SIAM J. Optim.
Vol. 11, No. 1, pp. 199–222

A FEASIBLE BFGS INTERIOR POINT ALGORITHM FOR SOLVING CONVEX MINIMIZATION PROBLEMS*

PAUL ARMAND†, JEAN CHARLES GILBERT‡, AND SOPHIE JAN-JÉGOU§

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. We also show that it is globally convergent to the analytic center of the primal-dual optimal set when μ tends to 0 and strict complementarity holds.

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

AMS subject classifications. 90Cxx, 90C25, 90C51, 90C53

PII. S1052623498344720

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

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

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. To simplify the presentation and to avoid complicated notation, the case when linear equality constraints are present is discussed at the end of the paper. Since we assume that the components of c 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 $f, -c_{(1)}, \dots, -c_{(m)}$ is *strongly* convex. The reason for this latter hypothesis will be clarified below. Since the algorithm belongs to the class of interior point (IP) methods, it may be well suited for problems with many inequality constraints. It is also more efficient when the number of variables remains small or medium, say, fewer than 500, because it updates $n \times n$ matrices by a quasi-Newton (qN) formula. For problems with more variables, limited memory BFGS updates [39] can be used, but we will not consider this issue in this paper.

Our motivation is based on practical considerations. During the last 15 years much progress has been realized on IP methods for solving linear or convex minimization problems (see the monographs [29, 10, 38, 44, 23, 42, 47, 49]). For nonlinear convex problems, these algorithms assume that the second derivatives of the functions used to define the problem are available (see [43, 35, 36, 12, 38, 26]). In practice, how-

*Received by the editors September 15, 1998; accepted for publication (in revised form) January 26, 2000; published electronically August 3, 2000.
http://www.siam.org/journals/siopt/11-1/34472.html

†LACO, Faculté des Sciences, 123 av. A. Thomas, 87060 Limoges Cedex, France (Paul.Armand@unilim.fr).
‡INRIA Rocquencourt, BP 105, 78153 Le Chesnay Cedex, France (Jean-Charles.Gilbert@inria.fr).
§MIP, UFR MIG, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex 4, France (jan@mip.ups-tlse.fr).

ever, it is not uncommon to find situations where this requirement cannot be satisfied, in particular for large scale engineering problems (see [27] 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, 19, 3, 28], the computing time needed to evaluate Hessians or Hessian-vector products may be so large that IP algorithms using second derivatives may be unattractive.

This situation is familiar in unconstrained optimization. In that case, qN techniques, which use first derivatives only, have proved to be efficient, even when there are millions of variables (see [32, 20] and [9] for an example in meteorology). This fact motivates the present paper, in which we explore the possibility of combining the IP approach and qN techniques. Our ambition remains modest, however, since we confine ourselves to the question of whether the elegant BFGS theory for unconstrained convex optimization [41, 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 necessarily convex. We postpone this more difficult subject for future research (see [21, 48] 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 [17], for example) as follows: there exists a vector of multipliers $\lambda \in \mathbb{R}^m$ such that

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

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 $C = \operatorname{diag}(c_{(1)}, \ldots, c_{(m)})$ 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 [24, 36, 35, 11, 12, 1, 26, 25, 15, 7, 5]). It computes iteratively approximate solutions of the perturbed optimality system

$$\begin{cases} \nabla f(x) - \nabla c(x)\lambda = 0, \\ C(x)\lambda = \mu e, \\ (c(x), \lambda) > 0 \end{cases} \tag{1.2}$$

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 aspect of the problem, inherent in 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 $\Lambda = \operatorname{diag}(\lambda_{(1)}, \dots, \lambda_{(m)})$. 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 [16, 33, 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, [46]).

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 [14, 17] 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 forced to remain 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 [45, 1, 18])

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

to control the change in λ. This function is also used in [30, 31] as a potential function for nonlinear complementarity problems. 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

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

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 the starting matrix for the next outer iteration. If this looks attractive, it has also the inconvenience of restricting 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 qN 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 will come back to this issue in the discussion section.

A large part of the paper is devoted to the analysis of the qN 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, 13, 40]. In section 5, the overall algorithm is presented and it is shown that the sequence of outer iterates is globally convergent, in the sense that it is bounded and that its accumulation points are primal-dual solutions of problem (1.1). If, in addition, strict complementarity holds, the whole sequence of outer iterates converges to the analytic center of the primal-dual optimal set.

