A Feasible BFGS Interior Point Algorithm for Solving Strongly Convex Minimization Problems

Paul Armand, Jean Charles Gilbert, Sophie Jan-Jégou

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A feasible BFGS interior point algorithm for solving strongly convex minimization problems

Paul ARMAND, Jean Charles GILBERT, Sophie JAN-JÉGOU

N° 3500

15 septembre 1998

THÈME 4

Rapport de recherche

ISSN 0249-6399

A feasible BFGS interior point algorithm for solving strongly convex minimization problems

Paul Armand, Jean Charles Gilbert, Sophie Jan-Jégou

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet Estime

Rapport de recherche n° 3500 — 15 septembre 1998 — 26 pages

Abstract: We propose a BFGS primal-dual interior point method for minimizing a convex function on a convex set defined by equality and inequality constraints. The algorithm generates feasible iterates and consists in computing approximate solutions of the optimality conditions perturbed by a sequence of positive parameters μ converging to zero. We prove that it converges q-superlinearly for each fixed μ and that it is globally convergent when $\mu \to 0$.

Key-words: BFGS quasi-Newton approximations, constrained optimization, convex programming, interior point algorithm, line-search, primal-dual method, superlinear convergence.

(Résumé : tsvp)

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : 01 39 63 55 11 - International : +33 1 39 63 55 11
Télécopie : (33) 01 39 63 53 30 - International : +33 1 39 63 53 30

Une méthode de points intérieurs admissibles avec mises-à-jour de BFGS pour résoudre un problème d'optimisation fortement convexe

Résumé : Nous proposons une méthode de points intérieurs primale-duale pour minimiser une fonction convexe sur un ensemble convexe défini par des contraintes d'égalité et d'inégalité. L'algorithme génère des itérés admissibles et consiste à calculer des solutions approchées des conditions d'optimalité perturbées par une suite de paramètres μ tendant vers zéro. Nous montrons la convergence q-superlinéaire des itérés pour tout μ fixé et la convergence globale lorsque $\mu \to 0$.

Mots-clé : Algorithmes de points intérieurs, approximation quasi-newtonienne, convergence superlinéaire, formule de BFGS, méthode primale-duale, optimisation avec contraintes, programmation convexe, recherche linéaire.

A Feasible BFGS Interior Point Algorithm for Solving Strongly Convex Minimization Problems

Paul ARMAND[†], Jean Charles GILBERT[‡], and Sophie JAN-JÉGOU[§]

September 15, 1998

Abstract: We propose a BFGS primal-dual interior point method for minimizing a convex function on a convex set defined by equality and inequality constraints. The algorithm generates feasible iterates and consists in computing approximate solutions of the optimality conditions perturbed by a sequence of positive parameters μ converging to zero. We prove that it converges q-superlinearly for each fixed μ and that it is globally convergent when $\mu \to 0$.

Key words: BFGS quasi-Newton approximations, constrained optimization, convex programming, interior point algorithm, line-search, primal-dual method, superlinear convergence.

1 Introduction

We consider the problem of minimizing a smooth convex function on a convex set defined by inequality constraints. The problem is written as follows

$$\begin{cases} \min f(x) \\ c(x) \geq 0, \end{cases} \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is the function to minimize and $c(x) \geq 0$ means that each component $c_{(i)} : \mathbb{R}^n \to \mathbb{R}$ $(1 \leq i \leq m)$ of c must be nonnegative at the solution. Since we assume that these components are *concave*, the feasible set of this problem is convex. The algorithm proposed in this paper and the convergence analysis require that f and c are differentiable and that at least one of the functions f, $-c_{(1)}$, ..., $-c_{(m)}$ is *strongly* convex. The reason of this latter hypothesis will be discussed below.

Our motivation is based on practical considerations. During the last 15 years much progress has been realized on interior point (IP) methods for solving linear or convex minimization problems (see the monographs [27, 17, 31, 36, 21, 34, 39]). For nonlinear convex problems, these algorithms assume that the second derivatives of the functions used to define the problem are available (see [35, 29, 30, 19, 31, 24]).

[†]LACO - Département de Mathématiques - 123, av. A. Thomas - 87060 Limoges Cedex (France); e-mail: armand@unilim.fr.

[‡]INRIA Rocquencourt, BP 105, 78153 Le Chesnay Cedex (France); e-mail: Jean-Charles.Gilbert@inria.fr.

[§]LAO - UFR MIG - Université Paul Sabatier - 118, route de Narbonne - 31062 Toulouse Cedex 4 (France); e-mail: jan@cict.fr.

In practice, however, it is not uncommon to find situations where this requirement cannot be satisfied, in particular for large scale engineering problems (see [25] for an example, which partly motivates this study and deals with the estimation of parameters in a three phase flow in a porous medium). Despite the possible use of computational differentiation techniques [8, 15, 3, 26], the computing time and effort needed to evaluate Hessians or Hessian-vector products may be so important that IP algorithms using second derivatives may be unattractive.

This situation is familiar in unconstrained optimization. In that case, quasi-Newton (qN) techniques, which use first derivatives only, have proved to be efficient, even when there are millions of variables (see [9] for an example in meteorology). This fact motivates the present paper, in which we explore the possibility to combine the IP approach and qN techniques. Our ambition remains modest, however, since we confine ourselves to the question to know whether the elegant BFGS theory for unconstrained convex optimization [33, 6] is still valid when inequality constraints are present. For the applications, it would be desirable to have a qN-IP algorithm in the case when f and $-c$ are nonlinear and not necessary convex. We postpone this more difficult subject for a future research (see [16, 40] for possible approaches).

Provided the constraints satisfy some qualification assumptions, the Karush-Kuhn-Tucker (KKT) optimality conditions of Problem (1.1) can be written (see [14] for example): there exists a vector of multipliers $\lambda \in \mathbb{R}^m$ such that

$$\left\{\begin{array}{l} \nabla f(x) - \nabla c(x)\lambda = 0 \\ C(x)\lambda = 0 \\ (c(x), \lambda) \geq 0, \end{array}\right.$$

where $\nabla f(x)$ is the gradient of f at x (for the Euclidean scalar product), $\nabla c(x)$ is a matrix whose columns are the gradients $\nabla c_{(i)}(x)$, and $C = \operatorname{diag}(c_{(1)}, \dots, c_{(m)})$ is the diagonal matrix, whose diagonal elements are the components of c. The Lagrangian function associated with Problem (1.1) is defined on $\mathbb{R}^n \times \mathbb{R}^m$ by

$$\ell(x, \lambda) = f(x) - \lambda^\top c(x).$$

Since f is convex and each component $c_{(i)}$ is concave, for any fixed $\lambda \geq 0$, $\ell(\cdot, \lambda)$ is a convex function from $\mathbb{R}^n$ to $\mathbb{R}$. When f and c are twice differentiable, the gradient and Hessian of ℓ with respect to x are given by

$$\nabla_x \ell(x, \lambda) = \nabla f(x) - \nabla c(x)\lambda \quad \text{and} \quad \nabla^2_{xx} \ell(x, \lambda) = \nabla^2 f(x) - \sum_{i=1}^m \lambda_{(i)} \nabla^2 c_{(i)}(x).$$

Our primal-dual IP approach is rather standard (see [22, 30, 29, 18, 19, 1, 24, 23, 12, 7, 5]). It computes iteratively approximate solutions of the perturbed optimality system

$$\left\{\begin{array}{l} \nabla f(x) - \nabla c(x)\lambda = 0 \\ C(x)\lambda = \mu e \\ (c(x), \lambda) > 0, \end{array}\right. \tag{1.2}$$

for a sequence of parameters $\mu > 0$ converging to zero. In (1.2), $e = (1 \ \cdots \ 1)^\top$ is the vector of all ones whose dimension will be clear from the context. The last inequality means that all the components of both $c(x)$ and λ must be positive. By perturbing the complementarity equation of the KKT conditions with the parameter μ, the combinatorial problem inherent to the determination of the active constraints or the zero multipliers is avoided. We use the word *inner* to qualify those iterations that are used to find an approximate solution of (1.2) for fixed μ, while an *outer iteration* is the collection of inner iterations corresponding to the same value of μ.

The Newton step for solving the first two equations in (1.2) with fixed μ is the solution $d = (d^x, d^\lambda) \in \mathbb{R}^n \times \mathbb{R}^m$ of the linear system

$$\begin{pmatrix} M & -\nabla c(x) \\ \Lambda \nabla c(x)^\top & C(x) \end{pmatrix} \begin{pmatrix} d^x \\ d^\lambda \end{pmatrix} = \begin{pmatrix} -\nabla f(x) + \nabla c(x)\lambda \\ \mu e - C(x)\lambda \end{pmatrix}, \tag{1.3}$$

in which $M = \nabla^2_{xx}\ell(x, \lambda)$ and $\Lambda = \operatorname{diag}(\lambda_{(1)}, \ldots, \lambda_{(m)})$. This direction is sometimes called the primal-dual step, since it is obtained by linearizing the primal-dual system (1.2), while the primal step is the Newton direction for minimizing in the primal variable x the *barrier function*

$$\varphi_\mu(x) := f(x) - \mu \sum_{i=1}^{m} \log c_{(i)}(x),$$

associated with (1.1) (the algorithms in [13, 28, 4] are in this spirit). The two problems are related since, after elimination of λ, (1.2) represents the optimality conditions of the unconstrained *barrier problem*

$$\begin{cases} \min \varphi_\mu(x) \\ c(x) > 0. \end{cases} \tag{1.4}$$

As a result, an approximate solution of (1.2) is also an approximate minimizer of the barrier problem (1.4). However, algorithms using the primal-dual direction have been shown to present a better numerical efficiency (see for example [38]).

