UNIVERSITE DE LIMOGES

XLIM

UMR CNRS 6172

Département Mathématiques-Informatique

Xlim institut de recherche

A Local Convergence Property of Primal-Dual Methods for Nonlinear Programming

Paul Armand & Joël Benoist

Rapport de recherche n° 2006-05
Déposé le 30 janvier 2006

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

A local convergence property of primal-dual methods for nonlinear programming

Paul Armand[†] and Joël Benoist[†]

January 30, 2006

Abstract. We prove a new local convergence property of a primal-dual method for solving nonlinear optimization problem. Following a standard interior point approach, the complementarity conditions of the original primal-dual system are perturbed by a parameter which is driven to zero during the iterations. The sequence of iterates is generated by a linearization of the perturbed system and by applying the fraction to the boundary rule to maintain strict feasibility of the iterates with respect to the nonnegativity constraints. The analysis of the rate of convergence is carried out by considering a linear or a superlinear arbitrary decreasing sequence of perturbation parameters. We show that, if the perturbation parameters converge to zero linearly or superlinearly and once an iterate belongs to a neighborhood of convergence of the Newton method applied to the original system, then the whole sequence of iterates converges and asymptotically follows the central trajectory in a natural way.

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

1 Introduction

We consider nonlinear programming problems of the form

$$\begin{cases} \min f(x), \\ c(x) = 0, \\ x \geq 0, \end{cases}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ are twice continuously differentiable functions. 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 minimizaion problem. The first order optimality conditions can be written

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

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

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

[†]Université de Limoges (France) ; e-mail: `paul.armand@unilim.fr, joel.benoist@unilim.fr`.

where $X = \mathrm{diag}(x_1, \ldots, x_n)$, $Z = \mathrm{diag}(z_1, \ldots, z_n)$ and e is the vector of all ones of dimension n.

Primal-dual interior point methods apply a sequence of Newton-type iterations to a perturbation of the first order conditions

$$F(w) = \mu\tilde{e}, \tag{1.1}$$

where $\mu > 0$ and $\tilde{e} = \begin{pmatrix} 0 & 0 & e^\top \end{pmatrix}^\top$. During the iterations, the parameter μ is progressively decreased to zero, while the components of v are kept sufficiently positive. Under standard assumptions and for sufficiently small $\mu > 0$, the solution of (1.1) defines a smooth curve $\mu \mapsto w(\mu)$, called the *central trajectory*. The endpoint $w^* := w(0)$ is a solution of the optimality conditions. The parameter μ is called the *barrier parameter*, because (1.1) can be interpreted as the optimality conditions of a penalty barrier problem [4].

The purpose of this paper is to analyse the local convergence of the following general primal-dual scheme. Given an arbitrary decreasing sequence $\{\mu_k\}$ converging to zero and an initial guess w_0, at iteration k a Newton iterate w_k^+ is computed by solving a linearization of (1.1)

$$F'(w_k)(w_k^+ - w_k) + F(w_k) = \mu_{k+1}\tilde{e}. \tag{1.2}$$

To keep the components of v_{k+1} sufficiently positive, a step length is computed by means of the *fraction to the boundary* rule: let α_k be the greatest $\alpha \in (0, 1]$ such that

$$v_k + \alpha(v_k^+ - v_k) \geq (1 - \tau_k)v_k, \tag{1.3}$$

where $\tau_k \in (0, 1)$ (the inequality is understood componentwise). The new iterate is then set according to

$$w_{k+1} = w_k + \alpha_k(w_k^+ - w_k). \tag{1.4}$$

In this paper we show that once an iterate belongs to a neighborhood of convergence of the Newton method applied to the original system, then the whole sequence $\{w_k\}$ converges to w^* and asymptotically follows the central trajectory in a natural manner, however on condition that the barrier parameter converges at most superlinearly to zero.

More precisely we show the following result. Under standard nondegeneracy assumptions, it is well known that $F'(w^*)$ is nonsingular [4]. It follows that there exist $\mu^* > 0$ and a neighborhood W^* such that for all initial values $\mu_0 \in (0, \mu^*)$ and $w_0 \in W^*$, a sequence defined by (1.2) and (1.4), *with* $\alpha_k = 1$ *for all* k, is well defined and converges to w^*. But such a sequence is *not* necessarily feasible with respect to the nonnegativity constraints, so that the convergence property of the Newton method cannot be directly applied to the convergence analysis of the above primal-dual scheme. Here we show that for all initial values $\mu_0 \in (0, \mu^*)$ and $w_0 \in W^*$, such that $v_0 > 0$, if $\{\mu_k\}$ converges at most superlinearly to zero, then the sequence

$\{w_k\}$ generated by the method (1.2)–(1.4) converges to w^*. Our main contribution lies in the analysis of the convergence rate of the sequence $\{w_k\}$. We successively examine the case of a linear convergence and the case of a superlinear convergence of $\{\mu_k\}$. In both cases we show that the unit step length $\alpha_k = 1$ is asymptotically accepted in (1.3) and that

$$w_k = w(\mu_k) + o(\mu_k).$$

This equality shows in particular that the rates of convergence of both sequences $\{w_k\}$ and $\{\mu_k\}$ are the same.