2. The algorithm for solving the barrier problem. The Euclidean or ℓ_2 norm is denoted by $\|\cdot\|$. We recall that a function $\xi : \mathbb{R}^n \to \mathbb{R}$ is said to be *strongly convex* with modulus $\kappa > 0$, if for all $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ one has $\xi(y) \geq \xi(x) + \nabla\xi(x)^\top(y - x) + \kappa\|y - x\|^2$ (for other equivalent definitions, see, for example, [22, Chapter IV]). Our minimal assumptions are the following.

Assumption 2.1. (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 the functions f, $-c_{(1)}, \ldots, -c_{(m)}$ 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 2.1(i) was motivated in section 1. Assumption 2.1(ii), also called the (strong) Slater condition, is necessary for the well-posedness 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 [22, Theorem VII.2.3.2], for example). These assumptions have the following clear consequence.

LEMMA 2.2. *Suppose that Assumption* 2.1 *holds. 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 [16, 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*

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

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

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

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

Proof. By Assumption 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 [14, 17]). 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 $\kappa > 0$ such that

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

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 Assumption* 2.1 *holds. 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 preceding two 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* $z \neq \hat{z}_\mu$, *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])

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

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

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 $d := (d^x, d^\lambda)$, the 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*:

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

   - 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 [50] (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

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

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:*

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

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

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 convergence analysis of BFGS, the path to q-superlinear convergence traditionally leads through r-linear convergence (see [41, 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

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

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 Assumption* 2.1 *holds. Then, the level set* $\mathcal{L}_1^{\text{PD}}$ *is compact and there exist positive constants* K_1 *and* K_2 *such that*

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

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 $z = (x, \lambda) \in \mathcal{L}_1^{\text{PD}}$. By Assumption 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 $z \in \mathcal{L}_1^{\text{PD}}$.

What we have just proven implies that $\{\varphi_\mu(x) : z = (x, \lambda) \in \mathcal{L}_1^{\text{PD}}\}$ 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 $z = (x, \lambda) \in \mathcal{L}_1^{\text{PD}}$. Hence the λ-components of the z's in $\mathcal{L}_1^{\text{PD}}$ are bounded and bounded away from zero.

We have shown that $\mathcal{L}_1^{\text{PD}}$ 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. The function ψ_μ is nonconvex, for example, when $f(x) = x^2$ is minimized on the half-line of nonnegative real numbers.

PROPOSITION 3.2. *Suppose that Assumption* 2.1 *holds. Then, there is a constant* $a > 0$ *such that for any* $z \in \mathcal{L}_1^{\text{PD}}$

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

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 as

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

From Assumption 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

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

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 an *open* neighborhood $\mathcal{N} \subset \mathcal{Z}$ of $\hat{z}_\mu$ such that

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

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 $\hat z_\mu$, $\psi_\mu(\hat z_\mu) \ge \psi_\mu(z) + \nabla\psi_\mu(z)^\top(\hat z_\mu - z)$. With the Cauchy–Schwarz inequality and the inequality on the left of (3.2), one gets

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

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

To extend the validity of (3.2) for all $z \in \mathcal{L}_1^{\mathrm{PD}}$, it suffices to note that, by virtue of Lemma 2.5, the ratios

$$\frac{\psi_\mu(z) - \psi_\mu(\hat z_\mu)}{\|z - \hat z_\mu\|^2} \quad \text{and} \quad \frac{\psi_\mu(z) - \psi_\mu(\hat z_\mu)}{\|\nabla\psi_\mu(z)\|^2}$$

are well defined and continuous on the compact set $\mathcal{L}_1^{\mathrm{PD}} \setminus \mathcal{N}$. Since $\hat z_\mu$ is the unique minimizer of ψ_μ on $\mathcal{L}_1^{\mathrm{PD}}$ (Lemma 2.5), the ratios are respectively bounded away from zero and bounded above on $\mathcal{L}_1^{\mathrm{PD}} \setminus \mathcal{N}$, by some positive constants K_1' and K_2'. The conclusion of the proposition now follows by taking $a = \min(a', K_1', 1/K_2')$. ☐

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 $\{(\gamma_k, \delta_k)\}_{k\ge1}$, satisfying for all $k \ge 1$*

$$\gamma_k^\top\delta_k \ge a_1\|\delta_k\|^2 \quad \textit{and} \quad \gamma_k^\top\delta_k \ge a_2\|\gamma_k\|^2, \tag{3.3}$$

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 $k \ge 1$,

$$b_1 \le \cos\theta_j \quad \textit{and} \quad b_2 \le \frac{\|M_j\delta_j\|}{\|\delta_j\|} \le b_3 \tag{3.4}$$

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

The assumptions (3.3) 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 to 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 (3.3).