In our algorithm for solving (1.2) or (1.4) approximately, a search direction d is computed as a solution of (1.3), in which M is now a positive definite symmetric matrix approximating $\nabla^2_{xx}\ell(x, \lambda)$ and updated by the BFGS formula (see [11, 14] for material on qN techniques). By eliminating d^λ from (1.3) we obtain

$$(M + \nabla c(x) C(x)^{-1} \Lambda \nabla c(x)^\top) d^x = -\nabla f(x) + \mu \nabla c(x) C(x)^{-1} e = -\nabla \varphi_\mu(x). \tag{1.5}$$

Since the iterates will be maintained strictly feasible, i.e., $(c(x), \lambda) > 0$, the positive definiteness of M implies that d^x is a descent direction of φ_μ at x. Therefore, to force convergence of the inner iterates, a possibility could be to force the decrease of

φ_μ at each iteration. However, since the algorithm also generates dual variables λ, we prefer to add to φ_μ the function (see [37, 1])

$$\mathcal{V}(x,\lambda) := \lambda^\top c(x) - \mu \sum_{i=1}^m \log\left(\lambda_{(i)} c_{(i)}(x)\right)$$

to control the change in λ. Even though the map $(x,\lambda) \mapsto \varphi_\mu(x) + \mathcal{V}(x,\lambda)$ is not necessarily convex, we will show that it has a unique minimizer, which is the solution of (1.2), and that it decreases along the direction $d = (d^x, d^\lambda)$. Therefore, this primal-dual merit function can be used to force the convergence of the pairs (x,λ) to the solution of (1.2), using line-searches. It will be shown that the additional function $\mathcal{V}$ does not prevent unit step-sizes from being accepted asymptotically, which is an important point for the efficiency of the algorithm.

Let us stress the fact that our algorithm is not a standard BFGS algorithm for solving the barrier problem (1.4), since it is the Hessian of the Lagrangian that is approximated by the updated matrix M, not the Hessian of φ_μ. This is motivated by the following arguments. First, the difference between $\nabla^2_{xx}\ell(x, \mu C(x)^{-1}e)$ and

$$\nabla^2\varphi_\mu(x) = \nabla^2 f(x) + \mu \sum_{i=1}^m \left(\frac{1}{c_{(i)}(x)^2} \nabla c_{(i)}(x) \nabla c_{(i)}(x)^\top - \frac{1}{c_{(i)}(x)} \nabla^2 c_{(i)}(x) \right), \quad (1.6)$$

involves first derivatives only. Since these derivatives are considered to be available, they need not be approximated. Second, the Hessian $\nabla^2_{xx}\ell$, which is approximated by M, is independent of μ and does not become ill-conditioned as μ goes to zero. Third, the approximation of $\nabla^2_{xx}\ell$ obtained at the end of an outer iteration can be used as starting matrix for the next outer iteration. If this looks attractive, it has also the inconvenient to restrict the approach to (strongly) convex functions, as we now explain.

After the computation of the new iterates $x_+ = x + \alpha d^x$ and $\lambda_+ = \lambda + \alpha d^\lambda$ (α is the step-size given by the line-search), the matrix M is updated by the BFGS formula using two vectors δ and γ. Since we want the new matrix M_+ to be an approximation of $\nabla^2_{xx}\ell(x_+, \lambda_+)$ and because it satisfies the quasi-Newton equation $M_+\delta = \gamma$ (a property of the BFGS formula), it makes sense to define δ and γ by

$$\delta := x_+ - x \quad \text{and} \quad \gamma := \nabla_x \ell(x_+, \lambda_+) - \nabla_x \ell(x, \lambda_+).$$

The formula is well defined and generates stable positive definite matrices provided these vectors satisfy $\gamma^\top\delta > 0$. This inequality, known as the curvature condition, expresses the strict monotonicity of the gradient of the Lagrangian between two successive iterates. In unconstrained optimization, it can always be satisfied by using the Wolfe line-search provided the function to minimize is bounded below. If this is a reasonable assumption in unconstrained optimization, it is no longer the case when constraints are present, since the optimization problem may be perfectly well defined even when ℓ is unbounded below. Now assuming this hypothesis on the boundedness of ℓ would have been less restrictive than assuming its strong convexity, but it is not satisfactory. Indeed, with a bounded below Lagrangian, the curvature condition can be satisfied by the Wolfe line-search as in unconstrained optimization, but near the solution the information on $\nabla^2_{xx}\ell$ collected in the matrix M could come from a region far from the optimal point, which would prevent q-superlinear convergence of the iterates. Because of this observation, we assume that f or one of the functions $-c_{(i)}$ is strongly convex, so that the Lagrangian becomes a strongly convex function of x for any fixed $\lambda > 0$. With this assumption, the curvature condition will be satisfied independently of the kind of line-search techniques actually used in the algorithm. The question whether the present theory can be adapted to convex problems, hence including linear programming, is puzzling. We come back on this issue in the discussion section.

A large part of the paper is devoted to the analysis of the quasi-Newton algorithm for solving the perturbed KKT conditions (1.2) with fixed μ. The algorithm is detailed in the next section, while its convergence speed is analyzed in Sections 3 and 4. In particular, it is shown that, for fixed $\mu > 0$, the primal-dual pairs (x, λ) converge q-superlinearly toward a solution of (1.2). The tools used to prove convergence are essentially those of the BFGS theory [6, 10, 32]. In Section 5, the overall algorithm is presented and it is shown that the sequence of outer iterates is globally convergent.

2 The algorithm for solving the barrier problem

Our minimal assumptions are the following.

Assumptions 2.1. (i) The functions f and $-c_{(i)}$ $(1 \le i \le m)$ are convex and differentiable from $\mathbb{R}^n$ to $\mathbb{R}$ and at least one of them is strongly convex. (ii) The set of strictly feasible points for Problem (1.1) is nonempty, i.e., there exists $x \in \mathbb{R}^n$ such that $c(x) > 0$.

Assumption (i) was motivated in Section 1. Assumption (ii), also called the (strong) Slater condition, is necessary for the wellposedness of a *feasible* interior point method. With the convexity assumption, it is equivalent to the fact that the set of multipliers associated with a given solution is nonempty and compact (see [20, Theorem VII.2.3.2] for example). These assumptions have the following clear consequence.

Lemma 2.2. *Suppose that Assumptions* 2.1 *hold. Then, the solution set of Problem* (1.1) *is nonempty and bounded.*

By Lemma 2.2, the level sets of the logarithmic barrier function φ_μ are compact, a fact that will be used frequently. It is a consequence of [13, Lemma 12], which we recall for completeness.

Lemma 2.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex continuous function and $c : \mathbb{R}^n \to \mathbb{R}^m$ be a continuous function having concave components. Suppose that the set $\{x \in \mathbb{R}^n : c(x) > 0\}$ is nonempty and that the solution set of Problem* (1.1) *is nonempty and bounded. Then, for any $\alpha \in \mathbb{R}$ and $\mu > 0$, the set*

$$\{x \in \mathbb{R}^n : c(x) > 0,\ f(x) - \mu \sum_{i=1}^{m} \log c_{(i)}(x) \le \alpha\}$$

is compact (*and possibly empty*).

Let x_1 be the first iterate of our *feasible* IP algorithm, hence satisfying $c(x_1) > 0$, and define the level set

$$\mathcal{L}_1^{\mathrm{P}} := \{x \in \mathbb{R}^n : c(x) > 0 \text{ and } \varphi_\mu(x) \le \varphi_\mu(x_1)\}.$$

Lemma 2.4. *Suppose that Assumptions* 2.1 *hold. Then, the barrier problem* (1.4) *has a unique solution, which is denoted by $\hat{x}_\mu$.*

Proof. By Assumptions 2.1, Lemma 2.2 and Lemma 2.3, $\mathcal{L}_1^{\mathrm{P}}$ is nonempty and compact, so that the barrier problem (1.4) has at least one solution. This solution is also unique, since φ_μ is strictly convex on $\{x \in \mathbb{R}^n : c(x) > 0\}$. Indeed, by Assumption 2.1 (i), $\nabla^2 \varphi_\mu(x)$ given by (1.6) is positive definite. □

To simplify the notation we denote by

$$z := (x, \lambda)$$

a typical pair of primal-dual variables and by $\mathcal{Z}$ the set of strictly feasible z's:

$$\mathcal{Z} := \{z = (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m : (c(x), \lambda) > 0\}.$$

The algorithm generates a sequence of pairs (z, M), where $z \in \mathcal{Z}$ and M is a positive definite symmetric matrix. Given a pair (z, M), the next one (z_+, M_+) is obtained as follows. First

$$z_+ := z + \alpha d,$$

where $\alpha > 0$ is a step-size and $d = (d^x, d^\lambda)$ is the unique solution of (1.3). The uniqueness comes from the positivity of $c(x)$ and from the positive definiteness of M (for the unicity of d^x, use (1.5)). Next, the matrix M is updated into M_+ by the BFGS formula

$$M_+ := M - \frac{M\delta\delta^\top M}{\delta^\top M \delta} + \frac{\gamma\gamma^\top}{\gamma^\top \delta}, \tag{2.1}$$

where γ and δ are given by

$$\delta := x_+ - x \quad \text{and} \quad \gamma := \nabla_x \ell(x_+, \lambda_+) - \nabla_x \ell(x, \lambda_+). \tag{2.2}$$

This formula gives a symmetric positive definite matrix M_+, provided M is symmetric positive definite and $\gamma^\top \delta > 0$ (see [11, 14]). This latter condition is satisfied because of the strong convexity assumption. Indeed, since at least one of the functions f or $-c_{(i)}$ is strongly convex, for any fixed $\lambda > 0$, the function $x \mapsto \ell(x, \lambda)$ is strongly convex, that is, there exists a constant $a > 0$ such that

$$a\|x - x'\|^2 \le \left(\nabla_x \ell(x, \lambda) - \nabla_x \ell(x', \lambda)\right)^\top (x' - x), \quad \text{for all } x \text{ and } x'.$$

Since α sizes the displacement in x and λ, the merit function used to estimate the progress to the solution must depend on both x and λ. We follow an idea of Anstreicher and Vial [1] and add to φ_μ a function forcing λ to take the value $\mu C(x)^{-1} e$. The merit function is defined for $z = (x, \lambda) \in \mathcal{Z}$ by

$$\psi_\mu(z) := \varphi_\mu(x) + \mathcal{V}(z),$$

where

$$\mathcal{V}(z) = \lambda^\top c(x) - \mu \sum_{i=1}^m \log\left(\lambda_{(i)} c_{(i)}(x)\right).$$

Note that

$$\nabla \psi_\mu(z) = \begin{pmatrix} \nabla f(x) - 2\mu \nabla c(x) C(x)^{-1} e + \nabla c(x) \lambda \\ c(x) - \mu \Lambda^{-1} e \end{pmatrix}. \tag{2.3}$$

Using ψ_μ as a merit function is reasonable provided the problem

$$\begin{cases} \min \psi_\mu(z) \\ z \in \mathcal{Z} \end{cases} \tag{2.4}$$

has for unique solution the solution of (1.2) and the direction $d = (d^x, d^\lambda)$ is a descent direction of ψ_μ. This is what we check in Lemmas 2.5 and 2.6 below.

