Computational Optimization and Applications, 17, 23–35, 2000

Local Convergence of the Affine-Scaling Interior-Point Algorithm for Nonlinear Programming

L.N. VICENTE lvicente@mat.uc.pt
Departamento de Matemática, Universidade de Coimbra, 3001-454 Coimbra, Portugal

Received May 28, 1998; Accepted September 24, 1999

Abstract. This paper addresses the local convergence properties of the affine-scaling interior-point algorithm for nonlinear programming. The analysis of local convergence is developed in terms of parameters that control the interior-point scheme and the size of the residual of the linear system that provides the step direction. The analysis follows the classical theory for quasi-Newton methods and addresses q-linear, q-superlinear, and q-quadratic rates of convergence.

Keywords: interior-point methods, affine scaling, local convergence, nonlinear programming

1. Introduction

Interior-point methods have been intensively and successfully applied to linear programming problems, linear complementarity problems, convex programming problems and other related classes of problems. For more general classes of problems, the application and analysis of interior-point methods is complicated by the presence of nonlinearity and nonconvexity. In the following paragraphs, we will survey the research carried in the field of interior-point methods for nonlinear and nonconvex optimization problems.

The local convergence theory for primal-dual interior-point methods has been established by El-Bakry et al. [14], Martinez et al. [21], and Yamashita and Yabe [29]. A few authors have considered primal-dual interior-point algorithms for which they proved global convergence (see the work by Argaez and Tapia [1], Conn et al. [8], and Yamashita [28]). The application of these algorithms to discretized optimal control problems has also been subject of study in the papers by Battermann and Heinkenschloss [2], Leibfritz and Sachs [19], Vicente [26], and Wright [27]. The recent papers by Gay, Overton, and Wright [16] and Vanderbei and Shanno [25] introduce and test globalization strategies for primal-dual interior-point algorithms.

In the papers cited above, the step direction for the interior-point method is defined in the primal variables, in the multipliers corresponding to equality constraints and in the multipliers corresponding to inequality constraints. Other authors (Forsgren and Gill [15], Byrd et al. [5], and references therein) investigated interior-point methods where the direction is defined only in the first two set of variables and an approximation is used to the multipliers corresponding to the inequality constraints.

On the other hand, affine-scaling interior-point methods for nonlinear optimization were developed by Coleman and Li (see, e.g., [3, 6, 12]) for minimization problems with simple bounds. The Coleman-Li affine scaling incorporates dual information and relates to the Dikin-Karmarkar affine scaling (see, e.g., [13, 18, 22, 23]). One attractive feature of affine-scaling interior-point methods is that they can be appropriately tailored to specific classes of problems. They have been applied to discretized optimal control problems by Dennis et al. [10] and to infinite dimensional control problems by Ulbrich and Ulbrich [24]. They have also been applied to other classes of problems like quadratic programming and nonlinear minimization subject to linear inequality constraints, but also to general nonlinear programming (Coleman and Li [7], Das [9], and Li [20]). One other attractive aspect of affine-scaling interior-point methods is that they exhibit strong local and global convergence properties: In many of the papers cited above the affine-scaling scheme has been combined with the trust-region strategy and the resulting interior-point algorithm converges globally to points satisfying first-order and second-order necessary conditions. The paper by Heinkenschloss et al. [17] combines the scaling with a projection and establishes superlinear and quadratic convergence without the strict complementarity assumption.

The paper by Vicente [26] gives a unified perspective of primal-dual and affine-scaling interior-point algorithms and introduces reduced primal-dual interior-point methods.

As far as the author is concerned, there is no general analysis of local convergence for affine-scaling interior-point algorithms like the analysis given in the aforementioned papers [14, 21, 29] for primal-dual interior-point methods. Our intention is to fill this gap in the current paper by providing a local convergence analysis of the affine-scaling interior-point algorithm for nonlinear programming when second-order derivatives are replaced by quasi-Newton updates and linear systems are solved inexactly. We do not present any analysis of global convergence or polynomiality. We start in Section 2 by describing the local version of the affine-scaling interior-point algorithm for nonlinear programming. The analysis will follow the approach given by Yamashita and Yabe [29] for primal-dual interior-point algorithms, which in turn relies on the theory developed by Broyden et al. [4] and Dennis and Moré [11] for quasi-Newton methods. However, the technical results needed for the analysis are obtained differently from [29] and they will be the subject of a careful study in Section 3. The results for linear, superlinear, and quadratic convergence are stated in Section 4.

2. The affine-scaling interior-point algorithm

Consider a nonlinear programming problem written in the form

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) = 0, \\ & x \geq 0, \end{array} \tag{1}$$

where $x \in \mathbb{R}^n$, $f : \Omega \to \mathbb{R}$, $g : \Omega \to \mathbb{R}^m$, n and m are positive integers satisfying $n > m$, and Ω is an open set of $\mathbb{R}^n$. We will assume that the functions f and g are twice Lipschitz continuously differentiable in Ω.