Local convergence analysis of interior point algorithms for nonlinear problems has been proposed in [2, 7, 9] for affine scaling methods and [1, 5, 6, 3, 8, 9, 10, 11, 12] for primal-dual methods. Note that our analysis does *not* assume that the iterates remain in a central trajectory neighborhood by satisfying some proximity measure of the form $\|F(w_k) - \mu_k \tilde{e}\| \leq \varepsilon_k$, where ε_k is of the same order as μ_k, as it is done in [1, 5, 6]. Note also that our analysis differs from those of [3, 8, 11, 12] for which it is assumed that $\mu_{k+1} = o(\|F(w_k)\|)$ or $\mu_{k+1} = O(\|F(w_k)\|^2)$ to get a superlinear or a quadratic rate of convergence. Note that in the case of a quadratic convergence the iterates do not asymptotically follow the central trajectory.

The paper is organized as follows. In Section 2 we detail our notation, assumptions and the main result. Some preliminary results on the solution of (1.1) are given in Section 3. Some bounds on the step length are introduced in Section 4 and the convergence of $\{w_k\}$ is proved in Section 5. In Section 6 we prove some technical lemmas that are the keys of the convergence rate analysis given is Section 7.

2 Notation, assumptions and main result

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 Euclidean 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 k. We use similar symbols with vector arguments, in which case they are understood normwise.

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 $\nabla c_i(x)$. Throughout the paper, we assume that the original minimization problem has a local solution $x^* \in \mathbb{R}^n$. Let $\mathcal{A} := \{i : x_i^* = 0\}$ be the set of indices of active inequality constraints. We also assume that the following four assumptions are satisfied.

Assumption 2.1 The functions f and c are twice continuously differentiable and their second derivatives are Lipschitz continuous over an open neighborhood of x^*.

Assumption 2.2 The linear independence constraint qualification holds at x^*, i.e., $\{\nabla c_i(x^*),\ i = 1, \ldots, m\} \cup \{e_i,\ i \in \mathcal{A}\}$ 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 the first order optimality conditions.

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

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

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^*. For w close to w^* and $\mu^+ > 0$, we can then define the Newton iterate at (w, μ^+) by

$$w^+ = w - F'(w)^{-1}(F(w) - \mu^+ \tilde{e}). \tag{2.1}$$

We use the notation $w = (x, y, z)$ and $v = (x, z)$ to denote the different components of w and use similar notation for subscripted vectors. For example, v_k^+ is the vector (x_k^+, z_k^+) where $w_k^+ = (x_k^+, y_k^+, z_k^+)$.

We give now conditions on the choice of the parameters μ_k and τ_k used in the method (1.2)-(1.4). We assume that there exist constants $\beta > 0$, $0 < \gamma < 1$, $0 < \sigma < 1$ and $0 < \tau < \frac{1}{2}$ such that, for all $k \geq 0$,

$$\beta \mu_k^{1+\sigma} \leq \mu_{k+1} \leq \gamma \mu_k \tag{2.2}$$

and

$$1 - \tau \frac{\mu_{k+1}}{\mu_k} \leq \tau_k < 1. \tag{2.3}$$

Condition (2.2) means that the sequence $\{\mu_k\}$ is decreasing and converges to zero at least linearly with convergence ratio γ, but at most q-superlinearly with order $1 + \sigma$. In particular, the rate of convergence of μ_k is not quadratic and we have

$$\mu_k^2 = o(\mu_{k+1}). \tag{2.4}$$

To analyse the rate of convergence of $\{w_k\}$, we will examine successively the two extreme cases for generating the sequence $\{\mu_k\}$:

$$\mu_{k+1} = \gamma \mu_k \tag{2.5}$$

and

$$\mu_{k+1} = \beta \mu_k^{1+\sigma}. \tag{2.6}$$

We are now in position to state the main result of the paper. The proof of the first part is given in Section 5 and the proof of the second part is given in Section 7.

Theorem 2.5 *There exist $\delta^* > 0$ and $\mu^* > 0$ such that for all sequence $\{\mu_k\}$ satisfying $\mu_0 \in (0, \mu^*)$ and (2.2), for all sequence $\{\tau_k\}$ satisfying (2.3) and for all $w_0 \in B(w^*, \delta^*)$ with $v_0 > 0$, the method (1.2)–(1.4) generates a sequence $\{w_k\}$ such that for all $k \geq 0$, $w_k \in B(w^*, \delta^*)$ with $v_k > 0$ and $\{w_k\}$ converges to w^*.*

In addition, if the sequence $\{\mu_k\}$ is generated according to either (2.5) or (2.6), then $\alpha_k = 1$ for sufficiently large k, which implies that $w_{k+1} = w_k^+$, and

$$w_k = w(\mu_k) + o(\mu_k).$$

In particular both sequences $\{w_k\}$ and $\{\mu_k\}$ have the same rate of convergence.

3 Preliminary results

The two following lemmas are well-known results (see e.g. [4]).