Lemma 2.5. *Suppose that Assumptions* 2.1 *hold. Then, Problem* (2.4) *has a unique solution* $\hat{z}_\mu := (\hat{x}_\mu, \hat{\lambda}_\mu)$, *where* $\hat{x}_\mu$ *is the unique solution of the barrier problem* (1.4) *and* $\hat{\lambda}_\mu$ *has its ith component defined by* $(\hat{\lambda}_\mu)_{(i)} := \mu / c_{(i)}(\hat{x}_\mu)$. *Furthermore,* ψ_μ *has no other stationary point than* $\hat{z}_\mu$.

Proof. By optimality of the unique solution $\hat{x}_\mu$ of the barrier problem (1.4)

$$\varphi_\mu(\hat{x}_\mu) \le \varphi_\mu(x), \quad \text{for any } x \text{ such that } c(x) > 0.$$

On the other hand, since $t \to t - \mu \log t$ is minimized at $t = \mu$ and since $\mu = c_{(i)}(\hat{x}_\mu)(\hat{\lambda}_\mu)_{(i)}$ for all index i, we have

$$\mathcal{V}(\hat{z}_\mu) \le \mathcal{V}(z), \quad \text{for any } z \in \mathcal{Z}.$$

Adding up the two preceding inequalities gives $\psi_\mu(\hat{z}_\mu) \le \psi_\mu(z)$ for all $z \in \mathcal{Z}$. Hence $\hat{z}_\mu$ is a solution of (2.4).

It remains to show that $\hat{z}_\mu$ is the unique stationary point of ψ_μ. If z is stationary, it satisfies

$$\left\{\begin{array}{rcl} \nabla f(x) - 2\mu\nabla c(x)\, C(x)^{-1}e + \nabla c(x)\,\lambda &=& 0 \\ c(x) - \mu\Lambda^{-1}e &=& 0. \end{array}\right.$$

Canceling λ from the first equality gives $\nabla f(x) - \mu\nabla c(x)C(x)^{-1}e = 0$, and thus $x = \hat{x}_\mu$ is the unique minimizer of the convex function φ_μ. Now, $\lambda = \hat{\lambda}_\mu$ by the second equation of the system above. □

Lemma 2.6. *Suppose that $z \in \mathcal{Z}$ and that M is symmetric positive definite. Let $d = (d^x, d^\lambda)$ be the solution of* (1.3). *Then*

$$\nabla\psi_\mu(z)^\top d = -(d^x)^\top(M+\nabla c(x)\Lambda C(x)^{-1}\nabla c(x)^\top)d^x - \|C(x)^{-1/2}\Lambda^{-1/2}(C(x)\lambda-\mu e)\|^2,$$

so that d is a descent direction of ψ_μ at a point $z \ne \hat{z}_\mu$, meaning that $\nabla\psi_\mu(z)^\top d < 0$.

Proof. We have, $\nabla\psi_\mu(z)^\top d = \nabla\varphi_\mu(x)^\top d^x + \nabla\mathcal{V}(z)^\top d$. Using (1.5),

$$\nabla\varphi_\mu(x)^\top d^x = -(d^x)^\top(M + \nabla c(x)C(x)^{-1}\Lambda\nabla c(x)^\top)d^x,$$

which is nonpositive. On the other hand, when d satisfies the second equation of (1.3), one has (see [1]):

$$\begin{array}{rcl} \nabla\mathcal{V}(z)^\top d &=& (\nabla c(x)\lambda - \mu\nabla c(x)\, C(x)^{-1}e)^\top d^x + (c(x) - \mu\Lambda^{-1}e)^\top d^\lambda \\ &=& (e - \mu C(x)^{-1}\Lambda^{-1}e)^\top(\Lambda\nabla c(x)^\top d^x + C(x)d^\lambda) \\ &=& -(\mu e - C(x)\lambda)^\top C(x)^{-1}\Lambda^{-1}(\mu e - C(x)\lambda) \\ &=& -\|C(x)^{-1/2}\Lambda^{-1/2}(C(x)\lambda - \mu e)\|^2, \end{array}$$

which is also nonpositive. The formula of $\nabla\psi_\mu(z)^\top d$ given in the statement of the lemma follows from this calculation. Furthermore $\nabla\psi_\mu(z)^\top d < 0$, if $z \ne \hat{z}_\mu$. □

We can now state precisely one iteration of the algorithm used to solve the perturbed KKT system (1.2). The constants $\omega \in\,]0,1[$ and $0 < \tau < \tau' < 1$ are given independently of the iteration index.

ALGORITHM A_μ for solving (1.2) (one iteration)

0. At the beginning of the iteration, the current iterate $z = (x, \lambda) \in \mathcal{Z}$ is supposed available, as well as a positive definite matrix M approximating the Hessian of the Lagrangian $\nabla^2_{xx}\ell(x,\lambda)$.
1. Compute $d := (d^x, d^\lambda)$, solution of the linear system (1.3).
2. Compute a step-size α by means of a backtracking line search.

   2.0. Set $\alpha = 1$.

2.1. Test the *sufficient decrease condition*:

$$\psi_\mu(z+\alpha d) \le \psi_\mu(z) + \omega\alpha\nabla\psi_\mu(z)^\top d. \tag{2.5}$$

2.2. If (2.5) is not satisfied, choose a new trial step-size α in $[\tau\alpha, \tau'\alpha]$ and go to Step 2.1. If (2.5) is satisfied, set $z_+ := z + \alpha d$.

3. Update M by the BFGS formula (2.1) where γ and δ are given by (2.2).

By Lemma 2.6, d is a descent direction of ψ_μ at z, so that a step-size $\alpha > 0$ satisfying (2.5) can be found. In the line-search, it is implicitly assumed that (2.5) is not satisfied if $z + \alpha d \notin \mathcal{Z}$, so that $(c(x_+), \lambda_+) > 0$ holds for the new iterate z_+.

We conclude this section with a result that gives the contribution of the line-search to the convergence of the sequence generated by Algorithm A_μ. It is in the spirit of a similar result given by Zoutendijk [41] (for a proof, see [6]). We say that a function is $C^{1,1}$ if it has Lipschitz continuous first derivatives. We denote the level set of ψ_μ determined by the first iterate $z_1 = (x_1, \lambda_1) \in \mathcal{Z}$ by

$$\mathcal{L}_1^{\mathrm{PD}} := \{z \in \mathcal{Z} : \psi_\mu(z) \le \psi_\mu(z_1)\}.$$

Lemma 2.7. *If ψ_μ is $C^{1,1}$ on an open convex neighborhood of the level set $\mathcal{L}_1^{\mathrm{PD}}$, there is a positive constant K, such that for any $z \in \mathcal{L}_1^{\mathrm{PD}}$, if α is determined by the line-search in Step 2 of algorithm A_μ, one of the following two inequalities holds:*

$$\psi_\mu(z+\alpha d) \le \psi_\mu(z) - K|\nabla\psi_\mu(z)^\top d|,$$

$$\psi_\mu(z+\alpha d) \le \psi_\mu(z) - K\frac{|\nabla\psi_\mu(z)^\top d|^2}{\|d\|^2}.$$

It is important to mention here that this result holds even though ψ_μ may not be defined for all positive step-sizes along d, so that the line-search may have to reduce the step-size in a first stage to enforce feasibility.

3 The global and r-linear convergence of Algorithm A_μ

In the BFGS theory, the track to the q-superlinear convergence traditionally goes through the r-linear convergence (see [33, 6]). In this section, we show that the iterates generated by Algorithm A_μ converge to $\hat{z}_\mu = (\hat{x}_\mu, \hat{\lambda}_\mu)$, the solution of (1.2), with that convergence speed. We use the notation

$$\hat{C}_\mu := \mathrm{diag}(c_{(1)}(\hat{x}_\mu), \dots, c_{(m)}(\hat{x}_\mu)) \quad \text{and} \quad \hat{\Lambda}_\mu := \mathrm{diag}((\hat{\lambda}_\mu)_{(1)}, \dots, (\hat{\lambda}_\mu)_{(m)}).$$

Our first result shows that, because the iterates (x, λ) remain in the level set $\mathcal{L}_1^{\mathrm{PD}}$, the sequence $\{(c(x), \lambda)\}$ is bounded and bounded away from zero.

Lemma 3.1. *Suppose that Assumptions* 2.1 *hold. Then, the level set* $\mathcal{L}_1^{\rm PD}$ *is compact and there exist positive constants* K_1 *and* K_2 *such that*

$$K_1 \le (c(x), \lambda) \le K_2, \quad \textit{for all } z \in \mathcal{L}_1^{\rm PD}.$$

Proof. Since $\lambda^\top c(x) - \mu \sum_i \log(\lambda_{(i)} c_{(i)}(x))$ is bounded below by $m\mu(1 - \log \mu)$, there is a constant $K_1' > 0$ such that $\varphi_\mu(x) \le K_1'$ for all $z = (x, \lambda) \in \mathcal{L}_1^{\rm PD}$. By Assumptions 2.1 and Lemma 2.3, the level set $\mathcal{L}' := \{x : c(x) > 0,\ \varphi_\mu(x) \le K_1'\}$ is compact. By continuity, $c(\mathcal{L}')$ is also compact, so that $c(x)$ is bounded and bounded away from zero for all $z \in \mathcal{L}_1^{\rm PD}$.

What we have just proven implies that $\{\varphi_\mu(x) : z = (x, \lambda) \in \mathcal{L}_1^{\rm PD}\}$ is bounded below, so that there is a constant $K_2' > 0$ such that $\lambda^\top c(x) - \mu \sum_i \log(\lambda_{(i)} c_{(i)}(x)) \le K_2'$ for all $z = (x, \lambda) \in \mathcal{L}_1^{\rm PD}$. Hence the λ-components of the z's in $\mathcal{L}_1^{\rm PD}$ are bounded and bounded away from zero.

We have shown that $\mathcal{L}_1^{\rm PD}$ is included in a compact set. Now, it is itself compact by continuity of ψ_μ. □

The next proposition is crucial for the technique we use to prove global convergence (see [6]). It claims that the proximity of a point z to the unique solution of (2.4) can be measured by the value of $\psi_\mu(z)$ or the norm of its gradient $\nabla\psi_\mu(z)$. In unconstrained optimization, the corresponding result is a direct consequence of strong convexity. Here, ψ_μ is not necessarily convex, but the result can still be established by using Lemma 2.5 and Lemma 3.1.