2.1. Motivation

If a point x is a local minimizer for problem (1) and if it satisfies a given constraint qualification (like the regularity condition to be described later), then x verifies the Karush-Kuhn-Tucker first-order necessary conditions, i.e., there exist $\lambda \in \mathbb{R}^m$ and $z \in \mathbb{R}^n$ such that

$$\nabla_x \ell(x, \lambda) - z = 0, \tag{2}$$

$$g(x) = 0, \tag{3}$$

$$x_i z_i = 0, \quad i = 1, \ldots, n, \tag{4}$$

$$x, z \geq 0, \tag{5}$$

where $\ell(x, \lambda) = f(x) + \lambda^\top g(x)$ and $\nabla_x \ell(x, \lambda) = \nabla f(x) + \nabla g(x)\lambda$.

The affine-scaling algorithm is based on the definition of the diagonal matrix $D(x, \lambda)$ whose diagonal elements are given by:

$$(D(x, \lambda))_{ii} = \begin{cases} (x_i)^{\frac{1}{2}} & \text{if } (\nabla_x \ell(x, \lambda))_i \geq 0, \\ 1 & \text{if } (\nabla_x \ell(x, \lambda))_i < 0, \end{cases}$$

for $i = 1, \ldots, n$.

Given the definition of this diagonal matrix, we can eliminate the multipliers z from the first-order necessary conditions. In fact, a point x satisfies the first-order necessary conditions if and only if there exists $\lambda \in \mathbb{R}^m$ such that

$$D(x, \lambda)^2 \nabla_x \ell(x, \lambda) = 0, \tag{6}$$

$$g(x) = 0, \tag{7}$$

$$x \geq 0. \tag{8}$$

The vector function $D(x, \lambda)^2 \nabla_x \ell(x, \lambda)$ is continuous, but not differentiable if $(\nabla_x \ell(x, \lambda))_i = 0$ for some $i \in \{1, \ldots, n\}$. If $(\nabla_x \ell(x, \lambda))_i \neq 0$, we will differentiate the i-th function in (6) using the product rule. For that purpose we introduce the diagonal matrix $E(x, \lambda)$ whose diagonal elements are the product of the derivative of the diagonal elements of $D(x, \lambda)^2$ and the components of $\nabla_x \ell(x, \lambda)$:

$$(E(x, \lambda))_{ii} = \begin{cases} (\nabla_x \ell(x, \lambda))_i & \text{if } (\nabla_x \ell(x, \lambda))_i > 0, \\ 0 & \text{otherwise}, \end{cases}$$

for $i = 1, \ldots, n$. If $(\nabla_x \ell(x, \lambda))_i = 0$, we formally apply the product rule assuming that the derivative of $(D(x, \lambda)^2)_{ii}$ is zero.

Given these considerations, the Newton step for (6)–(7) is computed from the solution of the linear system

$$\begin{pmatrix} D(x, \lambda)^2 \nabla^2_{xx} \ell(x, \lambda) + E(x, \lambda) & D(x, \lambda)^2 \nabla g(x) \\ \nabla g(x)^\top & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta \lambda \end{pmatrix} = -\begin{pmatrix} D(x, \lambda)^2 \nabla_x \ell(x, \lambda) \\ g(x) \end{pmatrix}, \tag{9}$$

where $\nabla^2_{xx} \ell(x, \lambda) = \nabla^2 f(x) + \sum_{i=1}^m \lambda_i \nabla^2 g_i(x)$.

With the definitions of the matrices $D(x, \lambda)$ and $E(x, \lambda)$ we can characterize the strict complementarity condition and the second-order sufficient conditions in terms of the pair of variables (x, λ). We define first the set of indices $\mathcal{A}(x)$:

$$\mathcal{A}(x) = \{i \in \{1, \ldots, n\} : x_i = 0\}.$$

The strict complementarity condition is satisfied at a point x, with corresponding multipliers λ and z satisfying the first-order necessary conditions (2)–(5), if

$$z_i > 0 \quad \text{for all } i \in \mathcal{A}(x)$$

or, equivalently, if

$$(E(x, \lambda))_{ii} > 0 \quad \text{for all } i \in \mathcal{A}(x). \tag{10}$$

The second-order sufficient conditions are given by (2)–(5) and the positive definiteness of $\nabla^2_{xx}\ell(x, \lambda)$ on the subspace

$$\{d \in \mathrm{I\!R}^n : \nabla g(x)^\top d = 0, \quad d_i \geq 0 \quad \text{if } i \in \mathcal{A}(x), \quad \text{and} \\ d_i = 0 \quad \text{if } i \in \mathcal{A}(x) \text{ and } z_i > 0\}.$$

If the pair (x, λ) satisfies strict complementarity (see (10)), the second-order sufficient conditions are equivalent to (6)–(8) and the positive definiteness of

$$D(x, \lambda)\nabla^2_{xx}\ell(x, \lambda)D(x, \lambda) + E(x, \lambda). \tag{11}$$

on the null space of $\nabla g(x)^\top D(x, \lambda)$.