Lemma 3.1 *There exist $\delta > 0$, $L > 0$ and $K > 0$ such that for all $w, w' \in B(w^*, \delta)$,*

$$\|F'(w) - F'(w')\| \leq L\|w - w'\|$$

and

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

Lemma 3.2 *There exist $\bar{\delta} > 0$, $\bar{\mu} > 0$ and a continuously differentiable function $w(\cdot) : (-\bar{\mu}, \bar{\mu}) \to \mathbb{R}^{n+m+n}$ such that*

$$(w, \mu) \in B(w^*, \bar{\delta}) \times (-\bar{\mu}, \bar{\mu}) \quad \text{and} \quad F(w) = \mu \tilde{e}$$

if and only if

$$\mu \in (-\bar{\mu}, \bar{\mu}) \quad \text{and} \quad w(\mu) = w.$$

There exists $C > 0$, such that for all μ, $\mu' \in (-\bar{\mu}, \bar{\mu})$,

$$\|w(\mu) - w(\mu')\| \leq C|\mu - \mu'|.$$

In addition, let $w(\mu) := (x(\mu), y(\mu), z(\mu))$, then for all $i \in \{1, \ldots, n\}$,

$$x_i^* z_i'(0) + x_i'(0) z_i^* = 1.$$

The next lemma analyses some properties of the Newton iterate. Though its proof is rather standard, we give it to explicitly show the dependence of the radii of convergence δ^* and μ^* relatively to the problem data.

Lemma 3.3 *There exist $M > 0$, $\delta^* > 0$ and $\mu^* > 0$ such that for all $w \in B(w^*, \delta^*)$ and for all μ, $\mu^+ \in [0, \mu^*)$ with $\mu^+ \leq \mu$, the Newton iterate defined by (2.1) satisfies*

$$\|w^+ - w(\mu^+)\| \leq M(\|w - w(\mu)\|^2 + \mu^2) \tag{3.1}$$

Moreover, we have

$$\|w^+ - w^*\| \leq \frac{1}{2}\|w - w^*\| + \frac{\mu^+}{\mu^*}\frac{\delta^*}{2}. \tag{3.2}$$

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

Proof. Let us define $M := KL\max\{1, C^2\}$, $\delta^* := \min\{\delta, \bar{\delta}, 1/(2KL)\}$ and $\mu^* := \min\{\bar{\mu}, (\sqrt{1 + 2KL\delta^*} - 1)/(2KLC)\}$, where the constants K, L, C, δ, $\bar{\delta}$ and $\bar{\mu}$ are defined in Lemmas 3.1 and 3.2. Let $w \in B(w^*, \delta^*)$ and $\mu^+ \in [0, \mu^*)$. By Lemma 3.2, one has $F(w(\mu^+)) - \mu^+\tilde{e} = 0$. It follows that

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

Taking the norm on both sides, we obtain

$$\|w^+ - w(\mu^+)\| \leq \frac{KL}{2}\|w - w(\mu^+)\|^2.$$

Using Lemma 3.2, for all $\mu \in [\mu^+, \mu^*)$ we have

$$\begin{aligned}
\|w^+ - w(\mu^+)\| &\leq \frac{KL}{2}(\|w - w(\mu)\| + \|w(\mu) - w(\mu^+)\|)^2 \\
&\leq \frac{KL}{2}(\|w - w(\mu)\| + C(\mu - \mu^+))^2 \\
&\leq M(\|w - w(\mu)\|^2 + \mu^2),
\end{aligned}$$

where we use the inequality $(a+b)^2 \leq 2(a^2 + b^2)$ for real numbers a and b.

In the same manner we have

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

which proves (3.2) by the choice of the values δ^* and μ^*. □

4 Bounds on the step length

Let δ^* and μ^* be the threshold values defined in Lemma 3.3. Recalling the definition of the constant a in Assumption 2.4, we implicitly have

$$\delta^* < a.$$

Let us define

$$b := \frac{1}{a - \delta^*}.$$

Lemma 4.1 *For all $w \in B(w^*, \delta^*)$ and $\mu^+ \in (0, \mu^*)$, if $v > 0$ then the Newton iterate defined by (2.1) satisfies*

$$v + \frac{v^+ - v}{1 + b\|w^+ - w\|} \geq 0.$$

Proof. Let $w \in B(w^*, \delta^*)$ such that $v > 0$. Let us define

$$t := \frac{1}{1 + b\|w^+ - w\|}.$$

It suffices to prove that for all $i \in \{1, \ldots, n\}$,

$$z_i + t(z_i^+ - z_i) \geq 0. \tag{4.1}$$

By a symmetric argument the same property will hold for the components x_i.

Let $i \in \{1, \ldots, n\}$. If $z_i^+ \geq 0$, then (4.1) is satisfied because $z_i > 0$ and $0 < t < 1$. Suppose then that $z_i^+ < 0$. Using the formula of the Jacobian $F'(w)$ and the definition of the Newton iterate (2.1), we have

$$z_i(x_i^+ - x_i) + x_i(z_i^+ - z_i) = \mu^+ - x_i z_i,$$

and thus

$$x_i^+ - x_i = \frac{\mu^+}{z_i} - \frac{x_i z_i^+}{z_i}.$$

We then deduce that

$$0 < -\frac{x_i z_i^+}{z_i} < x_i^+ - x_i. \tag{4.2}$$

Let us define $t_i = \frac{z_i}{z_i - z_i^+}$. One has $0 < t_i < 1$ and using the notation $v_i = (x_i, z_i)$ we have

$$\frac{1 - t_i}{t_i} = \frac{-z_i^+}{z_i} = \frac{((x_i z_i^+ / z_i)^2 + (z_i^+)^2)^{1/2}}{\|v_i\|}$$

