta(\mu_k^{\tau_k - 1}) \quad i = 1, \ldots, n + m + p, \tag{6.41}$$

for k sufficiently large, which implies that the iterates w_{k+1} converge componentwise Q-superlinearly to w^*.

Proof. First note that (6.40) and the convergence of μ_{k+1} to zero implies that $\{w_{k+2}\}_{k\in\mathcal{K}}$ also converges to w^*. Hence we may reapply (6.40) with $\mathcal{K}$ replaced by $\mathcal{K} \cup \{k+1\}_{k\in\mathcal{K}}$ and obtain that $\{w_{k+3}\}_{k\in\mathcal{K}}$ also converges to w^*. Applying this argument inductively, we obtain that the complete sequence $\{w_k\}$ converges to w^* and therefore that $\mathcal{K}$ may be identified with the set of all positive integers. Our assumptions then yield that Theorem 6 holds and that (6.26) is used for all k sufficiently large. Moreover, the estimate (6.39) also holds for all k sufficiently large, implying that $|[w_{k+1}]_i - [w^*]_i| = \Theta(\mu_k)$ since we noted that, without loss of generality, we have $[\dot{w}(0)]_i \neq 0$ for $i = 1, \ldots, n + m + p$. This proves that

$$\left|\frac{[w_{k+2}]_i - [w^*]_i}{[w_{k+1}]_i - [w^*]_i}\right| = \Theta\left(\frac{\mu_{k+1}}{\mu_k}\right) \quad i = 1, \ldots, n + m + p,$$

which then gives (6.41) because of (6.28). The componentwise Q-superlinear convergence of the iterates to w^* then follows from the convergence of μ_k to zero and the inequality $\tau_k \geq 1 + \epsilon_\tau$. □

Theorem 9 has the following consequence.

Corollary 10 *Under the assumptions of Theorem 9, suppose that*

$$\tau_k = \frac{2}{1 + \gamma_{k+1}} - \epsilon_\tau \tag{6.42}$$

with γ_k satisfying (5.3) and

$$\lim_{k\to\infty} \gamma_k = 0.$$

Then, for any $\sigma \in (1, 2 - \epsilon_\tau)$, there exists a constant $q_\sigma > 0$ such that

$$|[w_{k+2}]_i - [w^*]_i| \leq q_\sigma |[w_{k+1}]_i - [w^*]_i|^\sigma \quad i = 1, \ldots, n + m + p.$$

Proof. First recall that Corollary 8 and the fact that $[\dot{w}(0)]_i \neq 0$ for $i = 1, \ldots, n + m + p$ implies that $|[w_{k+1}]_i - [w^*]_i| = \Theta(\mu_k)$ for all k sufficiently large. This and (6.28) yield that

$$\frac{|[w_{k+2}]_i - [w^*]_i|}{|[w_{k+1}]_i - [w^*]_i|^\sigma} = \Theta(\mu_k^{\tau_k - \sigma}) \quad i = 1, \ldots, n + m + p.$$

Our assumptions and the fact that, for any $\sigma \in (1, 2 - \epsilon_\tau)$,

$$\tau_k = \frac{2}{1 + \gamma_{k+1}} - \epsilon_\tau \geq \sigma$$

for k sufficiently large then implies the desired result. □

Interior point methods of the type studied above are thus likely to achieve a rate of convergence that is in practice as fast as that of exterior-penalty methods. In addition, the rate of convergence implied by our theory is governed by ϵ_τ, and Corollary 10 shows that this rate can be made as close to quadratic as we wish by choosing ϵ_τ sufficiently close to zero in (6.42). Note that Corollary 8 also holds for [10, Lemma 5.13] in connection with exterior penalty methods, and thus that componentwise Q-superlinear convergence also occurs in that case.

Remark 6.1. Most of the qualitative observations made in this paper essentially remain true in the purely primal case. When considering the primal approach, one has to replace (2.2) by

$$\Psi_p(x, y; \mu) = \begin{bmatrix} \nabla_x \mathcal{L}_p(x, y; \mu) \\ Ax - b \end{bmatrix}$$

where $\nabla_x \mathcal{L}_p(x, y; \mu) = \nabla_x f(x) + A^T y - \mu J^T(x) C^{-1}(x) e$, since the left-hand side of (4.7) is always identically zero. The primal case is analyzed both for interior and exterior penalty functions by Dussault [5]. From the quantitative point of view, one obtains two-step superlinear convergence in the primal case as opposed to one-step superlinear convergence in the primal-dual case, using the same sort of Newton-like extrapolation step. In the primal case, the extrapolation step itself is not enough to satisfy the termination tolerances, and one has to perform an additional Newton step. The result obtained in [5] is that the updating rule for the barrier parameter has to satisfy $\mu_{k+1} = \Omega(\mu_k^{4/3})$, thereby limiting the speed of convergence. ■

7 An application

We now show that the theoretical framework developed above can be applied. In particular we first examine why it applies to the method proposed in [2]. This algorithm consists in an inner iteration imbricated in an outer one. We do not describe the inner minimization here, but refer the interested reader to [2] for a discussion. Suffice it to say that it uses a trust-region algorithm with a primal-dual model of the log-barrier function (1.3). The stopping conditions for this inner iteration are exactly (4.5)-(4.8) augmented by the requirement that

$$\lambda^{\min}_{M_{k+1}} \left[V(x_{k+1}, z_{k+1}) \right] \geq -\epsilon^{\mathrm{E}}(\mu_k), \tag{7.1}$$

for some forcing function $\epsilon^{\mathrm{E}}(\mu)$. This additional condition is meant to enforce convergence to a second-order critical point of the log-barrier function.