Finally, we address the regularity condition. A feasible point x is regular if the matrix

$$\begin{pmatrix} \nabla g(x) & I_{\mathcal{A}(x)} \end{pmatrix}$$

has full column rank, where $I_{\mathcal{A}(x)}$ is a submatrix of the identity formed by columns corresponding to indices in $\mathcal{A}(x)$. For the local convergence of the algorithm addressed in this paper, we need the two following facts:

1. If x is a regular point, then the matrix $D(x, \lambda)\nabla g(x)$ has full column rank.
2. If the regular point x, with corresponding multipliers λ, is such that the matrix (11) is positive definite, then the matrix

$$\begin{pmatrix} D(x, \lambda)\nabla^2_{xx}\ell(x, \lambda)D(x, \lambda) + E(x, \lambda) & D(x, \lambda)\nabla g(x) \\ \nabla g(x)^\top D(x, \lambda) & 0 \end{pmatrix} \tag{12}$$

   is nonsingular.

The proofs are given in [26, Prop. 3.3]. Note that the matrix (12) is obtained from the linear system (9) that defines the Newton step and the change of variables $\widetilde{\Delta x} = D(x, \lambda)^{-1}\Delta x$.

We end this section with the assumptions on problem (1) needed for the analysis. Let Ω be an open set of $\mathrm{I\!R}^n$ and x_* a point in Ω.

Assumption 2.1

1. The functions f and g are twice Lipschitz continuously differentiable in Ω.
2. The point x_* (with corresponding multipliers λ_*) is regular, verifies the strict complementarity condition, and satisfies the second-order sufficient conditions.

2.2. *Algorithm and notation*

We describe next the main steps of the affine-scaling interior-point algorithm. We use H_k to represent a symmetric approximation to $\nabla^2_{xx}\ell_k$. The vectors e and $\hat{e}$ are given by

$$e = (1, \dots, 1)^\top \in \mathbb{R}^n \quad \text{and} \quad \hat{e} = (e^\top, 0, \dots, 0)^\top \in \mathbb{R}^{n+m}.$$

We use subscripted indices to represent the evaluation of a function at a particular point of the sequences $\{x_k\}$ and $\{\lambda_k\}$. The vector and matrix norms used are the ℓ_2 norms, $\|\cdot\|_F$ is the Frobenius matrix norm, and $\|\cdot\|_M$ is a given matrix norm.

Algorithm 2.1 (Affine-scaling interior-point algorithm)

1. Choose an initial point (x_0, λ_0) with $x_0 > 0$.
2. For $k = 0, 1, \dots$ do

   2.1 Compute an approximate solution $(\Delta x_k, \Delta\lambda_k)$ to the linear system
$$\begin{pmatrix} D_k^2 H_k + E_k & D_k^2 \nabla g_k \\ \nabla g_k^\top & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} = -\begin{pmatrix} D_k^2 \nabla_x \ell_k - \mu_k e \\ g_k \end{pmatrix}, \tag{13}$$
   given the approximation H_k to the Hessian matrix $\nabla^2_{xx}\ell(x_k, \lambda_k)$ and $\mu_k > 0$. (μ_k is a perturbation parameter for centralization purposes, see [14, 26, 30].)

   2.2 Set $\alpha_k = \tau_k \min_{i=1,\dots,n}\{1, \min\{-\frac{(x_k)_i}{(\Delta x_k)_i} : (\Delta x_k)_i < 0\}\}$, where $\tau_k \in [\hat{\tau}, 1]$ and $\hat{\tau} \in (0, 1)$.

   2.3 Set the new iterates:
$$x_{k+1} = x_k + \alpha_k \Delta x_k, \quad \lambda_{k+1} = \lambda_k + \Delta\lambda_k.$$

For the analysis, it is convenient to use the following notations:

$$w_k = \begin{pmatrix} x_k \\ \lambda_k \end{pmatrix}, \quad \Delta w_k = \begin{pmatrix} \Delta x_k \\ \Delta\lambda_k \end{pmatrix},$$
$$\Lambda_k = \begin{pmatrix} \alpha_k I_n & 0 \\ 0 & I_m \end{pmatrix}, \quad w_{k+1} = w_k + \Lambda_k \Delta w_k,$$

and

$$A_k = \begin{pmatrix} D_k^2 H_k + E_k & D_k^2 \nabla g_k \\ \nabla g_k^\top & 0 \end{pmatrix}.$$

2.3. Inexactness

The linear system (13) can be solved inexactly, meaning that:

$$\begin{pmatrix} D_k^2 H_k + E_k & D_k^2 \nabla g_k \\ \nabla g_k^\top & 0 \end{pmatrix} \begin{pmatrix} \Delta x_k \\ \Delta \lambda_k \end{pmatrix} = - \begin{pmatrix} D_k^2 \nabla_x \ell_k - \mu_k e \\ g_k \end{pmatrix} + \begin{pmatrix} r_k^1 \\ r_k^2 \end{pmatrix}, \tag{14}$$

where

$$r_k = \begin{pmatrix} r_k^1 \\ r_k^2 \end{pmatrix}$$

is the residual vector. The analysis in this paper determines how fast the norm of the residual r_k must go to zero. We will also impose asymptotic conditions on the norm of the vector

$$s_k^1 = \begin{pmatrix} \frac{(r_k^1)_1}{(x_k)_1} \\ \vdots \\ \frac{(r_k^1)_n}{(x_k)_n} \end{pmatrix}.$$