Now using (4.2) we obtain

$$\begin{aligned}\frac{1-t_i}{t_i} &\le \frac{((x_i^+-x_i)^2+(z_i^+-z_i)^2)^{1/2}}{\|v_i\|}\\ &\le \frac{\|w-w^+\|}{\|v_i\|}.\end{aligned}\tag{4.3}$$

The definition of a in Assumption 2.4 implies

$$a-\delta^* \le \|v_i^*\| - \|v_i - v_i^*\| \le \|v_i\|.\tag{4.4}$$

From inequalities (4.3) and (4.4), we deduce that

$$\frac{1-t_i}{t_i} \le b\|w-w^+\|,$$

and thus

$$t \le t_i.$$

Since $z_i^+ - z_i < 0$, the last inequality implies that

$$z_i + t(z_i^+ - z_i) \ge z_i + t_i(z_i^+ - z_i) = 0,$$

from which (4.1) follows. □

We deduce two bounds on the step length computed by the fraction to the boundary rule. The first one will be used to prove the convergence of the sequence $\{w_k\}$ and the second one is used in the analysis of the rate of convergence.

Corollary 4.2 *Let $\mu^+ \in (0,\mu^*)$ and $w \in B(w^*,\delta^*)$ such that $v>0$. Let w^+ be the Newton iterate defined by (2.1). Let α be the greatest value in $(0,1]$ such that*

$$v + \alpha(v^+ - v) \ge (1-\tau_v)v,$$

where $\tau_v \in (0,1)$. Then the following inequalities hold:

$$\alpha \ge \tau_v \frac{a-\delta^*}{a+\delta^*}$$

and

$$1-\alpha \le 1-\tau_v + b\|w-w^+\|.$$

Proof. By using Lemma 4.1 and $0<\tau_v<1$, we have

$$\begin{aligned}(1-\tau_v)v &\le (1-\tau_v)v + \tau_v(v + \frac{v^+-v}{1+b\|w-w^+\|})\\ &= v + \tau_v \frac{v^+-v}{1+b\|w-w^+\|}.\end{aligned}$$

The definition of α implies that

$$\frac{\tau_v}{1 + b\|w - w^+\|} \leq \alpha$$

By assumption and Lemma 3.3 one has $\|w - w^+\| \leq 2\delta^*$. Recalling that $b = \frac{1}{a - \delta^*}$, the first ineqality is proved. Using

$$\begin{aligned} 1 - \alpha &\leq 1 - \frac{\tau_v}{1 + b\|w - w^+\|} \\ &= 1 - \tau_v + \tau_v \frac{b\|w - w^+\|}{1 + b\|w - w^+\|} \end{aligned}$$

and $\tau_v \in (0, 1)$, the second inequality is proved. □

5 Convergence of $\{w_k\}$

In this section we give a proof of the convergence of the sequence $\{w_k\}$ generated by the primal-dual algorithm (1.2)–(1.4).

Proof of the first part of Theorem 2.5. Let δ^* and μ^* be the threshold values defined in Lemma 3.3. Let $\mu_0 \in (0, \mu^*)$ and let $w_0 \in B(w^*, \delta^*)$. Let $\{\mu_k\}$ and $\{\tau_k\}$ be sequences of positive numbers satisfying (2.2) and (2.3). Lemma 3.3 implies that the sequence $\{w_k\}$ is well defined and that $\|w_k - w^*\| < \delta^*$ for all $k \geq 0$. Let us prove that $\{w_k\}$ converges to w^*. The second inequality in (2.2) and condition (2.3) imply that

$$\bar{\tau} := 1 - \gamma\tau \leq \tau_k < 1.$$

for all $k \geq 0$. Using this bound and Corollary 4.2 we obtain

$$\alpha_k \geq \bar{\alpha} := \bar{\tau}\frac{a - \delta^*}{a + \delta^*}.$$

Using (1.4) and (3.2), we have

$$\begin{aligned} \|w_{k+1} - w^*\| &= \|\alpha_k w_k^+ + (1 - \alpha_k)w_k - w^*\| \\ &\leq \alpha_k\|w_k^+ - w^*\| + (1 - \alpha_k)\|w_k - w^*\| \\ &\leq \alpha_k(\frac{1}{2}\|w_k - w^*\| + \frac{\delta^*}{2\mu^*}\mu_{k+1}) + (1 - \alpha_k)\|w_k - w^*\| \\ &\leq (1 - \frac{\bar{\alpha}}{2})\|w_k - w^*\| + \frac{\delta^*}{2\mu^*}\mu_{k+1}. \end{aligned}$$

Recalling that the sequence $\{w_k\}$ is bounded and taking the limit superior in the last inequality, we deduce that $\limsup \|w_k - w^*\| = 0$. It follows that the whole sequence $\{w_k\}$ converges to w^*. □

6 Technical lemmas

From now on we assume that the data defining the algorithm are chosen so that the first part of Theorem 2.5 is realized. In this section we prove a fundamental property of the distance of the iterates to the central path, that is

$$\|w_k - w(\mu_k)\|^2 = o(\mu_{k+1}).$$

This property is the key of the rate convergence analysis of $\{w_k\}$.

Lemma 6.1 *There exists $E > 0$ such that for all $k \geq 0$*

$$e_{k+1} \leq e_k^2 + \frac{\mu_{k+1}}{\mu_k} e_k + E\mu_{k+1},$$

where $e_k := \frac{E}{C}\|w_k - w(\mu_k)\|$.