Proposition 3.2. *Suppose that Assumptions* 2.1 *hold. Then, there is a constant* $a > 0$ *such that for any* $z \in \mathcal{L}_1^{\rm PD}$

$$a\|z - \hat{z}_\mu\|^2 \le \psi_\mu(z) - \psi_\mu(\hat{z}_\mu) \le \frac{1}{a}\|\nabla\psi_\mu(z)\|^2. \tag{3.1}$$

Proof. Let us show that ψ_μ is strongly convex in a neighborhood of $\hat{z}_\mu$. Using (2.3) and the fact that $\hat{C}_\mu \hat{\lambda}_\mu = \mu e$, the Hessian of ψ_μ at $\hat{z}_\mu$ can be written:

$$\nabla^2\psi_\mu(\hat{z}_\mu) = \begin{pmatrix} \nabla^2_{xx}\ell(\hat{x}_\mu, \hat{\lambda}_\mu) + 2\mu\nabla c(\hat{x}_\mu)\, \hat{C}_\mu^{-2}\nabla c(\hat{x}_\mu)^\top & \nabla c(\hat{x}_\mu) \\ \nabla c(\hat{x}_\mu)^\top & \frac{1}{\mu}\hat{C}_\mu^2 \end{pmatrix}.$$

From Assumptions 2.1, for fixed $\lambda > 0$, the Lagrangian is a strongly convex function in the variable x. It follows that its Hessian with respect to x is positive definite at $(\hat{x}_\mu, \hat{\lambda}_\mu)$. Let us show that the above matrix is also positive definite. Multiplying the matrix on both sides by a vector $(u, v) \in \mathbb{R}^n \times \mathbb{R}^m$ gives

$$\begin{aligned} &u^\top\nabla^2_{xx}\ell(\hat{x}_\mu, \hat{\lambda}_\mu)u + 2\mu u^\top\nabla c(\hat{x}_\mu)\, \hat{C}_\mu^{-2}\nabla c(\hat{x}_\mu)^\top u + 2u^\top\nabla c(\hat{x}_\mu)v + \frac{1}{\mu}v^\top\hat{C}_\mu^2 v = \\ &\quad u^\top\nabla^2_{xx}\ell(\hat{x}_\mu, \hat{\lambda}_\mu)u + \mu u^\top\nabla c(\hat{x}_\mu)\hat{C}_\mu^{-2}\nabla c(\hat{x}_\mu)^\top u + \|\mu^{1/2}\hat{C}_\mu^{-1}\nabla c(\hat{x}_\mu)^\top u + \mu^{-1/2}\hat{C}_\mu v\|^2. \end{aligned}$$

Since $\nabla^2_{xx}\ell(\hat{x}_\mu, \hat{\lambda}_\mu)$ is positive definite and $c(\hat{x}_\mu) > 0$, this quantity is nonnegative. If it vanishes, one deduces that $u = 0$ and next that $v = 0$. Hence $\nabla^2\psi_\mu(\hat{z}_\mu)$ is positive definite.

Let us now prove a local version of the proposition: there exist a constant $a' > 0$ and a neighborhood V of $\hat{z}_\mu$ such that

$$a'\|z - \hat{z}_\mu\|^2 \leq \psi_\mu(z) - \psi_\mu(\hat{z}_\mu) \leq \frac{1}{a'}\|\nabla\psi_\mu(z)\|^2, \quad \text{for all } z \in V. \tag{3.2}$$

The inequality on the left comes from the fact that $\nabla\psi_\mu(\hat{z}_\mu) = 0$ and the strong convexity of ψ_μ near $\hat{z}_\mu$. For the inequality on the right, we first use the local convexity of ψ_μ: for an arbitrary z near $\hat{z}_\mu$, $\psi_\mu(\hat{z}_\mu) \geq \psi_\mu(z) + \nabla\psi_\mu(z)^\top(\hat{z}_\mu - z)$. With the Cauchy-Schwarz inequality and the inequality on the left of (3.2), one gets

$$\psi_\mu(z) - \psi_\mu(\hat{z}_\mu) \leq \|\nabla\psi_\mu(z)\| \left(\frac{\psi_\mu(z) - \psi_\mu(\hat{z}_\mu)}{a'}\right)^{\frac{1}{2}}.$$

Simplifying and squaring give the inequality on the right of (3.2).

We now extend the validity of (3.2) for all $z \in \mathcal{L}_1^{\mathrm{PD}}$ and start with the left-hand side inequality: there exists a constant $K_1' > 0$, such that

$$K_1'\|z - \hat{z}_\mu\|^2 \leq \psi_\mu(z) - \psi_\mu(\hat{z}_\mu), \quad \text{for all } z \in \mathcal{L}_1^{\mathrm{PD}}. \tag{3.3}$$

If this claim was false, there would exist a sequence $\{z_k\} \subset \mathcal{L}_1^{\mathrm{PD}}$ such that

$$\frac{1}{k}\|z_k - \hat{z}_\mu\|^2 > \psi_\mu(z_k) - \psi_\mu(\hat{z}_\mu), \quad \forall k \geq 1.$$

Since $\{z_k\}$ is bounded (Lemma 3.1), $\psi_\mu(z_k) \to \psi_\mu(\hat{z}_\mu)$ and, since $\hat{z}_\mu$ is the unique minimizer of ψ_μ (Lemma 2.5), $z_k \to \hat{z}_\mu$. Since $z_k \in V$ for k large, the inequality above would be in contradiction with (3.2) for k large.

Similarly, let us show that there exists a constant $K_2' > 0$, such that

$$\psi_\mu(z) - \psi_\mu(\hat{z}_\mu) \leq K_2'\|\nabla\psi_\mu(z)\|^2, \quad \text{for all } z \in \mathcal{L}_1^{\mathrm{PD}}. \tag{3.4}$$

If this claim was false, there would exist a sequence $\{z_k\} \subset \mathcal{L}_1^{\mathrm{PD}}$ such that

$$\psi_\mu(z_k) - \psi_\mu(\hat{z}_\mu) > k\|\nabla\psi_\mu(z_k)\|^2, \quad \forall k \geq 1.$$

Since $\{\psi_\mu(z_k)\}$ is bounded, $\nabla\psi_\mu(z_k) \to 0$. By the boundedness of $\{z_k\}$ and Lemma 2.5, $z_k \to \hat{z}_\mu$ and the inequality above would be in contradiction with (3.2) for k large.

The conclusion of the proposition now follows from (3.3) and (3.4) by taking $a = \min(K_1', 1/K_2')$. □

The proof of the r-linear convergence rests on the following lemma, which is part of the theory of BFGS updates. It can be stated independently of the present context (see Byrd and Nocedal [6]). We denote by θ_k the angle between $M_k\delta_k$ and δ_k:

$$\cos\theta_k := \frac{\delta_k^\top M_k \delta_k}{\|M_k\delta_k\|\,\|\delta_k\|},$$

and by $\lceil\cdot\rceil$ the roundup operator: $\lceil x\rceil = i$, when $i-1 < x \le i$ and $i \in \mathbb{N}$.

Lemma 3.3. *Let $\{M_k\}$ be positive definite matrices generated by the BFGS formula using pairs of vectors $\{(\gamma_k, \delta_k)\}_{k\ge 1}$, satisfying for all $k \ge 1$*

$$\gamma_k^\top \delta_k \ge a_1\|\delta_k\|^2 \quad \textit{and} \quad \gamma_k^\top \delta_k \ge a_2\|\gamma_k\|^2, \tag{3.5}$$

where $a_1 > 0$ and $a_2 > 0$ are independent of k. Then, for any $r \in]0,1[$, there exist positive constants b_1, b_2, and b_3, such that for any index $k \ge 1$,

$$b_1 \le \cos\theta_j \quad \textit{and} \quad b_2 \le \frac{\|M_j\delta_j\|}{\|\delta_j\|} \le b_3, \tag{3.6}$$

for at least $\lceil rk\rceil$ indices j in $\{1,\dots,k\}$.

The assumptions (3.5) made on γ_k and δ_k in the above lemma are satisfied in our context. The first one is due to the strong convexity of one of the functions f, $-c_{(1)}$, $\dots$, $-c_{(m)}$, and the fact that λ is bounded away from zero (Lemma 3.1). When f and c are $C^{1,1}$, the second one can be deduced from the Lipschitz inequality, the boundedness of λ (Lemma 3.1) and the first inequality in (3.5).

Theorem 3.4. *Suppose that Assumptions* 2.1 *hold and that f and c are $C^{1,1}$ functions. Then, Algorithm A_μ generates a sequence $\{z_k\}$ converging to $\hat z_\mu$ r-linearly, meaning that* $\limsup_{k\to\infty}\|z_k-\hat z_\mu\|^{1/k} < 1$. *In particular*

$$\sum_{k\ge 1}\|z_k - \hat z_\mu\| < \infty.$$

Proof. We denote by K_1', K_2', $\dots$ positive constants (independent of the iteration index). We also use the notation

$$c_j := c(x_j) \quad \text{and} \quad C_j := \operatorname{diag}(c_{(1)}(x_j),\dots,c_{(m)}(x_j)).$$

The bounds on $(c(x),\lambda)$ given by Lemma 3.1 and the fact that f and c are $C^{1,1}$ imply that ψ_μ is $C^{1,1}$ on an open convex neighborhood of the level set $\mathcal{L}_1^{\mathrm{PD}}$, namely on

$$\left(c^{-1}\left(\left]\frac{K_1}{2}, 2K_2\right[^m\right) \times \left]\frac{K_1}{2}, 2K_2\right[^m\right) \cap \mathcal{O},$$

where $\mathcal{O}$ is an open bounded convex set containing $\mathcal{L}_1^{\mathrm{PD}}$.