2.4. Differentiability

The step Δw_k can be seen as a Newton step on a system of Lipschitz continuously differentiable nonlinear equations. For this purpose, we consider a point $w_* = (x_*, \lambda_*)$ in the conditions of the Assumptions 2.1. Of importance for this discussion is the fact that w_* satisfies the strict complementarity condition (10) and the first-order necessary conditions (6)–(8). We define a diagonal matrix $D[x_*, \lambda_*, k](x, \lambda)$ with diagonal elements given by

$$(D[x_*, \lambda_*, k](x, \lambda))_{ii} = \begin{cases} (x_i)^{\frac{1}{2}} & \text{if } (\nabla_x \ell(x_*, \lambda_*))_i > 0, \\ (x_i)^{\frac{1}{2}} & \text{if } (\nabla_x \ell(x_*, \lambda_*))_i = 0 \text{ and } (\nabla_x \ell_k)_i \geq 0, \\ 1 & \text{if } (\nabla_x \ell(x_*, \lambda_*))_i = 0 \text{ and } (\nabla_x \ell_k)_i < 0, \\ 1 & \text{if } (\nabla_x \ell(x_*, \lambda_*))_i < 0, \end{cases}$$

for $i = 1, \ldots, n$. Given this definition, we can easily deduce the three following facts:

1. The vector function $D[x_*, \lambda_*, k](x, \lambda) \nabla_x \ell(x, \lambda)$ is Lipschitz continuously differentiable on the variables x and λ. The definition of the vector function $D[x_*, \lambda_*, k](\cdot, \cdot)$ depends on (x_*, λ_*) and (x_k, λ_k). However, the definition of the i-th principal diagonal element of $D[x_*, \lambda_*, k](x, \lambda)$ is independent of (x, λ).
2. If w_k is sufficiently close to w_*, then

$$D[x_*, \lambda_*, k](x_k, \lambda_k) = D(x_k, \lambda_k).$$

To simplify notation, we define

$$D_{*,k}(x,\lambda) = D[x_*, \lambda_*, k](x,\lambda) \quad \text{and} \quad D_{*,k} = D_{*,k}(x_k, \lambda_k).$$

Thus we can write

$$\begin{pmatrix} D_{*,k}^2 H_k + E_k & D_{*,k}^2 \nabla g_k \\ \nabla g_k^\top & 0 \end{pmatrix} \begin{pmatrix} \Delta x_k \\ \Delta \lambda_k \end{pmatrix} = -\begin{pmatrix} D_{*,k}^2 \nabla_x \ell_k - \mu_k e \\ g_k \end{pmatrix} + \begin{pmatrix} r_k^1 \\ r_k^2 \end{pmatrix}. \tag{15}$$

Introducing the notation

$$A_{*,k} = \begin{pmatrix} D_{*,k}^2 H_k + E_k & D_{*,k}^2 \nabla g_k \\ \nabla g_k^\top & 0 \end{pmatrix}$$

and

$$F_{*,k}(w) = F_{*,k}(x,\lambda) = \begin{pmatrix} D_{*,k}(x,\lambda)^2 \nabla_x \ell(x,\lambda) \\ g(x) \end{pmatrix},$$

we rewrite the quasi-Newton step Δw_k as

$$A_{*,k} \Delta w_k = -F_{*,k}(w_k) + \mu_k \hat{e} + r_k. \tag{16}$$

3. If $r_k = 0$, $H_k = \nabla_{xx}^2 \ell(x_k, \lambda_k)$, and w_k is sufficiently close to w_*, then Δw_k is the Newton step for

$$\begin{aligned} D[x_*, \lambda_*, k](x,\lambda) \nabla_x \ell(x,\lambda) - \mu_k e &= 0, \\ g(x) &= 0, \end{aligned}$$

at $w = w_k$. Thus

$$\nabla F_{*,k}(w) = \nabla F_{*,k}(x,\lambda) = \begin{pmatrix} D_{*,k}^2(x,\lambda) \nabla_{xx}^2 \ell(x,\lambda) + E(x,\lambda) & D_{*,k}^2(x,\lambda) \nabla g(x) \\ \nabla g^\top(x) & 0 \end{pmatrix}$$

and the Newton step Δw_k satisfies

$$\nabla F_{*,k}(w_k) \Delta w_k = -F_{*,k}(w_k) + \mu_k \hat{e} + r_k.$$

3. Technical lemmas

The set of active indices at x_* is defined as

$$\mathcal{A}(x_*) = \{i \in \{1, \dots, n\} : (x_*)_i = 0\}.$$

Lemma 3.1. *There exist positive numbers* ϵ, κ_1, κ_2, *and* κ_3 *independent of* k, *such that if* $\|w_k - w_*\| \leq \epsilon$, $\kappa_1\|\Delta w_k\| \leq 1$, *and* $\kappa_2\|\Delta w_k\| + \kappa_3\|s_k^1\| \leq 1$, *then*

$$|1-\alpha_k| \leq |1-\tau_k| + \tau_k\big(\kappa_2\|\Delta w_k\| + \kappa_3\big\|s_k^1\big\|\big).$$

Proof: If $i \notin \mathcal{A}(x_*)$ and $\epsilon \leq \frac{(x_*)_i}{2}$ then

$$-\frac{(\Delta x_k)_i}{(x_k)_i} \leq \kappa_1\|\Delta w_k\|,$$

where $\kappa_1 = \max\{\frac{2}{(x_*)_i} : i \notin \mathcal{A}(x_*)\}$.