Proof. Let $k \geq 0$. Using (1.4), Lemmas 3.2 and 3.3, then the second inequality of Corollary 4.2, we have

$$\begin{aligned}
\|w_{k+1} - w(\mu_{k+1})\| &= \|\alpha_k w_k^+ + (1-\alpha_k)w_k - w(\mu_{k+1})\| \\
&\leq \alpha_k\|w_k^+ - w(\mu_{k+1})\| + (1-\alpha_k)\|w_k - w(\mu_{k+1})\| \\
&\leq \|w_k^+ - w(\mu_{k+1})\| + (1-\alpha_k)(\|w_k - w(\mu_k)\| + C\mu_k) \\
&\leq M(\|w_k - w(\mu_k)\|^2 + \mu_k^2) \\
&\quad + (1 - \tau_k + b\|w_k - w_k^+\|)(\|w_k - w(\mu_k)\| + C\mu_k).
\end{aligned}$$

Recalling (2.3), we also have

$$\begin{aligned}
1 - \tau_k + b\|w_k - w_k^+\| &\leq \tau\frac{\mu_{k+1}}{\mu_k} + b(\|w_k - w(\mu_k)\| + \|w(\mu_k) - w(\mu_{k+1})\| \\
&\quad + \|w(\mu_{k+1}) - w_k^+\|) \\
&\leq \tau\frac{\mu_{k+1}}{\mu_k} + b\|w_k - w(\mu_k)\| + bC\mu_k \\
&\quad + bM(\|w_k - w(\mu_k)\|^2 + \mu_k^2).
\end{aligned}$$

Combining the previous inequalities, we obtain

$$\begin{aligned}
\|w_{k+1} - w(\mu_{k+1})\| &\leq (M + b + bM(\|w_k - w(\mu_k)\| + C\mu_k))\|w_k - w(\mu_k)\|^2 \\
&\quad + (\tau\frac{\mu_{k+1}}{\mu_k} + 2bC\mu_k + bM\mu_k^2)\|w_k - w(\mu_k)\| \\
&\quad + M\mu_k^2 + \tau C\mu_{k+1} + bC^2\mu_k^2 + bMC\mu_k^3.
\end{aligned}$$

Using (2.4), we deduce that

$$\begin{aligned}
\|w_{k+1} - w(\mu_{k+1})\| &\leq \frac{E}{C}\|w_k - w(\mu_k)\|^2 + (\tau + o(1))\frac{\mu_{k+1}}{\mu_k}\|w_k - w(\mu_k)\| \\
&\quad + (\tau C + o(1))\mu_{k+1},
\end{aligned}$$

where $E := (M + b + 2bM\delta^* + C\mu^*)C$. There exists k_0 such that for all $k \geq k_0$

$$\|w_{k+1} - w(\mu_{k+1})\| \leq \frac{E}{C}\|w_k - w(\mu_k)\|^2 + \frac{\mu_{k+1}}{\mu_k}\|w_k - w(\mu_k)\| + C\mu_{k+1}.$$

By eventually increasing the value of the constant E, the above inequality holds for all $k \geq 0$ and thus the lemma is proved. $\square$

Lemma 6.2 *Let $\{e_k\}$ be a sequence of positive numbers converging to zero and let $\{\mu_k\}$ be a decreasing sequence of positive numbers generated either by the recurrence (2.5) or (2.6). Assume that there exists $E > 0$, such that for $k \geq 0$*

$$e_{k+1} \leq e_k^2 + \frac{\mu_{k+1}}{\mu_k} e_k + E\mu_{k+1}. \tag{6.1}$$

Then

$$e_k^2 = o(\mu_{k+1}).$$

The proof is given in Appendix A

As an immediate consequence of Lemmas 6.1 and 6.2, we have the following result.

Corollary 6.3 *If the sequence $\{\mu_k\}$ is generated according to either (2.5) or (2.6), then $\|w_k - w(\mu_k)\|^2 = o(\mu_{k+1})$.*

7 Rate of convergence analysis

We first give two asymptotic properties of the iterates.

Lemma 7.1 *If the sequence $\{\mu_k\}$ is generated according to either (2.5) or (2.6), then*

$$\|w_k^+ - w(\mu_{k+1})\| = o(\mu_{k+1}).$$

Proof. Let $k \geq 0$. According to Lemma 3.3, Corollary 6.3 and (2.4) we have

$$\|w_k^+ - w(\mu_{k+1})\| \leq M(\|w_k - w(\mu_k)\|^2 + \mu_k^2) = o(\mu_{k+1}).$$

$\square$

Lemma 7.2 *Let assumptions of Lemma 7.1 hold. For sufficiently large k, $v_k^+ > 0$.*

Proof. Let $k \geq 0$. By virtue of Lemma 3.2, we have

$$v(\mu_{k+1}) = v^* + \mu_{k+1}v'(0) + o(\mu_{k+1}).$$

By Lemma 7.1, we deduce that

$$v_k^+ = v^* + \mu_{k+1} v'(0) + o(\mu_{k+1}).$$

Let $(x_k^+)_i$ be a component of v_k^+ (the reasoning is identical for the components $(z_k^+)_i$). From Lemma 3.2, we have either $x_i^* > 0$ or $x_i^* = 0$ and $x_i'(0) > 0$. Using $\mu_{k+1} > 0$ and the fact the number of components is finite, we obtain the result. $\square$

Proof of the second part of Theorem 2.5. It suffices to show that $\alpha_k = 1$ for large k. Indeed, in that case we have $w_{k+1} = w_k^+$ and by Lemma 7.1 we then have $\|w_k - w(\mu_k)\| = o(\mu_k)$ for large k.

