Laboratoire d'Arithmétique,
Calcul formel et d'Optimisation

UMR CNRS 6090

Interpretation of interior point methods as damped Newton methods

Paul Armand
Joël Benoist
Dominique Orban

Rapport de recherche n° 2005-05
Déposé le 28 septembre 2005

Université de Limoges, 123 avenue Albert Thomas, 87060 Limoges Cedex
Tél. (33) 5 55 45 73 23 - Fax. (33) 5 55 45 73 22 - laco@unilim.fr
http://www.unilim.fr/laco/

Université de Limoges, 123 avenue Albert Thomas, 87060 Limoges Cedex
Tél. (33) 5 55 45 73 23 - Fax. (33) 5 55 45 73 22 - laco@unilim.fr
http://www.unilim.fr/laco/

Interpretation of nonlinear interior methods as damped Newton methods

Paul Armand,[†] Joël Benoist[†] and Dominique Orban[‡]

Abstract. We propose a unified framework for the update of the barrier parameter in interior-point methods for nonlinear programming. The original primal-dual system is augmented to incorporate explicitly an updating function. We analyze local convergence properties and recover known updating strategies as special cases. We report numerical experiments on nonlinear problems and compare our results to a state-of-the-art interior-point implementation.

Key words. constrained optimization, interior point method, nonlinear programming, primal-dual method, barrier method

AMS subject classification. 65K05, 90C06, 90C26, 90C30, 90C51

1 Introduction

We consider nonlinear minimization problems of the form

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

where $f : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ are twice continuously differentiable. Let $(y, z) \in \mathbb{R}^m \times \mathbb{R}^n$ denote a vector of Lagrange multipliers associated to the constraints. Let $w = (x, y, z)$, $v = (x, z)$ and define $\ell(w) = f(x) + y^\top c(x) - z^\top x$ the Lagrangian function associated to problem (1.1). The first order optimality conditions of problem (1.1) can be written

$$F(w) = 0 \quad \text{and} \quad v \geq 0 \tag{1.2}$$

where $F : \mathbb{R}^{2n+m} \to \mathbb{R}^{2n+m}$ is defined by

$$F(w) = \begin{pmatrix} \nabla_x \ell(w) \\ c(x) \\ XZe \end{pmatrix}, \tag{1.3}$$

where $e = (1 \cdots 1)$ is the vector of all ones. We use X to denote the diagonal matrix $\operatorname{diag}(x_1, \ldots, x_n)$ and employ similar notation for other quantities.

Primal-dual interior point methods apply a Newton-type method to a sequence of perturbations of (1.2)

$$F(w) = \mu \tilde{e}, \tag{1.4}$$

[†]Université de Limoges (France) ; e-mail: `Paul.Armand@unilim.fr, Joel.Benoist@unilim.fr`.
[‡]Ecole Polytechnique de Montréal (Canada) ; e-mail: `Dominique.Orban@polymtl.ca`.

where $\tilde{e} = \begin{pmatrix} 0 & 0 & e^\top \end{pmatrix}^\top$ and $\mu > 0$ is held fixed at some value μ_k, to identify w_k meeting a stopping criterion of the form

$$\|F(w_k) - \mu_k \tilde{e}\| \le \varepsilon_k, \tag{1.5}$$

for some tolerance $\varepsilon_k > 0$, while enforcing $v_k > 0$. The process is referred to as the *inner* iteration. The values of μ_k and ε_k are next decreased to ensure that $\{\mu_k\} \downarrow 0$ and $\{\varepsilon_k\} \downarrow 0$ and a new series of inner iterations begins. The variable μ is called the barrier parameter, because (1.4) can be interpreted as the optimality conditions of the penalty barrier problem

$$\begin{cases} \min & f(x) - \mu \sum_{i=1}^{n} \log x_i, \\ \text{s.t.} & c(x) = 0. \end{cases} \tag{1.6}$$

Under standard assumptions on (1.1), as $\mu \downarrow 0$, the solutions $x(\mu)$ of (1.6) trace a smooth trajectory, called the central path, converging to a solution x^* of (1.1) [6].

With the intent of finding an appropriate starting point for the next sequence of inner iterations and to guarantee fast local convergence, an extrapolation step is often performed from w_k. One possible extrapolation step, originally devised in a purely primal framework, is obtained by linearizing the perturbed optimality conditions at (w_k, μ_k) [4]. It is now known that, in the primal-dual framework, this step is equivalent to a Newton step for (1.4) using the *new* value μ_{k+1} [8]

$$F'(w_k)d_k + F(w_k) = \mu_{k+1}\tilde{e}. \tag{1.7}$$

If w_k satisfies (1.5) and if μ_k and ε_k are appropriately updated, it can be shown that for sufficiently small values of μ_{k+1}, the step d_k given by (1.7) preserves positivity of v and that $w_k + d_k$ readily satisfies the stopping conditions of the new barrier subproblem. This implies that a single step is asymptotically sufficient per update of μ and results in Q-subquadratic local convergence [2, 8, 9] .

For schemes based on proximity to the central path as enforced by (1.5), under standard regularity conditions, the local convergence rate of $\{w_k\}$ to w^* is directly tied to the convergence rate of $\{\mu_k\}$ to zero. The positivity requirement $v > 0$ imposes that μ cannot decrease too fast in the sense that

$$\mu_{k+1} \ge \kappa \mu_k^{2-\epsilon}$$

for some constant $\kappa > 0$ and some $\epsilon > 0$ as small as desired. Various rules for updating μ can be found in the literature. They range from the simplest, but globally convergent, monotonic decreases of μ, rules that draw inspiration from linear programming to rules of a different nature which are based on a measure of centrality. Most algorithms decrease μ linearly and will attempt to recover fast local convergence by switching to a rule which decreases μ superlinearly as soon as it is declared sufficiently small [14]. However, the notion *sufficiently small* is problem dependend and it remains unclear how to wisely switch from one rule to the other. In [7], the authors use a rule which mimics rules based on the duality gap in linear and convex programming but enforces that $\{\mu_k\}$ be decreasing. Numerically, the monotonicity of $\{\mu_k\}$ can be a source of difficulties, or limitating, and more dynamic rules are investigated, often at the expense of global convergence. An example is given by [5] where the barrier parameter is allowed to increase occasionally. Recently, the authors of [11] incorporated such a rule into a safeguarded globally convergent framework with a heuristic switch. At variance, the authors of [12] note the importance of keeping individual complementarity pairs clustered. They measure deviation from centrality and update μ accordingly in a manner similar to [10]. Unfortunately, this practice is not warranted by any global convergence theory and can cause failure if μ_k diverges or becomes too small prematurely.

In this paper, we propose a new interpretation of interior point iterations and a framework that unifies updating rules for the barrier parameter. Consider the system

$$\begin{pmatrix} F(w) - \mu\tilde{e} \\ \theta(\mu) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{1.8}$$

where θ is a scalar function defined on $\mathbb{R}_+$, derivable, increasing, strictly convex and such that $\theta(0) = 0$. Clearly, the vector (w^*, μ^*) is a solution of (1.8) if and only if $\mu^* = 0$ and $F(w^*) = 0$. To solve (1.2), the algorithm scheme that we consider, is based on the application of a sequence of Newton iterations to (1.8), while maintaining positivity of v. A Newton step from (w_k, μ_k) is the solution of the linear system

$$\begin{pmatrix} F'(w_k) & -\tilde{e} \\ 0 & \theta'(\mu_k) \end{pmatrix} \begin{pmatrix} d_k^w \\ d_k^\mu \end{pmatrix} + \begin{pmatrix} F(w_k) - \mu_k\tilde{e} \\ \theta(\mu_k) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{1.9}$$

The first equation of (1.9) may be rewritten

$$F'(w_k)d_k^w + F(w_k) = (\mu_k + d_k^\mu)\tilde{e}, \tag{1.10}$$

which is just the system (1.7) with $\mu_{k+1} = \mu_k + d_k^\mu$. The value of d_k^μ is obtained from the second equation of (1.9)

$$d_k^\mu = -\frac{\theta(\mu_k)}{\theta'(\mu_k)}. \tag{1.11}$$

This point of view shows that both the decrease of the barrier parameter and the extrapolation step can be interpreted as only one Newton step.

A first consequence of this formulation is that the tools of the convergence analysis of Newton's method can be directly applied to a local convergence analysis of the interior point algorithm. This point of view is developed in Section 5.