If $i \in \mathcal{A}(x_*)$ and ϵ is sufficiently small, then from the assumption (10) on strict complementarity we know that $(E_k)_{ii} \geq (E_*)_{ii}/2$. On the other hand, from the first equation in (15)

$$(\Delta x_k)_i = -\frac{\big(D_{*,k}^2 H_k \Delta x_k\big)_i}{(E_k)_{ii}} - \frac{\big(D_{*,k}^2 \nabla g_k \Delta\lambda_k\big)_i}{(E_k)_{ii}} - \frac{\big(D_{*,k}^2 \nabla_x \ell_k\big)_i}{(E_k)_{ii}} + \frac{\mu_k}{(E_k)_{ii}} + \frac{\big(r_k^1\big)_i}{(E_k)_{ii}}.$$

Thus,

$$-\frac{(\Delta x_k)_i}{(x_k)_i} = \frac{(H_k \Delta x_k)_i}{(E_k)_{ii}} + \frac{(\nabla g_k \Delta\lambda_k)_i}{(E_k)_{ii}} - 1 - \frac{\mu_k}{(x_k)_i (E_k)_{ii}} - \frac{\big(r_k^1\big)_i}{(x_k)_i (E_k)_{ii}}.$$

Since

$$\frac{\mu_k}{(x_k)_i (E_k)_{ii}} > 0$$

we get

$$-\frac{(\Delta x_k)_i}{(x_k)_i} \leq 1 + \kappa_2\|\Delta w_k\| + \kappa_3\big\|s_k^1\big\|,$$

where κ_2 and κ_3 are positive constants independent of k. A simple derivation yields

$$|1-\alpha_k| \leq |1-\tau_k| + \tau_k\left|1 - \frac{\alpha_k}{\tau_k}\right|. \tag{17}$$

If $\alpha_k = \tau_k$ then $|1-\alpha_k| \leq |1-\tau_k|$ and the proof is completed. If $\alpha_k < \tau_k$, then the value of α_k is determined by an index i for which $(\Delta x_k)_i < 0$. In this case, we have two situations. Either $i \notin \mathcal{A}(x_*)$, in which case

$$-\frac{(x_k)_i}{(\Delta x_k)_i} \geq \frac{1}{\kappa_1\|\Delta w_k\|} \geq 1 \geq 1 - \kappa_2\|\Delta w_k\| - \kappa_3\big\|s_k^1\big\|, \tag{18}$$

or $i \in \mathcal{A}(x_*)$, in which case

$$-\frac{(x_k)_i}{(\Delta x_k)_i} \geq \frac{1}{1 + \kappa_2 \|\Delta w_k\| + \kappa_3 \left\|s_k^1\right\|} \geq 1 - \kappa_2 \|\Delta w_k\| - \kappa_3 \left\|s_k^1\right\|. \tag{19}$$

The proof is completed by combining inequality (17) with the definition of α_k, and the two inequalities (18) and (19). □

From this lemma and the form of the quasi-Newton step Δw_k given by (16), we can establish

$$|1 - \alpha_k| \leq |1 - \tau_k| + \tau_k \big(\kappa_2 \left\|A_{*,k}^{-1}\right\| (\|F_{*,k}\| + \mu_k \|\hat{e}\| + \|r_k\|) + \kappa_3 \left\|s_k^1\right\|\big), \tag{20}$$

provided $A_{*,k}$ is nonsingular. This bound on $1 - \alpha_k$ is determinant for the analysis since $I - \Lambda_k$ appears in the formula for $w_{k+1} - w_*$:

$$\begin{aligned} w_{k+1} - w_* &= w_k - \Lambda_k A_{*,k}^{-1} (F_{*,k}(w_k) - \mu_k \hat{e} - r_k) - w_* \\ &= (I - \Lambda_k)(w_k - w_*) + \Lambda_k A_{*,k}^{-1} (F_{*,k}(w_*) - F_{*,k}(w_k) - A_{*,k}(w_* - w_k)) \\ &\quad + \Lambda_k A_{*,k}^{-1} (\mu_k \hat{e} + r_k). \end{aligned} \tag{21}$$

The matrix $A_{*,k}$ will be nonsingular and its norm bounded if w_k is sufficiently close to w_* and H_k is sufficiently close to $\nabla^2_{xx} \ell(x_k, \lambda_k)$, cf. Lemma 3.2. The analysis for local convergence consists of bounding $\|w_{k+1} - w_*\|$ in terms of $\|w_k - w_*\|$ (for q-linear and q-superlinear convergence) or $\|w_k - w_*\|^2$ (for q-quadratic convergence). From the expressions (20) and (21), we observe that these bounds will depend on the following quantities:

$$\begin{aligned} &|1 - \tau_k|,\ \mu_k,\ \|r_k\|,\ \left\|s_k^1\right\|, \\ &\|F_{*,k}(w_k) - F_{*,k}(w_*) - \nabla F_{*,k}(w_*)(w_k - w_*)\|, \\ &\|(\nabla F_{*,k}(w_*) - A_{*,k})(w_k - w_*)\|. \end{aligned}$$

We can monitor the sizes of $|1 - \tau_k|$, μ_k, $\|r_k\|$, and $\|s_k^1\|$, forcing these quantities to satisfy specific asymptotic conditions.