THEOREM 3.4. *Suppose that Assumption* 2.1 *holds 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,*

$$\sum_{k\ge1}\|z_k - \hat z_\mu\| < \infty.$$

Proof. We denote by K_1', K_2', ... 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 some open convex neighborhood of the level set $\mathcal{L}_1^{\mathrm{PD}}$, for example, on

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

where $\mathcal{O}$ is an open bounded convex set containing $\mathcal{L}_1^{\mathrm{PD}}$ (this set $\mathcal{O}$ is used to have ∇c bounded on the given neighborhood).

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

$$\psi_\mu(z_{k+1}) \le \psi_\mu(z_k) - K_1' |\nabla \psi_\mu(z_k)^\top d_k| \tag{3.5}$$

or

$$\psi_\mu(z_{k+1}) \le \psi_\mu(z_k) - K_1' \frac{|\nabla \psi_\mu(z_k)^\top d_k|^2}{\|d_k\|^2}. \tag{3.6}$$

Let us now apply Lemma 3.3: fix $r \in]0,1[$ and denote by J the set of indices j for which (3.4) holds. 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 (3.5) or (3.6) gives for some positive constant K_6' and for any $j \in J$

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

The end of the proof is standard (see [41, 6]). Using Proposition 3.2, for $j \in J$,

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

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

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

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 $k \ge 1$

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

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 $z_k \ne \hat{z}_\mu$). 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 $\hat{z}_\mu$.

Given two sequences of vectors $\{u_k\}$ and $\{v_k\}$ in some normed spaces and a positive number β, we write $u_k \ge o(\|v_k\|^\beta)$, if there exists a sequence of $\{\epsilon_k\} \subset \mathbb{R}$ such that $\epsilon_k \to 0$ and $u_k \ge \epsilon_k \|v_k\|^\beta$ for all k.

PROPOSITION 4.1. *Suppose that Assumption* 2.1 *holds 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*

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

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

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

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, $\nabla \psi_\mu(z_k)^\top d_k \leq -K \|d_k\|^2$. 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 [13] 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 slightly different from (4.1).

PROPOSITION 4.2. *Suppose that Assumption* 2.1 *holds 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 $\mathcal{M}d_k$ from 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 an $o(\|d_k\|)$ 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 $z_{k+1} - \hat{z}_\mu = o(\|z_k - \hat{z}_\mu\|)$. □

For proving the q-superlinear convergence of the sequence $\{z_k\}$, we need the following result from the BFGS theory (see [40, 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*

$$\gamma_k^\top \delta_k > 0 \quad \text{for all } k \geq 1 \qquad \text{and} \qquad \sum_{k\geq 1} \frac{\|\gamma_k - M\delta_k\|}{\|\delta_k\|} < \infty, \tag{4.7}$$

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

$$(M_k - M)\delta_k = o(\|\delta_k\|). \tag{4.8}$$

By using this lemma, we will see that the BFGS formula gives the estimate

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

Note that the above estimate implies (4.5), from which the q-superlinear convergence of $\{z_k\}$ will follow.

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 Assumption* 2.1 *holds 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 (4.7), we use a Taylor expansion, assuming that k is large enough (f and c are C^2 near $\hat{x}_\mu$):

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

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

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

Hence the series in (4.7) converges by Theorem 3.4. Therefore, by (4.8) with $M = \hat{M}_\mu$ and the fact that δ_k is parallel to d_k^x,

$$(M_k - \hat{M}_\mu) d_k^x = o(\|d_k^x\|). \tag{4.9}$$

By the estimate (4.9) and Proposition 4.1, the unit step-size is accepted when k is large enough. The q-superlinear convergence of $\{z_k\}$ 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

$$\left\{ \begin{array}{l} \nabla f(\hat{x}) - \nabla c(\hat{x})\hat{\lambda} = 0, \\ C(\hat{x})\hat{\lambda} = 0, \\ (c(\hat{x}), \hat{\lambda}) \ge 0. \end{array} \right. \tag{5.1}$$

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 an *outer* iteration the collection of inner iterations for solving (1.2) for a fixed value of μ. We index the outer iterations by superscripts $j \in \mathbb{N}\backslash\{0\}$.