The global convergence of the resulting minimization procedure (inner and outer minimizations together) to weak second-order critical points is guaranteed under standard assumptions [2, Theorem 4.12]. These assumptions are slightly different from those used here: in particular, strict complementarity slackness is not required, and approximate second-order derivatives of the objective function and constraints are allowed, while we concentrate here on the case where they are exact. However, this convergence result depends on three additional conditions, namely that

$$\lim_{k \to \infty} \frac{\epsilon^{\mathrm{D}}(\mu_k) \sqrt{\mu_k}}{\min_i c_i(x_{k+1})} = 0, \tag{7.2}$$

that the tolerance $\epsilon^{\mathrm{C}}(\mu)$ is asymptotically of the form $\mathcal{O}(\mu)$, and that the barrier function is bounded from below on $\mathcal{F}$ for all generated iterates and all small μ. Note that (5.1) implies that $\epsilon^{\mathrm{C}}(\mu_k)$ is asymptotically of the order of μ_k, and that (5.2) implies that condition (7.2) is satisfied because of (5.5) and (5.8). As a consequence we see that the global convergence theory for the particular implementation described in [2] is not upset by our choice of stopping tolerances. Moreover, the boundedness of the log-barrier function is guaranteed here, because of AS2, Theorem 3 and the fact that our analysis only considers convergent subsequences.

In order to apply our rate of convergence results, we finally have to verify that introducing condition (7.1) in the set of stopping criterion for the inner minimization algorithm does not affect our conclusions. Fortunately, we may deduce from Theorem 5 that the matrix $V(x, z)$ is asymptotically second-order sufficient and thus ultimately that condition (7.1) will automatically be satisfied at the iterates generated by Algorithm 4.1 that are sufficiently close to a local solution w^*, among which the iterates w_{k+1} with $k \in \mathcal{K}$. Consequently, Theorem 9 and Corollary 10 apply for the algorithm proposed by [2].

It is tempting, although technically difficult, to attempt applying our results to other primal-dual methods for nonlinear optimization. In particular, the methods of Gay, Overton and Wright [7] and Byrd, Liu and Nocedal [1] seem natural candidates. However, a fully unified theory appears to require more work. In particular, besides the fact that these methods handle the full nonlinear program, including nonlinear equality constraints, and allow infeasible iterates with respect to those constraints, they also differ from our framework in further respects. The method of Gay, Overton and Wright uses a watchdog technique to allow a possible non-monotone behavior of the sequence of values of a log-barrier based merit function, while our technique does not impose any condition on this sequence. The method of Byrd, Liu and Nocedal uses slack variables to transform general inequalities into bound constraints. Both methods impose the same accuracy requirement for (4.7) and (4.8) while our approach differentiates between those two components of the optimality conditions (see (5.1)–(5.3)). Moreover, the rules to update the barrier parameter differ in both cases from those considered here.

8 Conclusion

In this paper, we have studied the local convergence properties of primal-dual interior point algorithms for minimizing a general, nonconvex, objective function subject to linear equality constraints and nonconvex inequality constraints, of which the method proposed by Conn, Gould, Orban and Toint [2] is a prime example. Our analysis is inspired by those of [3, 5, 10, 21]. The theoretical results show a convergence rate for barrier methods that is essentially as fast as that previously obtained for exterior penalty methods [10]. These results rely on a suitable extrapolation of the central path from the current iterate w_{k+1} which leads to an asymptotically acceptable w_{k+2}, *i.e.* a point which immediately satisfies the tolerance requirements corresponding to the updated barrier parameter μ_{k+1}. This is shown to imply a componentwise Q-superlinear convergence rate, and one asymptotically has to solve, in each outer iteration, a single linear system whose coefficient matrix is that of the Newton equations at w_{k+1}. It is worth emphasizing that the results presented here hold independently of the exact inner minimization procedure used, provided it ensures that (4.5)–(4.8) are satisfied. The componentwise Q-superlinear convergence of w_k to w^* also holds independently of any particular updating rule for the variable γ_k used in (5.2), provided it satisfies (5.3). Moreover, if the sequence $\{\gamma_k\}$ converges to zero, then the rate of convergence can be made as close to quadratic as desired by choosing ϵ_τ sufficiently small in (6.42). A consequence is that we may alternatively view the results of the present paper as giving conditions on the stopping criterion of any barrier subproblem solver that ensure componentwise Q-superlinear convergence of the outer iterates. This parallels the results of [13] for linear complementarity problems.