A second consequence is that a control of the decrease of the barrier parameter can be introduced in a natural manner; decreasing μ rather conservatively when far from a solution where the extrapolation step can be a poor direction of improvement, and more enthusiastically when closer to a local solution. We explore this possibility in a linesearch framework. We study the introduction of the steplength computed by a linesearch, in the update formula of the barrier parameter. A first candidate steplength is given by the so-called *fraction to the boundary* rule: the largest $\alpha \in (0, 1]$ ensuring

$$v_k + \alpha d_k^v \geq (1 - \tau_k) v_k, \tag{1.12}$$

where $\tau_k \in (0, 1)$. A natural choice for the next value of the barrier parameter is then

$$\mu_{k+1} = \mu_k + \alpha_k d_k^\mu.$$

A second candidate steplength is given by a sufficient decrease condition on a merit function. At last, we explore the possibility of updating μ at each inner iteration. This results in an algorithm which is not warranted by a global convergence theory but we hope to illustrate however that it possesses some stabilizing properties. The benefits of damping the update of μ are illustrated by some numerical experiments in Section 8.

The paper is structured as follows. Section 2 describes the notation and assumptions used throughout the text. Section 3 gives general properties on the sequence defined by (1.11) under various assumptions of the function θ. Section 4 presents a number of preliminary results concerning local solutions to equation (1.4) and properties of a Newton iteration. Section 5 studies local convergence properties of the sequences defined by (1.9) but ignores the positivity constraints. The latter are addressed in Section 6 where conditions on the tolerance ε_k and on the decreasing rate of μ_k are derived to ensure that a single inner iteration is asymptotically sufficient. Finally, Section 7 presents three algorithm schemes and gives a general convergence result. Results of some numerical experiments are discussed in Section 8.

2 Notation and assumptions

Vector inequalities are understood componentwise. Given two vectors $x, y \in \mathbb{R}^n$, their Euclidean scalar product is denoted by $x^\top y$ and the associated ℓ_2 norm is $\|x\| = (x^\top x)^{1/2}$. The open Euclidian ball centered at x with radius $r > 0$ is denoted by $B(x, r)$, that is $B(x, r) := \{y : \|x - y\| < r\}$.

For two nonnegative scalar sequences $\{a_k\}$ and $\{b_k\}$ converging to zero, we use the Landau symbols $a_k = o(b_k)$ if $\lim_{k\to\infty} a_k/b_k = 0$ and $a_k = O(b_k)$ if there exists a constant $c > 0$, such that $a_k \leq c b_k$ for all sufficiently large k. We use similar symbols with vector arguments, in which case they are understood normwise. We write $a_k = \Theta(b_k)$ if both $a_k = O(b_k)$ and $b_k = O(a_k)$ hold. We also write $a_k \sim b_k$ if $\lim_{k\to\infty} a_k/b_k = 1$. For a positive function $\epsilon : \mathbb{R}_+ \to \mathbb{R}_+$, we write $\epsilon(t) = o(t)$ if $\lim_{t\downarrow 0} \epsilon(t)/t = 0$, $\epsilon(t) = O(t)$ if $\limsup_{t\downarrow 0} \epsilon(t)/t < +\infty$ and $\epsilon(t) = \Theta(t)$ if both $0 < \liminf_{t\downarrow 0} \epsilon(t)/t$ and $\limsup_{t\downarrow 0} \epsilon(t)/t < +\infty$ hold.

We denote by e_i the vectors of the canonical basis of $\mathbb{R}^n$. For all $x \in \mathbb{R}^n$, $\nabla c(x)$ denotes the transpose of the Jacobian matrix of c at x, i.e., the $n \times m$ matrix whose i-th column is $\nabla c_i(x)$. Let $x^* \in \mathbb{R}^n$ be a local solution of problem (1.1). Let $\mathcal{A} := \{i : x_i^* = 0\}$ be the set of indices of active inequality constraints. Throughout the paper, we assume that the following assumptions are satisfied.

Assumption 2.1 The functions f and c are twice continuously differentiable over an open neighborhood of x^*;

Assumption 2.2 The linear independence constraint qualification holds at x^*, i.e., $\{\nabla c_i(x^*),\ i = 1, \dots, m\} \cup \{e_i,\ i \in \mathcal{A}\}$ is a linearly independent set of vectors.

Note that Assumptions 2.1 and 2.2 imply that there exists a strictly feasible point, i.e., $\bar{x} \in \mathbb{R}$ such that $c(\bar{x}) = 0$ and $\bar{x} > 0$, and that there exists a unique vector of Lagrange multipliers $(y^*, z^*) \in \mathbb{R}^{m+n}$ such that $w^* = (x^*, y^*, z^*)$ is a solution of (1.2).

Assumption 2.3 The strong second-order sufficiency condition is satisfied at w^*, i.e., $u^\top \nabla^2_{xx} \ell(w^*) u > 0$ for all $u \neq 0$ satisfying $\nabla c(x^*)^\top u = 0$ and $u_i = 0$ for all $i \in \mathcal{A}$.

Assumption 2.4 Strict complementarity holds at w^*, that is

$$\min\{x_i^* + z_i^* : i = 1 \dots n\} > 0.$$

Under Assumptions 2.1-2.4, the Jacobian of F, defined by

$$F'(w) = \begin{pmatrix} \nabla^2_{xx}\ell(w) & \nabla c(x) & -I \\ \nabla c(x)^\top & 0 & 0 \\ Z & 0 & X \end{pmatrix},$$

is uniformly nonsingular over a neighbourhood of w^*. This fact will allow us to state strong results on the solution to (1.4) as $\mu \downarrow 0$.

3 The function θ and the update of μ

In this section we give some examples of function θ that can be used in equation (1.8) and establish a connection between θ and the sequence $\{\mu_k\}$ defined by $\mu_{k+1} = \mu_k + d_k^\mu$ where d_k^μ is given by (1.11). We start by recalling our basic assumption on θ.

Assumption 3.1 The function $\theta : \mathbb{R}_+ \to \mathbb{R}$ is continuously derivable, increasing, strictly convex and such that $\theta(0) = 0$.

Some properties of the sequence $\{\mu_k\}$ implied by the nature of the function θ are summarized in the next result.

Proposition 3.2 *Suppose that Assumption 3.1 holds. Let the sequence defined by* $\mu_0 > 0$ *and*

$$\mu_{k+1} = \mu_k - \frac{\theta(\mu_k)}{\theta'(\mu_k)}. \qquad (3.1)$$

(*i*) *The sequence* $\{\mu_k\}$ *is decreasing and converges to zero.*

(*ii*) *Suppose that* θ *is* p *times derivable, with* $\theta'(0) = \ldots = \theta^{(p-1)}(0) = 0$ *and* $\theta^{(p)}(0) \neq 0$ *(possibly* ∞*). Suppose also that the following limit exists:*

$$a := \lim_{t \downarrow 0} \frac{t\theta^{(p)}(t)}{\theta^{(p-1)}(t)}.$$

Then $\{\mu_k\}$ *converges linearly with convergence ratio* $1 - \frac{1}{p-1+a}$.

(*iii*) *If* $\theta'(0) > 0$ *then* $\{\mu_k\}$ *converges q-superlinearly to zero.*

(*iv*) *If* θ *is twice derivable on* $(0, \epsilon)$ *for some* $\epsilon > 0$ *and if* $\lim_{t \downarrow 0} \theta''(t) = +\infty$, *then* $\mu_k^2 = o(\mu_{k+1})$.

Proof. It follows from Assumption 3.1 that $\theta(\mu) > 0$ and $\theta'(\mu) > 0$ for all $\mu > 0$, therefore the sequence $\{\mu_k\}$ is decreasing. Using the strict convexity of θ, one has

$$\mu_{k+1}\theta'(\mu_k) = \theta'(\mu_k)\mu_k - \theta(\mu_k) > -\theta(0) = 0.$$

By induction, we then obtain $\mu_k > 0$ for all k and thus the sequence $\{\mu_k\}$ converges. Taking the limit in (3.1) and using the fact that zero is the unique solution of $\theta(\mu) = 0$, we deduce that $\{\mu_k\}$ converges to zero, which proves (*i*).

Point (*ii*) is proved by using $(p-1)$ times the extended mean value theorem. There exists $0 < \zeta_k < \mu_k$ such that

$$\begin{aligned}
\frac{\mu_{k+1}}{\mu_k} &= 1 - \frac{\theta(\mu_k)}{\mu_k\theta'(\mu_k)} \\
&= 1 - \frac{\theta^{(p-1)}(\zeta_k)}{(p-1)\theta^{(p-1)}(\zeta_k) + \zeta_k\theta^{(p)}(\zeta_k)} \\
&= 1 - \frac{1}{(p-1) + \zeta_k\theta^{(p)}(\zeta_k)/\theta^{(p-1)}(\zeta_k)}.
\end{aligned}$$

Taking the limit $k \to \infty$, we obtain the result.