The term $\|F_{*,k}(w_k) - F_{*,k}(w_*) - \nabla F_{*,k}(w_*)(w_k - w_*)\|$ is bounded by a constant times $\|w_k - w_*\|^2$. If $H_k = \nabla^2_{xx} \ell(x_k, \lambda_k)$ then $(\nabla F_{*,k}(w_*) - A_{*,k})(w_k - w_*) = 0$ and the q-quadratic convergence is achievable. In the case where H_k is an approximation to $\nabla^2_{xx} \ell(x_k, \lambda_k)$, we can expect q-linear or q-superlinear convergence. The following lemma is important for the q-linear convergence since it determines that $A_{*,k}$ is close to $\nabla F_{*,k}(w_*)$ provided w_k is sufficiently close to w_* and H_k is sufficiently close to $\nabla^2_{xx} \ell(x_k, \lambda_k)$.

Lemma 3.2. *There exist positive numbers ϵ and δ such that if $\|w_k - w_*\| \leq \epsilon$ and $\|H_k - \nabla^2_{xx} \ell(x_*, \lambda_*)\| \leq \delta$, then $A_{*,k}$ is nonsingular,*

$$\left\|A_{*,k}^{-1}\right\| \leq \kappa_4,$$

and

$$\|A_{*,k} - \nabla F_{*,k}(w_*)\| \leq \kappa_5(\delta + \epsilon),$$

where κ_4 and κ_5 are positive constants independent of k.

Proof: We have

$$A_{*,k} - \nabla F_{*,k}(w_*) = \begin{pmatrix} D_{*,k}(w_k)^2 H_k - D_{*,k}(w_*)^2 \nabla^2_{xx}\ell_* + E_k - E_* & D_{*,k}(w_k)^2 \nabla g_k - D_{*,k}(w_*)^2 \nabla g_* \\ \nabla g_k^\top - \nabla g_*^\top & 0 \end{pmatrix}.$$

Now, if we add and subtract $D_{*,k}(w_k)^2 \nabla^2_{xx}\ell_*$ in the 1, 1 block and $D_{*,k}(w_k)^2 \nabla g_*$ in the 1, 2 block, we obtain

$$\begin{aligned} \|A_{*,k} - \nabla F_{*,k}(w_*)\|_F^2 \leq\ & \|D_{*,k}(w_k)^2\|_F^2 \left\| H_k - \nabla^2_{xx}\ell_* \right\|_F^2 \\ & + \|D_{*,k}(w_k)^2 - D_{*,k}(w_*)^2\|_F^2 \left\| \nabla^2_{xx}\ell_* \right\|_F^2 \\ & + \|E_k - E_*\|_F^2 + \|D_{*,k}(w_k)^2\|_F^2 \|\nabla g_k - \nabla g_*\|_F^2 \\ & + \|D_{*,k}(w_k)^2 - D_{*,k}(w_*)^2\|_F^2 \|\nabla g_*\|_F^2 \\ & + \left\| \nabla g_k^\top - \nabla g_*^\top \right\|_F^2. \end{aligned}$$

Since

$$\|D_{*,k}(w_k)^2 - D_{*,k}(w_*)^2\|_F^2 \leq \|x_k - x_*\|^2, \qquad \|E_k - E_*\|_F^2 \leq \|x_k - x_*\|^2,$$

and $\nabla g(x)$ is Lipschitz continuous, we get

$$\|A_{*,k} - \nabla F_{*,k}(w_*)\|^2 \leq \kappa_5^2(\delta^2 + \epsilon^2) \leq \kappa_5^2(\delta^2 + \epsilon^2 + 2\delta\epsilon),$$

where κ_5 is positive and independent of k. The proof is complete since we know, from fact 2 (in Section 2.1), that the matrix $\nabla F_{*,k}(w_*)$ is nonsingular. □

4. Local convergence

The results in this section rely on the classical theory of quasi-Newton methods (see the papers by Broyden, Dennis, and Moré [4] and Dennis and Moré [11]) and correspond to the results that Yamashita and Yabe [29] obtained for the local version of the primal-dual interior-point algorithm. The proofs are similar and are omitted. The first result is the q-linear convergence of the affine-scaling interior-point algorithm. We require the approximation H_k to the Hessian to satisfy the bounded deterioration property (22). In the following theorems, if $\{a_k\}$ and $\{b_k\}$ are sequences of positive numbers, then $a_k = \mathcal{O}(b_k)$ is a notation for $\limsup_{k\to+\infty} a_k/b_k < +\infty$ and $a_k = o(1)$ represents $\limsup_{k\to+\infty} a_k = 0$.