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. Starting from z_1^j, use Algorithm A_μ until $z^j := (x^j, \lambda^j)$ satisfies

   $$\|\nabla f(x^j) - \nabla c(x^j)\lambda^j\| \le \epsilon^j \quad \text{and} \quad \|C(x^j)\lambda^j - \mu^j e\| \le \epsilon^j. \tag{5.2}$$

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 to maneuver, while Theorem 5.1 gives us a global convergence result for such a general algorithm.

THEOREM 5.1. *Suppose that Assumption* 2.1 *holds 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 a 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

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

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 $i = 1, \ldots, m$. It follows that $\hat{x} + \mathbb{R}_+ y \subset \{x : c(x) \ge 0,\ f(x) \le f(\hat{x})\}$ (see, for example, [22, 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 $\{\lambda^j\}_{j\in J'}$ for some subsequence 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^*). Dividing the two inequalities in (5.2) by $\|\lambda^j\|$ and taking limits when $j \to \infty$, $j \in J'$, we deduce that $\nu^* \ge 0$, $\nabla c(x^*)\nu^* = 0$, and $(\nu^*)^\top c(x^*) = 0$. Using the concavity of the components $c_{(i)}$, one has

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

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 $\|\nu^*\| = 1$. $\square$

In the rest of this section, we give conditions under which the whole sequence $\{z^j\}$ converges to a particular point called the analytic center of the primal-dual optimal set. This actually occurs when the following two conditions hold: strict complementarity and a proper choice of the forcing sequence ϵ^j in Algorithm A, which has to satisfy the estimate

$$\epsilon^j = o(\mu^j),$$

meaning that $\epsilon^j/\mu^j \to 0$ when $j \to \infty$.

Let us first recall the notion of analytic center of the optimal sets, which under Assumption 2.1 is uniquely defined (see Monteiro and Zhou [37], for related results). We denote by $\mathrm{opt}(P)$ and $\mathrm{opt}(D)$ the sets of primal and dual solutions of problem (1.1). The analytic center of $\mathrm{opt}(P)$ is defined as follows. If $\mathrm{opt}(P)$ is reduced to a single point, its analytic center is precisely that point. Otherwise, $\mathrm{opt}(P)$ is a convex set with more than one point. In that case, f is not strongly convex and, by Assumption 2.1(i), at least one of the constraint functions, $-c_{(i_0)}$ say, is strongly convex. It follows that the index set

$$B := \{i : \text{there exists } \hat{x} \in \mathrm{opt}(P) \text{ such that } c_{(i)}(\hat{x}) > 0\}$$

is nonempty (it contains i_0). The analytic center of $\mathrm{opt}(P)$ is then defined as the unique solution of the following problem:

$$\max_{\substack{\hat{x}\in\mathrm{opt}(P)\\ c_B(\hat{x})>0}} \left(\sum_{i\in B} \log c_{(i)}(\hat{x}) \right). \tag{5.3}$$

The fact that this problem is well defined and has a unique solution is the matter of Lemma 5.2 below. Similarly, if $\mathrm{opt}(D)$ is reduced to a single point, its analytic center is that point. In case of multiple dual solutions, the index set

$$N := \{i : \text{there exists } \hat{\lambda} \in \mathrm{opt}(D) \text{ such that } \hat{\lambda}_{(i)} > 0\}$$

is nonempty (otherwise $\mathrm{opt}(D)$ would be reduced to $\{0\}$). The analytic center of $\mathrm{opt}(D)$ is then defined as the unique solution of the following problem:

$$\max_{\substack{\hat{\lambda}\in\mathrm{opt}(D)\\ \hat{\lambda}_N>0}} \left(\sum_{i\in N} \log \hat{\lambda}_{(i)} \right). \tag{5.4}$$