Point (*iii*) follows from $\theta(\mu_k) = \theta'(\mu_k)\mu_k + o(\mu_k)$ and taking the limit in

$$\frac{\mu_{k+1}}{\mu_k} = 1 - \frac{\theta(\mu_k)/\mu_k}{\theta'(\mu_k)}.$$

Using the Taylor-Lagrange formula, for all $k \geq 0$, there exists $\zeta_k \in (0, \mu_k)$ such that

$$0 = \theta(0) = \theta(\mu_k) - \theta'(\mu_k)\mu_k + \frac{1}{2}\theta''(\zeta_k)\mu_k^2.$$

It follows that $\mu_k^2/\mu_{k+1} = 2\theta'(\mu_k)/\theta''(\zeta_k)$, therfore $\mu_k^2 = o(\mu_{k+1})$, which proves (iv). $\square$

A first instance of θ is given by

$$\theta_{\mathrm{L}}(\mu) = \mu^\sigma \quad \text{for } \mu \geq 0, \tag{3.2}$$

for some constant $\sigma > 1$. The barrier parameter update is given by

$$\mu_{k+1} = (1 - \frac{1}{\sigma})\mu_k.$$

The decrease of μ to zero is linear with ratio $1 - \frac{1}{\sigma}$. Note that the rate of convergence can also be deduced from part (ii) of Proposition 3.2 with $p - 1 = \lfloor \sigma \rfloor$.

A second example is given by

$$\theta_{\mathrm{S}}(\mu) = \frac{\mu}{(b^\gamma - \mu^\gamma)^{\frac{1}{\gamma}}} \quad \text{for } \mu \in [0, b), \tag{3.3}$$

for some parameters $b > 0$ and $0 < \gamma < 1$. It is not difficult to verify that this function is strongly convex and that $\theta_{\mathrm{S}}'(0) = \frac{1}{b}$. The update formula is given by

$$\mu_{k+1} = \frac{1}{b^\gamma}\mu_k^{1+\gamma},$$

which indicates that the rate of convergence of μ_k to zero is q-superlinear. Such a rule is used in, e.g., [2, p. 52]. Note that the rate of convergence is not quadratic, as implied by part (iv) of Proposition 3.2.

A third example of function θ can be given by imposing local Lipschitz continuity of θ' at zero. For example,

$$\theta_{\mathrm{Q}}(\mu) = \mu^\sigma + \beta\mu, \quad \text{for } \mu \geq 0, \tag{3.4}$$

for some parameters $\sigma \geq 2$ and $\beta > 0$. In that case, the rate of convergence of the barrier parameter is quadratic.

4 Preliminary results

In this section we give some useful results about the solution of equation (1.4).

Lemma 4.1 *Under Assumptions 2.1–2.2, the Jacobian F' is locally Lipschitzian about w^*, i.e., there exist $\delta_L > 0$ and $L > 0$ such that, for all $w_1, w_2 \in B(w^*, \delta_L)$,*

$$\|F'(w_1) - F'(w_2)\| \leq L\|w_1 - w_2\|.$$

The newt two lemmas are consequences of the nonsingularity of $F'(w^*)$ (see e.g.[6]).

Lemma 4.2 *Under Assumptions 2.1–2.4, there exist* $\delta_K > 0$ *and* $K > 0$ *such that, for all* $w \in B(w^*, \delta_K)$ *the Jacobian* $F'(w)$ *is nonsingular and*

$$\frac{1}{K} \leq \|F'(w)^{-1}\| \leq K.$$

Lemma 4.3 *Under Assumptions 2.1–2.4, there exist* $\delta_C > 0$, $\mu_C > 0$ *and a continuously differentiable function* $w(\cdot) : (-\mu_C, \mu_C) \to \mathbb{R}^{n+m+n}$ *such that, for all* $w \in B(w^*, \delta_C)$ *and* $|\mu| < \mu_C$,

$$F(w) = \mu\tilde{e} \quad \textit{if and only if} \quad w(\mu) = w.$$

We have the expansion

$$w(\mu) = w^* + w'(0)\mu + o(\mu)$$

where the tangent vector to the central path $w(\cdot)$ *at* $\mu = 0$ *is given by* $w'(0) = F'(w^*)^{-1}\tilde{e} \neq 0$. *As a consequence, there exists* $C > 0$, *such that for all* μ_1, $\mu_2 \in (-\mu_C, \mu_C)$,

$$\|w(\mu_1) - w(\mu_2)\| \leq C|\mu_1 - \mu_2|.$$

The following result states a property of the central path and its tangent vector. It is a consequence of the strict complementarity assumption.

Lemma 4.4 *Suppose that Assumptions 2.1–2.4 hold. Let* $w(\cdot)$ *be the implicit function defined in Lemma 4.3. Let* $w(\mu) := (x(\mu), y(\mu), z(\mu))$, *then*

$$x_i^* z_i'(0) + x_i'(0) z_i^* = 1. \tag{4.1}$$

Proof. By virtue of Lemma 4.3, for all $i = 1, \ldots, n$, we have

$$x_i(\mu) = x_i^* + \mu x_i'(0) + o(\mu) \quad \text{and} \quad z_i(\mu) = z_i^* + \mu z_i'(0) + o(\mu).$$

By using $x_i^* z_i^* = 0$ and $F(w(\mu)) = \mu\tilde{e}$, we deduce that

$$x_i(\mu) z_i(\mu) = (x_i^* z_i'(0) + x_i'(0) z_i^*)\mu + o(\mu) = \mu.$$

Diving both sides by μ and taking the limit $\mu \to 0$, we obtain (4.1). □

The last two lemmas analyse some basic properties of the Newton step generated by the primal-dual method.

Lemma 4.5 *Under Assumptions 2.1–2.4, there exist* $\delta^* > 0$, $\mu^* > 0$ *and* $M_1 > 0$ *such that, for all* $w \in B(w^*, \delta^*)$ *and* $|\mu| < \mu^*$, *the Newton iterate, defined by*

$$N(w,\mu) = w - F'(w)^{-1}(F(w) - \mu\tilde{e}),$$

satisfies

$$\|N(w,\mu) - w(\mu)\| \leq M_1 \|w - w(\mu)\|^2. \tag{4.2}$$

and

$$\|N(w,\mu) - w^*\| \leq \frac{1}{2}\|w - w^*\| + \frac{\delta^*}{2\mu^*}\mu. \tag{4.3}$$

In particular, $N(w,\mu) \in B(w^*, \delta^*)$.

Proof. Let us define

$$M_1 := \frac{1}{2}KL,\ \delta^* := \min\{\delta_L, \delta_K, \delta_C, \frac{1}{4M_1}\} \text{ and } \mu^* := \min\{1, \mu_C, \frac{\delta^*}{2C(2M_1C+1)}\},$$

where the constants are defined in Lemmas 4.1-4.3.

Let $(w,\mu) \in B(w^*, \delta^*) \times (-\mu^*, \mu^*)$. By Lemma 4.3, one has $F(w(\mu)) - \mu\tilde{e} = 0$. It follows that

$$\begin{aligned}
&N(w,\mu) - w(\mu) \\
&\quad = \quad w - w(\mu) - F'(w)^{-1}(F(w) - \mu\tilde{e}) \\
&\quad = \quad -F'(w)^{-1}(F(w) - \mu\tilde{e} - F'(w)(w - w(\mu))) \\
&\quad = \quad -F'(w)^{-1}\int_0^1 (F'(w(\mu) + t(w - w(\mu))) - F'(w))(w - w(\mu))\,\mathrm{d}t.
\end{aligned}$$

Taking the norm on both sides, using the Cauchy-Schwarz inequality and Lemmas 4.1 and 4.2, we obtain (4.2).

Using (4.2), Lemma 4.3 and the definition of the constants M_1, δ^* and μ^*, we have

$$\begin{aligned}
\|N(w,\mu) - w^*\| &\leq \|N(w,\mu) - w(\mu)\| + \|w(\mu) - w^*\| \\
&\leq M_1\|w - w(\mu)\|^2 + \|w(\mu) - w^*\| \\
&\leq 2M_1(\|w - w^*\|^2 + \|w(\mu) - w^*\|^2) + \|w(\mu) - w^*\| \\
&\leq 2M_1\delta^*\|w - w^*\| + (2M_1C\mu + 1)C\mu \\
&\leq \frac{1}{2}\|w - w^*\| + \frac{\delta^*}{2\mu^*}\mu \\
&< \delta^*,
\end{aligned}$$

which proves (4.3) and concludes the proof. □

Lemma 4.6 *Suppose that Assumptions 2.1–2.4 hold. Let δ^* and μ^* be the radii of convergence defined in Lemma 4.5. Then, there exists $M_2 > 0$ such that, for all $w \in B(w^*, \delta^*)$ and $0 \le \mu^+ \le \mu < \mu^*$, the Newton iterate*

$$N(w, \mu^+) = w - F'(w)^{-1}(F(w) - \mu^+ \tilde{e}),$$

satisfies

$$\|N(w, \mu^+) - w(\mu^+)\| \le M_2(\|w - w(\mu)\|^2 + \mu^2). \tag{4.4}$$

Proof. Let C, M_1, δ^* and μ^* be the constants defined in Lemmas 4.3 and 4.5. Using Lemmas 4.3 and 4.5, one has

$$\begin{aligned} \|N(w, \mu^+) - w(\mu^+)\| &\le M_1\|w - w(\mu^+)\|^2 \\ &\le M_1(\|w - w(\mu)\| + \|w(\mu) - w(\mu^+)\|)^2 \\ &\le M_1(\|w - w(\mu)\| + C(\mu - \mu^+))^2 \\ &\le 2M_1 \max\{1, C^2\}(\|w - w(\mu)\|^2 + \mu^2). \end{aligned}$$

The result follows with $M_2 = 2M_1 \max\{1, C^2\}$. □

5 Local convergence analysis

The first result is a direct consequence of the convergence properties of Newton's method. Note that it is not assumed that the iterates remain strictly feasible during the iterations (i.e., $v_k \not> 0$). Note also that the results are stated without any assumption on the proximity of the iterates to the central path.