To show that the unit step length is accepted by the fraction the boundary rule, it suffices to prove that

$$v_k^+ \geq \tau \frac{\mu_{k+1}}{\mu_k} v_k$$

for large k. Indeed, by (2.3) we have $\tau \frac{\mu_{k+1}}{\mu_k} v_k \geq (1 - \tau_k) v_k$ and therefore the above inequality implies that (1.3) is satisfied with $\alpha = 1$. We will prove that

$$\frac{(x_k^+)_i}{\mu_{k+1}} \geq \tau \frac{(x_k)_i}{\mu_k}, \tag{7.1}$$

holds for all $i \in \{1, \ldots, n\}$. For obvious symmetric reasons, the reasoning is identical for the components z_i.

Let $i \in \{1, \ldots, n\}$. Consider first the case $x_i^* > 0$. Since $x_k \to x^*$ we have

$$\lim_{k \to \infty} \frac{(x_k^+)_i}{(x_k)_i} = \frac{x_i^*}{x_i^*} = 1.$$

Using $\mu_{k+1} < \mu_k$ and $0 < \tau < 1$ we deduce that (7.1) holds for large k.

Consider now the case $x_i^* = 0$. By strict complementarity and by using the property of the tangent to the central trajectory at zero stated in Lemma 3.2, we have $x_i'(0) > 0$. Lemmas 3.2 and 7.1 imply

$$\begin{aligned} (x_k^+)_i &= (x_k^+)_i - x_i(\mu_{k+1}) + x_i(\mu_{k+1}) \\ &= \mu_{k+1} x_i'(0) + o(\mu_{k+1}). \end{aligned}$$

It follows that

$$\lim_{k \to \infty} \frac{(x_k^+)_i}{\mu_{k+1}} = x_i'(0). \tag{7.2}$$

The proof will be complete if we show that

$$\frac{x_i'(0)}{1 - \tau} \geq \limsup_{k \to \infty} \frac{(x_k)_i}{\mu_k}. \tag{7.3}$$

Indeed, using $\frac{\tau}{1-\tau} < 1$ (which follows from $\tau < \frac{1}{2}$), (7.2) and (7.3) imply (7.1).

By Lemma 7.1 we have

$$\begin{aligned} |(x_{k+1})_i - x_i(\mu_{k+1})| &= |\alpha_k(x_k^+)_i + (1-\alpha_k)(x_k)_i - x_i(\mu_{k+1})| \\ &\le \alpha_k|(x_k^+)_i - x_i(\mu_{k+1})| + (1-\alpha_k)|(x_k)_i - x_i(\mu_{k+1})| \\ &\le o(\mu_{k+1}) + (1-\alpha_k)(|(x_k)_i - x_i(\mu_k)| + |x_i(\mu_k) - x_i(\mu_{k+1})|) \end{aligned}$$

Let $\varepsilon > 0$. Using a first order expansion of $x_i(\cdot)$ at $\mu = 0$, $x_i^* = 0$ and $\mu_{k+1} < \mu_k$, one has

$$|x_i(\mu_k) - x_i(\mu_{k+1})| \le \mu_k(x_i'(0) + \varepsilon)$$

for large k. Let us define

$$q_k := \frac{|(x_k)_i - x_i(\mu_k)|}{\mu_k}.$$

The previous inequalities imply

$$q_{k+1} \le \varepsilon + (1-\alpha_k)\frac{\mu_k}{\mu_{k+1}}(q_k + x_i'(0) + \varepsilon),$$

for sufficiently large k. Lemma 7.2 implies that

$$(1-\tau_k)v_k + \tau_k v_k^+ \ge (1-\tau_k)v_k.$$

From the definition of α_k we then have $\alpha_k \ge \tau_k$ and by (2.3) we deduce that

$$(1-\alpha_k)\frac{\mu_k}{\mu_{k+1}} \le \tau.$$

The sequence $\{q_k\}$ then satisfies the inequality

$$q_{k+1} \le \tau q_k + \tau x_i'(0) + (1+\tau)\varepsilon.$$

Let $p_k := q_k - \frac{1}{1-\tau}(\tau x_i'(0) + (1+\tau)\varepsilon)$, then $p_{k+1} \le \tau p_k$. Since $\tau \in (0, \frac{1}{2})$, one has $\limsup_{k\to\infty} p_k \le 0$ and thus

$$\limsup_{k\to\infty} q_k \le \frac{1}{1-\tau}(\tau x_i'(0) + (1+\tau)\varepsilon).$$

The above inequality holds for any $\varepsilon > 0$, it follows that

$$\limsup_{k\to\infty} q_k \le \frac{\tau}{1-\tau} x_i'(0).$$

Finally, by writing $(x_k)_i = (x_k)_i - x_i(\mu_k) + x_i(\mu_k)$, we have

$$\frac{(x_k)_i}{\mu_k} \le q_k + \frac{x_i(\mu_k)}{\mu_k}.$$

Therefore

$$\limsup_{k\to\infty} \frac{(x_k)_i}{\mu_k} \le \frac{\tau}{1-\tau} x_i'(0) + x_i'(0),$$

which implies (7.3). □

8 Conclusion

In this paper we have proved a new result of local convergence for primal-dual methods in nonlinear optimization. Our analysis shows that it is not necessary to enforce the realization of the perturbed optimality conditions with a precision of the same order as the barrier parameter. If the sequence of barrier parameters converges linearly or superlinearly to zero, then the iterates naturally follow the central trajectory with an equal rate of convergence. This result can be obviously used in a globalization framework for which the Newton step is asymptotically accepted. The extension to the solution of nonlinear complementarity problems by means of interior point methods is straightforward.