LEMMA 5.2. *Suppose that Assumption* 2.1 *holds. If* $\mathrm{opt}(P)$ (*resp.,* $\mathrm{opt}(D)$) *is not reduced to a singleton, then problem* (5.3) (*resp.,* (5.4)) *has a unique solution.*

Proof. Consider first problem (5.3) and suppose that $\mathrm{opt}(P)$ is not a singleton. We have seen that B is nonempty. By the convexity of the set $\mathrm{opt}(P)$ and the concavity of the functions $c_{(i)}$, there exists $\hat{x} \in \mathrm{opt}(P)$ such that $c_B(\hat{x}) > 0$. Therefore the feasible set in (5.3) is nonempty. On the other hand, let $\hat{x}_0$ be a point satisfying the constraints in (5.3). Then the set

$$\left\{ \hat{x} : \hat{x} \in \mathrm{opt}(P),\ c_B(\hat{x}) > 0, \text{ and } \sum_{i\in B} \log c_i(\hat{x}) \geq \sum_{i\in B} \log c_i(\hat{x}_0) \right\}$$

is nonempty, bounded (Lemma 2.2), and closed. Therefore, problem (5.3) has a solution. Finally, by Assumption 2.1(i), we know that there is an index $i_0 \in B$ such that $-c_{(i_0)}$ is strongly convex. It follows that the objective in (5.3) is strongly concave and that problem (5.3) has a unique solution.

By similar arguments and the fact that the objective function in (5.4) is strictly concave, it follows that problem (5.4) has a unique solution. ∎

By complementarity (i.e., $C(\hat{x})\hat{\lambda} = 0$) and convexity of problem (1.1), the index sets B and N do not intersect, but there may be indices that are neither in B nor in N. It is said that problem (1.1) has the *strict complementarity* property if $B \cup N = \{1, \ldots, n\}$. This is equivalent to the existence of a primal-dual solution satisfying strict complementarity.

THEOREM 5.3. *Suppose that Assumption* 2.1 *holds and that f and c are $C^{1,1}$ functions. Suppose also that problem* (1.1) *has the strict complementarity property and that the sequence $\{\epsilon^j\}$ in Algorithm A satisfies the estimate $\epsilon^j = o(\mu^j)$. Then the sequence $\{z^j\}$ generated by Algorithm A converges to the point $\hat{z}_0 := (\hat{x}_0, \hat{\lambda}_0)$, where $\hat{x}_0$ is the analytic center of the primal optimal set and $\hat{\lambda}_0$ is the analytic center of the dual optimal set.*

Proof. Let $(\hat{x}, \hat{\lambda})$ be an arbitrary primal-dual solution of (1.1). Then $\hat{x}$ minimizes $\ell(\cdot, \hat{\lambda})$ and $\hat{\lambda}^\top c(\hat{x}) = 0$, so that

$$f(\hat{x}) = \ell(\hat{x}, \hat{\lambda}) \leq \ell(x^j, \hat{\lambda}) = f(x^j) - \hat{\lambda}^\top c(x^j).$$

Using the convexity of $\ell(\cdot, \lambda^j)$ and the stopping criterion (5.2) of the inner iterations in Algorithm A, one has

$$\begin{aligned} f(\hat{x}) - (\lambda^j)^\top c(\hat{x}) &= \ell(\hat{x}, \lambda^j) \\ &\geq \ell(x^j, \lambda^j) + \nabla_x \ell(x^j, \lambda^j)^\top (\hat{x} - x^j) \\ &= f(x^j) - (\lambda^j)^\top c(x^j) - \epsilon^j \|x^j - \hat{x}\| \\ &\geq f(x^j) - m\mu^j - m^{\frac{1}{2}}\epsilon^j - \epsilon^j \|x^j - \hat{x}\|, \end{aligned}$$

because $(\lambda^j)^\top c(x^j) = m\mu^j + e^\top (C(x^j)\lambda^j - \mu^j e) \leq m\mu^j + m^{\frac{1}{2}}\epsilon^j$. By Theorem 5.1, there is a constant C_1 such that $m^{\frac{1}{2}} + \|x^j - \hat{x}\| \leq C_1$. Then, adding the corresponding sides of the two inequalities above leads to

$$\hat{\lambda}_N^\top c_N(x^j) + (\lambda_B^j)^\top c_B(\hat{x}) = \hat{\lambda}^\top c(x^j) + (\lambda^j)^\top c(\hat{x}) \leq m\mu^j + C_1 \epsilon^j. \tag{5.5}$$