Therefore, by the line-search and Lemma 2.7, there is a positive constant K_1' such that either

$$\psi_\mu(z_{k+1}) \le \psi_\mu(z_k) - K_1' |\nabla\psi_\mu(z_k)^\top d_k| \tag{3.7}$$

or

$$\psi_\mu(z_{k+1}) \le \psi_\mu(z_k) - K_1' \frac{|\nabla\psi_\mu(z_k)^\top d_k|^2}{\|d_k\|^2}. \tag{3.8}$$

Let us now apply Lemma 3.3: fix $r \in]0,1[$ and denote by J the set of indices j for which (3.6) hold. Using Lemma 2.6 and the bounds from Lemma 3.1, one has for $j \in J$

$$\begin{aligned}
|\nabla\psi_\mu(z_j)^\top d_j| &= (d_j^x)^\top (M_j + \nabla c_j \Lambda_j C_j^{-1} \nabla c_j^\top) d_j^x + \|C_j^{-1/2}\Lambda_j^{-1/2}(C_j\lambda_j - \mu e)\|^2 \\
&\ge (d_j^x)^\top M_j d_j^x + K_2^{-2}\|C_j\lambda_j - \mu e\|^2 \\
&\ge \frac{b_1}{b_3}\|M_j d_j^x\|^2 + K_2^{-2}\|C_j\lambda_j - \mu e\|^2 \\
&\ge K_2' \left(\|M_j d_j^x\|^2 + \|C_j\lambda_j - \mu e\|^2\right).
\end{aligned}$$

Let us denote by K_4' a positive constant such that $\|\nabla c(x)\| \le K_4'$, for all $x \in \mathcal{L}_1^{\mathrm{PD}}$. By using (2.3), (1.5), and the inequality $(a+b+c)^2 \le 3(a^2+b^2+c^2)$, we obtain

$$\begin{aligned}
&\|\nabla\psi_\mu(z_j)\|^2 \\
&= \|\nabla_x\psi_\mu(z_j)\|^2 + \|\nabla_\lambda\psi_\mu(z_j)\|^2 \\
&= \| -(M_j + \nabla c_j C_j^{-1}\Lambda_j \nabla c_j^\top) d_j^x + \nabla c_j(\lambda_j - \mu C_j^{-1} e)\|^2 + \|c_j - \mu\Lambda_j^{-1} e\|^2 \\
&\le \left(\|M_j d_j^x\| + K_1^{-1}K_2 {K_4'}^2 \|d_j^x\| + K_1^{-1}K_4'\|C_j\lambda_j - \mu e\|\right)^2 + K_1^{-2}\|C_j\lambda_j - \mu e\|^2 \\
&\le 3\left(1 + \frac{K_1^{-2}K_2^2 {K_4'}^4}{b_2^2}\right)\|M_j d_j^x\|^2 + K_1^{-2}(3{K_4'}^2 + 1)\|C_j\lambda_j - \mu e\|^2 \\
&\le K_3' \left(\|M_j d_j^x\|^2 + \|C_j\lambda_j - \mu e\|^2\right)
\end{aligned}$$

and also, by (1.3),

$$\begin{aligned}
\|d_j\|^2 &= \|d_j^x\|^2 + \|d_j^\lambda\|^2 \\
&= \|d_j^x\|^2 + \|\mu C_j^{-1} e - \lambda_j - C_j^{-1}\Lambda_j \nabla c_j^\top d_j^x\|^2 \\
&\le \|d_j^x\|^2 + 2\|C_j^{-1}\Lambda_j \nabla c_j^\top d_j^x\|^2 + 2\|C_j^{-1}(C_j\lambda_j - \mu e)\|^2 \\
&\le \frac{1 + 2K_1^{-2}K_2^2 {K_4'}^2}{b_2^2}\|M_j d_j^x\|^2 + 2K_1^{-2}\|C_j\lambda_j - \mu e\|^2 \\
&\le K_5' \left(\|M_j d_j^x\|^2 + \|C_j\lambda_j - \mu e\|^2\right).
\end{aligned}$$

Combining these inequalities with (3.7) or (3.8) gives for some positive constant K_6' and for any $j \in J$:

$$\psi_\mu(z_{j+1}) \leq \psi_\mu(z_j) - K_6' \|\nabla \psi_\mu(z_j)\|^2.$$

The end of the proof is standard (see [33, 6]). Using Proposition 3.2, for $j \in J$:

$$\begin{aligned} \psi_\mu(z_{j+1}) - \psi_\mu(\hat{z}_\mu) &\leq \psi_\mu(z_j) - \psi_\mu(\hat{z}_\mu) - K_6' \|\nabla \psi_\mu(z_j)\|^2 \\ &\leq \tau^{\frac{1}{r}} (\psi_\mu(z_j) - \psi_\mu(\hat{z}_\mu)), \end{aligned}$$

where $\tau := (1 - K_6' a)^r \in [0, 1[$. On the other hand, by the line-search, $\psi_\mu(z_{k+1}) - \psi_\mu(\hat{z}_\mu) \leq \psi_\mu(z_k) - \psi_\mu(\hat{z}_\mu)$, for any $k \geq 1$. By Lemma 3.3, $|[1, k] \cap J| \geq \lceil rk \rceil \geq rk$, so that the last inequality gives for any $k \geq 1$:

$$\psi_\mu(z_{k+1}) - \psi_\mu(\hat{z}_\mu) \leq K_7' \tau^k,$$

where K_7' is the positive constant $(\psi_\mu(z_1) - \psi_\mu(\hat{z}_\mu))$. Now, using the inequality on the left in (3.1), one has for all $k \geq 1$:

$$\|z_{k+1} - \hat{z}_\mu\| \leq \frac{1}{\sqrt{a}} (\psi_\mu(z_{k+1}) - \psi_\mu(\hat{z}_\mu))^{\frac{1}{2}} \leq \left(\frac{K_7'}{a} \right)^{\frac{1}{2}} \tau^{\frac{k}{2}},$$

from which the r-linear convergence of $\{z_k\}$ follows. □

4 The q-superlinear convergence of Algorithm A_μ

With the r-linear convergence result of the previous section, we are now ready to establish the q-superlinear convergence of the sequence $\{z_k\}$ generated by Algorithm A_μ. By definition, $\{z_k\}$ converges q-superlinearly to $\hat{z}_\mu$ if the following estimate holds:

$$z_{k+1} - \hat{z}_\mu = o(\|z_k - \hat{z}_\mu\|),$$

which means that $\|z_{k+1} - \hat{z}_\mu\| / \|z_k - \hat{z}_\mu\| \to 0$ (assuming $z_k \neq \hat{z}_\mu$). To get this result, f and c have to be a little bit smoother, namely twice continuously differentiable near $\hat{x}_\mu$. We use the notation

$$\hat{M}_\mu := \nabla^2_{xx} \ell(\hat{x}_\mu, \hat{\lambda}_\mu).$$

We start by showing that the unit step-size is accepted asymptotically by the line-search condition (2.5), provided the updated matrix M_k becomes good (or sufficiently large) in a sense specified by inequality (4.1) below and provided the iterate z_k is sufficiently close to the solution $\hat{z}_\mu$.

Proposition 4.1. *Suppose that Assumptions* 2.1 *hold and that* f *and* c *are twice continuously differentiable near* $\hat{x}_\mu$. *Suppose also that the sequence* $\{z_k\}$ *generated by Algorithm* A_μ *converges to* $\hat{z}_\mu$ *and that the positive definite matrices* M_k *satisfy the estimate*

$$(d_k^x)^\top \left(M_k - \hat{M}_\mu\right) d_k^x \geq o(\|d_k^x\|^2), \tag{4.1}$$

when $k \to \infty$. *Then the sufficient decrease condition* (2.5) *is satisfied with* $\alpha_k = 1$ *for* k *sufficiently large, provided that* $\omega < \frac{1}{2}$.

Proof. Observe first that the positive definiteness of $\hat{M}_\mu$ with (4.1) implies that

$$(d_k^x)^\top M_k d_k^x \geq K' \|d_k^x\|^2, \tag{4.2}$$

for some positive constant K' and sufficiently large k. Observe also that $d_k \to 0$ (for $d_k^x \to 0$, use (1.5), (4.2), and $\nabla\varphi_\mu(x_k) \to 0$). Therefore, for k large enough, z_k and $z_k + d_k$ are near $\hat{z}_\mu$ and one can expand $\psi_\mu(z_k + d_k)$ about z_k. A second order expansion gives for the left-hand side of (2.5):

$$\begin{aligned}
&\psi_\mu(z_k + d_k) - \psi_\mu(z_k) - \omega \nabla\psi_\mu(z_k)^\top d_k \\
&= (1-\omega)\nabla\psi_\mu(z_k)^\top d_k + \frac{1}{2} d_k^\top \nabla^2\psi_\mu(z_k) d_k + o(\|d_k\|^2) \\
&= \left(\frac{1}{2} - \omega\right) \nabla\psi_\mu(z_k)^\top d_k \\
&\quad + \frac{1}{2}\left(\nabla\psi_\mu(z_k)^\top d_k + d_k^\top \nabla^2\psi_\mu(z_k) d_k\right) + o(\|d_k\|^2).
\end{aligned} \tag{4.3}$$

We want to show that this quantity is negative for k large.

Our first aim is to show that $\left(\nabla\psi_\mu(z_k)^\top d_k + d_k^\top \nabla^2\psi_\mu(z_k) d_k\right)$ is smaller than a term of order $o(\|d_k\|^2)$. For this purpose, one computes

$$\begin{aligned}
&d_k^\top \nabla^2\psi_\mu(z_k) d_k \\
&= (d_k^x)^\top \nabla_{xx}^2 \ell(x_k, \tilde{\lambda}_k) d_k^x + 2\mu (d_k^x)^\top \nabla c_k C_k^{-2} \nabla c_k^\top d_k^x \\
&\quad + 2(d_k^x)^\top \nabla c_k d_k^\lambda + \mu (d_k^\lambda)^\top \Lambda_k^{-2} d_k^\lambda,
\end{aligned}$$

where $\tilde{\lambda}_k = 2\mu C_k^{-1} e - \lambda_k$. On the other hand, using

$$C_k^{-1/2}\Lambda_k^{-1/2}(C_k\lambda_k - \mu e) = -C_k^{-1/2}\Lambda_k^{1/2}\nabla c_k^\top d_k^x - C_k^{1/2}\Lambda_k^{-1/2} d_k^\lambda,$$

one gets from Lemma 2.6

$$\begin{aligned}
&\nabla\psi_\mu(z_k)^\top d_k \\
&= -(d_k^x)^\top M_k d_k^x - (d_k^x)^\top \nabla c_k C_k^{-1}\Lambda_k \nabla c_k^\top d_k^x - \|C_k^{-1/2}\Lambda_k^{-1/2}(C_k\lambda_k - \mu e)\|^2 \\
&= -(d_k^x)^\top M_k d_k^x - 2(d_k^x)^\top \nabla c_k C_k^{-1}\Lambda_k \nabla c_k^\top d_k^x - 2(d_k^x)^\top \nabla c_k d_k^\lambda - (d_k^\lambda)^\top C_k \Lambda_k^{-1} d_k^\lambda.
\end{aligned}$$