It remains an open question whether the second part of the theorem is still true for an arbitrary sequence of barrier parameters that converges at least superlinearly to zero. The question can be answered by proving Lemma 6.2 without assuming that the sequence of barrier parameters is of the form (2.5) or (2.6). To study the behavior of a primal-dual method closer to a real implementation, it could also be interesting to extend the convergence result when the solution of the Newton system is inexact and/or uses a quasi-Newton approximation of the hessian of the Lagrangian. The interesting case of degeneracy, when linear independence of active constraint gradients is replaced by the Mangasarian-Fromovitz constraint qualification, could also be investigated. Last but not least, the case of a feasible algorithm, when nonlinear inequality constraints are directly penalized in the barrier term, could also be studied. All of these questions suggest directions for future research.

A Appendix

Lemma A.1 *Let $\{e_k\}$ be a sequence of positive numbers converging to zero and let $\{\mu_k\}$ be a sequence generated by (2.6). Assume that there exist $c_1 > 0$ and $c_2 > 0$ such that for all $k \geq 0$*

$$e_{k+1} \leq c_1 e_k^2 + c_2 \mu_{k+1}. \tag{A.1}$$

Then for sufficiently large k, one has

$$e_{k+1} \leq (1 + c_1 c_2) c_2 \mu_{k+1}. \tag{A.2}$$

Proof. Let us begin by proving the result with $c_2 = 1$. There exists a subsequence $\mathcal{K}$ such that

$$e_k^2 \leq \mu_{k+1} \quad \text{for all } k \in \mathcal{K}. \tag{A.3}$$

Otherwise $\mu_{k+1} < e_k^2$ for large k and inequality (A.1) would imply that $e_{k+1} \leq (1 + c_1)e_k^2$, which means that the sequence $\{e_k\}$ would converge quadratically to zero. But for sufficiently large k we would have $\mu_{k+1} \leq \sqrt{\mu_{k+1}} \leq e_k$, which is not possible because $\{\mu_k\}$ converges at most superlinearly to zero.

From property (2.4), we can choose $k_0 \in \mathcal{K}$ such that for all $k \geq k_0$,

$$(1+c_1)^2 \frac{\mu_k^2}{\mu_{k+1}} < 1.$$

Let us prove (A.2) by induction on $k \geq k_0$. Using (A.1) and (A.3) our claim holds for $k = k_0$. Assume that it holds for a given $k \geq k_0$. Using (A.1), the induction hypothesis and the above inequality, we obtain

$$\begin{aligned} e_{k+2} &\leq c_1(1+c_1)^2 \mu_{k+1}^2 + \mu_{k+2} \\ &\leq (c_1(1+c_1)^2 \frac{\mu_{k+1}^2}{\mu_{k+2}} + 1)\mu_{k+2} \\ &\leq (c_1+1)\mu_{k+2} \end{aligned}$$

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

In the general case, for an arbitrary value $c_2 > 0$, it suffices to apply the result just proved to the sequence $\{e_k/c_2\}$. □

Proof of Lemma 6.2. Assume first that the sequence $\{\mu_k\}$ is generated by the recurrence (2.5), so that $\mu_k = \mu_0 \gamma^k$. Inequality (6.1) can be written as

$$e_{k+1} \leq e_k^2 + \gamma e_k + E\mu_0\gamma^{k+1}.$$

Let us take $\varepsilon = \frac{\sqrt{\gamma}-\gamma}{2}$. There exists $k_0 \geq 0$ such that for all $k \geq k_0$ we have $e_k \leq \varepsilon$, which implies

$$e_{k+1} \leq (\gamma+\varepsilon)e_k + E\mu_0\gamma^{k+1}. \tag{A.4}$$

Let us define $\bar{e}_{k_0} := e_{k_0}$ and $\bar{e}_{k+1} := (\gamma+\varepsilon)\bar{e}_k + E\mu_0\gamma^{k+1}$, so that $e_k \leq \bar{e}_k$ for all $k \geq k_0$. We then have $\bar{e}_k = c_1(\gamma+\epsilon)^k + c_2\gamma^k$ for some constants c_1 and c_2. It follows that

$$e_k^2/\mu_{k+1} \leq \bar{e}_k^2/\mu_{k+1} = O\Big(\big(\tfrac{1+\sqrt{\gamma}}{2}\big)^{2k}\Big),$$

for $k \geq k_0$ and thus $e_k^2 = o(\mu_{k+1})$.

Now assume that the sequence $\{\mu_k\}$ is generated by the recurrence (2.6). There exists a subsequence $\mathcal{K}$ such that

$$e_k \leq \beta\mu_k^\sigma \quad \text{for all } k \in \mathcal{K}.$$

Otherwise $\beta\mu_k^\sigma < e_k$ for large k and inequality (6.1) would imply that

$$e_{k+1} \leq 2e_k^2 + E\mu_{k+1}$$

for large k. By Lemma A.1 we would have $e_k \leq (1+2E)E\mu_k$ for large k and thus $\beta\mu_k^\sigma \leq (1+2E)E\mu_k$, which is not possible because $\sigma \in (0,1)$.