We pursue this by adapting an idea used by McLinden [34] to give properties of the limit points of the path $\mu \mapsto (\hat{x}_\mu, \hat{\lambda}_\mu)$. Let us define $\Gamma^j := C(x^j)\lambda^j - \mu^j e$. One has for all indices i

$$c_{(i)}(x^j) = \frac{\mu^j + \Gamma^j_{(i)}}{\lambda^j_{(i)}} \qquad \text{and} \quad \lambda^j_{(i)} = \frac{\mu^j + \Gamma^j_{(i)}}{c_{(i)}(x^j)}.$$

Substituting this in (5.5) and dividing by μ^j give

$$\sum_{i \in N} \frac{\hat{\lambda}_{(i)}}{\lambda^j_{(i)}} \frac{\mu^j + \Gamma^j_{(i)}}{\mu^j} + \sum_{i \in B} \frac{c_{(i)}(\hat{x})}{c_{(i)}(x^j)} \frac{\mu^j + \Gamma^j_{(i)}}{\mu^j} \leq m + C_1 \frac{\epsilon^j}{\mu^j}.$$

By assumptions, $\epsilon^j = o(\mu^j)$, so that $\Gamma^j_{(i)} = o(\mu^j)$. Now supposing that $(\hat{x}_0, \hat{\lambda}_0)$ is a limit point of $\{(x^j, \lambda^j)\}$ and taking the limit in the preceding estimate provide

$$\sum_{i \in N} \frac{\hat{\lambda}_{(i)}}{(\hat{\lambda}_0)_{(i)}} + \sum_{i \in B} \frac{c_{(i)}(\hat{x})}{c_{(i)}(\hat{x}_0)} \leq m.$$

Necessarily, $c_B(\hat{x}_0) > 0$ and $(\hat{\lambda}_0)_N > 0$. Observe now that, by strict complementarity, there are exactly m terms on the left-hand side of the preceding inequality. Hence, by the arithmetic-geometric mean inequality

$$\left(\prod_{i\in N} \frac{\hat{\lambda}_{(i)}}{(\hat{\lambda}_0)_{(i)}}\right)\left(\prod_{i\in B} \frac{c_{(i)}(\hat{x})}{c_{(i)}(\hat{x}_0)}\right) \le 1$$

or

$$\left(\prod_{i\in N} \hat{\lambda}_{(i)}\right)\left(\prod_{i\in B} c_{(i)}(\hat{x})\right) \le \left(\prod_{i\in N} (\hat{\lambda}_0)_{(i)}\right)\left(\prod_{i\in B} c_{(i)}(\hat{x}_0)\right).$$

One can take $\hat{\lambda}_N = (\hat{\lambda}_0)_N > 0$ or $c_B(\hat{x}) = c_B(\hat{x}_0) > 0$ in this inequality, so that

$$\prod_{i\in B} c_{(i)}(\hat{x}) \le \prod_{i\in B} c_{(i)}(\hat{x}_0) \qquad \text{and} \quad \prod_{i\in N} \hat{\lambda}_{(i)} \le \prod_{i\in N} (\hat{\lambda}_0)_{(i)}.$$

This shows that $\hat{x}_0$ is a solution of (5.3) and that $\hat{\lambda}_0$ is a solution of (5.4). Since the problems in (5.3) and (5.4) have unique solutions, all the sequence $\{x^j\}$ converges to $\hat{x}_0$ and all the sequence $\{\lambda^j\}$ converges to $\hat{\lambda}_0$. $\square$

6. Discussion. By way of conclusion, we discuss the results obtained in this paper, give some remarks, 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

$$\text{(6.1)} \qquad \begin{cases} \min f(x), \\ Ax = b, \\ c(x) \ge 0, \end{cases}$$

obtained by adding linear constraints to problem (1.1). In (6.1), A is a $p \times n$ matrix with $p < n$ and $b \in \mathbb{R}^p$ 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 an $n \times q$ 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

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

With this notation, problem (6.1) can be rewritten as the problem in $u \in \mathbb{R}^q$:

$$\text{(6.2)} \qquad \begin{cases} \min f(x_1 + Zu), \\ c(x_1 + Zu) \ge 0, \end{cases}$$

which has the form (1.1).