With these estimates, (4.1), and the fact that $\nabla^2_{xx}\ell(x_k,\tilde{\lambda}_k) \to \hat{M}_\mu$ and $C_k\lambda_k \to \mu e$, with Lemma 3.1 and the boundedness of $\{\nabla c_k\}$, (4.3) becomes

$$\begin{aligned}
&\psi_\mu(z_k+d_k) - \psi_\mu(z_k) - \omega\nabla\psi_\mu(z_k)^\top d_k \\
&= \left(\frac{1}{2}-\omega\right)\nabla\psi_\mu(z_k)^\top d_k \\
&\quad -\frac{1}{2}(d_k^x)^\top\left(M_k - \nabla^2_{xx}\ell(x_k,\tilde{\lambda}_k)\right)d_k^x + (d_k^x)^\top\nabla c_k\left(\mu C_k^{-2} - C_k^{-1}\Lambda_k\right)\nabla c_k^\top d_k^x \\
&\quad +\frac{1}{2}(d_k^\lambda)^\top\left(\mu\Lambda_k^{-2} - C_k\Lambda_k^{-1}\right)d_k^\lambda + o(\|d_k\|^2) \\
&\le \left(\frac{1}{2}-\omega\right)\nabla\psi_\mu(z_k)^\top d_k + o(\|d_k\|^2).
\end{aligned} \tag{4.4}$$

Since $\omega < \frac{1}{2}$, it is clear that the result will be proven if we show that, for some positive constant K and k large $\nabla\psi_\mu(z_k)^\top d_k \le -K\|d_k\|^2$. To show this, we use the last expression of $\nabla\psi_\mu(z_k)^\top d_k$ and an upper bound of $|(d_k^x)^\top\nabla c_k d_k^\lambda|$, obtained by the Cauchy-Schwartz inequality:

$$\begin{aligned}
2\left|(d_k^x)^\top\nabla c_k d_k^\lambda\right| &= 2\left|\left(C_k^{-1/2}\Lambda_k^{1/2}\nabla c_k^\top d_k^x\right)^\top\left(C_k^{1/2}\Lambda_k^{-1/2}d_k^\lambda\right)\right| \\
&\le 2\left\|C_k^{-1/2}\Lambda_k^{1/2}\nabla c_k^\top d_k^x\right\|\left\|C_k^{1/2}\Lambda_k^{-1/2}d_k^\lambda\right\| \\
&\le \frac{3}{2}(d_k^x)^\top\nabla c_k C_k^{-1}\Lambda_k\nabla c_k^\top d_k^x + \frac{2}{3}(d_k^\lambda)^\top C_k\Lambda_k^{-1}d_k^\lambda.
\end{aligned}$$

It follows that

$$\nabla\psi_\mu(z_k)^\top d_k \le -(d_k^x)^\top M_k d_k^x - \frac{1}{2}(d_k^x)^\top\nabla c_k C_k^{-1}\Lambda_k\nabla c_k^\top d_k^x - \frac{1}{3}(d_k^\lambda)^\top C_k\Lambda_k^{-1}d_k^\lambda.$$

Therefore, using (4.2) and Lemma 3.1, one gets

$$\nabla\psi_\mu(z_k)^\top d_k \le -K\|d_k\|^2.$$

for some positive constant K and k large. □

Proposition 4.1 shows in particular that the function $\mathcal{V}$, which was added to φ_μ to get the merit function ψ_μ, has the right curvature around $\hat{z}_\mu$, so that the unit step-size in both x and λ is accepted by the line-search.

In the following proposition, we establish a necessary and sufficient condition of q-superlinear convergence of the Dennis and Moré [10] type. The analysis assumes that the unit step-size is taken and that the updated matrix M_k is sufficiently good asymptotically in a manner given by the estimate (4.5), which is stronger than (4.1).

Proposition 4.2. *Suppose that Assumptions* 2.1 *hold and that f and c are twice differentiable at $\hat{x}_\mu$. Suppose that the sequence $\{z_k\}$ generated by Algorithm A_μ converges to $\hat{z}_\mu$ and that, for k sufficiently large, the unit step-size $\alpha_k = 1$ is accepted by the line-search. Then $\{z_k\}$ converges q-superlinearly towards $\hat{z}_\mu$ if and only if*

$$(M_k - \hat{M}_\mu)d_k^x = o(\|d_k\|). \tag{4.5}$$

Proof. Let us denote by $\mathcal{M}$ the nonsingular Jacobian matrix of the perturbed KKT conditions (1.2) at the solution $\hat{z}_\mu = (\hat{x}_\mu, \hat{\lambda}_\mu)$:

$$\mathcal{M} = \begin{pmatrix} \hat{M}_\mu & -\nabla c(\hat{x}_\mu) \\ \hat{\Lambda}_\mu \nabla c(\hat{x}_\mu)^\top & \hat{C}_\mu \end{pmatrix}.$$

A first order expansion of the right hand side of (1.3) about $\hat{z}_\mu$ and the identities $\nabla f(\hat{x}_\mu) = \nabla c(\hat{x}_\mu)\hat{\lambda}_\mu$ and $\hat{C}_\mu \hat{\lambda}_\mu = \mu e$ give

$$\begin{pmatrix} M_k & -\nabla c_k \\ \Lambda_k \nabla c_k^\top & C_k \end{pmatrix} \begin{pmatrix} d_k^x \\ d_k^\lambda \end{pmatrix} = -\mathcal{M}(z_k - \hat{z}_\mu) + o(\|z_k - \hat{z}_\mu\|).$$

Subtracting $\mathcal{M} d_k$ to both sides and assuming a unit step-size, we obtain

$$\begin{aligned} &\begin{pmatrix} M_k - \hat{M}_\mu & -(\nabla c_k - \nabla c(\hat{x}_\mu)) \\ \Lambda_k \nabla c_k^\top - \hat{\Lambda}_\mu \nabla c(\hat{x}_\mu)^\top & C_k - \hat{C}_\mu \end{pmatrix} \begin{pmatrix} d_k^x \\ d_k^\lambda \end{pmatrix} \\ &\qquad = -\mathcal{M}(z_{k+1} - \hat{z}_\mu) + o(\|z_k - \hat{z}_\mu\|). \end{aligned} \tag{4.6}$$

Suppose now that $\{z_k\}$ converges q-superlinearly. Then, the right hand side of (4.6) is of order $o(\|z_k - \hat{z}_\mu\|)$, so that

$$(M_k - \hat{M}_\mu)d_k^x + o(\|d_k^\lambda\|) = o(\|z_k - \hat{z}_\mu\|).$$

Then (4.5) follows from the fact that, by the q-superlinear convergence of $\{z_k\}$, $z_k - \hat{z}_\mu = O(\|d_k\|)$.

Let us now prove the converse. By (4.5), the left hand side of (4.6) is a $o(\|d_k\|)$ and due to the nonsingularity of $\mathcal{M}$, (4.6) gives $z_{k+1} - \hat{z}_\mu = o(\|z_k - \hat{z}_\mu\|) + o(\|d_k\|)$. With a unit step-size, $d_k = (z_{k+1} - \hat{z}_\mu) - (z_k - \hat{z}_\mu)$, so that we finally get $z_{k+1} - \hat{z}_\mu = o(\|z_k - \hat{z}_\mu\|)$. □

Interestingly enough, the dual step intervenes in the right hand side of (4.5), while M_k is a primal quantity. We are also interested in the superlinear convergence of the pair $z_k = (x_k, \lambda_k)$, not just x_k. In fact, we will see that the BFGS formula gives the strongest estimate

$$(M_k - \hat{M}_\mu)d_k^x = o(\|d_k^x\|). \tag{4.7}$$

It is quite normal that d_k^λ does not intervene in the BFGS estimate (4.7), because the update of M_k uses primal quantities only. Now, (4.7) implies (4.5) and therefore the q-superlinear convergence of $\{z_k\}$.

For proving the q-superlinear convergence of the sequence $\{z_k\}$, we need the following result from the BFGS theory (see [32, Theorem 3] and [6]).

Lemma 4.3. *Let $\{M_k\}$ be a sequence of matrices generated by the BFGS formula from a given symmetric positive definite matrix M_1 and pairs (γ_k, δ_k) of vectors verifying*

$$\gamma_k^\top \delta_k > 0, \text{ for all } k \geq 1 \qquad \text{and} \qquad \sum_{k\geq 1} \frac{\|\gamma_k - M\delta_k\|}{\|\delta_k\|} < \infty, \tag{4.8}$$

where M is a symmetric positive definite matrix. Then, the sequences $\{M_k\}$ and $\{M_k^{-1}\}$ are bounded and

$$(M_k - M)\delta_k = o(\|\delta_k\|). \tag{4.9}$$

A function ϕ, twice differentiable in a neighborhood of a point $x \in \mathbb{R}^n$, is said to have a *locally radially Lipschitzian* Hessian at x, if there exists a positive constant L such that for x' near x, one has

$$\|\nabla^2\phi(x) - \nabla^2\phi(x')\| \leq L\|x - x'\|.$$

Theorem 4.4. *Suppose that Assumptions* 2.1 *hold and that f and c are $C^{1,1}$ functions, twice continuously differentiable near $\hat{x}_\mu$ with locally radially Lipschitzian Hessians at $\hat{x}_\mu$. Suppose that the line-search in Algorithm A_μ uses the constant $\omega < \frac{1}{2}$. Then the sequence $\{z_k\} = \{(x_k, \lambda_k)\}$ generated by this algorithm converges to $\hat{z}_\mu = (\hat{x}_\mu, \hat{\lambda}_\mu)$ q-superlinearly and, for k sufficiently large, the unit step-size $\alpha_k = 1$ is accepted by the line-search.*

Proof. Let us start by showing that Lemma 4.3 with $M = \hat{M}_\mu$ can be applied. First, $\gamma_k^\top \delta_k > 0$, as this was already discussed after Lemma 3.3. For the convergence of the series in (4.8), we use a Taylor expansion, assuming that k is large enough (f and c are C^2 near $\hat{x}_\mu$):

$$\begin{aligned} \gamma_k - \hat{M}_\mu\delta_k &= \int_0^1 \left(\nabla_{xx}^2\ell(x_k + t\delta_k, \lambda_{k+1}) - \nabla_{xx}^2\ell(\hat{x}_\mu, \lambda_{k+1})\right)\delta_k \,\mathrm{d}t \\ &\quad + \left(\nabla_{xx}^2\ell(\hat{x}_\mu, \lambda_{k+1}) - \hat{M}_\mu\right)\delta_k \end{aligned}$$