Theorem 5.1 *Suppose that Assumptions 2.1–2.4 and 3.1 hold. Consider the sequences defined by the recurrence*

$$\mu_{k+1} = \mu_k + d_k^\mu \quad \textit{and} \quad w_{k+1} = w_k + d_k^w,$$

where d_k^μ is defined by (1.11) and d_k^w is the solution of (1.10). Let δ^* and μ^* be the threshold values defined in Lemma 4.5. For all $w_0 \in B(w^*, \delta^*)$ and all $0 < \mu_0 < \mu^*$, the sequence $\{\mu_k\}$ is decreasing and the sequence $\{(w_k, \mu_k)\}$ converges to $(w^*, 0)$. Moreover, if $\theta'(0) \neq 0$, then the rate of convergence of both sequences $\{(w_k, \mu_k)\}$ and $\{\mu_k\}$ is q-superlinear. In addition, if θ' is locally Lipschitzian at 0, the rate of convergence of both sequences is quadratic.

Proof. The convergence of the sequence $\{\mu_k\}$ follows from outcome (i) of Proposition 3.2. From (4.3) we have

$$\|w_{k+1} - w^*\| \le \frac{1}{2}\|w_k - w^*\| + \frac{\delta^*}{2\mu^*}\mu_{k+1},$$

and therefore $w_k \to w^*$ as $k \to \infty$.

Suppose that $\theta'(0) \neq 0$. By Lemmas 4.1 and 4.2 the Jacobian matrix

$$\begin{pmatrix} F'(w) & -\tilde{e} \\ 0 & \theta'(\mu) \end{pmatrix}$$

is nonsingular and locally Lipschitzian at $(w^*, 0)$. The sequence $\{(w_k, \mu_k)\}$ is then generated by the Newton method with a step defined by (1.9). The convergence and the rates of convergence of the sequences follow from the local convergence properties of the Newton method (see for example [1, Proposition 1.17]). □

Theorem 5.1 shows that, whenever $\theta'(0) \neq 0$, the sequence $\{w_k\}$ converges to w^* with a r-superlinear or r-quadratic rate of convergence, depending on whether θ' is locally Lipschitzian at zero or not. The next result shows that, provided that $\{\mu_k\}$ does not converge too fast to zero, both sequences $\{w_k\}$ and $\{\mu_k\}$ have the same rate of convergence.

Theorem 5.2 *Suppose that Assumptions 2.1–2.4 hold. Let δ^* and μ^* be the threshold values defined in Lemma 4.5. Suppose that $\{\mu_k\}$ is a sequence of positive scalars converging to zero such that $0 < \mu_0 < \mu^*$ and*

$$\liminf_{k\to\infty} \frac{\mu_{k+1}}{\mu_k^p} > 0, \tag{5.1}$$

for some parameter $p \in (1, 2)$. Consider the sequence defined by the recurrence

$$w_{k+1} = w_k + d_k^w,$$

where d_k^w is the solution of (1.10) with d_k^μ defined by $d_k^\mu = \mu_{k+1} - \mu_k$. Then for all $w_0 \in B(w^*, \delta^*)$, the sequence $\{w_k\}$ converges to w^* and

$$w_k = w(\mu_k) + o(\mu_k). \tag{5.2}$$

In particular, $v_k > 0$ for sufficiently large k and both sequences $\{w_k\}$ and $\{\mu_k\}$ have the same rate of convergence.

Proof. As in the proof of Theorem 5.1, from (4.3) we have

$$\|w_{k+1} - w^*\| \leq \frac{1}{2}\|w_k - w^*\| + \frac{\delta^*}{2\mu^*}\mu_{k+1},$$

and thus $w_k \to w^*$ when $k \to \infty$.

Let us prove now that (5.2) holds. Let M_2 be the constant defined in Lemma 4.6. Let us define $s_k := M_2\|w_k - w(\mu_k)\|$ and $\epsilon_k := M_2\mu_k$. Condition (5.1) implies that $\mu_k^2 = o(\mu_{k+1})$ and that there exist $c > 0$, $\beta \in (0, 1)$ and $\hat{k}$ such that

$$c\beta^{p^k} \leq \mu_k \quad \text{for all } k \geq \hat{k} \tag{5.3}$$

(see for example [1, Proposition 1.2]). From (4.4), one has

$$s_{k+1} \le s_k^2 + \epsilon_k^2. \tag{5.4}$$

Since $\epsilon_k^2 = o(\epsilon_{k+1})$, there exists an index $\bar{k} \ge 0$ such that $\epsilon_k^2 \le \frac{1}{2}\epsilon_{k+1}$ for all $k \ge \bar{k}$. Let us prove that there exists $k_0 \ge \bar{k}$ such that $s_{k_0} \le \epsilon_{k_0}$. By contradiction, suppose that $s_k > \epsilon_k$ for all $k \ge \bar{k}$. From (5.4) we would deduce that $s_{k+1} \le 2s_k^2$, therefore $\{s_k\}$ would converge quadratically to zero, which should imply that $s_k = O(\alpha^{2^k})$ for some $\alpha > 0$. Since we have supposed that $s_k > \epsilon_k$, we would then have $\epsilon_k = O(\alpha^{2^k})$, a contradiction with (5.3) knowing that $1 < p < 2$.

Let us prove now by induction on k that $s_k \le \epsilon_k$ for $k \ge k_0$. Our claim holds for $k = k_0$. Assume that it is true for a given $k \ge k_0$. By (5.4) and the induction hypothesis one has $s_{k+1} \le 2\epsilon_k^2 \le \epsilon_{k+1}$, and thus our claim is also true for $k+1$.

From (5.4), we can then deduce that $s_{k+1} \le 2\epsilon_k^2 = o(\epsilon_{k+1})$, which implies (5.2) and proves the first part of the theorem.

Let us show that $v_k > 0$ for sufficiently large k. From (5.2) and Lemma 4.3 we have

$$(x_k)_i = x_i^* + x_i'(0)\mu_k + o(\mu_k),$$

for all index i. By (4.1) we have either $x_i^* > 0$ or $x_i'(0) > 0$. It follows that $(x_k)_i > 0$ for sufficiently large k. The reasoning is similar for the components $(z_k)_i$.

At last, using (5.2) and Lemma 4.3, we have $w_k - w^* = w_k - w(\mu_k) + w(\mu_k) - w^* = w'(0)\mu_k + o(\mu_k)$, and thus $\|w_k - w^*\| \sim \|w'(0)\|\mu_k$. □

It is interesting to note that, at variance with other local convergence theorems, Theorem 5.2 does assumes neither strict feasibility of the iterates—w_0 might very well have a component $v_0 \not> 0$—nor proximity of the iterates to the central path. Strict feasibility and proximity however follow, as specified by (5.2), from the sub-quadratic convergence of the barrier parameter to zero. The iterates w_k asymptotically approach the central path tangentially since from (5.2) and Lemma 4.3,

$$\lim_{k\to\infty} \frac{w_k - w^*}{\mu_k} = \lim_{k\to\infty} \frac{w(\mu_k) - w^*}{\mu_k} = w'(0).$$

Asymptotic results having a similar flavour are given in [5, 13, 15, 16, 17] where μ is allowed to decrease quadratically or faster and where proximity to the central path is not enforced and feasibility is eventually not satisfied (the stabilized primal-dual method in [13] may produce iterates with nonpositive components). Quadratic convergence follows but takes place in a restrictive neighbourhood around w^*.