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

Assumption 6.1. (i) The real-valued functions f and $-c_{(i)}$ $(1 \le i \le m)$ are convex and differentiable on the affine subspace $X := \{x : Ax = b\}$ and at least one of the functions f, $-c_{(1)}, \ldots, -c_{(m)}$ is strongly convex on X. (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 apply if f and c are also $C^{1,1}$.

Feasible algorithms and qN techniques. In the framework of qN 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 q only, instead of order n for an infeasible algorithm solving problem (6.1) directly. When $q \ll n$, the qN 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 qN 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 semidefinite, provided M_k is already positive semidefinite. 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.

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 essential for the efficiency of the algorithm and would deserve a detailed numerical study.

From a theoretical viewpoint, it would be highly desirable to have an update rule that would allow the outer iterates of Algorithm A to converge q-superlinearly. Along the same lines, an interesting problem is to design an algorithm in which the barrier parameter would be updated at every step, while having q-superlinear convergence of the iterates. Such extensions would involve more difficult issues.

The global convergence result proved in this paper gives us some reasons to believe that it is not unreasonable to tackle these open questions.

Acknowledgments. We would like to thank the referees for their valuable comments. One of them has shown us a direct argument for the last part of the proof of Proposition 3.2, which is the one we have finally chosen to give in the paper. The other referee has brought McLinden's paper to our attention, which led us to Theorem 5.3.

REFERENCES

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

[2] A. Auslender, R. Cominetti, and M. Haddou, *Asymptotic analysis for penalty and barrier methods in convex and linear programming*, Math. Oper. Res., 22 (1997), pp. 43–62.

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

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

[5] R.H. Byrd, J.Ch. Gilbert, and J. Nocedal, *A Trust Region Method Based on Interior Point Techniques for Nonlinear Programming*, Technical report 2896, INRIA, Le Chesnay, France, 1996; Math. Programming, submitted.

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

[7] A.R. Conn, N.I.M. Gould, and Ph.L. Toint, *A Primal-Dual Algorithm for Minimizing a Non-convex Function Subject to Bound and Linear Equality Constraints*, Technical Report, Facultés Universitaires de Namur, Belgique, 1996.

[8] A. Griewank, *Evaluating Derivatives. Principles and Techniques of Algorithmic Differentiation,* Frontiers in Applied Mathematics 19, SIAM, Philadelphia, 2000.

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

[10] D. den Hertog, *Interior Point Approach to Linear, Quadratic and Convex Programming*, Math. Appl. 277, Kluwer, Dordrecht, The Netherlands, 1992.

[11] D. den Hertog, C. Roos, and T. Terlaky, *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, 1990.

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

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

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

[15] 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., 89 (1996), pp. 507–541.

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

[17] R. Fletcher, *Practical Methods of Optimization,* 2nd ed., John Wiley, Chichester, UK, 1987.

[18] A. Forsgren and P.E. Gill, *Primal-dual interior methods for nonconvex nonlinear programming*, SIAM J. Optim., 8 (1998), pp. 1132–1152.

[19] J.Ch. Gilbert, G. Le Vey, and J. Masse, *La différentiation automatique de fonctions représentées par des programmes*, Rapport de Recherche 1557, INRIA, Le Chesnay, France, 1991; also available online from http://www.inria.fr/RRRT/RR-1557.html; available via ftp at ftp://ftp.inria.fr/INRIA/publication/RR, RR-1557.ps.gz.

[20] J.Ch. Gilbert and C. Lemaréchal, *Some numerical experiments with variable-storage quasi-Newton algorithms*, Math. Programming, 45 (1989), pp. 407–435.

[21] J. Herskovits, *Feasible direction interior-point technique for nonlinear optimization*, J. Optim. Theory Appl., 99 (1998), pp. 121–146.

[22] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, Grundlehren Math. Wiss. 305–306, Springer-Verlag, Berlin, New York, 1993.

[23] B. Jansen, *Interior Point Techniques in Optimization—Complementarity, Sensitivity and Algorithms*, Appl. Optim. 6, Kluwer, Dordrecht, The Netherlands, 1997.

[24] F. Jarre, *On the method of analytic centers for solving smooth convex problems*, Lecture Notes in Math. 1405, Springer-Verlag, Berlin, 1989, pp. 69–85.