With the local radial Lipschitz continuity of $\nabla^2 f$ and $\nabla^2 c$ at $\hat{x}_\mu$ and the boundedness of $\{\lambda_{k+1}\}$, there exist positive constants K' and K'' such that

$$\begin{aligned} \|\gamma_k - \hat{M}_\mu\,\delta_k\| &\leq K'\|\delta_k\|\left(\int_0^1 \|x_k + t\,\delta_k - \hat{x}_\mu\|\mathrm{d}t + \|\lambda_{k+1} - \hat{\lambda}_\mu\|\right) \\ &\leq K'\|\delta_k\|\Bigg(\int_0^1 \Big((1-t)\|x_k - \hat{x}_\mu\| + t\|x_{k+1} - \hat{x}_\mu\|\Big)\,\mathrm{d}t \\ &\qquad\qquad + \|\lambda_{k+1} - \hat{\lambda}_\mu\|\Bigg) \\ &= K''\|\delta_k\|\Big(\|x_k - \hat{x}_\mu\| + \|z_{k+1} - \hat{z}_\mu\|\Big). \end{aligned}$$

Hence the series in (4.8) converges by Theorem 3.4. Therefore, by (4.9) with $M = \hat{M}_\mu$ and the fact that δ_k is parallel to d_k^x:

$$(M_k - \hat{M}_\mu)d_k^x = o(\|d_k^x\|). \tag{4.10}$$

By the estimate (4.10) and Proposition 4.1, the unit step-size is accepted when k is large enough. The q-superlinear convergence of $\{z_k\}$ now follows from Proposition 4.2. □

5 The overall primal-dual algorithm

In this section, we consider an overall algorithm for solving Problem (1.1). Recall from Lemma 2.2 that the set of primal solutions of this problem is nonempty and bounded. By the Slater condition (Assumption 2.1 (*ii*)), the set of dual solutions is also nonempty and bounded. Let us denote by $\hat{z} = (\hat{x}, \hat{\lambda})$ a primal-dual solution of Problem (1.1), which is also a solution of the necessary and sufficient conditions of optimality

$$\begin{cases} \nabla f(\hat{x}) - \nabla c(\hat{x})\hat{\lambda} = 0 \\ C(\hat{x})\hat{\lambda} = 0 \\ (c(\hat{x}), \hat{\lambda}) \geq 0. \end{cases} \tag{5.1}$$

Our overall algorithm for solving (1.1) or (5.1), called Algorithm A, consists in computing approximate solutions of the perturbed optimality conditions (1.2), for a sequence of μ's converging to zero. For each μ, the primal-dual Algorithm A_μ is used to find an approximate solution of (1.2). This is done by so-called *inner* iterations. Next μ is decreased and the process of solving (1.2) for the new value of μ is repeated. We call *outer* iteration the collection of inner iterations for solving (1.2) for a fixed value of μ. We index the outer iterations by superscripts $j \in \mathbb{N}^*$.

ALGORITHM A for solving Problem (1.1) (one outer iteration)

0. At the beginning of the jth outer iteration, an approximation $z_1^j := (x_1^j, \lambda_1^j) \in \mathcal{Z}$ of the solution $\hat{z}$ of (5.1) is supposed available, as well as a positive definite matrix M_1^j approximating the Hessian of the Lagrangian. A value $\mu^j > 0$ is given, as well as a precision threshold $\epsilon^j > 0$.
1. Use Algorithm A_μ, starting from z_1^j, to solve (1.2) with $\mu = \mu^j$ approximately, meaning that the algorithm stops as soon as an iterate $z^j := (x^j, \lambda^j)$ satisfies
$$\|\nabla f(x^j) - \nabla c(x^j)\lambda^j\| \leq \epsilon^j \quad \text{and} \quad \|C(x^j)\lambda^j - \mu^j e\| \leq \epsilon^j.$$
2. Choose a new starting iterate $z_1^{j+1} \in \mathcal{Z}$ for the next outer iteration, as well as a positive definite matrix M_1^{j+1}. Set the new parameters $\mu^{j+1} > 0$ and $\epsilon^{j+1} > 0$, such that $\{\mu^j\}$ and $\{\epsilon^j\}$ converge to zero when $j \to \infty$.

To start the $(j+1)$th outer iteration, a possibility is to take $z_1^{j+1} = z^j$ and $M_1^{j+1} = M^j$, the updated matrix obtained at the end of the jth outer iteration.

As far as the global convergence is concerned, how z^j, M^j and μ^j are determined is not important. Therefore, on that point, Algorithm A leaves the user much freedom of maneuver, while Theorem 5.1 gives us a global convergence result for such a general algorithm.

Theorem 5.1. *Suppose that Assumptions* 2.1 *hold and that f and c are $C^{1,1}$ functions. Then Algorithm A generates a bounded sequence $\{z^j\}$ and any limit point of $\{z^j\}$ is a primal-dual solution of Problem* (1.1).

Proof. By Theorem 3.4, any outer iteration of Algorithm A terminates with an iterate z^j satisfying the stopping criteria in Step 1. Therefore Algorithm A generates a sequence $\{z^j\}$. Since the sequences $\{\mu^j\}$ and $\{\epsilon^j\}$ converge to zero, any limit point of $\{z^j\}$ is solution of Problem (1.1). It remains to show that $\{z^j\}$ is bounded.

Let us first prove the boundedness of $\{x^j\}$. The convexity of the Lagrangian implies that

$$\ell(x^j, \lambda^j) + \nabla_x \ell(x^j, \lambda^j)^\top (x^1 - x^j) \le \ell(x^1, \lambda^j).$$

Using the positivity of λ^j and $c(x^1)$ and next the stopping criteria of Algorithm A, it follows that

$$\begin{aligned} f(x^j) &\le f(x^1) + (\lambda^j)^\top c(x^j) + \nabla_x \ell(x^j, \lambda^j)^\top (x^j - x^1) \\ &\le f(x^1) + o(1) + o(\|x^j - x^1\|). \end{aligned}$$

If $\{x^j\}$ is unbounded, setting $t^j := \|x^j - x^1\|$ and $y^j := \frac{x^j - x^1}{t^j}$, one can choose a subsequence J such that

$$\lim_{\substack{j \to +\infty \\ j \in J}} t^j = +\infty \quad \text{and} \quad \lim_{\substack{j \to +\infty \\ j \in J}} y^j = y \neq 0.$$

From the last inequality we deduce that

$$f'_\infty(y) := \lim_{\substack{j \to +\infty \\ j \in J}} \frac{f(x^1 + t^j y^j) - f(x^1)}{t^j} \le 0,$$

Moreover, since $c(x^j) > 0$, we have $(-c_{(i)})'_\infty(y) \le 0$, for $i = 1, \dots, m$. It follows that $\hat{x} + \mathbb{R}_+ y \subset \{x : c(x) \ge 0,\ f(x) \le f(\hat{x})\}$ (see for example [20, Proposition IV.3.2.5] or [2, formula (1)]). Therefore, the solution set of Problem (1.1) would be unbounded, which is in contradiction with what is claimed in Lemma 2.2.

To prove the boundedness of the multipliers, suppose that the algorithm generates an unbounded sequence of positive vectors $\{\lambda^j\}_{j \in J}$. The sequence $\{(x^j, \lambda^j / \|\lambda^j\|)\}_{j \in J}$ is bounded and thus has at least one limit point, say (x^*, ν^*). This one satisfies $\nu^* \geq 0$, $\nabla c(x^*)\nu^* = 0$ and $(\nu^*)^\top c(x^*) = 0$. Using the concavity of the components $c_{(i)}$, one has

$$c(x^*) + \nabla c(x^*)^\top (x^1 - x^*) \geq c(x^1) > 0,$$

where the inequality on the right follows from the strict feasibility of the first iterate. Multiplying by ν^*, we deduce that $(\nu^*)^\top c(x^1) = 0$, and thus $\nu^* = 0$, a contradiction with $\|\nu^*\| = 1$ $\square$

Note that the strong convexity hypothesis given in Assumption 2.1 (i) was used in the proof only for ensuring that Algorithm A_μ converges for any positive value of μ.

6 Discussion

By way of conclusion, we discuss the results obtained in this paper and raise some open questions.

Problems with linear constraints

The algorithm is presented with convex inequality constraints only, but it can also be used when linear constraints are present. Consider the problem

$$\left\{ \begin{array}{l} \min f(x) \\ Ax = b \\ c(x) \geq 0, \end{array} \right. \tag{6.1}$$

obtained by adding linear constraints to Problem (1.1). In (6.1), A is an $m \times n$ matrix with $m < n$ and $b \in \mathbb{R}^m$ is given in the range space of A.

Problem (6.1) can be reduced to Problem (1.1) by using a basis of the null space of the matrix A. Indeed, let x_1 be the first iterate, which is supposed to be strictly feasible in the sense that

$$Ax_1 = b \quad \text{and} \quad c(x_1) > 0.$$

Let us denote by Z a $n \times p$ matrix whose columns form a basis of the null space of A. Then, any point satisfying the linear constraints of (6.1) can be written

$$x = x_1 + Zu, \quad \text{with } u \in \mathbb{R}^p.$$

With this notation, Problem (6.1) can be re-written as the problem in $u \in \mathbb{R}^p$:

$$\left\{ \begin{array}{l} \min f(x_1 + Zu) \\ c(x_1 + Zu) \geq 0, \end{array} \right. \tag{6.2}$$

which has the form (1.1).

Thanks to this transformation, we can deduce from Assumptions 2.1 what are the minimal assumptions under which our algorithm for solving Problem (6.2) or, equivalently, Problem (6.1) will converge.

Assumptions 6.1. (i) The functions f and $-c_{(i)}$ $(1 \leq i \leq m)$ are convex and differentiable from $\mathbb{R}^n$ to $\mathbb{R}$ and at least one of them is strongly convex. (ii) There exists an $x \in \mathbb{R}^n$, such that $Ax = b$ and $c(x) > 0$.

With these assumptions, all the previous results apply. In particular, Algorithm A_μ converges r-linearly (if f and c are also $C^{1,1}$) and q-superlinearly (if f and c are also $C^{1,1}$, twice continuously differentiable near $\hat{x}_\mu$ with locally radially Lipschitzian Hessian at $\hat{x}_\mu$). Similarly, the conclusions of Theorem 5.1 applies if f and c are also $C^{1,1}$.