Reasons why one should use extrapolated steps in barrier-type methods are developed in [3, 12, 17], and an analysis similar to that developed in the present paper is developed in [1] where one-step superlinear convergence of an interior point primal-dual trust-region algorithm is exhibited. As superlinear convergence has already been observed in practice during tests on quadratic programs (see [2]), the authors believe that it will be equally worthwhile to experiment with the strategy sketched in this paper on highly nonlinear and high-dimensional optimization problems. It should be mentioned however that the extrapolation strategy is only likely to be numerically efficient in conjunction with a method that solves the Newton equations accurately, without suffering from any ill-conditioning that is not already present in problem NLP [5]. Furthermore, higher convergence rates analysis, achievable by taking a further Newton step from the extrapolated point, will be analyzed in a companion paper. In view of the analysis conducted in [21], one may reasonably hope that the results exhibited in the present paper remain essentially true when the linear independence constraint qualification is replaced by the weaker Mangasarian-Fromovitz constraint qualification. One may also hope to obtain interesting, yet similar, results when the strict complementarity condition is dropped. Relaxation of those assumptions is left for future work.

Acknowledgments

The authors are grateful to Jorge Nocedal for his helpful comments on an earlier draft of this paper.

References

[1] R. H. Byrd, G. Liu, and J. Nocedal, *On the local behavior of an interior point method for nonlinear programming*, in Numerical Analysis 1997, D. Griffiths and D. Higham, eds., Addison Wesley Longman, 1997, pp. 37–56. Proceedings of the Dundee 1997 Conference on Numerical Analysis.

[2] A. R. Conn, N. I. M. Gould, D. Orban, and P. L. Toint, *A primal-dual trust-region algorithm for non-convex nonlinear programming*, Mathematical Programming, http://dx.doi.org/10.1007/s101070000144 (2000).

[3] A. R. Conn, N. I. M. Gould, and P. L. Toint, *A note on using alternative second-order models for the subproblems arising in barrier function methods for minimization*, Numerische Mathematik, 68 (1994), pp. 17–33.

[4] ——, *On the number of inner iterations per outer iteration of a globally convergent algorithm for optimization with general nonlinear inequality constraints and simple bounds*, Computational Optimization and Applications, 7 (1997), pp. 41–70.

[5] J.-P. Dussault, *Numerical stability and efficiency of penalty algorithms*, SIAM Journal on Numerical Analysis, 32 (1995), pp. 296–317.

[6] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, J. Wiley and Sons, Chichester, England, 1968. Reprinted as *Classics in Applied Mathematics*, SIAM, Philadelphia, USA, 1990.

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

[8] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*, Academic Press, London, 1981.

[9] N. I. M. Gould, *On practical conditions for the existence and uniqueness of solutions to the general equality quadratic-programming problem*, Mathematical Programming, 32 (1985), pp. 90–99.

[10] ——, *On the convergence of a sequential penalty function method for constrained minimization*, SIAM Journal on Numerical Analysis, 26 (1989), pp. 107–126.

[11] R. Mifflin, *Convergence bounds for nonlinear programming algorithms*, Mathematical Programming, 8 (1975), pp. 251–271.

[12] S. G. Nash and A. Sofer, *Why extrapolation helps in barrier methods*, tech. rep., Operations Research and Engineering Department, George Mason University, Fairfax USA, September 1998.

[13] F. Potra, *Q-superlinear convergence of the iterates in primal-dual interior-point methods*, tech. rep., University of Maryland, Baltimore County, 1999.

[14] M. C. Villalobos, R. A. Tapia, and Y. Zhang, *The Behavior of Newton-Type Methods on Two Equivalent Systems from Linear Programming*, Technical Report CRPC-TR98770-S, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA, 1998.

[15] ——, *The Sphere of Convergence of Newton's Method on Two Equivalent Systems from Nonlinear Programming*, Technical Report TR9913, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA, 1999.

[16] M. H. Wright, *Interior methods for constrained optimization*, Acta Numerica, 1 (1992), pp. 341–407.

[17] ——, *Why a pure primal Newton barrier step may be infeasible*, SIAM Journal on Optimization, 5 (1995), pp. 1–12.

[18] ——, *The Interior-Point Revolution in Constrained Optimization*, Technical Report 98-4-09, Computing Sciences Research Center, Bell Laboratories, Murray Hill, New Jersey 07974, June 1998.

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

[20] S. J. Wright and F. Jarre, *The role of linear objective functions in barrier methods*, Preprint MCS-P485-1294, Argonne National Laboratory, Argonne, Illinois, USA, April 1998.

[21] S. J. Wright and D. Orban, *Local Convergence of the Newton/Log-Barrier Method for Degenerate Problems*, Technical Report ANL/MCS-P772-0799, Argonne National Laboratory, Argonne, Illinois, USA, July 1999. Submitted to Mathematics of Operations Research.

[22] H. Yabe and H. Yamashita, *Q-superlinear convergence of primal-dual interior point quasi-Newton methods for constrained optimization*, Journal of the Operations Research Society of Japan, 40 (1997), pp. 415–436.

[23] H. Yamashita and H. Yabe, *Superlinear and quadratic convergence of some primal-dual interior point methods for constrained optimization*, Mathematical Programming, Series A, 75 (1996), pp. 377–397.