[25] F. Jarre, *Interior-point methods for classes of convex programs*, in Interior Point Methods of Mathematical Programming, T. Terlaky, ed., Kluwer, Dordrecht, The Netherlands, 1996, chapter 7.

[26] F. Jarre and M.A. Saunders, *A practical interior-point method for convex programming*, SIAM J. Optim., 5 (1995), pp. 149–171.

[27] S. Jégou, *Estimation des perméabilités relatives dans des expériences de déplacements triphasiques en milieu poreux*, Ph.D. thesis, Université de Paris IX (Dauphine), Paris, 1997.

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

[29] M. Kojima, N. Megiddo, T. Noma, and A. Yoshise, *A Unified Approach to Interior Point Algorithms for Linear Complementarity Problems*, Lecture Notes in Comput. Sci. 538, Springer-Verlag, Berlin, 1991.

[30] M. Kojima, S. Mizuno, and T. Noma, *A new continuation method for complementarity problems with uniform P-functions*, Math. Programming, 43 (1989), pp. 107–113.

[31] M. Kojima, S. Mizuno, and T. Noma, *Limiting behavior of trajectories generated by a continuation method for monotone complementarity problems*, Math. Oper. Res., 15 (1990), pp. 662–675.

[32] D.C. Liu and J. Nocedal, *On the limited memory BFGS method for large scale optimization*, Math. Programming, 45 (1989), pp. 503–520.

[33] G.P. McCormick, *The projective SUMT method for convex programming problems*, Math. Oper. Res., 14 (1989), pp. 203–223.

[34] L. McLinden, *An analogue of Moreau's proximation theorem, with application to the nonlinear complementarity problem*, Pacific J. Math., 88 (1980), pp. 101–161.

[35] S. Mehrotra and J. Sun, *An interior point algorithm for solving smooth convex programs based on Newton's method*, in Mathematical Developments Arising from Linear Programming, J.C. Lagarias and M.J. Todd, eds., Contemp. Math. 114, AMS, Providence, RI, 1990, pp. 265–284.

[36] R. Monteiro and I. Adler, *An extension of Karmarkar-type algorithms to a class of convex separable programming problems with global linear rate of convergence*, Math. Oper. Res., 15 (1990), pp. 408–422.

[37] R.D.C. Monteiro and F. Zhou, *On the existence and convergence of the central path for convex programming and some duality results*, Comput. Optim. Appl., 10 (1998), pp. 51–77.

[38] Y. Nesterov and A.S. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM Stud. Appl. Math. 13, SIAM, Philadelphia, 1994.

[39] J. Nocedal, *Updating quasi-Newton matrices with limited storage*, Math. Comput., 35 (1980), pp. 773–782.

[40] M.J.D. Powell, *On the convergence of the variable metric algorithm*, J. Inst. Math. Appl., 7 (1971), pp. 21–36.

[41] M.J.D. Powell, *Some global convergence properties of a variable metric algorithm for minimization without exact line searches*, in Nonlinear Programming, R.W. Cottle and C.E. Lemke, eds., SIAM-AMS Proceedings 9, AMS, Providence, RI, 1976.

[42] C. Roos, T. Terlaky, and J.-Ph. Vial, *Theory and Algorithms for Linear Optimization—An Interior Point Approach*, John Wiley, Chichester, UK, 1997.

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

[44] T. Terlaky ed., *Interior Point Methods of Mathematical Programming*, Kluwer, Dordrecht, The Netherlands, 1996.

[45] J.-P. Vial, *Computational experience with a primal-dual interior-point method for smooth convex programming*, Optim. Methods Softw., 3 (1994), pp. 285–310.

[46] M.H. WRIGHT, *Why a pure primal Newton barrier step may be infeasible*, SIAM J. Optim., 5 (1995), pp. 1–12.

[47] S.J. WRIGHT, *Primal-Dual Interior-Point Methods*, SIAM, Philadelphia, 1997.

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

[49] Y. YE, *Interior Point Algorithms—Theory and Analysis*, Wiley-Intersci. Ser. Discrete Math. Optim., John Wiley, New York, 1997.

[50] G. ZOUTENDIJK, *Nonlinear programming, computational methods*, in Integer Nonlinear Programming, J. Abadie, ed., North-Holland, Amsterdam, 1970, pp. 37–86.