6 Maintaining the positivity and the proximity to the central path

Suppose that at the current iteration, the iterates w and μ satisfy the approximate optimality condition

$$\|F(w) - \mu\tilde{e}\| \leq \varepsilon(\mu), \tag{6.1}$$

for some tolerance $\varepsilon(\mu) > 0$. The Newton iterate for (1.8) is

$$w^+ = w - F'(w)^{-1}(F(w) - \mu^+\tilde{e}), \tag{6.2}$$

where $\mu^+ = \mu - \theta(\mu)/\theta'(\mu)$. We propose to analyse the conditions under which w^+ will satisfy the fraction to the boundary rule and the next stopping test, that is

$$v^+ \geq (1 - \tau(\mu))v, \tag{6.3}$$

for some parameter $\tau(\mu) \in (0, 1)$ and

$$\|F(w^+) - \mu^+\tilde{e}\| \leq \varepsilon(\mu^+). \tag{6.4}$$

The analysis proposed below uses some tools developed in [2].

Lemma 6.1 *Suppose that Assumptions 2.1–2.4 hold. Suppose also that $\varepsilon(\mu) = O(\mu)$. Let δ^* be defined in Lemma 4.5. Then, there exists $M_3 > 0$ such that, for all $w \in B(w^*, \delta^*)$ satisfying (6.1) and all $\mu > 0$ sufficiently small,*

$$\|w - w^*\| \leq M_3\mu$$

Proof. Let $w \in B(w^*, \delta^*)$. Since $F(w^*) = 0$ and $F'(w^*)$ is invertible (Lemma 4.2), we have

$$w - w^* = F'(w^*)^{-1}(F(w) - \int_0^1 (F'(w^* + t(w - w^*)) - F'(w^*))(w - w^*)\,\mathrm{d}t).$$

Taking the norm on both sides, using the Cauchy-Schwarz inequality and Lemmas 4.1 and 4.2, we obtain

$$\|w - w^*\| \leq K(\|F(w)\| + \frac{L}{2}\|w - w^*\|^2).$$

Recall that $\delta^* \leq \frac{1}{2KL}$ (see the definition of δ^* in the proof of Lemma 4.5), it follows that

$$\|w - w^*\| \leq \frac{4}{3}K\|F(w)\|.$$

Now using $\|\tilde{e}\| = \sqrt{n}$, (6.1) and $\varepsilon(\mu) = O(\mu)$, we then have

$$\begin{aligned}\|w - w^*\| &\leq \frac{4}{3}K(\|F(w) - \mu\tilde{e}\| + \sqrt{n}\mu) \\ &\leq \frac{4}{3}K(\varepsilon(\mu) + \sqrt{n}\mu) \\ &\leq M_3\mu.\end{aligned}$$

□

Lemma 6.2 *Suppose that Assumptions 2.1–2.4 hold. Suppose also that* $\varepsilon(\mu) = O(\mu)$. *Let* δ^* *and* μ^* *be defined in Lemma 4.5. Then, there exists* $M_4 > 0$ *such that, for all* $w \in B(w^*, \delta^*)$ *satisfying (6.1) and all* $0 < \mu^+ \leq \mu$ *sufficiently small, the Newton iterate* w^+ *defined by (6.2) satisfies*

$$\|w^+ - w(\mu^+)\| \leq M_4\mu^2. \tag{6.5}$$

Proof. By using Lemmas 6.1 and 4.3 we obtain

$$\begin{aligned}\|w - w(\mu)\| &\leq \|w - w^*\| + \|w^* - w(\mu)\| \\ &\leq (M_3 + C)\mu\end{aligned}$$

Using this last inequality in (4.4), we obtain (6.5) with a constant M_4 defined by $M_4 = M_2((M_3 + C)^2 + 1)$. □

Regarding positivity of the iterates, we distinguish the case where no fraction to the boundary rule is enforced, to the case where an explicit fraction to the boundary rule is present.

Theorem 6.3 *Suppose that Assumptions 2.1–2.4 hold. Suppose also that the parameters* μ^+ *and* $\varepsilon(\mu)$ *are chosen such that*

$$\mu^2 = o(\mu^+) \quad \text{and} \quad \varepsilon(\mu) = O(\mu).$$

Let δ^* *be defined in Lemma 4.5. Then for all* $w \in B(w^*, \delta^*)$ *satisfying (6.1) and such that* $v > 0$, *the Newton iterate* w^+ *defined by (6.2) is such that* $v^+ > 0$ *and* $F(w^+) - \mu^+\tilde{e} = o(\mu^+)$ *for sufficiently small* $\mu > 0$. *Moreover, if* $\tau(\mu)$ *satisfies*

$$\limsup_{\mu \downarrow 0} (1 - \tau(\mu))\frac{\mu + \varepsilon(\mu)}{\mu^+} < 1, \tag{6.6}$$

then (6.3) is satisfied for sufficiently small $\mu > 0$. *In addition, if* $\varepsilon(\mu) = \Theta(\mu)$, *then (6.4) is satisfied for sufficiently small* $\mu > 0$.

Proof. Let $0 < \kappa < 1$. We begin by proving that, for all index i

$$x_i^+ \geq \kappa \frac{\mu^+}{\mu + \varepsilon(\mu)} x_i.$$

Consider first the case where $x_i^* > 0$. Lemmas 6.1 and 6.2 imply that by taking $\mu > 0$ sufficiently small, w and w^+ can be arbitrarily close to w^*. It follows that

$$\lim_{\mu \downarrow 0} \frac{x_i^+}{x_i} = \frac{x_i^*}{x_i^*} = 1.$$

Using $\mu^+ < \mu + \varepsilon(\mu)$ and $0 < \kappa < 1$ we deduce that

$$\frac{x_i^+}{x_i} \geq \kappa \frac{\mu^+}{\mu + \varepsilon(\mu)}, \tag{6.7}$$

for sufficiently small $\mu > 0$.

Consider now the case where $x_i^* = 0$. By Lemma 4.4, we have $x_i'(0) > 0$. By using Lemma 6.2, Lemma 4.3 and $\mu^2 = o(\mu^+)$, we have

$$\begin{aligned} x_i^+ &= (x_i^+ - x_i(\mu^+)) + x_i(\mu^+) \\ &\geq -M_4\mu^2 + x_i'(0)\mu^+ + o(\mu^+) \\ &\geq x_i'(0)\mu^+ + o(\mu^+). \end{aligned}$$

Dividing both sides by x_i we obtain

$$\frac{x_i^+}{x_i} \geq \frac{x_i'(0)\mu^+ + o(\mu^+)}{x_i}.$$

From the definition of F and (6.1), we deduce that $x_i z_i \leq \mu + \varepsilon(\mu)$. Multiplying and dividing the previous equation by z_i yields

$$\frac{x_i^+}{x_i} \geq \frac{\mu^+}{\mu + \varepsilon(\mu)} (x_i'(0) z_i + o(1)).$$

Using Lemma 4.4, the term in parentheses tends to $x_i'(0) z_i^* = 1 - x_i^* z_i'(0) = 1$ as μ goes to zero. We then deduce that (6.7) also holds for sufficiently small $\mu > 0$ and for all i such that $x_i^* = 0$.

By the symmetric role of the variables x and z, the reasoning is the same for the components z_i. We then have

$$v^+ \geq \kappa \frac{\mu^+}{\mu + \varepsilon(\mu)} v. \tag{6.8}$$

Since we have supposed that $v > 0$, (6.8) implies that $v^+ > 0$.

If condition (6.6) is satisfied, it implies that there exists $\kappa \in (0,1)$ such that, for sufficiently small $\mu > 0$, we have

$$1 - \tau(\mu) \leq \kappa \frac{\mu^+}{\mu + \varepsilon(\mu)}.$$

From this last inequality and (6.8) we deduce that (6.3) holds.

Turning now to the residual at the updated iterate, by using Lemmas 4.1 and 6.2 and $\mu^2 = o(\mu^+)$, we have

$$\begin{aligned} \|F(w^+) - \mu^+ \tilde{e}\| &= \|F(w^+) - F(w(\mu^+)\| \\ &\leq L\|w^+ - w(\mu^+)\| \\ &\leq LM_4\mu^2 \\ &= o(\mu^+). \end{aligned}$$

If $\varepsilon(\mu) = \Theta(\mu)$, this last equation implies that for sufficiently small μ, (6.4) is satisfied. □

7 Practical algorithms

In this section, we are concerned with ways to incorporate the ideas presented above into practical algorithms. We first present two algorithms which embed steps in (w, μ) into a globally-convergent framework and give a convergence theorem. Next, we give a third algorithm, whose global convergence has not been established, but which is conceptually closer to intuition and possessed desirable numerical properties.