Theorem 4.1. *Suppose Assumptions* 2.1 *hold. Consider a sequence generated by Algorithm* 2.1 *where*

$$0 < \hat{\tau} \leq \tau_k \leq 1, \qquad \mu_k = \mathcal{O}\big(\|F_{*,k}(w_k)\|^{1+\xi_1}\big),$$
$$\|r_k\| = \mathcal{O}\big(\|F_{*,k}(w_k)\|^{1+\xi_1}\big), \quad \text{and} \quad \big\|s_k^1\big\| = \mathcal{O}\big(\|F_{*,k}(w_k)\|^{\xi_2}\big),$$

and $\{H_k\}$ *satisfies the bounded deterioration property*

$$\big\|H_{k+1} - \nabla^2_{xx}\ell_*\big\|_M \leq (1 + \beta_1\sigma_k)\big\|H_k - \nabla^2_{xx}\ell_*\big\|_M + \beta_2\sigma_k, \tag{22}$$

with

$$\sigma_k = \max\{\|w_{k+1} - w_*\|, \|w_k - w_*\|\}.$$

(*The constants* ξ_1, ξ_2, β_1, *and* β_2 *are positive.*)

For each $\nu \in (1 - \hat{\tau}, 1)$ *there exist an* $\epsilon(\nu) > 0$ *and a* $\delta(\nu) > 0$ *such that if* $\|w_0 - w_*\| \leq \epsilon(\nu)$ *and* $\|H_0 - \nabla^2_{xx}\ell_*\| \leq \delta(\nu)$, *the sequence* $\{w_k\}$ *is well defined, converges to* w_*, *and the rate is q-linear with constant* ν, *i.e.,*

$$\|w_{k+1} - w_*\| \leq \nu\|w_k - w_*\|.$$

The characterization of q-superlinearity is given by a Dennis-Moré condition (see (23) below).

Theorem 4.2. *Suppose Assumptions* 2.1 *hold. Consider a sequence* $\{w_k\}$ *generated by Algorithm* 2.1 *converging q-linearly to* w_*, *where*

$$1 - \tau_k = o(1), \qquad \mu_k = \mathcal{O}(\|F_{*,k}(w_k)\|),$$
$$\|r_k\| = \mathcal{O}(\|F_{*,k}(w_k)\|), \quad \text{and} \quad \big\|s_k^1\big\| = o(1).$$

The sequence $\{w_k\}$ *converges q-superlinearly to* w_* *if and only if*

$$\lim_{k\to+\infty} \frac{\|(A_k - \nabla F_{*,k}(w_*))(w_{k+1} - w_k)\|}{\|w_{k+1} - w_k\|} = 0.$$

It is easy to prove that

$$\lim_{k\to+\infty} \frac{\big\|\big(H_k - \nabla^2_{xx}\ell_*\big)(x_{k+1} - x_k)\big\|}{\|x_{k+1} - x_k\|} = 0 \tag{23}$$

implies

$$\lim_{k\to+\infty} \frac{\|(A_k - \nabla F_{*,k}(w_*))(w_{k+1} - w_k)\|}{\|w_{k+1} - w_k\|} = 0.$$

Finally, we state the q-quadratic convergence of the affine-scaling interior-point algorithm.

Theorem 4.3. *Suppose Assumptions* 2.1 *hold. Consider a sequence generated by Algorithm* 2.1 *where* $H_k = \nabla^2_{xx}\ell(x_k, \lambda_k)$,

$$1 - \tau_k = \mathcal{O}(\|F_{*,k}(w_k)\|), \qquad \mu_k = \mathcal{O}(\|F_{*,k}(w_k)\|^2),$$
$$\|r_k\| = \mathcal{O}(\|F_{*,k}(w_k)\|^2), \quad \textit{and} \quad \left\|s_k^1\right\| = \mathcal{O}(\|F_{*,k}(w_k)\|).$$

There exists $\epsilon > 0$ *such that if* $\|w_0 - w_*\| \leq \epsilon$, *then the sequence* $\{w_k\}$ *is well defined, converges to* w_*, *and the rate is q-quadratic, i.e.,*

$$\|w_{k+1} - w_*\| \leq \kappa_6 \|w_k - w_*\|^2, \tag{24}$$

where κ_6 *is positive and independent of* k.

Acknowledgment

Support for this work has been provided by Centro de Matemática da Universidade de Coimbra, FCT, and Praxis XXI 2/2.1/MAT/346/94.