Feasible algorithms and quasi-Newton techniques

In the framework of quasi-Newton methods, the property of having to generate feasible iterates should not be only viewed as a restriction limiting the applicability of a feasible algorithm. Indeed, in the case of Problem (6.2), if it is sometimes difficult to find a strictly feasible initial iterate, the matrix to update for solving this problem is of order p only, instead of order n for an infeasible algorithm solving Problem (6.1) directly. When $p \ll n$, the quasi-Newton updates will approach the reduced Hessian of the Lagrangian $Z^\top(\nabla^2\ell)Z$ more rapidly than the full Hessian $\nabla^2\ell$, so that a feasible algorithm is likely to converge more rapidly.

About the strong convexity hypothesis

Another issue concerns the extension of the present theory to convex problems, without the strong convexity assumption (Assumption 2.1 (i)).

Without this hypothesis, the class of problems to consider encompasses linear programming (f and c are affine). It is clear that for dealing properly with linear programs, our algorithm needs modifications, since then $\gamma_k = 0$ and the BFGS formula is no longer defined. Of course, it would be very ineffective to solve linear programs with the quasi-Newton techniques proposed in this paper ($M_k = 0$ is the desired matrix), but problems that are almost linear near the solution may be encountered, so that a technique for dealing with a situation where $\|\gamma_k\| \ll \|\delta_k\|$ can be of interest.

To accept $\gamma_k = 0$, one can look at the limit of the BFGS formula (2.1) when $\gamma_k \to 0$. A possible update formula could be

$$M_{k+1} := M_k - \frac{M_k \delta_k \delta_k^\top M_k}{\delta_k^\top M_k \delta_k}.$$

The updated matrix satisfies $M_{k+1}\delta_k = 0$ and is positive semi-definite, provided M_k is already positive semi-definite. The fact that M_{k+1} may be singular raises some difficulties, however. For example, the search direction d^x may no longer be defined (see formula (1.5), in which the matrix $M + \nabla c(x)C(x)^{-1}\Lambda\nabla c(x)^\top$ can be singular). Therefore, the present theory cannot be extended in a straightforward manner.

On the other hand, the strong convexity assumption may not be viewed as an important restriction, because a fictive strongly convex constraint can always be added. An obvious example of fictive constraint is "$x^\top x \le K$". If the constant K is large enough, the constraint is inactive at the solution, so that the solution of the original problem is not altered by this new constraint and the present theory applies.

Computational aspects

In unconstrained optimization, it is sometimes advantageous to update approximations of the inverse Hessian of the objective, in order to avoid the $O(n^3)$ operations used to solve the linear system defining the descent direction. In the present situation, it is not clear whether updating $W = M^{-1}$ instead of M would be of interest. Indeed, the linear system (1.5) defining d^x involves the matrix $M + \nabla c(x)C(x)^{-1}\Lambda\nabla c(x)^\top$, and not only M. It would be important to find a technique for solving this linear system in only $O(n^2)$ operations, knowing that M (or W or the Cholesky factors of M) is updated by the BFGS formula.

Better control of the outer iterations

Last but not least, the global convergence result of Section 5 is independent of the update rule of the parameters ϵ^j and μ^j. In practice, however, the choice of the decreasing values ϵ^j and μ^j is important for the efficiency of the algorithm. From a theoretical viewpoint, it would be highly desirable to have update rules for ϵ^j and μ^j that would allow the outer iterates of Algorithm A to converge q-superlinearly.

References

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

[2] A. Auslender, R. Cominetti, M. Haddou (1997). Asymptotic analysis for penalty and barrier methods in convex and linear programming. *Mathematics of Operations Research*, 22, 43–62.

[3] M. Berz, C. Bischof, G. Corliss, A. Griewank (editors) (1996). *Computational Differentiation: Techniques, Applications, and Tools.* SIAM, Philadelphia.

[4] J.F. Bonnans, C. Pola (1997). A trust region interior point algorithm for linearly constrained optimization. *SIAM Journal on Optimization*, 7, 717–731.

[5] R.H. Byrd, J.Ch. Gilbert, J. Nocedal (1996). A trust region method based on interior point techniques for nonlinear programming. Rapport de Recherche 2896, INRIA. (submitted to *Mathematical Programming*).

[6] R.H. Byrd, J. Nocedal (1989). A tool for the analysis of quasi-Newton methods with application to unconstrained minimization. *SIAM Journal on Numerical Analysis*, 26, 727–739.

[7] A.R. Conn, N.I.M. Gould, Ph.L. Toint (1996). A primal-dual algorithm for minimizing a non-convex function subject to bound and linear equality constraints. Technical report, Facultés Universitaires de Namur, Belgique.

[8] G. Corliss, A. Griewank (editors) (1991). *Automatic Differentiation of Algorithms: Theory, Implementation, and Application*, Proceedings in Applied Mathematics 53. SIAM, Philadelphia.

[9] P. Courtier, J.N. Thépaut, A. Hollingsworth (1994). A strategy for operational implementation of 4D-Var, using an incremental approach. *Quaterly Journal of the Royal Meteorological Society*, 120, 1367–1387.

[10] J.E. Dennis, J.J. Moré (1974). A characterization of superlinear convergence and its application to quasi-Newton methods. *Mathematics of Computation*, 28, 549–560.

[11] J.E. Dennis, R.B. Schnabel (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliffs.

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

[13] A.V. Fiacco, G.P. McCormick (1968). *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. John Wiley, New York.

[14] R. Fletcher (1987). *Practical Methods of Optimization* (second edition). John Wiley & Sons, Chichester.

[15] J.Ch. Gilbert, G. Le Vey, J. Masse (1991). La différentiation automatique de fonctions représentées par des programmes. Rapport de Recherche n° 1557, INRIA, BP 105, F-78153 Le Chesnay, France. `http://www.inria.fr/RRRT/RR-1557.html`; `ftp://ftp.inria.fr/INRIA/publication/RR`, `RR-1557.ps.gz`.

[16] J. Herskovits (1993). An interior point technique for nonlinear optimization. Technical report, Mechanical Engineering Program, COPPE, Federal University of Rio de Janeiro.

[17] D. den Hertog (1992). *Interior Point Approach to Linear, Quadratic and Convex Programming*. Mathematics and its Applications 277. Kluwer Academic Publishers, Dordrecht.

[18] D. den Hertog, C. Roos, T. Terlaky (1990). A potential reduction method for a class of smooth convex programming problems. Report 90-01, Faculty of Technical Mathematics and Informatics, Technische Universiteit Delft, Holland.

[19] D. den Hertog, C. Roos, T. Terlaky (1992). On the classical logarithmic barrier function method for a class of smooth convex programming problems. *Journal of Optimization Theory and Applications*, 73, 1–25.

[20] J.-B. Hiriart-Urruty, C. Lemaréchal (1993). *Convex Analysis and Minimization Algorithms*. Grundlehren der mathematischen Wissenschaften 305-306. Springer-Verlag.

[21] B. Jansen (1997). *Interior Point Techniques in Optimization – Complementarity, Sensitivity and Algorithms*. Applied Optimization 6. Kluwer Academic Publishers, Dordrecht.

[22] F. Jarre (1989). On the method of analytic centers for solving smooth convex problems. Lecture Notes in Mathematics 1405, pages 69–85. Springer-Verlag, Berlin.

[23] F. Jarre (1996). Interior-point methods for classes of convex programs. In T. Terlaky (editor), *Interior Point Methods of Mathematical Programming*, chapter 7. Kluwer Academic Press, Dordrecht.

[24] F. Jarre, M.A. Saunders (1995). A practical interior-point method for convex programming. *SIAM Journal on Optimization*, 5, 149–171.

[25] S. Jégou (1997). *Estimation des perméabilités relatives dans des expériences de déplacements triphasiques en milieu poreux*. PhD Thesis, Université de Paris IX (Dauphine), Paris.

[26] S. Jégou (1997). Using Maple for symbolic differentiation to solve inverse problems. *MapleTech*, 4, 32–40.

[27] M. Kojima, N. Megiddo, T. Noma, A. Yoshise (1991). *A Unified Approach to Interior Point Algorithms for Linear Complementarity Problems*. Lecture Notes in Computer Science 538. Springer-Verlag, Berlin.

[28] G.P. McCormick (1989). The projective SUMT method for convex programming problems. *Mathematics of Operations Research*, 14, 203–223.

[29] S. Mehrotra, J. Sun (1990). An interior point algorithm for solving smooth convex programs based on Newton's method. In J.C. Lagarias, M.J. Todd (editors), *Mathematical Developments Arising from Linear Programming*, Contemporary Mathematics 114, pages 265–284. AMS, Providence, RI.

[30] R. Monteiro, I. Adler (1990). An extension of Karmarkar-type algorithms to a class of convex separable programming problems with global linear rate of convergence. *Mathematics of Operations Research*, 15, 408–422.

[31] Y.E. Nesterov, A.S. Nemirovskii (1994). *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM Studies in Applied Mathematics 13. SIAM, Philadelphia.

[32] M.J.D. Powell (1971). On the convergence of the variable metric algorithm. *Journal of the Institute of Mathematics and its Applications*, 7, 21–36.

[33] M.J.D. Powell (1976). Some global convergence properties of a variable metric algorithm for minimization without exact line searches. In R.W. Cottle, C.E. Lemke (editors), *Nonlinear Programming*, SIAM-AMS Proceedings 9. American Mathematical Society, Providence, RI.

[34] C. Roos, T. Terlaky, J.-Ph. Vial (1997). *Theory and Algorithms for Linear Optimization – An Interior Point Approach.* John Wiley & Sons, Chichester.

[35] G. Sonnevend (1986). An "analytical center" for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming. Lecture Notes in Control and Information Sciences 84, pages 866–876. Springer-Verlag, New York.

[36] T. Terlaky (editor) (1996). *Interior Point Methods of Mathematical Programming.* Kluwer Academic Press, Dordrecht.

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

[38] M.H. Wright (1995). Why a pure primal Newton barrier step may be infeasible. *SIAM Journal on Optimization*, 5, 1–12.

[39] S.J. Wright (1997). *Primal-Dual Interior-Point Methods.* SIAM Publication, Philadelphia.

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

[41] G. Zoutendijk (1970). Nonlinear programming, computational methods. In J. Abadie (editor), *Integer and Nonlinear Programming*, pages 37–86. North-Holland, Amsterdam.

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399