We choose a scalar function θ satisfying Asssumption 3.1 and for all $\mu > 0$ we define

$$\mu^+ = \mu - \frac{\theta(\mu)}{\theta'(\mu)}.$$

We select two tolerance functions $\varepsilon : \mathbb{R}_+ \to \mathbb{R}_+$ and $\tau : \mathbb{R}_+ \to (0,1)$, such that

$$\mu^2 = o(\mu^+), \quad \varepsilon(\mu) = \Theta(\mu) \quad \text{and} \quad \limsup_{\mu \downarrow 0} \, (1 - \tau(\mu)) \frac{\mu + \varepsilon(\mu)}{\mu^+} < 1. \tag{7.1}$$

The function θ given by examples (3.2) and (3.3) meet the requirement $\mu^2 = o(\mu^+)$, but (3.4) does not, for in that case $\mu^+ = O(\mu^2)$. The other conditions in (7.1) are satisfied if, for instance, we choose $\varepsilon(\mu) = \rho\mu$ for some parameter $\rho > 0$ and $\tau(\mu) = \max\{\bar{\tau}, 1 - \mu\}$ for some $0 < \bar{\tau} < 1$. We describe our first algorithm as Algorithm 7.1.

Algorithm 7.1

Given an initial barrier parameter $\mu_0 > 0$, a tolerance $\varepsilon_0 := \varepsilon(\mu_0)$ and $w_0 := (x_0, y_0, z_0) \in \mathbb{R}^{2n+m}$ such that $v_0 > 0$ and $\|F(w_0) - \mu_0 \tilde{e}\| \leq \varepsilon_0$, set $k = 0$ and perform the following steps:

1. Compute d_k^μ from (1.11) and d_k^w from (1.10).
2. Choose $\tau_k := \tau(\mu_k) \in (0, 1)$ and compute α_k the largest value of $\alpha \in (0, 1]$ such that (1.12) holds. Set $\mu_{k+1} = \mu_k + \alpha_k d_k^\mu$ and $\bar{w}_k = w_k + \alpha_k d_k^w$.
3. Choose $\varepsilon_{k+1} := \varepsilon(\mu_{k+1}) < \varepsilon_k$. Starting from $\bar{w}_k$, apply a sequence of inner iterations to find w_{k+1} satisfying

$$\|F(w_{k+1}) - \mu_{k+1} \tilde{e}\| \leq \varepsilon_{k+1}.$$

4. Set $k := k + 1$ and go to Step 1.

Upon denoting the Newton iterates by

$$w_k^+ = w_k + d_k^w \quad \text{and} \quad \mu_k^+ = \mu_k + d_k^\mu.$$

where d_k^μ and d_k^w are defined by (1.11) and (1.10), equation (1.10) can be rewritten

$$F'(w_k)(w_k^+ - w_k) + F(w_k) = \mu_k^+ \tilde{e}.$$

We also have

$$\bar{w}_k = \alpha_k w_k^+ + (1 - \alpha_k) w_k \quad \text{and} \quad \mu_{k+1} = \alpha_k \mu_k^+ + (1 - \alpha_k) \mu_k.$$

Theorem 7.1 *Suppose that Assumptions 2.1–2.4 and 3.1 hold. Suppose also that the function $\theta(\cdot)$, $\varepsilon(\cdot)$ and $\tau(\cdot)$ satisfy the properties (7.1). Let $\{w_k\}$ and $\{\mu_k\}$ be the sequences generated by Algorithm 7.1. Suppose that $\{\mu_k\}$ converges to zero and that a subsequence of $\{w_k\}$ converges to w^*. The following conclusions hold:*

(i) $w_k \to w^*$ *when* $k \to \infty$,

(ii) $\alpha_k = 1$ *and* $w_{k+1} = w_k^+ = \bar{w}_k$ *for sufficiently large* k,

(iii) $\dfrac{\|w_{k+1} - w^*\|}{\|w_k - w^*\|} \sim \dfrac{\mu_{k+1}}{\mu_k}$.

Proof. Let δ^* and μ^* be defined in Lemma 4.5. Let $\bar{\mu}$ be a threshold value such that all the conclusions of Theorem 6.3 are satisfied. Let us define $\hat{\mu} = \min\{1, \mu^*, \bar{\mu}, \frac{\delta^*}{M_4+C}\}$, where M_4 and C are defined in Lemmas 6.2 and 4.3. By assumption, there exists an index $k_0 \geq 0$ such that, $w_{k_0} \in B(w^*, \delta^*)$ and $\mu_{k_0} < \hat{\mu}$. Let us prove that $w_k \in B(w^*, \delta^*)$ for all $k \geq k_0$. We proceed by induction on k. Our claim is true for $k = k_0$. Assume that it is true for a given $k \geq k_0$. Considering the assumptions that have been made and the induction hypothesis, Theorem 6.3 applies, therefore $\alpha_k = 1$ and $w_{k+1} = w_k^+$. By Lemmas 6.2 and 4.3, we have

$$\begin{aligned}
\|w_{k+1} - w^*\| &\leq \|w_{k+1} - w(\mu_{k+1})\| + \|w(\mu_{k+1}) - w^*\| \\
&\leq M_4\mu_k^2 + C\mu_{k+1} \\
&< (M_4 + C)\hat{\mu} \\
&< \delta^*,
\end{aligned}$$

and thus our claim is also true for $k + 1$.

By applying Lemma 6.1 one has $\|w_k - w^*\| \leq M_3\mu_k$ for all $k \geq k_0$, which proves conclusion (i).

Applying again Theorem 6.3, we have $\alpha_k = 1$ and $w_{k+1} = \bar{w}_k = w_k^+$ for all $k \geq k_0$, which proves point (ii).

Lemma 6.2 and the fact that $\mu_k^2 = o(\mu_{k+1})$ imply that $\|w_k - w(\mu_k)\| = o(\mu_k)$. Using again Lemma 4.3, it follows that $w_k - w^* = w_k - w(\mu_k) + w(\mu_k) - w^* = \mu_k w'(0) + o(\mu_k)$, from which conclusion (iii) follows. □

We will refer to the steps described in Step 2 of Algorithm 7.1 as *extrapolation* steps. Note that since the inner iteration in Step 3 of Algorithm 7.1 relies on a globally convergent method, global convergence of Algorithm 7.1 is ensured once the generated sequence $\{\mu_k\}$ converges to zero. This will be the case provided the sequence of steplengths $\{\alpha_k\}$ along the extrapolation is bounded away from zero. Once the iterates reach a region where the central path exists and once μ is sufficiently small, it is easy to show that there exists $\alpha_{\min} > 0$ such that $\alpha_k \geq \alpha_{\min} > 0$ for all sufficiently large k. Theorem 6.3 confirms this and goes even further to show that the unit step is asymptotically accepted. In a practical implementation, preventing α_k from converging to zero when far from a solution, and possibly in regions where no central path exists, can be achieved by imposing a minimal decrease in μ at each outer iteration. This has, however, not been necessary in the tests of §8.

From a numerical point of view, a danger with Algorithm 7.1 is that $\bar{w}_k$ may lie too close to the boundary, causing ill conditioning and potentially leading to a large number of inner iterations. Although this situation did not arise in the tests, of §8, it could certainly occur. To circumvent this possibility, we suggest performing a linesearch on a merit function along extrapolation steps in order to keep $\bar{w}_k$ safely away from the boundary. This results in the following algorithm, for which we only specify the difference with Algorithm 7.1.

ALGORITHM 7.2

2. Choose $\tau_k := \tau(\mu_k) \in (0,1)$ and compute $\tilde{\alpha}_k$ the greatest value of $\alpha \in (0,1]$ such that (1.12) holds. Starting with $\alpha = \tilde{\alpha}_k$ and from $\tilde{w}_k = w_k + \tilde{\alpha}_k d_k^w$, perform a backtracking linesearch on a merit function for which d_k^w is a descent direction from w_k. Let $\alpha_k > 0$ be the resulting steplength. Set $\mu_{k+1} = \mu_k + \alpha_k d_k^\mu$ and $\bar{w}_k = w_k + \alpha_k d_k^w$.

In our tests, the inner iterations are globalized by an exact penalty function and we use this merit function to identify μ_{k+1} and $\bar{w}_k$ at Step 2 of Algorithm 7.2. It will be made explicit in Section 8.

We now present a third algorithm which only takes steps in (w, μ) but is not embedded into any globally-convergent framework. It simply eliminates the sequences of inner iterations and only takes extrapolation steps. Quite surprisingly, this algorithm turned out to be more robust in practice. We describe it as Algorithm 7.3.