Let $k_0 \in \mathcal{K}$ such that for all $k \geq k_0$

$$2\beta + E\mu_k^{1-\sigma} \leq \beta^\sigma \mu_k^{\sigma(\sigma-1)}. \tag{A.5}$$

Let us prove by induction on $k \geq k_0$ that $e_k \leq \beta\mu_k^\sigma$. Our claim holds for $k = k_0$. Assume that it holds for a given $k \geq k_0$. Using (6.1), the induction hypothesis and (A.5), we obtain

$$\begin{array}{rcl} e_{k+1} & \leq & 2\beta^2\mu_k^{2\sigma} + E\beta\mu_k^{1+\sigma} \\ & \leq & (2\beta + E\mu_k^{1-\sigma})\beta\mu_k^{2\sigma} \\ & \leq & \beta^{\sigma+1}\mu_k^{\sigma(\sigma+1)} \\ & = & \beta\mu_{k+1}^\sigma \end{array}$$

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

Let us define the following sequence:

$$\sigma_0 = \sigma \quad \text{and} \quad \sigma_{j+1} = \frac{\sigma + \sigma_j}{\sigma + 1} \quad \text{for } j \geq 0.$$

It is clear that $\sigma < \sigma_j < 1$ for $j \geq 1$ and that $\{\sigma_j\} \uparrow 1$. Let us prove, by induction on $j \geq 0$ that there exists $c_j > 0$ such that $e_k \leq c_j\mu_k^{\sigma_j}$ for all $k \geq k_0$. Our claim holds for $j = 0$ with $c_0 = \beta$. Assume that it is true for a given $j \geq 0$. Using (6.1) and $\sigma + \sigma_j < \min\{2\sigma_j, 1 + \sigma\}$, for all $k \geq k_0$ we have

$$\begin{array}{rcl} e_{k+1} & \leq & c_j^2\mu_k^{2\sigma_j} + \beta c_j\mu_k^{\sigma+\sigma_j} + N\beta\mu_k^{1+\sigma} \\ & \leq & (c_j^2 + \beta c_j + N\beta)\mu_k^{\sigma+\sigma_j} \\ & = & c_{j+1}\mu_{k+1}^{\sigma_{j+1}}, \end{array}$$

where $c_{j+1} = (c_j^2 + \beta c_j + N\beta)/\beta^{\sigma_{j+1}}$, and thus our claim is true for $j+1$.

Let us choose $j \geq 0$ such that $2\sigma_j - (1+\sigma) > 0$. It follows that

$$\frac{e_k^2}{\mu_{k+1}} \leq \frac{c_j^2}{\beta}\mu_k^{2\sigma_j-(1+\sigma)}$$

for all $k \geq k_0$ and thus $e_k^2 = o(\mu_{k+1})$. □

References

[1] R. H. Byrd, G. Liu, and J. Nocedal, *On the local behaviour of an interior point method for nonlinear programming*, in Numerical analysis 1997 (Dundee), vol. 380 of Pitman Res. Notes Math. Ser., Longman, Harlow, 1998, pp. 37–56.

[2] T. F. Coleman and Y. Li, *On the convergence of interior-reflective Newton methods for nonlinear minimization subject to bounds*, Math. Programming, 67 (1994), pp. 189–224.

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

[4] A. V. Fiacco and G. P. McCormick, *Nonlinear programming: Sequential unconstrained minimization techniques*, John Wiley and Sons, Inc., New York-London-Sydney, 1968.

[5] N. I. M. Gould, D. Orban, A. Sartenaer, and P. L. Toint, *Superlinear convergence of primal-dual interior point algorithms for nonlinear programming*, SIAM J. Optim., 11 (2001), pp. 974–1002 (electronic).

[6] ——, *Componentwise fast convergence in the solution of full-rank systems of nonlinear equations*, Mathematical Programming, 92 (2002), pp. 481–508. ISMP 2000, Part 2 (Atlanta, GA).

[7] M. Heinkenschloss, M. Ulbrich, and S. Ulbrich, *Superlinear and quadratic convergence of affine-scaling interior-point Newton methods for problems with simple bounds without strict complementarity assumption*, Math. Program., 86 (1999), pp. 615–635.

[8] H. J. Martinez, Z. Parada, and R. A. Tapia, *On the characterization of Q-superlinear convergence of quasi-Newton interior-point methods for nonlinear programming*, Bol. Soc. Mat. Mexicana (3), 1 (1995), pp. 137–148.

[9] L. N. Vicente, *Local convergence of the affine-scaling interior-point algorithm for nonlinear programming*, Comput. Optim. Appl., 17 (2000), pp. 23–35.

[10] L. N. Vicente and S. J. Wright, *Local convergence of a primal-dual method for degenerate nonlinear programming*, Comput. Optim. Appl., 22 (2002), pp. 311–328.

[11] H. Yabe and H. Yamashita, *Q-superlinear convergence of primal-dual interior point quasi-Newton methods for constrained optimization*, J. Oper. Res. Soc. Japan, 40 (1997), pp. 415–436.

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