References

1. M. Argaez and R.A. Tapia, “On the global convergence of a modified augmented Lagrangian linesearch interior point Newton method for nonlinear programming,” Technical Report TR95-38, Department of Computational and Applied Mathematics, Rice University, 1995. Revised Sept. 1999.
2. A. Battermann and M. Heinkenschloss, “Preconditioners for Karush-Kuhn-Tucker systems arising in the optimal control of distributed systems,” in Optimal Control of Partial Differential Equations, W. Desch, F. Kappel, and K. Kunisch (Eds.) Birkhäuser Verlag: Basel, 1996.
3. M.A. Branch, T.F. Coleman, and Y. Li, “A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems,” SIAM J. Sci. Comput., vol. 21, pp. 1–23, 1999.
4. C.G. Broyden, J.E. Dennis, and J.J. Moré, “On the local and superlinear convergence of quasi-Newton methods,” J. Inst. Math. Appl., vol. 12, pp. 223–245, 1973.
5. R.H. Byrd, G. Liu, and J. Nocedal, “On the local behavior of an interior point method for nonlinear programming,” Numerical analysis 1997 (Dundee) Pitman Res. Notes Math. Sec., 380, pp. 37–56, Longman, 1998.
6. T.F. Coleman and Y. Li, “On the convergence of interior-reflective Newton methods for nonlinear minimization subject to bounds,” Math. Programming, vol. 67, pp. 189–224, 1994.
7. T.F. Coleman and Y. Li, “A trust region and affine scaling interior point method for nonconvex minimization with linear inequality constraints,” Math. Programming, vol. 88, pp. 1–31, 2000.
8. A.R. Conn, N. Gould, D. Orban, and Ph.L. Toint, “A primal-dual trust-region algorithm for non-convex nonlinear programming,” Math. Programming, vol. 87, pp. 215–249, 2000.
9. I. Das, “An interior point algorithm for the general nonlinear programming problem with trust region globalization,” Technical Report TR96-17, Department of Computational and Applied Mathematics, Rice University, 1996.
10. J.E. Dennis, M. Heinkenschloss, and L.N. Vicente, “Trust-region interior-point SQP algorithms for a class of nonlinear programming problems,” SIAM J. Control Optim., vol. 36, pp. 1750–1794, 1998.
11. J.E. Dennis and J.J. Moré, “A characterization of superlinear convergence and its application to quasi-Newton methods,” Math. Comput., vol. 28, pp. 549–560, 1974.

12. J.E. Dennis and L.N. Vicente, “Trust-region interior-point algorithms for minimization problems with simple bounds,” in Applied Mathematics and Parallel Computing, Festschrift for Klaus Ritter, H. Fisher, B. Riedmüller, and S. Schäffler (Eds.), Physica-Verlag, Springer-Verlag, 1996, pp. 97–107.
13. I.I. Dikin, “Iterative solution of problems of linear and quadratic programming,” Soviet Math. Dok., vol. 8, pp. 674–675, 1967.
14. A.S. El-Bakry, R.A. Tapia, T. Tsuchiya, and Y. Zhang, “On the formulation and theory of the Newton interior-point method for nonlinear programming,” J. Optim. Theory Appl., vol. 89, pp. 507–541, 1996.
15. A. Forsgren and P.E. Gill, “Primal-dual interior methods for nonconvex nonlinear programming,” SIAM J. Optim., vol. 8, pp. 1132–1152, 1998.
16. D.M. Gay, M.L. Overton, and M.H. Wright, “A primal-dual interior method for nonconvex nonlinear programming,” Proceedings of the 1996 International Conference on Nonlinear Programming, Beijing, China, pp. 31–56, Y. Yuan (Ed.), Kluwer, 1998.
17. 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. Programming, vol. 86, pp. 615–635, 1999.
18. N. Karmarkar, “A new polynomial-time algorithm for linear programming,” Combinatorica, vol. 2, pp. 373–395, 1984.
19. F. Leibfritz and E.W. Sachs, “Numerical solution of parabolic state constrained control problems using SQP- and interior-point-methods,” in Large Scale Optimization: State of the Art, W.W. Hager, D. Hearn, and P. Pardalos (Eds.), Kluwer, 1994, pp. 251–264.
20. Y. Li, “A trust region and affine scaling method for nonlinearly constrained minimization,” Technical Report CTC94TR198, Advanced Computing Research Institute, Cornell University, 1994.
21. 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,” Boletin de la Sociedad Matematica Mexicana, vol. 1, pp. 1–12, 1995.
22. G.P. McCormick, “The projective SUMT method for convex programming,” Math. Oper. Res., vol. 14, pp. 203–223, 1989.
23. C.M. Samuelson, “The Dikin-Karmarkar principle for steepest descent,” PhD Thesis, Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77251, USA, Tech. Rep. TR92-29, 1992.
24. M. Ulbrich and S. Ulbrich, “Superlinear convergence of affine-scaling interior-point Newton methods for infinite-dimensional nonlinear problems with pointwise bounds,” to appear in SIAM J. Control Optim.
25. R.J. Vanderbei and D.F. Shanno, “An interior-point algorithm for nonconvex nonlinear programming,” Comput. Optim. and Appl., vol, 13, pp. 231–252, 1999.
26. L.N. Vicente, “On interior-point Newton algorithms for discretized optimal control problems with state constraints,” Optim. Methods Softw., vol. 8, pp. 249–275, 1998.
27. S.J. Wright, “Interior point methods for optimal control of discrete-time systems,” J. Optim. Theory Appl., vol. 77, pp. 161–187, 1993.
28. H. Yamashita, “A globally convergent primal-dual interior-point method for constrained optimization,” Technical report, Mathematical Systems Institute, Japan, 1992.
29. H. Yamashita and H. Yabe, “Superlinear and quadratic convergence of some primal-dual interior point methods for constrained optimization,” Math. Programming, vol. 75, pp. 377–397, 1996.
30. Y. Zhang, R.A. Tapia, and J.E. Dennis, “On the superlinear and quadratic convergence of primal-dual interior point linear programming algorithms,” SIAM J. Optim., vol. 2, pp. 304–324, 1992.