ALGORITHM 7.3

Given an initial barrier parameter $\mu_0 > 0$, a tolerance $\varepsilon_0 := \varepsilon(\mu_0)$ and $w_0 := (x_0, y_0, z_0) \in \mathbb{R}^{2n+m}$ such that $v_0 > 0$ and $\|F(w_0) - \mu_0 \tilde{e}\| \leq \varepsilon_0$, set $k = 0$ and perform the following steps:

1. Compute d_k^μ from (1.11) and d_k^w from (1.10).
2. Choose $\tau_k := \tau(\mu_k) \in (0,1)$ and compute $\tilde{\alpha}_k$ the greatest value of $\alpha \in (0,1]$ such that (1.12) holds. Starting with $\alpha = \tilde{\alpha}_k$ and from $\tilde{w}_k = w_k + \tilde{\alpha}_k d_k^w$, perform a backtracking linesearch on a merit function for which d_k^w is a descent direction from w_k. Let α_k be the resulting steplength. Set $\mu_{k+1} = \mu_k + \alpha_k d_k^\mu$ and $w_{k+1} = w_k + \alpha_k d_k^w$.
3. Set $k := k + 1$ and go to Step 1.

Used as such, the merit function in Algorithm 7.3 mostly serves the purpose of staying safely away from the boundary of the feasible set. Obviously, this algorithm would be globally convergent using for instance a linesearch on the residual (1.4). However, this abandons the minimization aspect of the problem and merely attempts to identify a first-order critical point. Research on a global convergence framework for Algorithm 7.3 is under way.

8 Numerical experiments

In this section we describe our framework for numerical experiments and summarize the results. The three algorithms of Section 7 were implemented as modifications of the IPOPT software [14][1], an implementation of a globally-convergent algorithm for general smooth nonlinear programming. Nonlinear inequality constraints are transformed into equalities by the addition of slack variables, resulting in a problem of the form (1.1). Bounds on the variables and slacks are treated as in (1.6). The IPOPT implementation fixes a pre-determined sequence $\{\mu_k\}$ of barrier parameters which decreases to zero and each set of inner iterations consists in an SQP algorithm applied to the barrier subproblem. Global convergence is ensured in one of several ways, left as an option to the user. IPOPT was chosen for its excellent robustness, demonstrated numerical efficiency and availability.

Algorithms 7.1 and 7.2 depart from the IPOPT reference implementation mainly in the extrapolation step and the fact that cutting back this step influences the next value of μ. Step 3 of both algorithms is identical to the globalization implemented in IPOPT. We had to choose particular parameters in IPOPT and for completeness, we now describe them and mention what options differ from their default values and why. We stress that the following modifications were applied to *all* algorithms compared in this section.

1. Globalization in Step 3 consisted in a linesearch ensuring sufficient decrease in the non-differentiable exact merit function

$$\psi(x; \mu, \nu) = f(x) - \mu \sum_{i=1}^{n} \log x_i + \nu \|c(x)\|_2, \tag{8.1}$$

   in which $\mu > 0$ is fixed and $\nu > 0$ is dynamically adjusted. The default globalization mechanism in IPOPT is instead based on a two-dimensional filter. In order to exclude extra complications due to the management of the filter, and in particular the restoration phase, and for its more widespread understanding, the linesearch was chosen. Clearly, globalizing using the filter also makes perfect sense.

2. IPOPT makes provision for second-order correction steps. These aim to improve feasibility after the step just taken was rejected by the merit function when it appears that the reason for this is insufficient improvement in feasibility. We disabled second-order correction steps just after an extrapolation step since the latter is not concerned with descent. Not doing so interferes negatively with fast asymptotic convergence. All other possibilities of second-order corrections were left unchanged.

[1] version 2.3.x

3. After a barrier subproblem is solved, IPOPT resets ν to 10^{-6} in the merit function (8.1). We also ensured that it was reset to 10^{-6} after an extrapolation step.

4. We allowed for a barrier subproblem to be deemed solved if w_k readily satisfies the next barrier stopping test—with $\mu_k^+ = \mu_{k+1}$. No extrapolation step was imposed in such a case. Similarly, we allowed $\bar{w}_k$, resulting from an extrapolation step, to satisfy the barrier stopping test with μ_{k+1} and did not impose any inner iteration step, but rather moved on to the next barrier subproblem. It must be stressed that this is *not* the default in IPOPT, which unnecessarily *imposes* that at least one step be taken in each sequence of inner iterations. Contrary to what is stated in [14], imposing a step is sufficient but not necessary to guarantee fast local convergence [2, 8, 9]. In order to compare comparable algorithms, we relaxed IPOPT so as to not impose such a step. In practice this improved the number of iterations performed by the reference implementation noticeably. This relaxation allowed us to fix a problem in the original IPOPT implementation where satisfaction of the optimality conditions for a barrier subproblem with the smallest allowed value of μ did not imply satisfaction of the optimality conditions for the original problem [14, § 2.1, Equation (7)]. This caused the code to loop infinitely, performing no work and being unable to terminate, on some of our test problems. Again, fixing this minor inconsistency noticeably increased the reliability of the reference implementation.

5. For consistency with Newton's method and the usual theory of interior-point methods, we elected to take equal primal and dual steplengths. This is not the default in IPOPT.

We compared the three algorithms on all the problems from the COPS 3.0 collection [3] which possess at least one inequality constraint and/or at least one bound constraint. This results in 28 problems, all of which were used in their default dimension. In all cases, the initial barrier parameter was left at its default value in IPOPT, $\mu_0 = 0.1$. The update of μ was given by the function $\theta_S(\cdot)$ defined in (3.3) with $b = 1$ and $\gamma = 0.1$, yielding $\mu_k^+ = \mu_k^{1.1}$ for all k. Each problem was given a limit of 1000 iterations and 10 minutes of CPU time.

Generally speaking, a common occurence of numerical failure in practical interior point methods is the situation where, at the beginning of a sequence of inner iterations, μ_k and $\|F(w_k) - \mu_k \tilde{e}\|$ differ by several orders of magnitude. This situation may lead to a large number of inner iterations and eventually, failure may occur due to limited allowed CPU time or a maximal number of iterations being reached. Using the reference implementation of IPOPT, this situation occurs, among others, on instances of problem `dirichlet`. The behaviour of μ_k and $\|F(w_k) - \mu_k \tilde{e}\|$ is depicted in Fig. 1. After a few iterations, the two differ sufficiently to cause a large number of inner iterations and, in one of the two cases, the maximum CPU time to be reached.

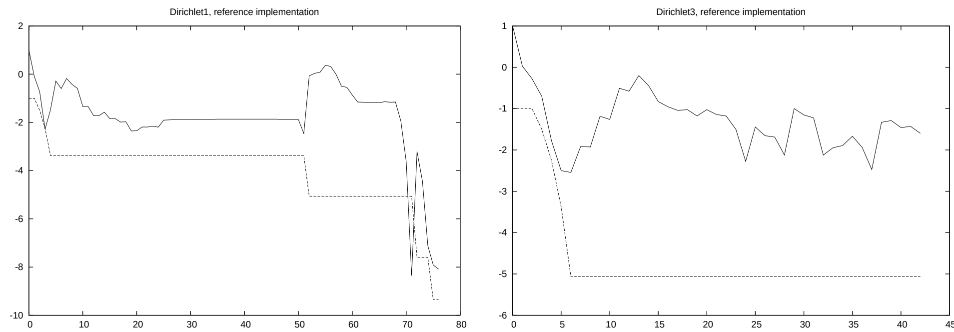

Figure 1: Log-plot of the evolution of μ_k (dashed line) and $\|F(w_k) - \mu_k \tilde{e}\|$ (solid line) using the reference implementation on two instances of problem `dirichlet`. On the left plot, the discrepancy is moderate and the method manages to identify an optimal solution. On the right plot, the discrepancy is too large and causes failure due to CPU time.

In contrast, Algorithm 7.3, by means of the steplength in the variable μ, appears to introduce a regularization which works towards avoiding a discrepancy such as that of Fig. 1. This is illustrated in Fig. 2.

Note that on the two plots of Fig. 2, the shape of the curves is indicative of superlinear convergence. Moreover, this fast convergence does not only occur asymptotically, but over a large portion of the iterations. The first problem is solved much more smoothly and in less than half as many iterations while the second problem no longer fails to be solved.

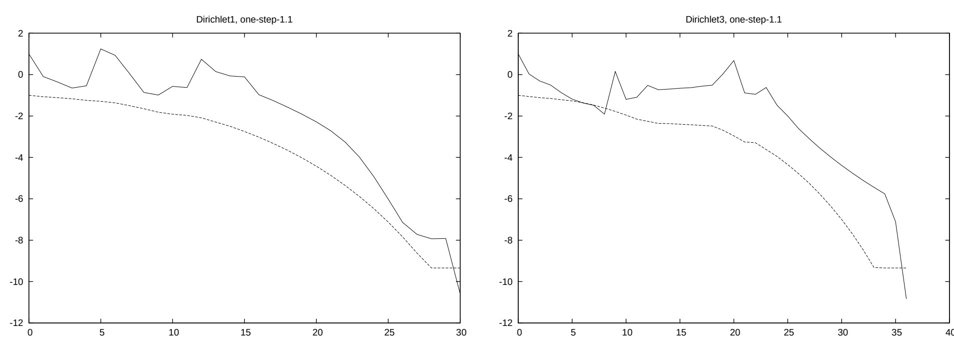

Figure 2: Log-plot of the evolution of μ_k (dashed line) and $\|F(w_k) - \mu_k \tilde{e}\|$ (solid line) using Algorithm 7.3 on the same two instances of problem `dirichlet` as in Fig. 1.

	IPOPT		EXTRAP		EXTRAP-LS		1-STEP-1.5		1-STEP-1.1	
Problem	it	#f	it	#f	it	#f	it	#f	it	#f
bearing	19	20	20	42	20	42	23	42	27	28
camshape	-45	-46	-45	-46	-45	-46	-44	-45	-43	-44
catmix	18	19	23	24	23	24	19	20	28	29
dirichlet1	76	337	58	139	58	139	-546	-7695	30	33
dirichlet2	64	343	36	37	36	37	64	267	34	36
dirichlet3	-42	-43	35	36	35	36	-41	-42	36	39
gasoil	352	2855	213	1717	213	1717	225	1771	219	1658
glider	-379	-390	-379	-390	-379	-390	-1000	-1001	-1000	-1001
lane_emden1	25	72	24	61	24	61	-24	-334	27	36
lane_emden2	-325	-4572	-324	-4509	-314	-4349	28	30	39	41
lane_emden3	45	54	45	51	45	51	-57	-103	24	26
marine	15	16	15	16	15	16	15	16	29	30
methanol	15	19	19	23	19	23	10	11	30	31
minsurf	80	130	98	249	87	183	41	50	211	486
pinene	16	17	16	17	16	17	8	9	23	24
polygon	12	17	12	17	12	17	9	10	26	27
robot	38	134	38	134	38	134	17	18	26	27
rocket	53	61	58	66	58	66	542	543	89	313
steering	14	16	14	16	14	16	9	10	23	25
tetra_duct12	10	11	9	10	9	10	9	10	24	25
tetra_duct15	10	11	10	11	10	11	9	10	24	25
tetra_duct20	9	10	9	10	9	10	6	7	21	22
tetra_gear	9	10	9	10	9	10	7	8	23	24
tetra_hook	9	10	9	10	9	10	7	8	23	24
torsion	15	27	15	28	15	28	17	36	25	26
triangle_deer	7	8	7	8	7	8	7	8	23	24
triangle_pacman	7	8	7	8	7	8	7	8	23	24
triangle_turtle	9	10	9	10	9	10	7	8	24	30

Table 1: Results on the COPS 3.0 collection. A negative number specifies a failure with the given number of iterations and/or function evaluations. When failure occurs with less than 1000 iterations, it is due to CPU time being exceeded.

Table 1 compares the number of iterations and number of function evaluations across all variants. Algorithm IPOPT is the reference implementation using the updating rule

$$\mu_{j+1} = \max\left\{\frac{\epsilon_{\text{tol}}}{11}, \min\left\{\mu_j/5, \mu_j^{1.5}\right\}\right\},$$

EXTRAP is Algorithm 7.1, EXTRAP-LS is Algorithm 7.2, 1-STEP-1.5 is Algorithm 7.3 using (3.3) with $b = 1$ and $\gamma = 0.5$ and 1-STEP-1.1 is Algorithm 7.3 using (3.3) with $b = 1$ and $\gamma = 0.1$. Both EXTRAP and EXTRAP-LS use (3.3) with $b = 1$ and $\gamma = 0.1$ during extrapolation steps. During an extrapolation step in Algorithm 7.2 and in Algorithm 7.3, the merit function used was the exact merit function (8.1) with barrier parameter $\mu = \mu_k + d_k^\mu$.

All variants fail on problems `camshape` and `glider` while `minsurf` seems best solved with an aggressive decrease of μ. The performance of Algorithm 1-STEP-1.1 on problems `steering` through `triangle_turtle` seems worse but it appeared in our tests that those problems are very efficiently solved with $\mu_0 = 10^{-9}$. We therefore attribute this decay to the initial value μ_0 rather than to the mechanism of Algorithm 7.3. Table 1 suggests that the function θ should, in general, not attempt to decrease μ too aggressively—rather, Newton's method itself should take care of the fast convergence. Moreover, the 1-STEP-1.1 variant solves two more problems than the reference implementation: `dirichlet3` and `lane_emden2`.

A natural concern regarding Algorithm 7.3 is that it might not identify a local minimizer. Fortunately, in our tests, it always identified a point at which the objective function value is comparable to or lower than that obtained with the reference implementation. The only significant differences are on problems

- `dirichlet1` for the 1-STEP-1.5 variant, which fails. The variant 1-STEP-1.1 finishes successfully with a value 10% lower than the remaining three variants,
- `glider` on which all variants fail,
- `dirichlet3` with a relative difference of order 10^{-3} caused by variants IPOPT and 1-STEP-1.5 which fail. The largest relative difference across the remaining variants if of order 10^{-9}.
- `lane_emden3` with a relative difference of order 0.3% caused by the variant 1-STEP-1.5 which fails. The largest relative difference across the remaining variants if of order 10^{-9}.

On all other problems, the largest relative variation in final objective value is of order 10^{-6}. Although all variants fail on `camshape`, the largest relative difference is of order 10^{-8}, which leads to believe that all algorithms are stuck in the same region.

9 Conclusions

We have presented a framework which attempts to give satisfactory theoretical ground to so-called *adaptive*, or *dynamic* updating rules for the barrier parameter in interior-point methods for nonlinear programming. In the absence of a notion of duality and of a meaning to a duality gap, basing such dynamic rules on observations arising from linear or convex quadratic programming lacks a theoretical justification. We hope to have shown in this paper that instead of attempting to *design an updating rule*, the power of Newton's method applied to an augmented primal-dual system provides the desired behaviour.

The present framework requires the choice of an updating function which determines the fastest allowed decrease of μ. The convexity of this function does not allow μ to increase.

It appears from our preliminary numerical tests that, in its more liberal form, this framework produces a much improved agreement between the barrier parameter and the barrier subproblem residual all along the iterations and a smaller number of iterations and function evaluations in many cases. There remains however much work to be done in order to devise globally-convergent algorithms based on the Newton/path-following scheme presented in this paper.

References

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

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

[3] E. D. Dolan, J. J. Moré, and T. S. Munson. Benchmarking optimization software with COPS 3.0. Technical Report ANL/MCS-273, Argonne National Laboratory, 2004.

[4] J.-P. Dussault. Numerical stability and efficiency of penalty algorithms. *SIAM Journal on Numerical Analysis*, 32(1):296–317, 1995.

[5] A. S. El-Bakry, R. A. Tapia, T. Tsuchiya, and Y. Zhang. On the formulation and theory of newton interior point methods for nonlinear programming. *Journal of Optimization Theory and Applications*, 89(3):507–541, 1996.

[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 Y. Yuan, editor, *Advances in Nonlinear Programming*, pages 31–56, Dordrecht, The Netherlands, 1998. Kluwer Academic Publishers.

[8] N. I. M. Gould, D. Orban, A. Sartenaer, and Ph. L. Toint. Superlinear Convergence of Primal-Dual Interior Point Algorithms for Nonlinear Programming. *SIAM Journal on Optimization*, 11(4):974–1002, 2001.

[9] N. I. M. Gould, D. Orban, A. Sartenaer, and Ph. L. Toint. Componentwise fast convergence in the solution of full-rank systems of nonlinear equation. *Mathematical Programming, Series B*, 92(3):481–508, 2002.

[10] S. Mehrotra. On the implementation of a primal-dual interior point method. *SIAM Journal on Optimization*, 2(4):575–601, 1992.

[11] J. Nocedal, A. Wächter, and R. A. Waltz. Adaptive barrier strategies for nonlinear interior methods. Technical Report, Northwestern Univeristy, IL, USA, 2005.

[12] R. J. Vanderbei and D. F. Shanno. An interior point algorithm for nonconvex nonlinear programming. *Computational Optimization and Applications*, 13(2):231–252, 1999.

[13] L. N. Vicente and S. J. Wright. Local convergence of a primal-dual method for degenerate nonlinear programming. *Computational Optimization and Applications*, 22(3):311–328, 2002.

[14] A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Research Report RC 23149, IBM T. J. Watson Research Center, NY, USA, 2004.

[15] 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(3):415–436, 1997.

[16] H. Yamashita and H. Yabe. Superlinear and quadratic convergence of some primal-dual interior point methods for constrained optimization. *Mathematical Programming, Series A*, 75(3):377–397, 1996.

[17] H. Yamashita, H. Yabe, and T. Tanabe. A globally and superlinearly convergent primal-dual interior point trust region method for large scale constrained optimization. *Mathematical Programming, Series A*, 102(1):111–151, 2005.